

AVANCES EN INFORMÁTICA Y AUTOMÁTICA

DECIMOCTAVA EDICIÓN



VNIVERSIDAD
D SALAMANCA



MÁSTER UNIVERSITARIO
EN SISTEMAS INTELIGENTES

Avances en Informática y Automática

Decimoctavo Workshop

Avances en Informática y Automática

Decimoctavo Workshop

Editores

Arturo Álvarez Sánchez
Rubén Herrero Pérez

Publicado en España por:

Departamento de Informática y Automática Facultad de Ciencias
Universidad de Salamanca
Plaza de los Caídos s/n
37008, Salamanca, España
Tel.: + 34 923294653
Fax: + 34 923294514
Web: <http://mastersi.usal.es>
Email: mastersi@usal.es

ISBN 978-84-09-83945-2

Editores:

Arturo Álvarez Sánchez
Rubén Herrero Pérez

Prólogo

El Máster Universitario en Sistemas Inteligentes de la Universidad de Salamanca tiene como misión fundamental introducir a sus estudiantes en el rigor de la investigación científica. En este contexto, el congreso organizado por el Departamento de Informática y Automática se consolida como el escenario idóneo para que los alumnos presenten los resultados de sus Trabajos de Fin de Máster (TFM), sometiénolos al análisis y debate propios de la comunidad académica.

La decimoctava edición del workshop “Avances en Informática y Automática”, celebrada durante el curso 2024 - 2025, ha destacado por su marcado carácter multidisciplinar. En esta ocasión, las investigaciones presentadas abarcan un espectro tecnológico de vanguardia, incluyendo:

- **Procesamiento de Lenguaje Natural y LLMs:** Desde la clasificación de textos en sectores como la hostelería hasta el ajuste fino de modelos para detectar visualizaciones engañosas.
- **Machine Learning y Predicción:** Aplicaciones prácticas en la previsión de demanda para retail, optimización energética en sistemas intralogísticos y control de alineamiento en sistemas láser de alta potencia.
- **Visión Artificial y Ciberseguridad:** Comparativas de modelos para la detección de violencia y el desarrollo de sistemas IDS/IPS mediante deep learning.
- **Gestión de Datos y Sociedad:** Análisis de espacios de datos en el sector Agrotech, desambiguación de autores en bases bibliográficas y herramientas de IA para facilitar el preprocesamiento de datos a usuarios no expertos.

Bajo la supervisión de investigadores de prestigio de la Universidad de Salamanca, este encuentro no solo valida la calidad técnica de los trabajos, sino que sirve de puente hacia la realización de futuras tesis doctorales. Los objetivos principales del evento se mantienen firmes:

- **Exposición:** Brindar a los estudiantes su primera experiencia formal en la difusión de resultados de investigación.
- **Intercambio:** Crear un foro de discusión donde converjan ideas de compañeros, docentes y expertos.
Retroalimentación: Facilitar críticas constructivas que orienten las futuras líneas de investigación de los egresados.
- **Colaboración:** Fortalecer el espíritu de trabajo conjunto y la sinergia entre diferentes áreas de conocimiento.

Organización

El *workshop* “Avances en Informática y Automática” está organizado por el Departamento de Informática y Automática de la Universidad de Salamanca.

Comité Organizativo

María Belén Curto Diego
Francisco J. Blanco Rodríguez
Luis A. Miguel Quintales
Ángel L. Sánchez Lázaro
Pastora I. Vega Cruz
María N. Moreno García
Llanos Navarra Laespada
Angélica González Arrieta
Francisco J. García Peñalvo
Vidal Moreno Rodilla
Roberto Therón Sánchez

Índice general

Uso de información secundaria heterogénea en sistemas de recomendación	1
<i>A. Álvarez Sánchez, M. N. Moreno García, D. M. Jiménez-Bravo</i>	
Machine Learning for Retail Demand Forecasting	23
<i>T. Bilel Seifeddine, Á. Lozano Murciego, M. N. Moreno García</i>	
Ajuste fino de modelos de lenguaje de gran tamaño para la detección de visualizaciones de datos engañosas basadas en elementos primitivos	47
<i>D. Marcelo Miranda Rodríguez, R. Therón Sánchez, A. Vázquez Ingelmo</i>	
Desambiguación de nombres de autor en bases de datos bibliográficas con pocos metadatos	71
<i>D. Uribe Esquivel, Á. F. Zazo Rodríguez</i>	
Espacios de datos y su aplicación en el sector Agrotech	94
<i>G. Milagros Miguel Jurado, B. Pérez Lancho, A. B. Gil González</i>	
Clasificación Automática de Textos de Hotelería Basada en Transformadores	113
<i>J. Jimenez Valido, M.A. González Arrieta, y G. Hernández</i>	
Aplicación de técnicas de lingüística computacional y visualización para identificar patrones en la creación de historias	119
<i>P. Díez Gastón, R. Therón Sánchez, A. Vázquez Ingelmo</i>	
Modelos de predicción para la mejora del control del alineamiento de un sistema láser de alta potencia	153
<i>J. D. Pisonero Carabias, M. A. González Arrieta, F. Galán Prado, G. Hernández González</i>	
Optimización energética de sistemas intralogísticos mediante técnicas de IA	174
<i>M. Jammal, J. P. Domínguez, J. M. Corchado Rodríguez</i>	
Visión e Inteligencia Artificial: Comparativa de Modelos en Detección de Violencia	203
<i>M. Hernández Lemos, V. Moreno Rodilla y B. Curto Diego</i>	

Incorporación de técnicas de IA para asistir al preprocesamiento de datos para usuarios no expertos	225
<i>M. de Castro Leira, A. García Holgado y A. Vázquez Ingelmo</i>	
Sistemas de IDS/IPS en redes utilizando deep learning	244
<i>Ó. de las Heras García, A. González Arrieta y Á. Bustos-Tabernero</i>	
Evaluación teórico-práctica de modelos de detección de objetos en imágenes	266
<i>R. Herrero Pérez, V. Félix López Batista y G. Villarrubia González</i>	

Uso de información secundaria heterogénea en sistemas de recomendación

Arturo Álvarez Sánchez, María N. Moreno García, and Diego M. Jiménez-Bravo

Universidad de Salamanca
Departamento de Informática y Automática
Facultad de Ciencias.
Plaza de los Caídos s/n
37008 Salamanca, España
{arturoas,mmg,dmjimenez}@usal.com

Resumen Los sistemas de recomendación son fundamentales en plataformas digitales, facilitando a los usuarios el descubrimiento de productos, contenido y servicios. Tradicionalmente, se han basado en métodos como el filtrado colaborativo o métodos basados en contenido, utilizando datos como interacciones usuario-ítem o metadatos de los productos. Con el auge del *big data*, ha aumentado la disponibilidad de información secundaria como datos contextuales, de comportamiento social y metadatos, que permite generar recomendaciones más personalizadas. Este estudio realiza un análisis comparativo entre sistemas tradicionales y modelos recientes que incorporan información secundaria, en particular características visuales extraídas de imágenes. Se examina si la integración de estos datos complejos mejora significativamente la calidad de las recomendaciones. Asimismo, se analizan los beneficios y desafíos que implica el uso del contenido visual. El objetivo es contribuir al desarrollo de sistemas de recomendación más robustos, adaptativos e innovadores, proporcionando además nuevas perspectivas para futuras investigaciones en este campo.

Keywords: Sistemas de recomendación, aprendizaje automático, información secundaria, visión artificial, aprendizaje profundo.

1. Introducción

La aparición de Internet y el auge de los sistemas web han dado lugar a una disponibilidad masiva de información y opciones para los usuarios. Esta abundancia, aunque beneficiosa, ha generado un problema creciente de sobrecarga informativa, dificultando la localización eficiente de contenidos relevantes. Para abordar esta problemática, los sistemas de recomendación han emergido como una solución clave,

proporcionando sugerencias personalizadas a partir del historial de interacción del usuario. Estos sistemas han demostrado ser especialmente eficaces en dominios como el comercio electrónico, el turismo y los servicios multimedia, donde han mejorado significativamente la experiencia del usuario y la toma de decisiones.

Dentro de los enfoques más utilizados destacan los métodos basados en contenido y los de filtrado colaborativo. Los primeros hacen uso de los atributos descriptivos de los ítems para recomendar productos similares a aquellos con los que el usuario ha interactuado previamente. Los segundos, por su parte, infieren preferencias a partir de patrones compartidos entre usuarios con gustos similares, utilizando matrices de valoraciones o interacciones. En ambos casos, las valoraciones pueden ser explícitas (introducidas directamente por el usuario) o implícitas (inferidas a partir de acciones como clics o compras), cada una con sus propias limitaciones en cuanto a la representación del interés real del usuario.

Pese a su utilidad, estos métodos enfrentan desafíos importantes. Uno de los más comunes es la dispersión de los datos, donde la mayoría de los usuarios han valorado solo una pequeña fracción del catálogo disponible. Otro es el problema del arranque en frío, que dificulta generar recomendaciones cuando se incorporan nuevos usuarios o ítems sin historial de interacciones. Para superar estas limitaciones, se ha recurrido al uso de información secundaria o *side information*, la cual complementa los datos primarios con información adicional sobre los usuarios (datos demográficos, relaciones sociales, etc.), sobre los ítems (géneros, descripciones, características técnicas, etc.) o información contextual (momento temporal, ubicación o dispositivo de acceso del usuario, entre otras), utilizada por los denominados sistemas de recomendación sensibles al contexto.

La literatura recoge numerosas propuestas que integran diferentes tipos de información secundaria. No obstante, ciertos tipos de datos no estructurados, como las imágenes asociadas a los ítems, han sido explorados en menor medida. Este tipo de información, al estar compuesta por características visuales complejas, presenta un desafío adicional que requiere el uso de técnicas avanzadas como la visión artificial y el aprendizaje profundo. En base a esto, se propone un estudio del comportamiento de diversos sistemas de recomendación, junto a información secundaria visual. El objetivo general es evaluar en qué medida mejora esta información secundaria adicional las recomendaciones de los sistemas propuestos.

El presente artículo se estructura de la siguiente manera: tras esta introducción, se presenta el estado del arte en relación con los sistemas de recomendación y el uso de información secundaria; después se describe la metodología empleada y el caso de estudio seleccionado; a continuación, se detallan los resultados obtenidos y su análisis;

finalmente, se discuten las conclusiones y posibles líneas futuras de investigación.

2. Estado del arte

Como ya se ha comentado, las técnicas tradicionales de recomendación, como el filtrado colaborativo y los métodos basados en el contenido, suelen enfrentarse a limitaciones debidas a la escasez de datos y al problema del arranque en frío. Para hacer frente a estos retos, la investigación reciente se ha centrado cada vez más en la incorporación de información secundaria contextual o datos auxiliares que enriquecen la matriz de interacción usuario-elemento. En esta sección se repasan los tipos de información secundaria que se suelen utilizar en los sistemas de recomendación y se explora cómo se integran en distintos paradigmas de recomendación.

La información secundaria utilizada en la mitigación de los problemas mencionados previamente puede ser de diversos tipos. Esta incluye metadatos y atributos de los ítems, así como datos derivados de redes sociales que permiten inferir preferencias basadas en contactos cercanos. Las etiquetas o categorías asociadas con los ítems, como los géneros de películas o tipos de música, también suelen reflejar los gustos de los usuarios. Sun et al. [1] proponen una categorización de la información secundaria basada en su estructura, dividiéndola en datos estructurados y no estructurados.

Dentro de esta clasificación destacan los datos no estructurados por su complejidad de uso, entre los cuales se incluyen datos textuales y visuales. Los datos textuales suelen ser extraídos de reseñas o descripciones, obtenidos mediante modelos de factores latentes o redes neuronales [2, 3], mientras que los visuales suelen provenir de imágenes asociadas a los ítems como portadas o carteles [4, 5], se obtienen mediante autoencoders, redes neuronales o modelos basados en transformadores como *Vision Transformers (ViT)*. Este último, propuesto por Dosovitskiy et al. [6], adapta la arquitectura *Transformer* al dominio visual, dividiendo las imágenes en fragmentos, incorporando codificaciones posicionales y procesándolos como secuencias, logrando un rendimiento competitivo con menor coste computacional en comparación con redes convolucionales tradicionales.

2.1 Métodos de recomendación que incorporan información secundaria

Existen diferentes técnicas que se emplean para integrar información secundaria en sistemas de recomendación, especialmente datos textuales y visuales no estructurados, para mejorar las recomendaciones en técnicas tradicionales basadas en filtrado colaborativo, contenido y aprendizaje profundo.

En cuanto a los modelos de filtrado colaborativo con información secundaria, la incorporación de características textuales y visuales ha demostrado mejorar las recomendaciones. De cara a las características textuales, los enfoques incluyen modelos basados en memoria que calculan similitud semántica o analizan sentimiento en reseñas [7], y modelos de factores latentes enriquecidos con representaciones semánticas de palabras [8], análisis de sentimiento [9], modelado de aspectos [10] y modelado de temas [11]. Para los datos visuales, las redes neuronales convolucionales se han utilizado para extraer características que se combinan con factores latentes, beneficiando sistemas en contextos de escasez de datos [4, 12].

Respecto a los métodos basados en contenido con información secundaria, se utilizan técnicas semánticas como *TF-IDF* (*Term Frequency-Inverse Document Frequency*), *embeddings* de palabras y conjuntos de sinónimos para representar texto, mejorando la evaluación y mitigando el problema del arranque en frío [13, 14]. Para las características visuales, se emplean algoritmos de visión por computadora y redes como *VGG19* para complementar la información textual y reforzar la retroalimentación del usuario [15, 16].

En cuanto a los modelos de aprendizaje profundo con información secundaria, estos integran dicha información para superar limitaciones de datos, utilizando *CNNs* (*Convolutional Neural Networks*), *RNNs* (*Recurrent Neural Networks*) y *autoencoders* para procesar texto [17, 18]. Las arquitecturas neuronales también se combinan con técnicas de factorización para extraer aspectos y sentimientos [19]. Modelos basados en *Transformers* y *LLMs* (Large Language Models), como *BERT* (*Bidirectional Encoder Representations from Transformers*) y *RoBERTa*, generan *embeddings* contextuales que mejoran recomendaciones y explicaciones [20, 21]. Para las características visuales, se utilizan redes entrenadas desde cero o preentrenadas, como *ResNet*, *VGG*, para extraer representaciones visuales [22, 23], junto con modelos *Transformers* especializados como *FashionBERT* [24]. Además, modelos híbridos como *VBPR* integran vectores visuales en la factorización de matrices para mejorar la clasificación y mitigar el arranque en frío [25].

3. Metodología

Siguiendo la sección anterior sobre el estado del arte, se define la metodología que servirá como base para los casos de estudio. Se ha diseñado un sistema estructurado en dos módulos diferenciados que lleva a cabo el proceso de recomendación en dos etapas, con el objetivo de optimizar el uso de la información secundaria disponible para generar recomendaciones finales más precisas y personalizadas. La estructura general se ilustra en la Figura 1.

Esta arquitectura modular permite una fácil adaptación a diferentes casos de estudio mediante la posibilidad de diferentes combi-

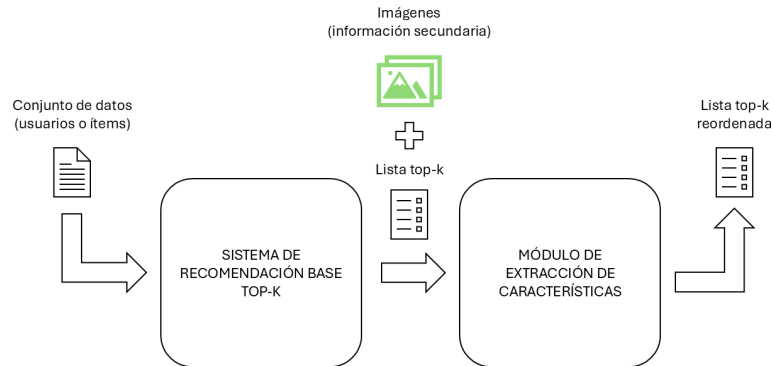


Figura 1: Estructura general de la metodología propuesta.

naciones y variaciones del sistema, dependiendo de los requerimientos específicos del estudio. Cada módulo se describe en detalle a continuación:

- Sistema de recomendación base top- k :** este módulo se basa en sistemas tradicionales de recomendación, específicamente en enfoques basados en contenido y de filtrado colaborativo, proporcionando una base sólida para la experimentación. El conjunto de datos empleado, dependiendo de la situación específica, se introduce al inicio del módulo y sirve como base para generar las recomendaciones. Esta configuración permite una amplia gama de experimentos, posibilitando la evaluación de distintos algoritmos y modelos. Un requisito común para todas las implementaciones es su capacidad para generar listas de recomendación top- k , es decir, listas que contengan los k ítems más relevantes para cada usuario según sus preferencias.
- Módulo de extracción de características y reordenamiento:** este módulo tiene como objetivo implementar un sistema para reordenar las listas de recomendación top- k previamente generadas por el sistema base de recomendación. El proceso de reordenamiento se basa en la similitud visual, basándose en extracción de características, entre las imágenes asociadas a los ítems recomendados y las asociadas a los ítems favoritos del usuario. En concreto, para cada ítem recomendado, el sistema compara su imagen asociada con las imágenes de los ítems favoritos y se calcula una puntuación promediando las similitudes por pares. La lista se reordena de forma que los ítems con mayor similitud visual aparezcan antes en la lista. Un enfoque potencial y reciente para realizar esta tarea son los *Transformers*, específicamente *Vision Transformers (ViT)*. *ViT*, como ya se ha comentado anteriormente, es una arquitectura *Transformer* diseñada para el procesamiento de imágenes, que divide cada imagen en fragmentos. Con esto en mente, se obtienen los vectores

de características que sirven como base para calcular la similitud visual y, posteriormente, reordenar la lista top- k .

4. Casos de estudio y experimentación

Con el objetivo de evaluar la eficacia de distintos sistemas de recomendación, se han diseñado casos de estudio que simulan escenarios reales de recomendación de películas. Se comparan enfoques de filtrado colaborativo y basados en contenido, así como el impacto de una etapa adicional de reordenación basada en similitud visual. A continuación se describen los enfoques base que se van a emplear junto a sus abreviaciones para agilizar futuras explicaciones:

- **TF-IDF**: modelo base basado en contenido empleando frecuencia de texto.
- **SE**: modelo base basado en contenido haciendo uso de *embeddings* de frases.
- **KNN**: modelo base basado en filtrado colaborativo aplicando el método de los vecinos más cercanos.
- **SVD**: modelo base basado en filtrado colaborativo usando el método de descomposición en valores singulares.

Cada usuario seleccionado ha calificado al menos 20 películas, permitiendo la construcción de conjuntos personalizados de entrenamiento y prueba. Las recomendaciones se generan a partir del conjunto de entrenamiento y se evalúan frente al conjunto de prueba mediante métricas top- k ($NDCG@k$, $MAP@k$, $MRR@k$).

De cara al reordenamiento visual se han empleado diversos modelos *Visual Transformer* (ViT), incluyendo variantes como *Data-efficient image Transformers* ($DeiT$), *DINOv2* y *Bidirectional Encoder representation from Image Transformers* ($BEiT$), con distintas arquitecturas y estrategias de entrenamiento (supervisadas y auto-supervisadas). Esta diversidad permite mitigar sesgos individuales y mejorar la robustez del sistema.

La evaluación experimental mostrada en la Figura 2 se estructura en cinco fases: (1) carga y preparación de datos, (2) inducción del modelo base, (3) evaluación de la recomendación inicial, (4) reordenación visual de la lista recomendada y (5) evaluación final. Esta secuencia permite analizar de forma comparativa el rendimiento de los sistemas antes y después de incorporar la información visual.

A continuación, se detallan los conjuntos de datos utilizados, junto a los casos de estudio definidos para esta investigación.

4.1 Conjuntos de datos utilizados

Para desarrollar los sistemas de recomendación propuestos, se utilizaron diversos conjuntos de datos seleccionados según el enfoque

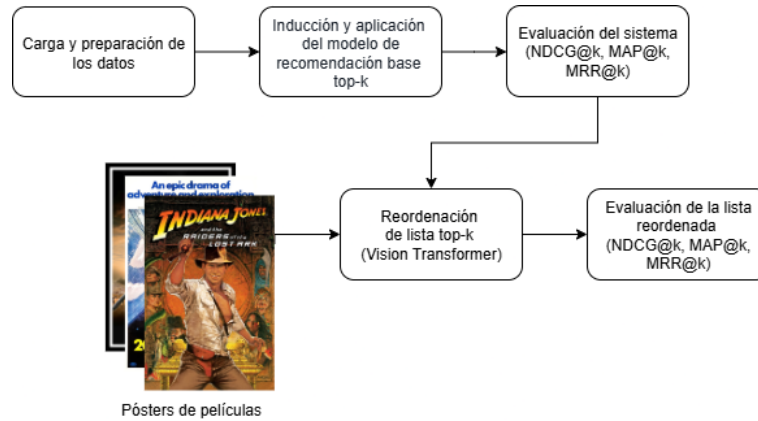


Figura 2: Esquema general del proceso experimental para el caso basado en contenido.

adoptado (basado en contenido o filtrado colaborativo). Para incorporar información visual complementaria, se recurrió a fuentes externas para obtener los recursos visuales asociados a cada ítem. La elección de los conjuntos de datos se fundamenta en su uso extendido en la investigación sobre sistemas de recomendación y la riqueza de su información.

La fuente principal es The Movies Dataset¹. Este conjunto de datos respalda los modelos basados en contenido mediante información textual y metadatos como:

- *movieId*: identificador único para cada película.
- *title*, *genres*, *overview*, y *tagline*: descriptores semánticos.
- *release_date*, *production_companies*, *runtime*: metadatos auxiliares.

Para incorporar las valoraciones de los usuarios, se utiliza el conjunto de datos MovieLens², en particular el subconjunto *ratings_small.csv* integrado dentro de The Movies Dataset. Este conjunto contiene:

- *userId*, *movieId*: identificadores que vinculan usuarios y películas,
- *rating*: puntuaciones otorgada por los usuario a las películas (de 0.5 a 5.0),
- *timestamp*: fechas de la valoraciones en formato epoch Unix.

Para enriquecer el proceso de recomendación con elementos visuales, se utiliza la API de acceso abierto TMDB³ para obtener pósters de películas. La integración se realiza mediante el archivo `links.csv`, que incluye los campos `movieId`, `imdbId` y `tmdbId` como campos de referencia cruzada entre conjuntos de datos.

¹ <https://www.kaggle.com/datasets/rounakbanik/the-movies-dataset>

² <https://grouplens.org/datasets/movielens/>

³ <https://developers.themoviedb.org/3>

4.2 Experimento 1: caso de estudio basado en contenido

Se desarrolla inicialmente un sistema de recomendación basado exclusivamente en la información textual de las películas (sinopsis), asumiendo que películas con sinopsis similares a los previamente valorados positivamente por un usuario podrían ser de su interés. Para ello se emplea el conjunto *The Movie Dataset*. Se exploran dos representaciones textuales: *TF-IDF* y *Sentence Embeddings* (modelo *SentenceTransformer*).

- **Modelo base:** el sistema emplea valoraciones históricas y descripciones textuales para recomendar películas no vistas. Incluye:
 1. **Vectorización:** mediante TF-IDF se generan vectores dispersos que destacan términos relevantes y permiten cálculos de similitud coseno. Con *Sentence Embeddings*, se obtienen vectores densos de baja dimensión que capturan semántica contextual.
 2. **Similitud:** se calcula la similitud coseno entre películas.
 3. **Selección del usuario:** se seleccionan 10 películas valoradas positivamente ($\text{rating} \geq 3$) por usuarios.
 4. **Recomendación:** se genera un vector de similitud agregado y se excluyen películas ya vistas. Se recomiendan las más similares.
- **Reordenación:** Se reordenan los ítems de la lista de recomendación utilizando los modelos descritos en el apartado 4.4

El sistema resulta robusto ante problemas de arranque en frío y dispersión, mejorado con una capa de reordenación visual que considera aspectos estéticos relevantes para el usuario.

4.3 Experimento 2: caso de estudio basado en filtrado colaborativo

Se implementa un sistema de recomendación basado únicamente en valoraciones de usuarios, asumiendo que usuarios con patrones similares tendrán gustos similares. Para ello se emplea el conjunto *ratings_small.csv*. Se emplean dos variantes: filtrado por vecinos (KNN) y factorización de matrices (SVD).

- **Modelo base:** el sistema emplea valoraciones históricas para predecir preferencias, con las siguientes fases:
 1. **Entrenamiento:** se transforma el *dataset* utilizando `Dataset.load_from_df` y `Reader` para definir el rango de puntuaciones.
 2. **Predicción:** se identifican usuarios activos y se usan 10 películas con valoraciones positivas como base para predicciones. En KNN, se emplea filtrado basado en ítems y similitud de Pearson ajustado. En SVD, se proyectan usuarios e ítems en un espacio latente, capturando patrones complejos incluso con datos dispersos.

3. **Recomendación:** se seleccionan las películas no vistas con mayor puntuación estimada para listas *top@10* o *top@5*.
4. **Conversión de identificadores:** se mapean los `movieId` a `tmdbId` mediante `links.csv` para interoperabilidad.

El sistema demuestra ser robusto frente a problemas de escasez de datos, mejorado, de nuevo, con una etapa adicional de reordenación visual que incorpora criterios visuales relevantes para el usuario.

4.4 Modelos de reordenación de las listas de recomendación

Ambos enfoques incorporan una etapa final de reordenación visual empleando modelos obtenidos con la librería de HuggingFace. A continuación, se detallan los modelos (base y con reordenación) que se han planteado para la evaluación junto a sus abreviaturas, con el fin de simplificar y agilizar la exposición:

- **Modelo+G:** modelo base con reordenamiento mediante el modelo *ViT* de *Google* (*google/vit-base-patch16-224-in21k*). Modelo supervisado basado en Vision Transformer, entrenado sobre ImageNet-21k para generar representaciones visuales generales.
- **Modelo+G-L:** modelo base con reordenamiento mediante el modelo *ViT Large* de *Google* (*google/vit-large-patch16-224-in21k*). Variante de mayor capacidad del ViT original, entrenado también sobre ImageNet-21k, con mejor captación de patrones visuales complejos.
- **Modelo+F:** modelo base con reordenamiento mediante el modelo *ViT* de *Facebook* (*facebook/deit-base-patch16-224*). Modelo eficiente basado en *DeiT*, que incorpora aprendizaje por destilación para reducir la dependencia de datos etiquetados.
- **Modelo+F-L:** modelo base con reordenamiento mediante el modelo *DINOv2 Large* de *Facebook* (*facebook/dinov2-large-imagenet1k-1-layer*). Modelo auto-supervisado basado en distilación sin etiquetas, entrenado para aprender representaciones visuales transferibles.
- **Modelo+M:** modelo base con reordenamiento mediante el modelo *BEiT* de *Microsoft* (*microsoft/beit-base-patch16-224*). Modelo auto-supervisado inspirado en BERT, que predice tokens visuales enmascarados para aprender representaciones semánticas.
- **Modelo+M-L:** modelo base con reordenamiento mediante el modelo *BEiT Large* de *Microsoft* (*microsoft/beit-large-patch16-224-pt22k-ft22k*). Variante de mayor escala con preentrenamiento y ajuste fino sobre ImageNet-22k, adecuada para tareas visuales complejas y contextuales.

5. Resultados

En esta sección se presenta la evaluación de los sistemas de recomendación, considerando los enfoques basados en contenido y filtrado colaborativo junto al reordenamiento visual.

5.1 Enfoque basado en TF-IDF

El modelo basado en *TF-IDF*, siendo uno de los más sencillos, sirve como línea base para evaluar las mejoras del reordenamiento. Las Tablas 1 y 2 junto a la Figura 3 muestran los valores obtenidos para cada métrica en las distintas configuraciones.

Modelo	NDCG@10	MAP@10	MRR@10
TF-IDF	0.215	0.019	0.177
TF-IDF+G	0.285	0.030	0.269
TF-IDF+G-L	0.291	0.030	0.282
TF-IDF+F	0.273	0.028	0.256
TF-IDF+F-L	0.321	0.035	0.317
TF-IDF+M	0.296	0.032	0.283
TF-IDF+M-L	0.320	0.034	0.316

Tabla 1: Resultados comparativos de los modelos *TF-IDF* en métricas top@10.

Modelo	NDCG@5	MAP@5	MRR@5
TF-IDF	0.193	0.018	0.168
TF-IDF+G	0.238	0.026	0.228
TF-IDF+G-L	0.250	0.028	0.244
TF-IDF+F	0.225	0.024	0.211
TF-IDF+F-L	0.254	0.028	0.250
TF-IDF+M	0.254	0.028	0.250
TF-IDF+M-L	0.254	0.028	0.250

Tabla 2: Resultados comparativos de los modelos *TF-IDF* en métricas top@5.

En la Tabla 1 se observa que todas las variantes del modelo base *TF-IDF* presentan mejoras significativas en las métricas NDCG@10, MAP@10 y MRR@10. La variante *TF-IDF+F-L* alcanza el mejor rendimiento global, logrando una mejora del 49.3% en NDCG@10, un incremento del 84.2% en MAP@10 y una mejora del 79.1% en MRR@10 en comparación con el modelo base. Le siguen de cerca las variantes *TF-IDF+M-L* y *TF-IDF+G-L*, cuyas mejoras también son notables, aunque ligeramente inferiores. Estos resultados evidencian el impacto positivo del reordenamiento basado en representaciones visuales empleando los modelos *ViT* sobre la posición de las recomendaciones del *ranking* generado por el modelo base.

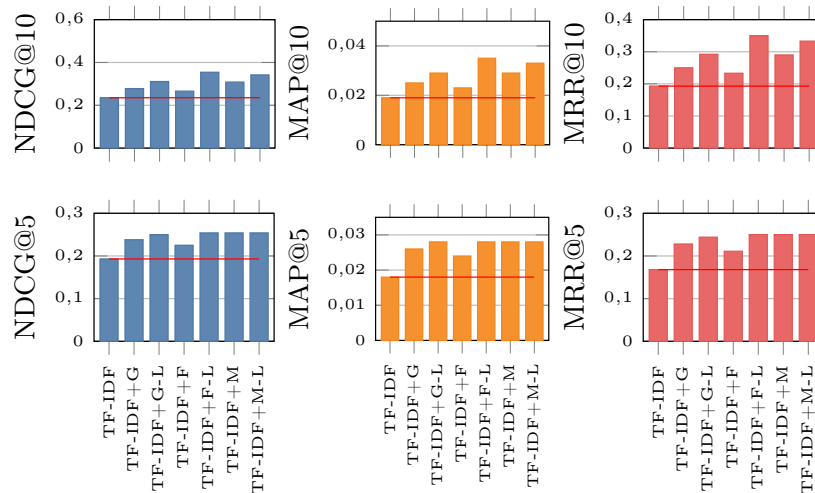


Figura 3: Comparación de métricas @10 (fila superior) y @5 (fila inferior) entre el modelo base ($TF-IDF$) y sus variantes con reordenamiento visual.

En cuanto a los resultados top@5 mostrados en la Tabla 2, se mantiene una tendencia similar, aunque con un comportamiento más homogéneo entre las mejores variantes. Las versiones $TF-IDF+F-L$, $TF-IDF+M$ y $TF-IDF+M-L$ alcanzan valores idénticos en todas las métricas, superando al modelo base con mejoras del 31.6%, 55.5% y 48.8% respectivamente. Este empate sugiere que, en rangos más acotados como el top 5, los distintos modelos de extracción visual aportan beneficios similares en términos de posicionamiento de los ítems en la lista. En conjunto, los resultados confirman que el uso de reordenamiento visual resulta efectivo para mejorar la calidad de las recomendaciones, especialmente en los primeros puestos del *ranking*, donde la relevancia de los ítems es más crítica.

5.2 Enfoque basado en *Sentence Embedding*

El segundo enfoque se basa en la técnica *Sentence Embedding* (SE). Las configuraciones evaluadas incluyen las variantes con reordenamiento visual mencionadas anteriormente. Las Tablas 3 y 4 junto a la Figura 4 muestran los valores obtenidos para cada métrica en las distintas configuraciones.

En la Tabla 3 se aprecia que la incorporación de información semántica al modelo *SE* genera mejoras notables en todas las métricas evaluadas. La variante $SE+G$ presenta el mejor desempeño general, con un incremento del 79.3% en NDCG@10, del 133.3% en MAP@10 y del 142.5% en MRR@10 respecto al modelo base. Estos aumentos son

Modelo	NDCG@10	MAP@10	MRR@10
SE	0.121	0.009	0.087
SE+G	0.217	0.021	0.211
SE+G-L	0.209	0.020	0.200
SE+F	0.181	0.016	0.164
SE+F-L	0.190	0.018	0.178
SE+M	0.190	0.018	0.178
SE+M-L	0.196	0.018	0.183

Tabla 3: Resultados comparativos de los modelos *SE* en métricas top@10.

Modelo	NDCG@5	MAP@5	MRR@5
SE	0.075	0.008	0.067
SE+G	0.100	0.010	0.100
SE+G-L	0.100	0.010	0.100
SE+F	0.081	0.008	0.075
SE+F-L	0.088	0.008	0.083
SE+M	0.100	0.010	0.100
SE+M-L	0.088	0.008	0.083

Tabla 4: Resultados comparativos de los modelos *SE* en métricas top@5.

significativamente mayores en comparación con otras variantes como *SE+F* o *SE+M*, cuyas mejoras, aunque positivas, son más modestas. Además, la versión *SE+G-L* no logra superar a *SE+G*, lo que indica que un mayor tamaño del modelo no garantiza mejoras adicionales en este contexto. En conjunto, estos resultados confirman que las representaciones visuales extraídas mediante modelos como *ViT* aportan un valor significativo al proceso de reordenamiento.

En la Tabla 4, el comportamiento de los modelos es más homogéneo, aunque también se destaca la eficacia de las variantes semánticas. Las versiones *SE+G*, *SE+G-L* y *SE+M* logran empates en todas las métricas, superando al modelo base con mejoras del 33.3% en NDCG@5, del 25.0% en MAP@5 y del 49.2% en MRR@5. Aunque los incrementos absolutos son pequeños debido a las magnitudes iniciales reducidas, el avance relativo es considerable. Estas cifras confirman que los beneficios del reordenamiento visual también se mantienen cuando se restringe la evaluación a las primeras cinco posiciones del *ranking*, lo que resulta especialmente relevante en aplicaciones donde la visibilidad de los ítems es limitada y la precisión en los primeros resultados es crítica.

5.3 Enfoque basado en vecinos más cercanos

El modelo *KNN* constituye la versión base sin información visual, sirviendo como línea base para evaluar los beneficios del reordena-

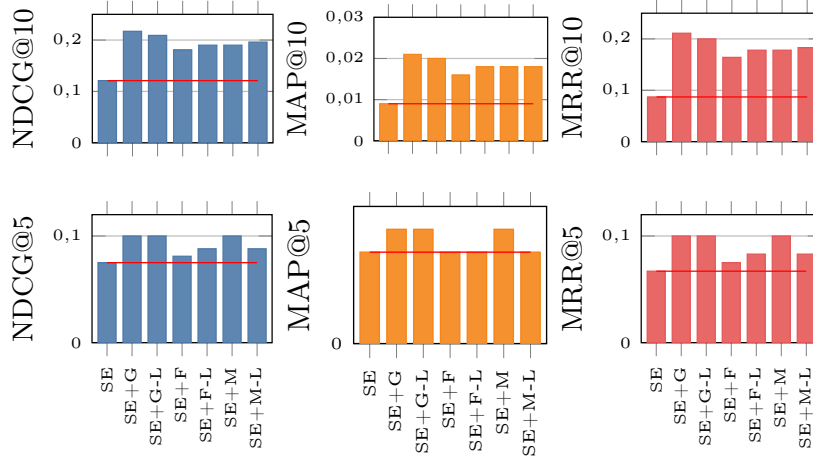


Figura 4: Comparación de métricas @10 (fila superior) y @5 (fila inferior) entre el modelo base (*SE*) y sus variantes con reordenamiento visual.

miento. Las configuraciones evaluadas incluyen las variantes con reordenamiento visual mencionadas anteriormente. La Tablas 5 y 6 junto a la Figura 5 muestran los valores obtenidos para cada métrica en las distintas configuraciones.

Modelo	NDCG@10	MAP@10	MRR@10
KNN	0.322	0.055	0.266
KNN+G	0.439	0.086	0.430
KNN+G-L	0.419	0.083	0.406
KNN+F	0.363	0.067	0.325
KNN+F-L	0.420	0.082	0.410
KNN+M	0.335	0.063	0.305
KNN+M-L	0.435	0.082	0.430

Tabla 5: Resultados comparativos de los modelos *KNN* en métricas top@10.

Los resultados revelan un impacto considerable del reordenamiento visual sobre el modelo *KNN*. La variante *KNN+G* lidera todas las métricas evaluadas, alcanzando un incremento del 36.3% en NDCG@10, un 56.3% en MAP@10 y un 61.6% en MRR@10. Le siguen muy de cerca las combinaciones *KNN+M-L* y *KNN+F-L*, lo cual indica que los modelos base de Google desempeñan mejor que los modelos grandes de Microsoft y Facebook. De esta forma, *KNN+M-L* mejora en

Modelo	NDCG@5	MAP@5	MRR@5
KNN	0.287	0.040	0.251
KNN+G	0.337	0.052	0.328
KNN+G-L	0.338	0.051	0.333
KNN+F	0.308	0.045	0.289
KNN+F-L	0.332	0.051	0.322
KNN+M	0.300	0.044	0.275
KNN+M-L	0.310	0.046	0.301

Tabla 6: Resultados comparativos de los modelos *KNN* en métricas $top@5$.

un 35.0 % la NDCG@10 y en un 61.6 % la MRR@10 respecto al modelo original. En conjunto, estos resultados vuelven a confirmar que las representaciones visuales extraídas mediante modelos como *ViT* aportan un valor significativo al proceso de reordenamiento.

En el caso del $top@5$, cuyos resultados se muestran en la Tabla 6, se observa una competencia más cerrada entre las variantes extendidas del modelo. Las versiones *KNN+G* y *KNN+G-L* destacan ligeramente con un incremento cercano al 17.8 % frente al modelo base. En términos de MRR@5, *KNN+G-L* alcanza el valor más alto con 0.333, un 32.7 % superior al modelo sin extensiones. Estas cifras confirman que el reordenamiento visual sigue prevaleciendo cuando se restringe la evaluación a las primeras cinco posiciones de la lista.

5.4 Enfoque basado en factorización de matrices

El segundo enfoque evaluado se basa en la técnica de *Singular Value Decomposition (SVD)*, que permite capturar relaciones latentes entre usuarios e ítems. Al igual que con *KNN*, se analiza el efecto del reordenamiento visual sobre las recomendaciones generadas por este modelo, utilizando las mismas variantes de *ViT*. Las Tablas 7 y 8 junto a la Figura 6 muestran los valores obtenidos para cada métrica en las distintas configuraciones.

Tal como se observa en los resultados $top@10$, las variantes del modelo *SVD* muestran mejoras sustanciales al incorporar características visuales mediante un esquema de reordenamiento. Las configuraciones *SVD+F-L* y *SVD+M-L* destacan particularmente, alcanzando incrementos cercanos al 40.7 % en NDCG@10 respecto al modelo base. En cuanto a MAP@10, la variante *SVD+M-L* obtiene la mayor ganancia, con una mejora del 52.2 %, mientras que en MRR@10 se eleva hasta un 66.2 % sobre la versión sin reordenamiento. Aunque las extensiones con *ViT* de Google (*SVD+G* y *SVD+G-L*) también presentan resultados positivos, las variantes basadas en modelos de Facebook y Microsoft obtienen los mayores beneficios. Estos resultados evidencian que

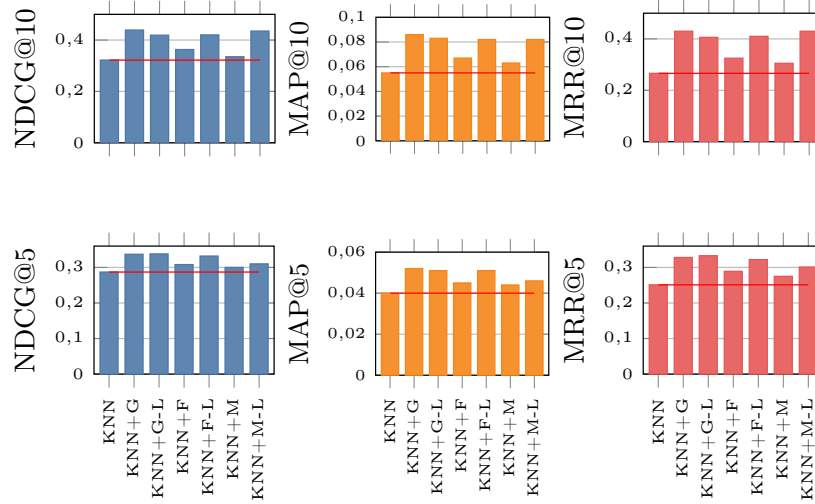


Figura 5: Comparación de métricas $NDCG$, MAP y MRR entre el modelo base (KNN) y sus variantes con reordenamiento visual, evaluadas en @10 (arriba) y @5 (abajo).

el modelo SVD , al capturar patrones latentes a partir de interacciones usuario-ítem, puede beneficiarse significativamente de la incorporación de señales visuales para refinar la ordenación final del *ranking*.

Los resultados para $top@5$ mantienen esta tendencia, aunque con diferencias más matizadas. La configuración $SVD+M-L$ obtiene la mejor puntuación en $NDCG@5$, con una mejora del 25.3% frente al modelo base. En $MAP@5$ y $MRR@5$, esta misma variante supera al modelo original en un 36.3% y 38.4% respectivamente. Aunque las mejoras absolutas son algo más contenidas debido al tamaño reducido del conjunto $top 5$, los resultados siguen mostrando una ventaja clara para los enfoques que integran representaciones visuales avanzadas. En conjunto, los hallazgos refuerzan el valor del reordenamiento visual en sistemas de recomendación, particularmente en escenarios donde la precisión en las primeras posiciones del *ranking* es esencial.

5.5 Comparativa entre casos de estudio basados en contenido y filtrado colaborativo

Este apartado tiene como objetivo comparar los resultados obtenidos en los apartados anteriores tanto en el sistema basado en contenido como en los basados en filtrado colaborativo. Las gráficas mostradas en la Figura 7 muestran una comparación directa de los valores de $NDCG$, MAP y MRR para listas $top@10$ y $top@5$, para cada una de las variantes utilizadas en todos los casos de estudio, permitiendo así

Modelo	NDCG@10	MAP@10	MRR@10
SVD	0.373	0.067	0.308
SVD+G	0.447	0.088	0.413
SVD+G-L	0.455	0.092	0.420
SVD+F	0.468	0.088	0.454
SVD+F-L	0.525	0.100	0.506
SVD+M	0.432	0.082	0.405
SVD+M-L	0.524	0.102	0.512

Tabla 7: Resultados comparativos de los modelos *SVD* en métricas $top@10$.

Modelo	NDCG@5	MAP@5	MRR@5
SVD	0.323	0.055	0.289
SVD+G	0.391	0.071	0.378
SVD+G-L	0.387	0.070	0.378
SVD+F	0.375	0.068	0.369
SVD+F-L	0.402	0.074	0.400
SVD+M	0.360	0.067	0.350
SVD+M-L	0.405	0.075	0.400

Tabla 8: Resultados comparativos de los modelos *SVD* en métricas $top@5$.

una visualización clara del impacto de las representaciones visuales en distintos contextos.

En ambos escenarios, $top@10$ y $top@5$, se observa un patrón consistente: los modelos de filtrado colaborativo, especialmente *SVD*, superan claramente a los enfoques basados en contenido en todas las métricas evaluadas. Asimismo, las extensiones que incorporan información secundaria visual generan mejoras sustanciales en todos los modelos, siendo más pronunciadas en los colaborativos. En concreto se observa como ciertos modelos basados en contenido, como *TF-IDF*, consiguen alcanzar modelos base basados en filtrado colaborativo más complejos, como *KNN*, empleando este reordenamiento visual.

De cara a los resultados $top@10$ permiten realizar una comparación directa entre los distintos enfoques evaluados, tanto de modelos basados en contenido (*TF-IDF* y *SE*) como de filtrado colaborativo (*KNN* y *SVD*). En líneas generales, las variantes del modelo *SVD* destacan como las más efectivas en $top@10$, alcanzando máximos de NDCG@10 y MRR@10 con *SVD+F-L* (0.525 y 0.506, respectivamente) y de MAP@10 con *SVD+M-L* (0.102). Estas cifras superan ampliamente a sus versiones base, con mejoras relativas de hasta 40% en NDCG@10 y más del 66% en MRR@10. En el caso de *KNN*, los resultados también son notables, especialmente con la incorporación de

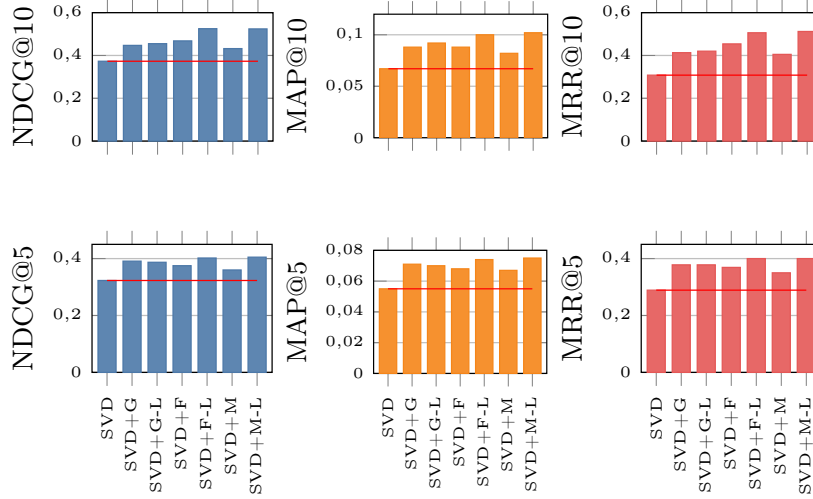


Figura 6: Comparación de métricas $NDCG$, MAP y MRR entre el modelo base (SVD) y sus variantes con reordenamiento visual, evaluadas en @10 (arriba) y @5 (abajo).

características semánticas y morfológicas, logrando valores cercanos a los de SVD y posicionándose como una alternativa competitiva. Por otro lado, los modelos basados en contenido presentan rendimientos más modestos, aunque también se benefician sustancialmente de las extensiones. $TF-IDF+F-L$ y $TF-IDF+M-L$ alcanzan mejoras relevantes, con incrementos del 49.3% en $NDCG@10$ y del 84.2% en $MAP@10$ respecto al modelo original. En el caso de SE , la mejora más significativa ocurre con $SE+G$, que duplica el rendimiento en varias métricas, aunque sin alcanzar los niveles obtenidos por los métodos colaborativos.

En cuanto a los resultados $top@5$ reafirma estas tendencias en un escenario más estricto como el $top@5$. Aquí, las mejores puntuaciones también provienen del modelo SVD , siendo $SVD+M-L$ la combinación más robusta, con valores de $NDCG@5 = 0.405$, $MAP@5 = 0.075$ y $MRR@5 = 0.400$. El modelo KNN muestra un comportamiento igualmente competitivo, especialmente con variantes semánticas como $KNN+G-L$, que alcanza un $MRR@5$ de 0.333, mejorando un 32.6% respecto al modelo base. Las diferencias entre modelos de contenido y colaborativos se mantienen: $TF-IDF+M-L$ y $TF-IDF+F-L$ logran empates técnicos con mejoras superiores al 30% en $NDCG@5$, mientras que SE continúa siendo el modelo más limitado en rendimiento absoluto, a pesar de que sus variantes enriquecidas, como $SE+G$ y $SE+M$, muestran mejoras relativas importantes. En conjunto, esta comparación cruzada permite concluir que los enfoques colaborativos, especialmente aquellos basados en descomposición matricial como SVD , son los más receptivos a la integración de información externa semántica, fonética y

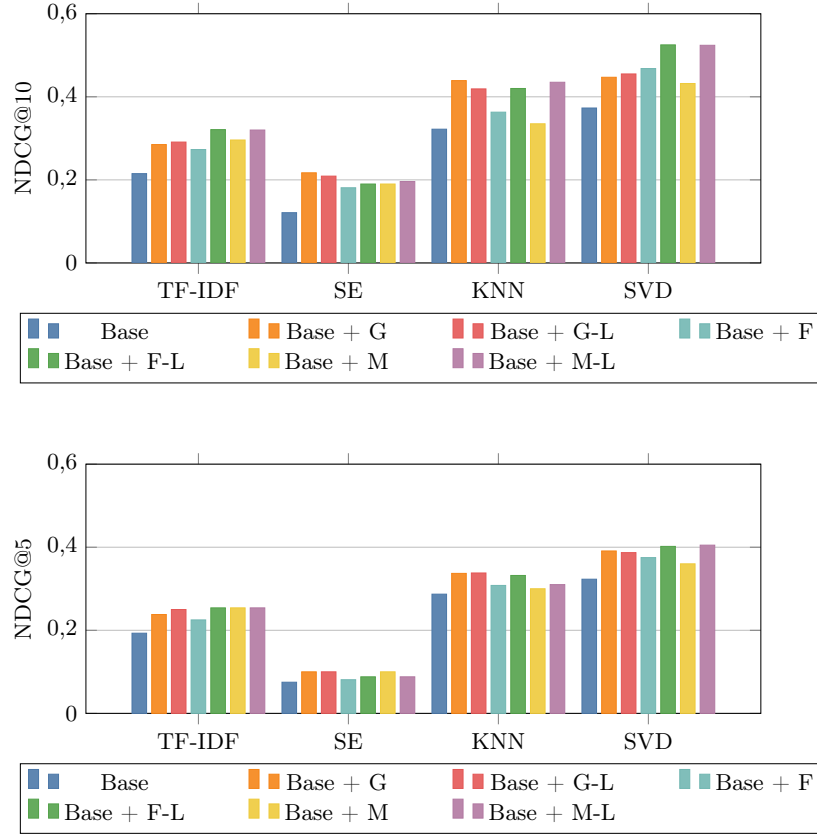


Figura 7: Comparación de $NDCG@10$ (arriba) y $NDCG@5$ (abajo) entre los enfoques *TF-IDF*, *SE*, *KNN* y *SVD* para las variantes base y con *ViT* de Google, Facebook y Microsoft.

morfológica, lo que los convierte en la opción más sólida para sistemas de recomendación que prioricen precisión en los primeros resultados.

Además, se observa que entre los distintos modelos ViT empleados para la extracción de características, los modelos de Google tienden a ofrecer los mejores resultados en combinación con sistemas base como *SE* y *KNN*, mientras que modelos como *DINOv2* y *BEiT-Large* destacan en configuraciones con *TF-IDF* y *SVD*. Esto sugiere que la efectividad del reordenamiento visual depende de la compatibilidad entre el modelo de extracción de características y el sistema de recomendación subyacente. Aun así, la calidad de las representaciones visuales sigue siendo un factor clave, ya que modelos entrenados con grandes conjuntos de datos como *ImageNet-21k* y arquitecturas profundas continúan mostrando un impacto significativo en la mejora del rendimiento en tareas de *ranking*.

6. Conclusiones

En este trabajo se ha propuesto y validado una metodología para sistemas de recomendación que integra información visual no estructurada, mediante imágenes asociadas a los ítems, con enfoques tradicionales de recomendación top- k . Una revisión del estado del arte fundamentó la selección de modelos base y justificó el uso de *Vision Transformers* (ViT) para extraer representaciones semánticas visuales relevantes, destacando su potencial para mejorar la personalización y precisión.

La arquitectura propuesta combina un sistema base de recomendación con un módulo visual que refina el orden de las listas top- k según la similitud semántica de las imágenes, facilitando la adaptación a distintos escenarios y la evaluación independiente del valor visual. La validación se realizó sobre conjuntos de datos cinematográficos reconocidos lo que permitió explorar la influencia de atributos visuales en sistemas tradicionales. Los resultados muestran mejoras significativas en todos los modelos evaluados, destacando que el reordenamiento visual permite a modelos basados en contenido alcanzar o superar el desempeño de modelos colaborativos base.

En resumen, este estudio cumple los objetivos planteados al establecer un marco metodológico para la combinación de sistemas clásicos con técnicas modernas de visión artificial, evidenciando su efectividad para potenciar la calidad de las recomendaciones.

6.1 Limitaciones y líneas de trabajo futuras

Este estudio presenta ciertas limitaciones que abren oportunidades claras para investigaciones futuras. En primer lugar, el uso de modelos visuales basados en *Vision Transformers* implica un elevado coste computacional, dificultando su aplicación en entornos con recursos limitados. Además, la eficacia del sistema depende en gran medida de la calidad y disponibilidad de las imágenes, siendo los pósters a menudo poco representativos o inexistentes. También se ha recurrido a modelos preentrenados sin ajuste fino (*fine-tuning*), lo que puede limitar la capacidad para captar características visuales más complejas.

Otro aspecto a mejorar es la integración multimodal, que actualmente se realiza mediante un reordenamiento posterior sin aprendizaje conjunto, limitando el potencial de la fusión estructurada-visual. Asimismo, la escasez de conjuntos de datos visuales diversos sugiere la necesidad de recopilar fuentes adicionales como imágenes de productos o material promocional, e incluso explorar otras modalidades como audio o texto enriquecido.

Como líneas futuras, se propone optimizar la eficiencia computacional y explorar arquitecturas *end-to-end* que integren la información multimodal de forma conjunta, y entrenar modelos visuales específicos

mediante *fine-tuning*. Además, sería valioso adaptar el sistema a contextos dinámicos y preferencias cambiantes, así como incorporar arquitecturas avanzadas como modelos basados en grafos, aprendizaje profundo o grandes modelos de lenguaje (LLMs), con el objetivo de mejorar la personalización y la eficacia general del sistema de recomendación.

Referencias

1. Z. Sun, Q. Guo, J. Yang, H. Fang, G. Guo, J. Zhang, and R. Burke, "Research commentary on recommendations with side information: A survey and research directions," *Electronic Commerce Research and Applications*, vol. 37, p. 100879, 2019.
2. A. Pujahari and D. S. Sisodia, "Ordinal consistency based matrix factorization model for exploiting side information in collaborative filtering," *Information Sciences*, vol. 643, p. 119258, 2023.
3. V. K. Sejwal and M. Abulaish, "A hybrid recommendation technique using topic embedding for rating prediction and to handle cold-start problem," *Expert Systems with Applications*, vol. 209, p. 118307, 2022.
4. R. Patel and P. Thakkar, "Ibcfaicdr: Auxiliary data-driven item-based collaborative filtering in cross-domain rss to address user cold start problem," *Results in Engineering*, vol. 24, p. 103257, 2024.
5. H. Ma, Z. Qi, X. Dong, X. Li, Y. Zheng, X. Meng, and L. Meng, "Cross-modal content inference and feature enrichment for cold-start recommendation," in *Proceedings of the International Joint Conference on Neural Networks (IJCNN)*, (New York, NY, USA), IEEE, 2023.
6. A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16×16 words: Transformers for image recognition at scale," in *Proceedings of the 9th International Conference on Learning Representations (ICLR) 2021*, (virtual conference), pp. 1–22, May 2021.
7. M. Terzi, M. Rowe, M.-A. Ferrario, and J. Whittle, "Text-based user-knn: Measuring user similarity based on text reviews," in *User Modeling, Adaptation, and Personalization*, (Cham), pp. 195–206, Springer International Publishing, 2014.
8. N. Heidari, P. Moradi, and A. Koochari, "An attention-based deep learning method for solving the cold-start and sparsity issues of recommender systems," *Knowledge-Based Systems*, vol. 256, p. 109835, 2022.
9. X. Xu, K. Dutta, and C. Ge, "Do adjective features from user reviews address sparsity and transparency in recommender systems?," *Electronic Commerce Research and Applications*, vol. 29, 04 2018.
10. Q. Guo, Z. Sun, J. Zhang, Q. Chen, and Y.-L. Theng, "Aspect-aware point-of-interest recommendation with geo-social influence," in *Adjunct Publication of the 25th Conference on User Modeling, Adaptation and Personalization, UMAP '17*, (New York, NY, USA), p. 17–22, Association for Computing Machinery, 2017.
11. J. McAuley and J. Leskovec, "Hidden factors and hidden topics: understanding rating dimensions with review text," in *Proceedings of the 7th ACM Conference on Recommender Systems, RecSys '13*, (New York, NY, USA), p. 165–172, Association for Computing Machinery, 2013.

12. W.-T. Chu and Y.-L. Tsai, "A hybrid recommendation system considering visual information for predicting favorite restaurants," *World Wide Web*, vol. 20, pp. 1313–1331, Nov 2017.
13. M. M. Bendouch, F. Frasincar, and T. Robal, "A visual-semantic approach for building content-based recommender systems," *Information Systems*, vol. 117, p. 102243, 2023.
14. C. Ma, Y. Sun, Z. Yang, H. Huang, D. Zhan, and J. Qu, "Content feature extraction-based hybrid recommendation for mobile application services," *Computers, Materials and Continua*, vol. 71, no. 3, pp. 6201–6217, 2022.
15. A. K. Jaiswal, H. Liu, and I. Frommholz, "Information foraging for enhancing implicit feedback in content-based image recommendation," in *Proceedings of the 11th Annual Meeting of the Forum for Information Retrieval Evaluation, FIRE '19*, (Kolkata, India), pp. 65–69, Association for Computing Machinery, Dec. 2019. Also available as arXiv preprint arXiv:2001.06765 (submitted 19 Jan 2020).
16. F. Deng, P. Ren, Z. Qin, G. Huang, and Z. Qin, "Leveraging image visual features in content-based recommender system," *Scientific Programming*, vol. 2018, pp. 1–8, 08 2018.
17. Y. Jiang, L. Guo, X. Chen, and J. X. Liu, "Making recommender systems more knowledgeable: A framework to incorporate side information," *CoRR*, vol. abs/2406.00615, pp. 1–15, Jun 2024.
18. P. Liang, W. Shi, and X. Zhang, "Remote sensing image classification based on stacked denoising autoencoder," *Remote Sensing*, vol. 10, no. 1, 2018.
19. D. Malitesta, G. Cornacchia, C. Pomo, F. A. Merra, T. Di Noia, and E. Di Sciascio, "Formalizing multimedia recommendation through multimodal deep learning," *ACM Transactions on Recommender Systems*, vol. 3, p. 1–33, Mar. 2025.
20. J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, "Bert: Pre-training of deep bidirectional transformers for language understanding," in *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, (Minneapolis, Minnesota), pp. 4171–4186, Association for Computational Linguistics, 2019.
21. Q. Ma, X. Ren, and C. Huang, "Xrec: Large language models for explainable recommendation," *CoRR*, vol. abs/2406.02377, pp. 1–11, Jun 2024.
22. B. A. Yilma, N. Aghenda, M. Romero, Y. Naudet, and H. Panetto, "Personalised visual art recommendation by learning latent semantic representations," in *2020 15th International Workshop on Semantic and Social Media Adaptation and Personalization (SMA)*, pp. 1–6, 2020.
23. K. Kandoi, R. R. N. S. G. Singh, R. Bediya, K. Mishra, and S. K. Singh, "Comparative analysis of deep learning models for fashion recommendation in e-commerce," in *2023 IEEE Fifth International Conference on Advances in Electronics, Computers and Communications (ICAEECC)*, pp. 1–6, 2023.
24. D. Gao, L. Jin, B. Chen, M. Qiu, P. Li, Y. Wei, Y. Hu, and H. Wang, "Fashionbert: Text and image matching with adaptive loss for cross-modal retrieval," in *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '20*, (New York, NY, USA), p. 2251–2260, Association for Computing Machinery, 2020.

25. R. He and J. McAuley, “Vbpr: Visual bayesian personalized ranking from implicit feedback,” in *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, AAAI’16, pp. 144–150, AAAI Press, 2016.

Machine Learning for Retail Demand Forecasting

Taref Bilel Seifeddine, Álvaro Lozano Murciego, María N. Moreno
García

Universidad de Salamanca
Departamento de Informática y Automática
Facultad de Ciencias.
Plaza de los Caídos s/n
37008 Salamanca, España
{idu065621,mmg,loza}@usal.es

Resumen La previsión de la demanda en el sector minorista es un proceso esencial para optimizar la gestión de inventarios, mejorar la planificación operativa y aumentar la rentabilidad. Este trabajo se centra en el uso de series temporales como herramienta fundamental para identificar patrones históricos, tendencias y estacionalidades que influyen directamente en las ventas. En la era actual, caracterizada por una abundancia de datos, las técnicas de aprendizaje automático se han consolidado como soluciones eficaces para abordar la complejidad del análisis predictivo. Estas técnicas permiten el procesamiento de grandes volúmenes de información y la adaptación a cambios dinámicos en el comportamiento del consumidor y factores externos como promociones o eventos especiales. Existe una distinción clave entre modelos univariantes, que se basan únicamente en los valores históricos de la serie objetivo, y modelos multivariantes, que incorporan variables externas relevantes. Esta incorporación de información contextual mejora significativamente la precisión de las predicciones. El objetivo principal de este trabajo es presentar un enfoque metodológico claro, robusto y replicable que combine técnicas de análisis de series temporales y aprendizaje automático para la predicción de la demanda minorista, aportando una solución flexible que puede aplicarse en diversos entornos comerciales.

Palabras clave: series temporales, demanda minorista, aprendizaje automático, modelos univariantes, modelos multivariantes, predicción

1. Introducción

En un entorno comercial cada vez más competitivo, globalizado y digitalizado, anticiparse a la demanda futura se ha vuelto un factor estratégico clave para las empresas del sector minorista (*retail*). Una

predicción precisa no solo permite optimizar la planificación del inventario y la producción, sino que también reduce riesgos de sobrestock o roturas de stock, mejora la eficiencia logística y contribuye a una mejor experiencia del cliente.

Tradicionalmente, el *forecasting* se ha abordado mediante métodos estadísticos como medias móviles, suavizado exponencial o modelos ARIMA (*Autoregressive Integrated Moving Average*). Estos enfoques han demostrado ser efectivos en contextos estables, con patrones temporales claros. No obstante, el comportamiento del consumidor actual está influido por múltiples factores dinámicos como promociones, eventos, estacionalidad o incluso fenómenos inesperados como pandemias, lo que exige técnicas más flexibles y capaces de capturar relaciones no lineales. En este sentido, el aprendizaje automático (*machine learning*) ha emergido como una solución eficaz para el análisis predictivo, especialmente en problemas complejos como la predicción de demanda en *retail*. Su capacidad para procesar grandes volúmenes de datos, identificar patrones ocultos y adaptarse a entornos cambiantes lo convierte en una herramienta poderosa para el modelado de series temporales, permitiendo capturar tanto tendencias como estacionalidades y fluctuaciones puntuales.

Dependiendo de la información disponible, la predicción de series temporales puede abordarse desde dos enfoques: el univariante, que se basa únicamente en el historial de la variable objetivo (por ejemplo, las ventas semanales), y el multivariante, que incorpora variables adicionales como fechas especiales, indicadores económicos o datos de calendario. El presente trabajo analiza y compara el comportamiento de modelos tradicionales y modernos aplicados a la predicción de la demanda en el sector minorista, empleando dos conjuntos de datos reales. Se utilizan técnicas como AutoARIMA, NaiveDrift y Prophet para el enfoque univariante, y modelos como LightGBM, CatBoost y linear regression con covariables para el enfoque multivariante.

Este trabajo tiene como objetivo principal estudiar y comparar distintos modelos de predicción de la demanda semanal en el sector *retail*, utilizando tanto enfoques clásicos de series temporales como técnicas modernas de aprendizaje automático. Y Para alcanzar este objetivo general, se plantean los siguientes objetivos específicos:

1. Analizar el comportamiento de la demanda en el sector *retail* a partir de datos históricos de ventas semanales.
2. Explorar distintos modelos de predicción basados en series temporales, tanto univariantes como multivariantes, incluyendo enfoques clásicos y de *machine learning*.
3. Diseñar un *pipeline* dinámico de generación de modelos de *forecasting*, adaptable a diferentes conjuntos de datos y escenarios sin necesidad de ajustes manuales.

4. Comparar el rendimiento de los modelos aplicando distintas estrategias de validación temporal, como *rolling window* y *expanding window*, en múltiples horizontes de predicción.
5. Incorporar covariables como mes, año, semana del año o días festivos para evaluar su impacto en la precisión de los modelos multivariantes.
6. Evaluar el desempeño de los modelos mediante diferentes métricas como MAPE, SMAPE, RMSE y MSE, identificando los más precisos y robustos.
7. Generar visualizaciones e informes automáticos que faciliten la interpretación de resultados y la toma de decisiones en entornos reales del comercio minorista.

2. Estado del arte

2.1 Introducción al *Forecasting*

La predicción de series temporales, conocida como forecasting, es una herramienta clave para anticipar eventos futuros a partir de datos históricos. Su objetivo es asistir en la toma de decisiones estratégicas mediante la modelización del comportamiento pasado de una serie temporal, ya sea de forma autorregresiva o integrando variables externas [5]. Esta técnica permite proyectar el valor futuro de una variable con base en valores pasados (lags).

En el contexto minorista, el forecasting resulta esencial para anticipar la demanda, mejorar la eficiencia operativa, aumentar la satisfacción del cliente y maximizar la rentabilidad. La planificación adecuada permite ajustar la producción y evitar tanto el exceso de inventario como el desabastecimiento. Gracias a los avances tecnológicos y a la integración de sistemas ERP (*Enterprise Resource Planning*), se ha logrado una mayor precisión en las previsiones mediante el uso de datos actualizados y conectados en tiempo real [3].

Los métodos de forecasting se clasifican en dos grandes categorías: cualitativos [14] y cuantitativos [9]. Los métodos cualitativos, como la técnica Delphi o la opinión de expertos, se basan en la experiencia y juicio humano. Por su parte, los métodos cuantitativos utilizan datos históricos y técnicas estadísticas como el método ingenuo [3], medias móviles [12], suavizado exponencial [3] y proyecciones de tendencia [9], siendo estos últimos el foco principal del presente trabajo. Como muestra la siguiente figura 1 estos métodos.

Otra dimensión importante es el horizonte de predicción, que puede ser de un solo paso (*single-step*) o de múltiples pasos (*multi-step*), y se implementa mediante estrategias como la recursiva, directa o de salida múltiple. La elección del enfoque depende del contexto y del tipo de

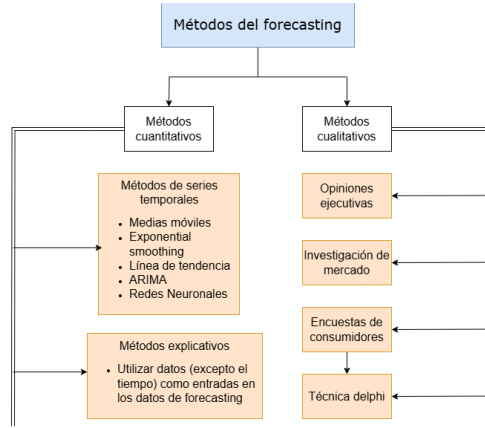


Figura 1: Métodos de *forecasting*.

modelo utilizado [5]

Finalmente, se incorpora el backtesting como técnica de validación retrospectiva para evaluar la precisión del modelo. Este puede implementarse mediante ventanas deslizantes o expansivas, según la naturaleza de los datos y los objetivos del análisis [19]

2.2 Tipos de series temporales

Las series temporales pueden clasificarse según diversas características estructurales que influyen en su análisis y modelado. Una serie temporal univariante consiste en una única variable observada a lo largo del tiempo. En cambio, una serie multivariante incluye múltiples variables registradas simultáneamente, lo que permite modelar interacciones entre variables. Aunque el análisis multivariante puede ofrecer mayor riqueza informativa, también requiere técnicas más complejas que el análisis univariante [1].

Otra distinción importante es la forma en que se registran los datos en el tiempo. Las series discretas contienen observaciones en intervalos regulares (como minutos, días o meses), y son comunes en contextos empresariales y económicos, como ventas diarias o precios bursátiles. Las series continuas, en cambio, registran datos en cualquier instante dentro de un intervalo y son típicas de disciplinas como la ingeniería o la medicina, donde se analizan señales eléctricas, fisiológicas o biomédicas que varían de manera continua, como el electrocardiograma (ECG) o la presión arterial [18].

Por otro lado, una serie estacionaria presenta propiedades estadísticas (media, varianza, autocorrelación) constantes a lo largo del tiempo. Por el contrario, una serie no estacionaria muestra cambios estructurales o tendencias, lo cual puede dificultar su modelado. Para el análisis predictivo, muchas técnicas estadísticas requieren datos estacionarios; por

ello, es común aplicar transformaciones como la diferenciación para estabilizar la serie antes de su modelado [9, 18]

2.3 Modelos de predicción de series temporales

Los modelos de predicción en series temporales son herramientas fundamentales para anticipar la demanda en el sector minorista. Una predicción precisa permite mejorar la organización de las operaciones, gestionar mejor el stock y tomar decisiones estratégicas informadas, lo que ayuda a minimizar costes, evitar escasez o exceso de inventario y responder con agilidad a los cambios del mercado. La literatura recoge enfoques clásicos estadísticos y enfoques modernos basados en técnicas de aprendizaje automático, incluyendo modelos híbridos.

Entre los modelos estadísticos tradicionales destaca el modelo Naive Drift [10], que proyecta una tendencia lineal entre el primer y último valor de la serie de entrenamiento, asumiendo una tasa de cambio constante. Es sencillo, fácilmente interpretable y usado como línea base para comparar modelos más complejos.

Uno de los enfoques más populares es ARIMA, diseñado para series no estacionarias mediante diferenciación. Este modelo combina tres componentes: autorregresivo (AR), integración (I) y media móvil (MA). Por otro lado, existe el modelo AutoARIMA [10], que automatiza la selección del modelo ARIMA(p,d,q) más adecuado según el valor del Criterio de Información de Akaike (AIC), su versión corregida (AICc) o el Criterio de Información Bayesiano (BIC). Este modelo explora configuraciones viables dentro de restricciones de clasificación específicas. Con el auge del aprendizaje automático, han surgido modelos que permiten capturar patrones complejos sin requerir supuestos estrictos de estacionariedad. Algoritmos como la regresión lineal [11], LightGBM [13] o CatBoost [17] han sido aplicados exitosamente a problemas de predicción temporal.

Finalmente, los modelos híbridos como Prophet [7] combinan componentes lineales, estacionales y efectos externos en una descomposición aditiva de la serie temporal.

2.4 Métricas de Evaluación en Predicción de Series Temporales

Analizar la efectividad de los modelos de predicción constituye una etapa fundamental en cualquier proceso de previsión, particularmente en el ámbito del *retail*. Elegir las métricas correctas no solo facilita la comparación objetiva entre los modelos, sino que también apoya la toma de decisiones estratégicas en áreas como la gestión de inventarios, logística o ventas. Para capturar diferentes dimensiones del error en las predicciones, se sugiere utilizar una mezcla de métricas tanto absolutas como relativas.

Las métricas absolutas, como el Error Cuadrático Medio (MSE) [8] y la Raíz del Error Cuadrático Medio (RMSE) [2], castigan de manera significativa los errores grandes al elevarlos al cuadrado, lo que las hace muy sensibles a los valores extremos. Por otro lado, las métricas relativas, tales como el Error Absoluto Porcentual Medio (MAPE) [4] y el Error Absoluto Porcentual Medio Simétrico (SMAPE) [16], presentan el error en forma de porcentaje, lo que simplifica la comparación entre diversas magnitudes de productos o series.

La elección de métricas depende de un equilibrio entre criterios técnicos y de negocio. En contextos empresariales, suelen priorizarse MAPE o SMAPE por su expresividad porcentual y facilidad de interpretación. Por otro lado, RMSE permite analizar la magnitud absoluta del error en las unidades originales, mientras que el MSE lo hace en unidades al cuadrado.

En cuanto a su formulación matemática, el MSE se define como el promedio del cuadrado de los errores entre valores reales y pronosticados. El RMSE corresponde a su raíz cuadrada, expresando el error en las mismas unidades que la variable predicha. El MAPE mide el promedio del error porcentual absoluto, mientras que el SMAPE ajusta este error dividiéndolo por la media entre los valores reales y predichos, lo que lo hace más robusto cuando los valores reales son bajos o cercanos a cero.

3. Metodología

En esta parte se expone el enfoque metodológico utilizado para examinar el caso de estudio, el cual se enfoca en anticipar la demanda en un ambiente de series temporales utilizando métodos de aprendizaje automatizado. La metodología sigue un procedimiento sistemático y que puede reproducirse, apoyado en el ciclo de vida del análisis predictivo, basado en técnicas científicas establecidas en la literatura. La Figura 2 ilustra de manera general las etapas que componen el proceso metodológico propuesto:

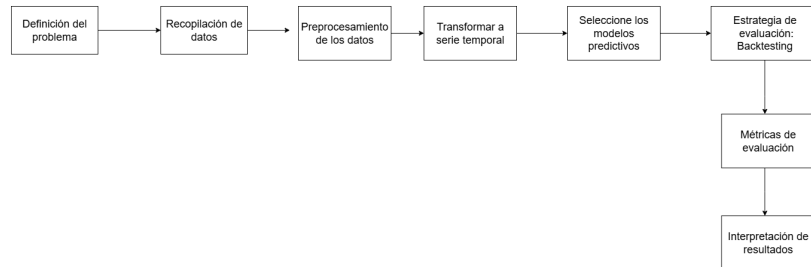


Figura 2: Diagrama general de la metodología seguida en el caso de estudio.

Este diagrama presenta un resumen del proceso metodológico, comenzando con la identificación del problema, seguido de la recolección y el análisis de los datos, la transformación de estos en series temporales, la elección de modelos de predicción, la evaluación utilizando técnicas como el *backtesting*, y la interpretación final de los resultados logrados. A continuación, se explica cada fase de manera detallada:

- La etapa inicial se centra en la identificación del problema, lo que incluye determinar la variable que se desea predecir, la frecuencia de los datos (por ejemplo, diaria o semanal) y el plazo de predicción deseado. Este paso es crucial para establecer el enfoque analítico y situar la solución dentro del contexto del comercio minorista.
- A continuación, se llevan a cabo la recolección y el preprocesamiento de los datos, que pueden tener su origen en una variedad de fuentes, tales como bases de datos públicas, dispositivos de sensores o archivos históricos. En situaciones que involucran múltiples variables, es esencial transformar los datos para que se conviertan en series univariantes que sean prácticas para el propósito comercial. Además, se ejecutan procesos de limpieza de datos para descartar registros que no son consistentes y se emplean métodos de imputación de datos para abordar los valores que faltan. En metodologías híbridas, se combinan variables externas como calendarios, días festivos o indicadores económicos.
- Los datos limpios se transforman en objetos de series temporales. Esta modificación garantiza una indexación temporal adecuada y una frecuencia constante, lo que hace más fácil el entrenamiento de modelos de forecasting.
- En la cuarta etapa, se eligen los modelos de predicción más apropiados, teniendo en cuenta el tipo de serie, el horizonte de predicción y la cantidad de datos disponibles. Se analizan tanto los modelos estadísticos como los de aprendizaje automático, evaluando su eficacia, capacidad de interpretación y costo computacional.
- Una vez seleccionados los modelos, se aplica una estrategia de evaluación basada en *backtesting*, que consiste en dividir la serie tem-

poral en ventanas de entrenamiento y prueba para simular predicciones en condiciones reales. Se utilizan tanto ventanas deslizantes como expansivas para comparar distintos escenarios.

- Posteriormente, se implementan métricas de evaluación numéricas tales como MSE, RMSE, MAPE y SMAPE, las cuales facilitan la comparación del desempeño de los modelos desde diferentes perspectivas (error absoluto, error porcentual, sensibilidad ante valores atípicos, entre otros).
- Finalmente, se lleva a cabo una evaluación de los resultados obtenidos a partir de las métricas y el análisis del error, lo que facilita la identificación de los modelos más efectivos según el tipo de serie, así como la determinación de si la adición de variables externas contribuye a una mejor precisión en la predicción.

4. Caso de estudio

Este caso de estudio busca generar conclusiones que respondan las siguientes cuestiones de investigación.(RQ):

- **RQ1:** ¿Qué modelos ofrecen el mejor rendimiento en cada conjunto de datos (Walmart y Suplementos)?
- **RQ2:** ¿La inclusión de variables externas mejora significativamente la precisión de las predicciones?
- **RQ3:** ¿Qué diferencias se observan entre los resultados obtenidos con modelos univariantes y multivariantes?
- **RQ4:** ¿El rendimiento de los modelos varía según el horizonte de predicción (4, 8 o 12 semanas)?
- **RQ5:** ¿Qué estrategia de validación (*expanding* vs *rolling window*) ofrece una evaluación más estable y realista?

4.1 Conjuntos de datos

Se emplearon dos conjuntos de datos representativos del ámbito minorista [?], ambos obtenidos desde la plataforma Kaggle. El primero contiene ventas semanales de 45 tiendas entre 2010 y 2012, incluyendo variables como identificador de tienda, fecha de la observación, ventas semanales (variable objetivo), indicador de semana con festivo, así como variables económicas y ambientales como temperatura, precio del combustible, IPC y desempleo. La frecuencia es semanal y la estructura temporal es constante.

El segundo conjunto de datos recopila ventas semanales por producto desde 2020 hasta 2025 [15], incorporando variables como fecha, nombre del producto, categoría, unidades vendidas, precio, ingresos (variable objetivo), descuentos, devoluciones, ubicación y plataforma. Su riqueza en variables explicativas lo hace idóneo para enfoques multivariantes

4.2 Preprocesamiento de los datos

Para el conjunto de Walmart, se aplicó una agregación temporal sumando las ventas semanales de todas las tiendas, generando una única serie temporal semanal representativa del comportamiento global. Se reestructuró el índice temporal y se adecuaron los formatos de fecha para su uso con bibliotecas como Darts.

El conjunto de suplementos fue tratado de forma similar, agregando las unidades vendidas por semana para generar una serie única representativa.

4.3 Enriquecimiento con variables externas

Ambos conjuntos fueron enriquecidos con variables derivadas del índice de fecha, como número de semana del año, mes y año. Además, se incorporó una variable binaria (*is_holiday*) generada con `workalendar` [6] para identificar semanas festivas. Esto permitió estructurar los datos como series multivariantes con contexto adicional.

4.4 Transformación a objetos de series temporales

Los datos se convirtieron a objetos tipo `TimeSeries` de Darts [20], garantizando compatibilidad con modelos y flujos de trabajo posteriores. Este formato permite aplicar procesos de normalización, segmentación temporal, visualización y entrenamiento de modelos de manera estandarizada.

4.5 Modelos predictivos

Se aplicaron tanto modelos univariantes como multivariantes para abordar el problema de predicción.

En el enfoque univariante, se entrenaron los modelos Naive Drift, AutoARIMA y Prophet, utilizando únicamente la variable objetivo (ventas o ingresos), sin añadir variables adicionales. El modelo Naive Drift sirve como línea base, asumiendo que la serie sigue una tendencia lineal simple. AutoARIMA selecciona automáticamente los parámetros del modelo ARIMA más adecuado, optimizando criterios como AIC o BIC. Por su parte, Prophet es un modelo aditivo desarrollado por Meta, que maneja bien estacionalidades múltiples y anomalías, y está diseñado para ser interpretable y robusto en entornos empresariales.

En el enfoque multivariante, se incorporaron covariables contextuales y se aplicaron modelos como regresión lineal, LightGBM y CatBoost, entrenando con todas las variables enriquecidas disponibles. La regresión lineal ofrece una interpretación sencilla de las relaciones entre variables. LightGBM es un modelo basado en árboles de decisión optimizado para eficiencia y velocidad, mientras que CatBoost está especialmente diseñado para manejar variables categóricas de forma efectiva, siendo robusto ante sobreajuste y eficaz incluso con datos desequilibrados.

4.6 Estrategia de evaluación: *Backtesting*

Para garantizar una evaluación rigurosa y realista del rendimiento de los modelos predictivos, se implementaron dos estrategias de validación temporal adaptadas a la naturaleza secuencial de las series temporales: la **ventana expandible** (*expanding window*) [5] y la **ventana deslizante** (*rolling window*) [5]. A diferencia de la validación cruzada convencional, estas metodologías respetan la estructura temporal de los datos, evitando el uso de información futura durante el entrenamiento.

Se definieron tres horizontes de predicción: **4, 8 y 12 semanas**, lo que permitió analizar la evolución del rendimiento en distintos escenarios de planificación. En ambos enfoques, se utilizó un intervalo de actualización (*stride*) de 1 semana, generando predicciones de manera continua en el tiempo.

La ventana expandible simula un entorno de aprendizaje continuo, donde el modelo se reentrena añadiendo progresivamente nuevas observaciones al conjunto de entrenamiento. Este enfoque resulta útil para evaluar cómo mejora el rendimiento del modelo a medida que dispone de más datos históricos. Se parte de una ventana inicial, se realiza la predicción para uno de los tres horizontes definidos, y se incorporan las nuevas observaciones al conjunto de entrenamiento, repitiendo el proceso con un stride de una semana. Esta estrategia permite observar la evolución del modelo con datos crecientes, reflejar entornos de aprendizaje incremental y ofrecer una evaluación longitudinal útil para despliegues en producción.

Por otro lado, en la ventana deslizante, el tamaño del conjunto de entrenamiento se mantiene constante, desplazándose a lo largo del tiempo. En cada iteración, se utiliza una ventana fija con los datos más recientes para realizar la predicción del siguiente horizonte. Esta técnica permite evaluar la capacidad de los modelos para adaptarse a dinámicas recientes del mercado, siendo especialmente útil en contextos donde los patrones de comportamiento cambian con frecuencia.

4.7 Selección de métricas de evaluación

Para evaluar el rendimiento de los modelos se emplearon métricas estándar ampliamente utilizadas en series temporales: RMSE, MSE, MAPE y SMAPE. La elección de estas métricas responde a la necesidad de captar tanto errores absolutos como relativos, y garantizar un análisis equilibrado entre precisión, claridad y robustez.

En particular, RMSE y MSE se emplearon por su capacidad para penalizar errores grandes, lo cual es crucial en contextos donde las desviaciones significativas pueden implicar costos logísticos o pérdidas económicas relevantes. Estas métricas son especialmente útiles para identificar modelos que fallan de forma drástica en ciertas semanas o productos, permitiendo una detección temprana de comportamientos anómalos.

Por otro lado, las métricas porcentuales como MAPE y SMAPE fueron seleccionadas por su facilidad de interpretación y su utilidad para la comparación relativa entre diferentes productos o períodos. Estas métricas son especialmente valoradas en contextos empresariales, ya que permiten comunicar los resultados a perfiles no técnicos y facilitan la evaluación del impacto económico de los errores de predicción.

5. Resultados

5.1 Resultados del Conjunto de Datos de Walmart (Modelos Univariantes)

En primer lugar, se evaluaron modelos univariantes sobre el conjunto de datos de Walmart [?], considerando únicamente la serie temporal de ventas semanales sin incluir variables externas. Se entrenaron tres modelos representativos: Naive Drift [20], AutoARIMA [10] y Prophet [7]. La evaluación se realizó utilizando dos estrategias de validación temporal: ventana deslizante (*rolling window*) y ventana expansiva (*expanding window*), con horizontes de predicción de 4, 8 y 12 semanas.

5.1.1 Ventana deslizante La siguiente tabla resume los valores obtenidos para cada modelo y horizonte en función de las métricas utilizadas 1.

Modelo	Horizonte	MAPE (%)	MSE	RMSE	SMAPE (%)
AutoARIMA	4	8.59	5.93×10^{13}	4.73×10^6	8.12
Prophet	4	9.74	6.53×10^{13}	5.21×10^6	9.29
Naive Drift	4	9.65	6.78×10^{13}	5.17×10^6	9.18
AutoARIMA	8	9.13	6.61×10^{13}	5.40×10^6	8.74
Prophet	8	11.48	8.99×10^{13}	6.42×10^6	11.16
Naive Drift	8	11.91	1.01×10^{14}	6.63×10^6	11.50
AutoARIMA	12	9.28	6.86×10^{13}	5.81×10^6	8.93
Prophet	12	12.85	1.15×10^{14}	7.46×10^6	12.74
Naive Drift	12	13.62	1.36×10^{14}	7.84×10^6	13.46

Tabla 1: Resultados de los modelos univariantes – (ventana deslizante)

En esta estrategia, se entrenaron tres modelos univariantes: Naive Drift, AutoARIMA y Prophet, sobre la serie temporal de ventas semanales del dataset de Walmart, con horizontes de predicción de 4, 8 y 12 semanas.

Los resultados evidenciaron que AutoARIMA fue el modelo más preciso y consistente en todos los horizontes de predicción. Para un horizonte corto (4 semanas), obtuvo los mejores valores en todas las métricas, con

un MAPE de 8.59 % y un SMAPE de 8.12 %, superando claramente a Prophet (SMAPE: 9.29 %) y a Naive Drift (SMAPE: 9.84 %).

A medida que se amplió el horizonte de predicción, AutoARIMA mantuvo su liderazgo: con un MAPE de 9.28 % y un SMAPE de 8.93 % a 12 semanas, lo que indica una gran capacidad de generalización. En cambio, Prophet y Naive Drift mostraron un deterioro notable. Naive Drift alcanzó un MAPE de 13.62 % y un SMAPE de 13.46 %, reflejando su poca adaptabilidad.

5.1.2 Ventana expansiva La siguiente tabla resume los valores obtenidos para cada modelo y horizonte en función de las métricas utilizadas

Modelo	Horizonte	MAPE (%)	MSE	RMSE	SMAPE (%)
AutoARIMA	4	7.08	4.44×10^{13}	4.34×10^6	7.13
Prophet	4	6.85	4.14×10^{13}	4.27×10^6	7.04
Naive Drift	4	8.74	6.08×10^{13}	4.84×10^6	8.32
AutoARIMA	8	7.01	4.58×10^{13}	4.78×10^6	7.15
Prophet	8	7.00	4.48×10^{13}	4.81×10^6	7.25
Naive Drift	8	10.37	7.95×10^{13}	6.04×10^6	9.95
AutoARIMA	12	6.88	4.57×10^{13}	5.08×10^6	7.07
Prophet	12	7.01	4.61×10^{13}	5.19×10^6	7.29
Naive Drift	12	10.89	8.94×10^{13}	6.70×10^6	10.55

Tabla 2: Resultados de los modelos univariantes (ventana expansiva) – Dataset Walmart

Bajo esta estrategia, Prophet obtuvo un mejor comportamiento en horizontes cortos, especialmente a 4 semanas, con un MAPE de 6.85 % y un SMAPE de 7.04 %, superando a AutoARIMA. No obstante, en horizontes más amplios (8 y 12 semanas), AutoARIMA recuperó la ventaja, con un rendimiento muy competitivo: MAPE de 6.88 % y SMAPE de 7.06 % en $h=12$.

Naive Drift mostró un rendimiento inestable en todos los horizontes. Su error aumentó hasta un MAPE de 10.89 % y un SMAPE de 10.55 %, lo que sugiere una limitación para capturar patrones más complejos en el largo plazo.

En resumen, AutoARIMA demostró ser el modelo más robusto en general, con Prophet como opción válida para predicciones de corto plazo.

5.2 Resultados del Conjunto de Datos de Walmart (Modelos multivariantes)

En segundo lugar, se examinaron modelos multivariantes en el conjunto de datos de Walmart, incluyendo no solo la serie de ventas

semanales en el tiempo, sino también variables externas pertinentes para potenciar el modelo predictivo. Se capacitaron tres modelos emblemáticos: Linear Regression (modelo *baseline*) [11], CatBoost [17] y LightGBM [13], ambos conocidos por su gran habilidad para gestionar datos tabulares con diversas variables explicativas. La evaluación se realizó utilizando las mismas estrategias de validación temporal que se usaron en el caso de los modelos univariantes: ventana deslizante y ventana expansiva y con los mismos horizontes de predicción de 4, 8 y 12 semanas.

5.2.1 Ventana deslizante La siguiente tabla resume los valores obtenidos 3.

Modelo	Horizonte	MAPE (%)	MSE	RMSE	SMAPE (%)
Linear Regression (baseline)	4	35.61	2.45×10^{15}	1.96×10^7	22.43
LightGBM	4	7.17	3.95×10^{13}	4.37×10^6	7.31
CatBoost	4	7.99	4.22×10^{13}	4.49×10^6	7.68
Linear Regression (baseline)	8	107.09	2.44×10^{17}	6.86×10^7	28.95
LightGBM	8	7.18	4.09×10^{13}	4.80×10^6	7.36
CatBoost	8	8.87	5.30×10^{13}	5.33×10^6	8.62
Linear Regression (baseline)	12	904.65	7.97×10^{19}	8.72×10^8	34.23
LightGBM	12	7.19	4.14×10^{13}	5.11×10^6	7.40
CatBoost	12	8.50	5.23×10^{13}	5.59×10^6	8.43

Tabla 3: Resultados de los modelos multivariantes con ventana deslizante – Dataset Walmart

En este enfoque se incorporaron variables contextuales (mes, semana del año, festivos, etc.), y se evaluaron tres modelos multivariantes: Regresión Lineal (baseline), LightGBM y CatBoost.

Los resultados fueron claros: LightGBM y CatBoost superaron ampliamente a la regresión lineal. Para el horizonte de 4 semanas, LightGBM fue el modelo más preciso, con un MAPE de 7.17% y un SMAPE de 7.31%, seguido por CatBoost con 7.99% y 7.68% respectivamente. En contraste, la regresión lineal tuvo un MAPE de 35.61%, demostrando una baja capacidad para capturar relaciones complejas.

Este patrón se mantuvo en los horizontes de 8 y 12 semanas. LightGBM siguió mostrando resultados sólidos, con errores que apenas variaron, lo cual indica alta estabilidad y fiabilidad. CatBoost también ofreció buena precisión, especialmente en $h=8$ (MAPE de 7.18%). La regresión lineal, en cambio, se degradó hasta un MAPE extremo de 904.65% en $h=12$, lo que representa una ineficacia total para forecasting a largo plazo.

5.2.2 Ventana expansiva La siguiente Tabla 4 resume los valores obtenidos

Modelo	Horizonte	MAPE (%)	MSE	RMSE	SMAPE (%)
Linear Regression (baseline)	4	10.99	8.57×10^{13}	6.06×10^6	10.51
LightGBM	4	7.77	3.96×10^{13}	4.54×10^6	7.80
CatBoost	4	5.49	2.67×10^{13}	3.10×10^6	5.32
Linear Regression (baseline)	8	14.24	1.64×10^{14}	8.26×10^6	13.74
LightGBM	8	7.87	4.08×10^{13}	4.90×10^6	7.90
CatBoost	8	5.92	3.27×10^{13}	3.60×10^6	5.82
Linear Regression (baseline)	12	16.96	3.48×10^{14}	1.04×10^7	15.63
LightGBM	12	7.96	4.15×10^{13}	5.23×10^6	7.99
CatBoost	12	5.72	3.20×10^{13}	3.80×10^6	5.72

Tabla 4: Resultados de los modelos **multivariantes** con ventana expansiva – *Dataset Walmart*

Con esta estrategia, CatBoost fue el modelo con mejor rendimiento global, especialmente en Walmart. Para un horizonte de 4 semanas, logró un MAPE de 5.49% y un SMAPE de 5.32%, consolidándose como la alternativa más precisa.

LightGBM también mostró un comportamiento consistente, aunque ligeramente inferior: MAPE de 7.77% y SMAPE de 7.80%. En horizontes más largos, mantuvo una variación mínima en los errores, con un MAPE de 7.19% en $h=12$, confirmando su estabilidad.

Por otro lado, la regresión lineal volvió a destacar negativamente. Sus errores siguieron siendo elevados, alcanzando un MAPE de 16.96% en $h=12$. Esto evidenció su incapacidad para adaptarse a escenarios multivariantes complejos.

5.3 Resultados del Conjunto de Datos de Suplementos (Modelos univariantes)

Esta sección presenta los resultados obtenidos al aplicar diferentes modelos de predicción al conjunto de datos de ventas de suplementos [15]. Se han considerado tanto modelos univariantes como multivariantes, evaluados en tres horizontes de predicción (4, 8 y 12 semanas), y utilizando dos estrategias de validación temporal: ventana deslizante (*rolling window*) y ventana expansiva (*expanding window*). El objetivo es comparar el rendimiento de los modelos bajo distintos esquemas temporales y niveles de complejidad en las predicciones.

5.3.1 Ventana deslizante La siguiente tabla resume los valores obtenidos para cada modelo y horizonte en función de las métricas utilizadas

Modelo	Horizonte	MAPE (%)	MSE	RMSE	SMAPE (%)
AutoARIMA	4	9.29	9.21×10^7	8.96×10^3	9.13
Prophet	4	10.13	1.10×10^8	9.77×10^3	9.99
NaiveDrift	4	11.31	1.38×10^8	1.06×10^4	11.16
AutoARIMA	8	9.41	9.47×10^7	9.28×10^3	9.25
Prophet	8	11.04	1.34×10^8	1.08×10^4	10.88
NaiveDrift	8	12.59	1.77×10^8	1.22×10^4	12.42
AutoARIMA	12	9.47	9.56×10^7	9.34×10^3	9.32
Prophet	12	11.99	1.63×10^8	1.17×10^4	11.83
NaiveDrift	12	13.98	2.28×10^8	1.34×10^4	13.82

Tabla 5: Resultados de modelos univariantes con ventana deslizante – Dataset Suplementos

En el enfoque de ventana deslizante, AutoARIMA se posicionó como el modelo más preciso y estable en los tres horizontes de predicción (4, 8 y 12 semanas). En el horizonte de 4 semanas, obtuvo un MAPE de 9.29 % y un SMAPE de 9.13 %, superando tanto a Prophet como a Naive Drift. Prophet presentó un rendimiento aceptable (MAPE 10.13 %, SMAPE 9.99 %), mientras que Naive Drift mostró el peor desempeño con un MAPE de 11.31 % y un SMAPE de 11.16 %.

A medida que se incrementa el horizonte a 8 semanas, AutoARIMA mantuvo su liderazgo (MAPE 9.41 %, SMAPE 9.25 %), mientras que Prophet y Naive Drift sufrieron un deterioro más pronunciado, alcanzando MAPEs de 10.88 % y 12.59 %, respectivamente.

En el horizonte de 12 semanas, todos los modelos vieron incrementos en el error, pero AutoARIMA mantuvo la ventaja (MAPE 9.47 %, SMAPE 9.32 %). Prophet mostró un descenso en estabilidad (MAPE 11.99 %, SMAPE 11.84 %), mientras que Naive Drift nuevamente fue el menos fiable (MAPE 13.98 %, SMAPE 13.82 %).

5.3.2 Ventana expansiva La siguiente tabla resume los valores obtenidos para cada modelo y horizonte en función de las métricas utilizadas

Modelo	Horizonte	MAPE (%)	MSE	RMSE	SMAPE (%)
AutoARIMA	4	8.67	8.20×10^7	8.37×10^3	8.61
Prophet	4	9.12	8.76×10^7	8.76×10^3	8.93
NaiveDrift	4	11.56	1.43×10^8	1.09×10^4	11.35
AutoARIMA	8	8.55	8.01×10^7	8.53×10^3	8.50
Prophet	8	9.09	8.79×10^7	9.06×10^3	8.91
NaiveDrift	8	11.79	1.50×10^8	1.13×10^4	11.56
AutoARIMA	12	8.54	7.98×10^7	8.62×10^3	8.49
Prophet	12	9.12	8.91×10^7	9.20×10^3	8.95
NaiveDrift	12	12.06	1.59×10^8	1.17×10^4	11.79

Tabla 6: Resultados de modelos univariantes con ventana expansiva – Dataset Suplementos

Con ventana expansiva, los tres modelos mejoraron ligeramente su rendimiento, siendo AutoARIMA nuevamente el más consistente. En el horizonte de 4 semanas, AutoARIMA logró un MAPE de 8.67 % y un SMAPE de 8.61 %, seguido por Prophet (MAPE 9.12 %, SMAPE 8.93 %) y Naive Drift (MAPE 11.56 %, SMAPE 11.35 %).

A 8 semanas, AutoARIMA mantuvo su liderazgo (MAPE 8.55 %, SMAPE 8.50 %) y Prophet presentó buen rendimiento (MAPE 9.09 %, SMAPE 8.91 %). Naive Drift continuó siendo el menos preciso (MAPE 11.79 %, SMAPE 11.56 %).

Finalmente, en el horizonte de 12 semanas, AutoARIMA mostró gran solidez (MAPE 8.54 %, SMAPE 8.49 %), mientras que Prophet decayó ligeramente (MAPE 9.12 %, SMAPE 8.95 %). Naive Drift registró los errores más altos (MAPE 12.06 %, SMAPE 11.79 %).

5.4 Resultados del Conjunto de Datos de Suplementos (Modelos multivariantes)

En esta sección se presentan los resultados obtenidos al aplicar modelos multivariantes (Linear Regression (modelo baseline) [11], CatBoost [17] y LightGBM [13]) al dataset de ventas de suplementos. A diferencia del enfoque univariante, el enfoque multivariante permite considerar simultáneamente múltiples factores que pueden influir en la demanda, lo que potencialmente mejora la precisión de las predicciones.

5.4.1 Ventana deslizante

La tabla 7 resume los valores obtenidos

Modelo	Horizonte	MAPE (%)	MSE	RMSE	SMAPE (%)
LinearRegression (baseline)	4	20.68	1.48×10^9	2.07×10^4	19.86
LightGBM	4	8.97	8.44×10^7	8.63×10^3	8.86
CatBoost	4	9.91	1.07×10^8	9.71×10^3	9.80
LinearRegression (baseline)	8	25.06	1.95×10^9	2.61×10^4	25.05
LightGBM	8	8.98	8.50×10^7	8.89×10^3	8.88
CatBoost	8	10.01	1.10×10^8	1.01×10^4	9.92
LinearRegression (baseline)	12	29.27	2.37×10^9	3.00×10^4	28.88
LightGBM	12	8.97	8.53×10^7	8.98×10^3	8.87
CatBoost	12	10.13	1.12×10^8	1.03×10^4	10.04

Tabla 7: Resultados de modelos multivariantes con ventana deslizante – Dataset Suplementos

En este enfoque, los modelos basados en árboles de decisión superaron ampliamente a la regresión lineal. En el horizonte de 4 semanas, CatBoost ofreció el mejor rendimiento (MAPE 9.91 %, SMAPE 9.80 %), seguido por LightGBM (MAPE 8.97 %, SMAPE 8.86 %). La regresión lineal presentó el peor resultado (MAPE 20.68 %, SMAPE 19.86 %).

A 8 semanas, LightGBM se posicionó como el más estable (MAPE 8.98 %, SMAPE 8.88 %), con CatBoost ligeramente por detrás (MAPE 10.01 %, SMAPE 9.96 %). La regresión lineal continuó mostrando debilidad con un MAPE de 25.06 % y SMAPE de 25.05 %.

En el horizonte de 12 semanas, CatBoost destacó por su precisión (MAPE 10.13 %, SMAPE 10.04 %), ligeramente superior a LightGBM (MAPE 8.97 %, SMAPE 8.87 %). La regresión lineal alcanzó su peor desempeño (MAPE 29.27 %, SMAPE 28.88 %).

5.4.2 Ventana expansiva La siguiente tabla 8 resume los valores obtenidos

Modelo	Horizonte	MAPE (%)	MSE	RMSE	SMAPE (%)
Regresión Lineal (baseline)	4	11.81	3.05×10^8	1.13×10^4	11.58
LightGBM	4	9.11	9.09×10^7	8.63×10^3	8.98
CatBoost	4	9.92	1.04×10^8	9.61×10^3	9.74
Regresión Lineal (baseline)	8	12.36	3.44×10^8	1.20×10^4	11.60
LightGBM	8	9.27	9.40×10^7	9.35×10^3	9.18
CatBoost	8	9.90	1.05×10^8	9.95×10^3	9.75
Regresión Lineal (baseline)	12	12.60	3.65×10^8	1.23×10^4	11.64
LightGBM	12	9.32	9.49×10^7	9.47×10^3	9.23
CatBoost	12	9.90	1.06×10^8	1.01×10^4	9.77

Tabla 8: Resultados de modelos multivariantes con ventana expansiva – Dataset Suplementos

Con esta estrategia, los modelos mejoraron su rendimiento general. En el horizonte de 4 semanas, LightGBM fue el más preciso (MA-

PE 9.11 %, SMAPE 8.98 %), apenas por encima de CatBoost (MAPE 9.92 %, SMAPE 9.74 %). La regresión lineal nuevamente mostró bajo rendimiento (MAPE 11.81 %, SMAPE 11.58 %).

En 8 semanas, LightGBM volvió a liderar (MAPE 9.27 %, SMAPE 9.18 %), seguido por CatBoost (MAPE 9.90 %, SMAPE 9.75 %), con la regresión lineal rezagada (MAPE 12.36 %, SMAPE 11.66 %).

A 12 semanas, LightGBM se mantuvo estable (MAPE 9.32 %, SMAPE 9.23 %) y superó ligeramente a CatBoost (MAPE 9.90 %, SMAPE 9.77 %). La regresión lineal volvió a quedar atrás con MAPE 12.60 % y SMAPE 11.64 %.

5.5 Comparación entre Conjuntos de Datos

Esta parte ofrece un análisis detallado del desempeño de los modelos utilizados para los dos grupos de datos abordados: Walmart y Suplementos [15]. El propósito es analizar el comportamiento de los modelos univariantes y multivariantes bajo diferentes estrategias de validación (ventana deslizante y ventana expansiva) basándose en dos métricas fundamentales: MAPE (Error Promedio Absoluto Medio) [4] y SMAPE (Error Promedio Absoluto Symmetrico) [16]. Comparación de modelos univariantes

5.5.1 Comparación de modelos univariantes (Ventana deslizante)

La tabla 9 muestra los resultados de ambos datasets

Modelo	Horizonte	MAPE Walmart	MAPE Suplementos	SMAPE Walmart	SMAPE Suplementos
AutoARIMA	4	8.59 %	9.29 %	8.12 %	9.13 %
AutoARIMA	8	9.13 %	9.41 %	8.74 %	9.25 %
AutoARIMA	12	9.28 %	9.47 %	8.94 %	9.32 %
NaiveDrift	4	9.65 %	11.31 %	9.18 %	11.16 %
NaiveDrift	8	11.91 %	12.59 %	11.50 %	12.42 %
NaiveDrift	12	13.62 %	13.98 %	13.46 %	13.82 %
Prophet	4	9.74 %	10.13 %	9.29 %	9.99 %
Prophet	8	11.48 %	11.04 %	11.16 %	10.88 %
Prophet	12	12.85 %	11.99 %	12.74 %	11.83 %

Tabla 9: Comparación de MAPE y SMAPE para modelos univariantes con ventana deslizante – Walmart vs Suplementos

El análisis de la Tabla 9 revela que el modelo AutoARIMA destaca como el más preciso en ambos conjuntos de datos y para todos los horizontes de predicción (4, 8 y 12 semanas), con valores de MAPE y SMAPE consistentemente por debajo del 10 %. En contraste, los modelos Naive Drift y Prophet presentan errores considerablemente más altos, especialmente en el caso de Walmart.

- Para Walmart, el modelo Naive Drift muestra un incremento del MAPE desde 9.65 % (h=4) hasta 13.62 % (h=12).

- Prophet alcanza su peor desempeño con un SMAPE de 12.74% en $h=12$.
- El conjunto de Suplementos, sin embargo, muestra errores más estables entre modelos y horizontes, lo que sugiere una menor sensibilidad a predicciones de largo plazo.

En Conclusión el AutoARIMA se posiciona como el modelo univariante más confiable y robusto, especialmente para horizontes largos. Naive Drift y Prophet son menos adecuados, sobre todo para Walmart.

5.5.2 Comparación de modelos univariantes(Ventana expansiva)

La tabla 10 muestra los resultados de ambos datasets

Modelo	Horizonte	MAPE Walmart	MAPE Suplementos	SMAPE Walmart	SMAPE Suplementos
AutoARIMA	4	7.08 %	8.67 %	7.13 %	8.61 %
AutoARIMA	8	7.01 %	8.55 %	7.15 %	8.50 %
AutoARIMA	12	6.88 %	8.54 %	7.07 %	8.49 %
NaiveDrift	4	8.74 %	11.56 %	8.32 %	11.35 %
NaiveDrift	8	10.37 %	11.79 %	9.95 %	11.56 %
NaiveDrift	12	10.89 %	12.06 %	10.55 %	11.79 %
Prophet	4	6.85 %	9.12 %	7.04 %	8.93 %
Prophet	8	7.00 %	9.09 %	7.25 %	8.91 %
Prophet	12	7.01 %	9.12 %	7.29 %	8.95 %

Tabla 10: Comparación de MAPE y SMAPE para modelos univariantes con ventana expansiva – Walmart vs Suplementos

Los resultados de la Tabla 10 indican un rendimiento muy competitivo entre AutoARIMA y Prophet, con diferencias mínimas. AutoARIMA mantiene su superioridad global en estabilidad y precisión.

- En Walmart, AutoARIMA logra el mejor MAPE (6.85%) y SMAPE (7.04%) para $h=12$.
- Prophet mejora considerablemente respecto a la ventana deslizante y logra superar a AutoARIMA en algunos horizontes cortos.
- Naive Drift sigue siendo el modelo menos fiable, con errores superiores al 10% en la mayoría de los casos

En Conclusión Aunque que el Prophet mejora notablemente, AutoARIMA sigue siendo el modelo más robusto y confiable bajo la estrategia de ventana expansiva, especialmente en horizontes más largos.

5.5.3 Comparación de modelos Multivariantes(Ventana deslizante)

La siguiente tabla 11 muestra los resultados de ambos datasets

Modelo	Horizonte	MAPE Suplementos	MAPE Walmart	SMAPE Suplementos	SMAPE Walmart
CatBoost	4	9.91 %	7.99 %	9.80 %	7.68 %
CatBoost	8	10.01 %	8.87 %	9.92 %	8.62 %
CatBoost	12	10.13 %	8.50 %	10.04 %	8.43 %
LightGBM	4	8.97 %	7.17 %	8.86 %	7.31 %
LightGBM	8	8.98 %	7.18 %	8.88 %	7.36 %
LightGBM	12	8.97 %	7.19 %	8.87 %	7.40 %
LinearRegression (baseline)	4	20.68 %	35.61 %	19.86 %	22.43 %
LinearRegression (baseline)	8	25.61 %	107.10 %	25.05 %	28.95 %
LinearRegression (baseline)	12	29.27 %	904.65 %	28.87 %	34.23 %

Tabla 11: Comparación de MAPE y SMAPE entre modelos multivariantes con ventana deslizante – Datasets Walmart y Suplementos

En la Tabla 11 se observa que los modelos basados en árboles de decisión (LightGBM y CatBoost) superan ampliamente a la regresión lineal, que actúa como modelo base.

- LightGBM logra los mejores resultados en el horizonte corto ($h=4$), con un MAPE de 7.17% y un SMAPE de 7.31% en Walmart, y 8.97% / 8.86% en Suplementos.
- CatBoost también muestra buen rendimiento, pero sus errores aumentan ligeramente con el horizonte.
- La regresión lineal, en cambio, exhibe un desempeño muy deficiente, con errores extremadamente altos, sobre todo en el dataset de Walmart (MAPE de 904.65% en $h=12$).

En Conclusión, El LightGBM es el modelo más recomendable en este esquema, seguido por CatBoost. La regresión lineal queda descartada por su baja capacidad predictiva

5.5.4 Comparación de modelos Multivariantes(Ventana expansiva) La siguiente tabla 12 muestra los resultados de ambos datasets

Modelo	Horizonte	MAPE Walmart	MAPE Suplementos	SMAPE Walmart	SMAPE Suplementos
CatBoost	4	5.49 %	9.92 %	5.32 %	9.74 %
CatBoost	8	5.92 %	9.90 %	5.82 %	9.75 %
CatBoost	12	5.72 %	9.90 %	5.72 %	9.77 %
LightGBM	4	7.77 %	9.11 %	7.80 %	8.98 %
LightGBM	8	7.87 %	9.27 %	7.90 %	9.18 %
LightGBM	12	7.96 %	9.32 %	7.99 %	9.23 %
LinearRegression (baseline)	4	10.99 %	11.81 %	10.51 %	11.58 %
LinearRegression (baseline)	8	14.24 %	12.36 %	13.74 %	11.60 %
LinearRegression (baseline)	12	16.96 %	12.60 %	15.63 %	11.64 %

Tabla 12: Comparación de MAPE y SMAPE para modelos multivariantes con ventana expansiva – Walmart vs Suplementos

La Tabla 12 confirma la supremacía de CatBoost en el dataset de Walmart, con MAPE y SMAPE por debajo del 6% en todos los horizontes.

- En Walmart, CatBoost logra 5.49% MAPE y 5.32% SMAPE para $h=4$.

- En el conjunto de Suplementos, LightGBM se comporta mejor que CatBoost, logrando su mejor MAPE en $h=4$ (9.11 %) y SMAPE (8.98 %).
- Nuevamente, la regresión lineal presenta los errores más elevados, especialmente en Walmart (hasta 16.96 % MAPE en $h=12$).

6. Conclusiones

Este estudio examinó el desempeño de modelos de predicción tanto univariantes como multivariantes, aplicándolos a dos conjuntos de datos reales (Walmart y Suplementos). Se evaluó su exactitud a través de las métricas MAPE, SMAPE, RMSE y MSE, empleando diversas estrategias de validación (ventana deslizante y ventana expansiva) así como diferentes horizontes de predicción (4, 8 y 12 semanas).

Los resultados obtenidos indican que los modelos multivariantes, en general, superaron a los univariantes en precisión, especialmente al integrar variables externas. Modelos tales como CatBoost y LightGBM lograron resultados destacados, evidenciando menores tasas de error porcentuales y una mayor consistencia a través de diversos horizontes de predicción. En particular, la técnica de ventana expansiva permitió una evaluación más precisa del comportamiento futuro de los modelos, sobre todo en situaciones donde las series muestran tendencias o patrones a largo plazo.

Además, se observó que el horizonte de predicción influye directamente en la precisión, siendo los errores más pronunciados en horizontes más largos. No obstante, ciertos modelos multivariantes como CatBoost y LightGBM mostraron mayor robustez a lo largo de los tres horizontes, manteniendo un rendimiento estable. Por otra parte, el rendimiento también varió según el conjunto de datos, lo que indica que no existe un único modelo dominante aplicable a cualquier contexto, sino que la elección del enfoque debe adaptarse a las características específicas del problema y los datos disponibles.

En resumen, los resultados respaldan el uso de modelos multivariantes enriquecidos con información contextual y validados mediante ventanas expansivas para mejorar la precisión de predicción en series temporales del ámbito comercial.

7. Trabajos futuros

Esta sección propone diversas líneas de investigación que podrían ser exploradas en futuros trabajos con el fin de mejorar la precisión y robustez de los modelos de predicción de series temporales aplicados en entornos reales:

1. **Nuevas variables externas:** Se sugiere incorporar nuevas fuentes de datos externas (por ejemplo, indicadores económicos, clima,

campañas promocionales o redes sociales) que puedan influir en la demanda, con el objetivo de enriquecer los modelos multivariantes y mejorar su precisión.

2. **Modelos de deep learning especializados:** Se plantea la posibilidad de emplear arquitecturas avanzadas como LSTM, GRU, Temporal Fusion Transformer o N-BEATS, capaces de capturar patrones temporales complejos y no lineales que los modelos tradicionales no detectan con facilidad.
3. **Optimización de hiperparámetros:** La aplicación de técnicas como *Grid Search*, *Random Search* o algoritmos evolutivos podría permitir un ajuste más fino de los modelos, mejorando así su rendimiento, especialmente en modelos como *CatBoost* y *LightGBM*.
4. **Evaluación en entornos reales de producción:** Se destaca la relevancia de aplicar los modelos en escenarios reales de forecasting, donde se puedan medir no solo las métricas de error, sino también su impacto en decisiones empresariales, eficiencia operativa y reducción de costes.
5. **Métodos de combinación de modelos (ensembles):** La integración de modelos univariantes y multivariantes mediante enfoques tipo *ensemble* podría mejorar la robustez general de las predicciones, aprovechando las fortalezas de ambos enfoques.
6. **Análisis de periodicidades y anomalías:** La incorporación de métodos específicos para detectar cambios estructurales, estacionalidades múltiples o eventos anómalos puede aumentar la adaptabilidad de los modelos ante situaciones imprevistas.

Referencias

1. Beeram, S. R. and Kuchibhotla, S. Time series analysis on univariate and multivariate variables: A comprehensive survey. En: *Communication Software and Networks: Proceedings of INDIA 2019*. Springer, pp. 119–126, 2020.
2. Chai, T. and Draxler, R. R. Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature. *Geoscientific Model Development*, Vol. 7, No. 3, pp. 1247–1250, 2014.
3. Chopra, S. and Meindl, P. *Supply Chain Management: Strategy, Planning, and Operation*. 6th ed. Upper Saddle River, New Jersey: Pearson Education, 2016.
4. PerMetrics Developers. Mean Absolute Percentage Error (MAPE) — PerMetrics Documentation. 2024. Accedido: 2025-06-16. URL: <https://permetrics.readthedocs.io/en/latest/pages/regression/MAPE.html>.
5. Skforecast Developers. Introduction to Forecasting - Skforecast Documentation. 2024. Accessed: July 7, 2025. URL: <https://skforecast.org/0.16.0/introduction-forecasting/introduction-forecasting#backtesting-with-refit-and-fixed-training-size-rolling-origin>.
6. Workalendar Developers. Workalendar: World calendars and holidays for Python. 2024. Accedido: 2025-06-03. URL: <https://github.com/workalendar/workalendar>.
7. Facebook Core Data Science Team. Prophet: Forecasting at scale. 2024. Accedido: 2025-05-28. URL: <https://facebook.github.io/prophet/>.
8. Hodson, T. O., Over, T. M. and Foks, S. S. Mean Squared Error, Deconstructed. *Journal of Advances in Modeling Earth Systems*, Vol. 13, No. 12, p. e2021MS002681, 2021. DOI: <https://doi.org/10.1029/2021MS002681>. URL: <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2021MS002681>.
9. Hyndman, R. J. and Athanasopoulos, G. *Forecasting: Principles and Practice*. OTexts, 2021. URL: <https://otexts.com/fpp3/>.
10. Hyndman, R. J. auto.arima function. *forecast package, version 8.24.0*. 2024. Accedido: 28 mayo 2025. URL: <https://www.rdocumentation.org/packages/forecast/versions/8.24.0/topics/auto.arima>.
11. James, G., Witten, D., Hastie, T. and Tibshirani, R. *An Introduction to Statistical Learning*. Vol. 112. Springer, 2013.
12. Karuppan, C. M., Dunlap, N. E., Waldrum, M. R. and Ferniany, W. Forecasting Demand. En: *Operations Management in Healthcare*. New York: Springer Publishing Company, pp. 57–104, 2021. DOI: 10.1891/9780826147721.0003. URL: <https://connect.springerpub.com/content/book/978-0-8261-4772-1/part/part01/chapter/ch03>.
13. Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q. and Liu, T. LightGBM: A highly efficient gradient boosting decision tree. En: *Advances in Neural Information Processing Systems*, Vol. 30, 2017.
14. Makridakis, S., Wheelwright, S. C. and Hyndman, R. J. *Forecasting: Methods and Applications*. 3rd ed. New York: John Wiley & Sons, 1998.
15. Mughal, Z. Supplement Sales Data. 2025. Accedido: 2025-06-10. URL: <https://www.kaggle.com/datasets/zahidmughal2343/supplement-sales-data>.

16. PerMetrics Documentation. Symmetric Mean Absolute Percentage Error (SMAPE). 2023. Accedido: 2025-05-31. URL: <https://permetrics.readthedocs.io/en/latest/pages/regression/SMAPE.html>.
17. Prokhorenkova, L., Gusev, G., Vorobev, A., Dorogush, A. V. and Gulin, A. CatBoost: Unbiased boosting with categorical features. En: *Advances in Neural Information Processing Systems*, Vol. 31, 2018.
18. R. Rosca, E. Stationary and non-stationary time series. *The USV Annals of Economics and Public Administration*, Vol. 10, No. 1, pp. 177–186, 2011.
19. Rodrigo, J. A. Backtesting in time series forecasting with skforecast. 2024. Accedido: 2025-07-07. URL: https://joaquinamatrodrigo.github.io/skforecast/0.12.0/user_guides/backtesting.
20. Unit8 SA. NaiveDrift - Darts Documentation. 2024. https://unit8co.github.io/darts/generated_api/darts.models.forecasting.baselines.html#darts.models.forecasting.baselines.NaiveDrift. Accessed: 2025-05-28.

Ajuste fino de modelos de lenguaje de gran tamaño para la detección de visualizaciones de datos engañosas basadas en elementos primitivos

Diego Marcelo Miranda Rodriguez, Roberto Therón Sánchez, and
Andrea Vázquez Ingelmo

Universidad de Salamanca
Departamento de Informática y Automática
Facultad de Ciencias.
Plaza de los Caídos s/n
37008 Salamanca, España
{diegomarcelomr, theron, andreavazquez}@usal.es

Resumen En la era actual de sobreinformación visual, las visualizaciones de datos se han consolidado como una herramienta clave para la comprensión y comunicación de fenómenos complejos. Sin embargo, su uso inapropiado o intencionadamente sesgado puede inducir a interpretaciones erróneas, afectando la percepción pública, la toma de decisiones y la integridad del conocimiento. Este Trabajo de Fin de Máster aborda dicha problemática desde una perspectiva innovadora, explorando el potencial de los modelos de lenguaje de gran tamaño (LLMs), específicamente aquellos con capacidades multimodales, para detectar visualizaciones de datos engañosas. El enfoque se centra en los elementos primitivos que componen los gráficos, integrando técnicas de ajuste fino y estrategias avanzadas de ingeniería de *prompts*. Los resultados demuestran que, con el ajuste adecuado, estos modelos pueden no solo identificar visualizaciones manipuladas, sino también argumentar sus hallazgos de forma explicativa y transparente. Se concluye que los LLMs ajustados representan una herramienta prometedora para fortalecer la alfabetización visual y combatir la desinformación gráfica en distintos ámbitos, como la educación, el periodismo y la investigación científica.

Keywords: Fine-Tuning, LLMs, Multimodal Model, Misleading, Prompts

1. Introducción

Hoy en día nos encontramos en una era digital muy avanzada, por ello la visualización de datos es una herramienta fundamental

para comunicar información compleja de forma clara, concisa y efectiva. Sin embargo, cuando en esta se emplea con poca transparencia o con intención manipuladora, puede inducir a conclusiones erróneas. Las visualizaciones engañosas, que incluyen prácticas como escalas truncadas, ejes dobles o distorsión de proporciones, representan una amenaza creciente a la interpretación veraz de la información. Detectarlas de forma automática y confiable es, por tanto, un desafío al que hoy en día investigadores, informáticos y la misma población se encuentra. Tradicionalmente, el análisis de visualizaciones engañosas se ha centrado en enfoques basados en reglas, que aplican reglas explícitas como, por ejemplo: "si el eje no empieza en cero, marcar como posible engaño". Estos sistemas muestran resultados efectivos en escenarios concretos, pero presentan limitaciones significativas: son rígidos, requieren mantenimiento constante, y no se adaptan fácilmente a nuevos tipos de engaño o a estilos gráficos inesperados. En contraste, LLMs (Large Language Models) han evolucionado hacia sistemas multimodales que integran capacidades de procesamiento de texto e imágenes. Investigaciones recientes demuestran que estos LLMs multimodales pueden interpretar gráficos complejos, reconocer patrones engañosos y razonar críticamente sobre ellos. Este avance abre una vía innovadora y más robusta para la detección automática de visualizaciones engañosas. Una contribución destacada en este campo es el estudio de Lo y Qu, titulado "How Good (Or Bad) Are LLMs at Detecting Misleading Visualizations?", presentado en IEEE VIS 2024. Usando modelos multimodales y un conjunto de visualizaciones con fallos predefinidos (escalas truncadas, gráficas 3D, ejes dobles, etc.), realizaron experimentos con nueve tipos de *prompts* diseñados para instruir a los modelos en la detección. Sus resultados evidenciaron que estos modelos no solo son capaces de comprender el contenido gráfico, sino también de argumentar críticamente sobre su veracidad. El trabajo amplía el análisis desde cinco tipos de engaños iniciales hasta 21, ofreciendo una sólida base metodológica para el uso de LLMs en el ámbito de la visualización de datos [1].

Mientras tanto, Guocong Li, presentan un enfoque complementario con su artículo "From Misleading Queries to Accurate Answers: A Three-Stage Fine-Tuning Method for LLMs", aceptado en ACL 2025. Su propuesta no aborda directamente visualizaciones, pero introduce un esquema de ajuste fino de tres etapas que resulta altamente adaptable al análisis multimodal. En la primera etapa, el modelo aprende a detectar contenido engañoso; en la segunda, a corregir dicho contenido; y finalmente, a responder con precisión a partir de la información limpia. Esta metodología, probada en tareas de Question Answering (QA) y mejora significativamente la precisión del modelo frente a entradas inapropiadas o malintencionadas [2].

Por otra parte, la investigación FlowLLM: Large Language Model Driven Flow Visualization, publicada en Visual Informatics en mayo de 2025 [3], explora cómo ajustar un LLM con datos de visualización de

flujo para generar análisis visuales más ricos. Aunque enfocada en un tipo específico de visualización (flujos), ofrece una visión valiosa sobre la preparación de datos gráficos y técnicas de entrenamiento para modelos adaptados a la interpretación visual.

El presente Trabajo de Fin de Máster propone un enfoque aplicado para abordar la necesidad del ajuste fino (*fine-tuning*) de modelos de lenguaje de gran tamaño con el objetivo de dotarlos de la capacidad de detectar visualizaciones de datos engañosas, haciendo hincapié en los elementos primitivos que conforman los gráficos. Entendemos por elementos primitivos las componentes básicas de una visualización según la gramática de gráficos: marcas geométricas (barras, líneas, puntos, sectores, etc.) y canales visuales (posición, longitud, área, color, etc.) [4]. Muchas visualizaciones engañosas alteran estos elementos fundamentales como, por ejemplo, cuando usan áreas desproporcionadas para representar valores, manipulando colores para sesgar la interpretación, invirtiendo ejes, omitiendo referencias o alterando escalas. Al centrar la detección en estos componentes básicos, se busca identificar patrones de engaño reconocibles sin dejar de lado el contexto de los datos [5].

1.1 Motivación

Las visualizaciones de datos engañosas, como ejes truncados, escalas inconsistentes o gráficos 3D distorsionados, han incrementado la propagación de información errónea en medios públicos, investigaciones y entornos educativos. Los Modelos de Lenguaje de Gran Tamaño multimodales han demostrado capacidades superiores para “razonar” sobre gráficos y detectar si una visualización es engañosa, siendo capaces de identificar errores gráficos y explicarlos de forma argumentada. Sin embargo, estos modelos suelen verse vulnerables ante representación visual distorsionada, lo que impulsa la necesidad de un ajuste fino dirigido a su robustez frente a engaños visuales.

1.2 Objetivos

1.2.1 Objetivo General Investigar el potencial de los modelos de lenguaje de gran tamaño (LLMs) para detectar visualizaciones de datos engañosas mediante el análisis de sus elementos primitivos y desarrollar un marco de trabajo para la detección automática.

1.2.2 Objetivos Específicos

- Analizar los elementos primitivos de visualización: Investigar cómo los ejes, las etiquetas, los colores y los tipos de gráficos contribuyen a la creación de visualizaciones engañosas.
- Creación y etiquetado de un conjunto de datos: Construir un conjunto de datos que contenga visualizaciones tanto engañosas como

precisas, categorizándolas según problemas comunes como ejes truncados, escalas engañosas, uso inapropiado de colores y omisiones de datos.

- Ajuste fino de LLMs: Implementar un proceso de ajuste fino para LLMs multimodales con el fin de mejorar su capacidad de reconocer visualizaciones engañosas, enfocándose en los elementos primitivos de los gráficos.
- Evaluación de los métodos de detección: Evaluar el desempeño de los LLMs ajustados en la detección de visualizaciones engañosas utilizando *benchmarks* establecidos y nuevos conjuntos de prueba.

2. Revisión Sistemática de Literatura

La revisión sistemática de la literatura, específicamente la de Kitchenham [6] y Kitchenham y Charters [7], es una metodología de investigación que permite identificar, evaluar y sintetizar de manera rigurosa y transparente los estudios más relevantes sobre una temática específica. A diferencia de una revisión narrativa, este enfoque sigue un protocolo estructurado que garantiza la reproducibilidad del proceso y la objetividad en la selección de fuentes. En el presente Trabajo de Fin de Máster, se ha realizado una revisión sistemática con el objetivo de comprender el estado actual de la investigación en torno al uso de modelos de lenguaje de gran tamaño (LLMs), especialmente los multimodales, en la detección de visualizaciones de datos engañosas.

2.1 Preguntas de investigación

- **RQ1:** ¿Qué modelos de lenguaje de gran tamaño (LLMs) se han utilizado para analizar o interpretar las visualizaciones de los datos?
- **RQ2:** ¿Qué técnicas de ajuste fino se han empleado para mejorar la capacidad de los LLMs en la detección de visualizaciones engañosas?
- **RQ3:** ¿Qué conjuntos de datos se han usado para entrenar o evaluar LLMs en tareas relacionadas con visualizaciones engañosas?
- **RQ4:** ¿Qué estrategias de *prompts* se han desarrollado para mejorar la detección de engaños visuales en LLMs?
- **RQ5:** ¿Cómo se ha evaluado el rendimiento de los LLMs en la tarea de detección de visualizaciones engañosas? ¿Qué tan eficaces son?

2.2 Marco PICOC

- **P (Población o problemas de interés):** Investigadores y profesionales que aborden temas como la interpretación, comprensión o detección de visualizaciones de datos, especialmente si involucran gráficos engañosos.
- **I (Intervención):** Aplicación de modelos de lenguaje (LLMs), particularmente multimodales, para detectar y razonar sobre visualizaciones.
- **C (Comparación):** Comparación entre LLMs sin ajuste fino y modelos ajustados específicamente; comparación entre diferentes estrategias de *prompting* o *benchmarks*.
- **O (Resultados esperados):** Rendimiento de los LLMs (precisión, robustez, capacidad de detección de errores, capacidad de razonamiento visual-crítico).
- **C (Contexto):** Artículos publicados entre el año 2015 y 2025.

2.3 Criterios de inclusión y exclusión

Una vez definidas las preguntas de investigación y los objetivos se deben definir los criterios de Inclusión e exclusión. El objetivo que buscan los criterios mencionados serán descartar los documentos encontrados durante la fase de búsqueda que carezcan de ciertas condiciones mínimas establecidas. A continuación, se evidencian los criterios de inclusión:

- **IC1:** Se consideran artículos escritos en inglés o español.
- **IC2:** Artículos científicos revisadas por pares.
- **IC3:** Estudios que involucren LLMs o modelos de visión multimodal aplicados a gráficos o visualizaciones.
- **IC4:** Estudios que reporten resultados cuantitativos o cualitativos sobre la detección o análisis de gráficos engañosos.
- **IC5:** Los estudios deben estar publicados entre 2015 y 2025.

Se procede a evidenciar los criterios de exclusión:

- **EC1:** Estudios centrados únicamente en visualización general sin análisis de veracidad o engaño.
- **EC2:** Estudios donde los LLMs no sean el objeto de intervención o comparación.
- **EC3:** Artículos con referencias incompletas o mal documentadas.
- **EC4:** Artículos, reseñas, editoriales sin validación empírica.
- **EC5:** Artículos que sólo aborden generación de gráficos sin evaluación crítica.

2.4 Criterios de calidad

Los criterios de calidad nos permiten evaluar la rigurosidad metodológica de los estudios seleccionados para el mapeo. Dixon-Woods *et al.* [8] proponen 5 criterios en forma de una lista de preguntas que por su amplitud pueden ser de uso general para aplicar criterios de calidad

- **CC1:** ¿Los propósitos y objetivos de la investigación están claramente establecidos?.
- **CC2:** ¿El diseño de la investigación está claramente especificado y es apropiado para los fines y objetivos de la investigación?.
- **CC3:** ¿Se proporciona una explicación clara del proceso por el cual se obtuvieron los hallazgos?.
- **CC4:** ¿Se muestran suficientes datos para respaldar las interpretaciones y conclusiones?.
- **CC5:** ¿El método de análisis es apropiado y está adecuadamente explicado?.

2.5 Cadena de búsqueda

Para este estudio, fue necesario crear una cadena de búsqueda específica para identificar y seleccionar los estudios relevantes. Dicha cadena fue la siguiente:

```
("large language model"OR "LLM"OR "multimodal model") AND ("misleading visualization"OR "deceptive chart"OR "data visualization analysis"OR "visual literacy") AND ("fine-tuning"OR "prompt engineering"OR "robustness"OR "interpretation")
```

La búsqueda se llevó a cabo entre mayo y junio de 2025, en distintas bases de datos académicas reconocidas por su cobertura en el ámbito de la inteligencia artificial y la visualización de datos: Web of Science (WoS), Scopus, arXiv y IEEE Xplore.

2.6 Discusión de la revisión sistemática de literatura

Los resultados de la revisión evidencian que el uso de modelos de lenguaje multimodal ha crecido significativamente en el análisis crítico de visualizaciones. Los modelos han demostrado capacidades prometedoras para interpretar datos visuales, detectar errores de diseño y razonar sobre afirmaciones derivadas de gráficos. Sin embargo, la mayoría de los modelos aún son vulnerables a técnicas de engaño como ejes truncados, escalas inconsistentes o efectos visuales manipuladores. Se ha observado que el rendimiento mejora considerablemente con estrategias como el ajuste fino especializado en gráficos (*fine-tuning*), el uso de datos sintéticos con anotaciones detalladas, y la ingeniería de *prompts* conversacionales o críticos.

2.7 Conclusión de la revisión sistemática

La revisión sistemática ha permitido constatar que los LLMs multimodales constituyen una herramienta con gran potencial para la detección de visualizaciones engañosas. Aunque aún existen limitaciones técnicas, las investigaciones actuales ofrecen bases sólidas para desarrollar sistemas asistidos por Inteligencia Artificial (IA) que promuevan la alfabetización visual y combatan la desinformación gráfica. Los avances en ajuste fino, entrenamiento con datos visuales reales y diseño de *prompts* efectivos, sugieren que un enfoque híbrido, visión computacional y validación numérica, será clave en el futuro de esta línea de investigación.

3. Metodología

3.1 Definición de conceptos, herramientas y librerías usadas

A grandes rasgos, las técnicas de visualización engañosa pueden agruparse en dos categorías, la de engaños visuales directos y engaños contextuales [1]. En la primera categoría se ubican las alteraciones de los elementos gráficos primitivos, que hacen que la imagen por sí misma induzca a error. Entre ellos se cuentan:

- Truncamiento o expansión inapropiada de ejes:** Consiste en no iniciar el eje Y en cero en gráficos de barras (truncamiento), o elegir rangos no equivalentes en ejes duales, generando comparaciones visuales distorsionadas. También incluye invertir ejes o escalar los ejes de forma que acentúen o minimicen tendencias. Estas prácticas no cumplen guías básicas de diseño; sin embargo, un estudio reciente reveló que solo 11 % de los gráficos engañosos en medios sociales empleaban efectivamente estas violaciones obvias a las guías, lo que sugiere que, aunque son notorias, representan la punta del iceberg en la desinformación visual [9].

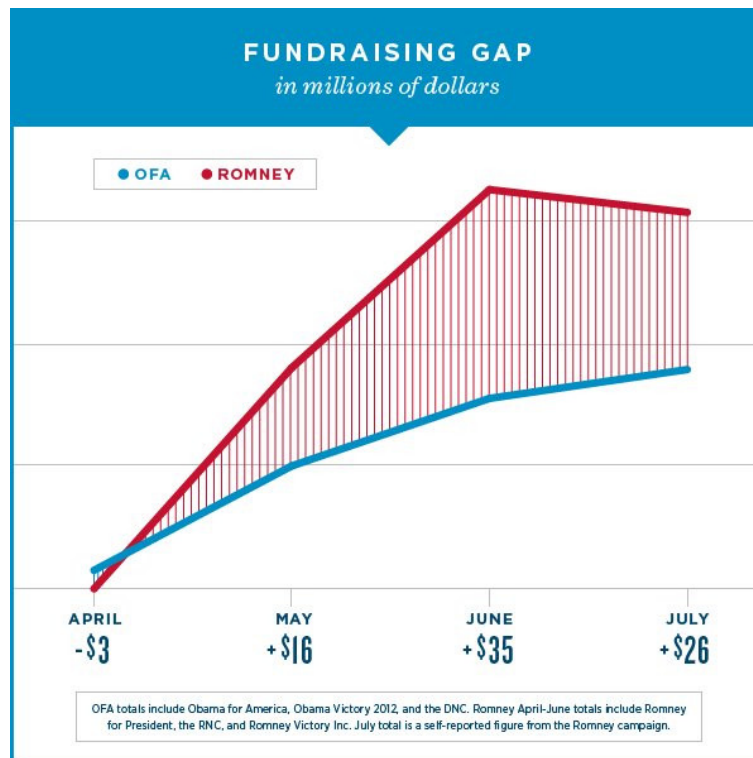


Figura 1: Trucamiento en los ejes

- Uso de dimensiones o formas inadecuadas:** Por ejemplo, utilizar áreas o volúmenes para codificar valores unidimensionales sin ajustar adecuadamente puede engañar. Es famoso el caso de dibujos de figuras (pictogramas) donde duplicar un valor acaba cuadruplicando el área de la imagen. Los colores también pueden emplearse de forma engañosa: el uso de escalas de color no lineales o la elección

de paletas que resalten ciertas categorías por encima de otras inducen sesgos. Igualmente, el abuso de gráficos 3D o con perspectiva inclinada tiende a complicar la comparación correcta de magnitudes [5].

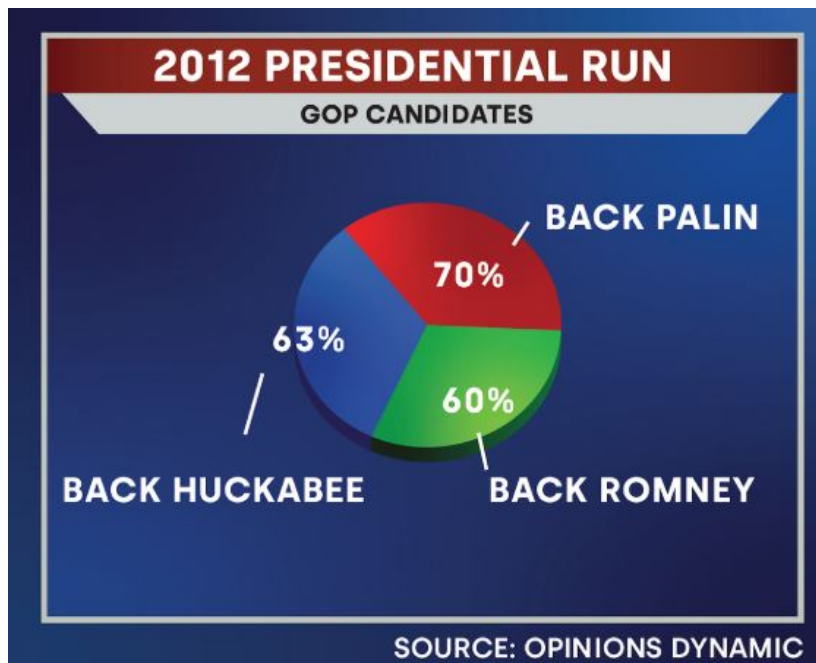


Figura 2: Dimensiones inadecuadas

- **Sobrecarga o trucos de diseño:** Incluye la inclusión de efectos visuales que distraen o confunden o la presentación de datos acumulados sin contexto (por ejemplo, acumulando porcentajes que suman más de 100 % en un gráfico de sectores). Aunque muchos de estos trucos hacen el gráfico “vistoso”, tienden a oscurecer la información o incluso a sugerir [10].

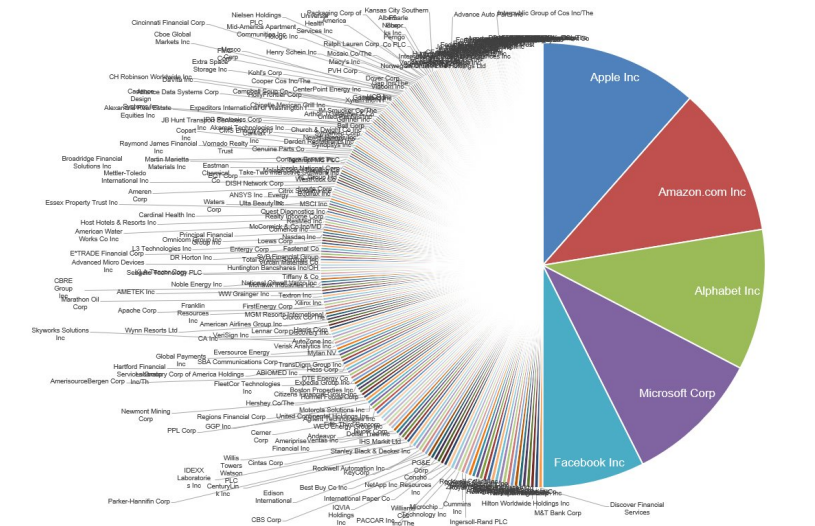


Figura 3: Sobrecarga de Datos

Herramientas y librerías utilizadas:

- Google Colab:** Colab fue diseñado para que investigadores de IA y ciencia de datos compartieran experimentos reproducibles y explicaciones de técnicas y es por ello que también ha dado buenos resultados en esta investigación. Su principal ventaja es que proporciona a los usuarios recursos computacionales suficientes para ejecutar técnicas modernas de IA de forma interactiva y evita que tengan que configurar paquetes de software y dependencias por separado [11].
- Python:** Es un lenguaje de programación fácil de aprender y ampliamente utilizado para crear diversos elementos, como plataformas web, aplicaciones, sistemas de inteligencia artificial, entre otros. La sencilla sintaxis de Python, similar a la del inglés, hace que la comprensión del código y la creación de *scripts* sean notablemente intuitivos. Asimismo, se utiliza para una gran cantidad de tareas: automatizar trabajos repetitivos, examinar una gran cantidad de datos, crear juegos, supervisar servidores web. Además, la flexibilidad de Python se ve reforzada por la disponibilidad de bibliotecas, que son depósitos de código preexistente, lo que permite la expansión de su funcionalidad, lo que permite realizar prácticamente cualquier tarea utilizando Python. Es por ello que se utilizó este lenguaje para la creación del software.
- Transformers:** Librería de python, en el presente trabajo es importante ya que ayuda a cargar, entrenar e inferir con modelos pre-entrenados (NLP, visión, audio e incluso multimodales). Incluye *pipeline* y *trainer* para facilitar su uso y rendimiento a escala [12]

- **Torch/Torchvision:** Librería de python que maneja tensores, modelos de *deep learning* y procesamiento de imágenes. Indispensable para ejecutar entrenamiento y acelerarlo con GPU.
- **SentencePiece:** Librería de python, la cual es una herramienta de tokenización que segmenta el texto en unidades (*tokens*) de forma eficiente y compatible con modelos que usan byte-pair-encoding (BPE) o *tokens* sub-palabra [13].
- **BitsAndBytes:** Librería de python, la cual permite cuantización en 8 o 4 bits, reduciendo el uso de memoria y haciendo posible el *fine-tuning* de modelos grandes con menos recursos.
- **Accelerate:** Librería de python que simplifica el entrenamiento distribuido y la inferencia con *mixed-precision*. Se integra con bitsandbytes para usar cuantización sin errores.
- **PEFT (Parameter-Efficient Fine-Tuning):** Librería de python que implementa técnicas como LoRA y QLoRA, que permiten ajustar solo partes pequeñas del modelo, ahorrando tiempo y recursos computacionales [14].
- **Decord:** Librería de python, la cual ayuda a la optimización de carga de secuencias de imágenes o multimedia.
- **Pillow:** Librería de python de manejo básico de imágenes (lectura, redimensionado, conversión), necesario para los trabajos con modelos que procesan imágenes.

3.2 Limitaciones

- Limitaciones computacionales
- Restricciones del entorno Colab
- Limitaciones de disponibilidad de datos
- Subjetividad y ambigüedad del problema

3.3 Evaluación preliminar de algoritmos descartados

Durante la fase exploratoria del proyecto se llevó a cabo una revisión y prueba preliminar de distintos modelos de lenguaje de gran tamaño (LLMs) con el propósito de identificar aquellos con el mayor potencial para ser ajustados finamente y aplicados a la detección de visualizaciones de datos engañosas, específicamente desde una perspectiva basada en elementos visuales primitivos. Esta fase resultó importante, para delimitar el enfoque metodológico, pues permitió analizar las capacidades reales de cada modelo en cuanto a la comprensión multimodal, así como su adaptabilidad a tareas específicas mediante ajuste fino. Entre los modelos considerados en esta evaluación se encuentran “*microsoft/git-base*” y “*MiniCPM-V-2_6-int4*”, seleccionados inicialmente por su disponibilidad pública, soporte en la plataforma Hugging Face, y documentación técnica que sugería compatibilidad con tareas similares. El

modelo “microsoft/git-base” está diseñado para tareas multimodales, y se ha empleado en diversos escenarios donde se requiere la integración de información visual y textual. No obstante, en las pruebas realizadas dentro del marco de esta investigación, dicho modelo presentó limitaciones significativas. Específicamente, se observó un bajo desempeño en la extracción semántica de información relevante desde las visualizaciones de datos, así como dificultades para identificar relaciones entre elementos visuales primitivos y el texto asociado. Estas deficiencias impactaron negativamente en la calidad de las inferencias generadas, y en consecuencia, en su utilidad práctica para los objetivos. Asimismo, las respuestas brindadas del modelo eran insuficientes.

3.4 Selección y análisis del modelo

Para el presente trabajo se ha utilizado como modelo el “*MiniCPM-V-2_6*” el cual está en la plataforma “Hugging Face” [15]. Como mencionan Ait *et al.* [16] la plataforma Hugging Face es una de las más populares. Ya que se centra no solo en facilitar el intercambio de conjuntos de datos, sino también, modelos preentrenados y aplicaciones creadas con ellos. Además, se están añadiendo cada vez más funciones colaborativas, para facilitar la creación de estos artefactos dentro de la propia plataforma. Con más de 100 000 repositorios y en rápido crecimiento, Hugging face se está convirtiendo en una fuente prometedora de datos. Por ello, se le ha considerado como una fuente prometedora para el presente trabajo. Es importante mencionar que el modelo fue elaborado por la empresa openbmb que es la abreviatura de sus siglas en inglés “Open Lab for Big Model Base” [17] y como lo mencionan en su web el objetivo de OpenBMB es construir la base de modelos y el conjunto de herramientas para modelos de lenguaje preentrenados a gran escala y con ello ayudan a acelerar el proceso de entrenamiento, ajuste e inferencia para modelos grandes y reducir las barreras para su uso. Ellos brindan el código de manera abierta junto con desarrolladores de todo el mundo para promover los ecosistemas de código abierto de modelos de lenguaje preentrenados. Con respecto al *dataset*, se descargaron gráficos en formato de imágenes que cumplan con que sean engañosas y no engañosas, esto quiere decir que algunas imágenes han sido alteradas para transmitir un mensaje erróneo y la otra parte son imágenes no alteradas. Asimismo, se tomó el *dataset* de los autores Masry *et al.* [18] haciéndole algunas modificaciones al JSON para adaptarlo mejor al presente trabajo, gracias a ello se obtuvo más de dieciocho mil imágenes.

3.4.1 Implementación

Para llevar a cabo la implementación del modelo ajustado, se desarrolló un entorno de trabajo en Google Colab Pro utilizando la infraestructura

de GPU proporcionada por la plataforma. El objetivo principal fue realizar el *fine-tuning* del modelo multimodal *openbmb/MiniCPM-V-2_6*, adaptándolo específicamente a la tarea de detección de visualizaciones de imágenes engañosas.

La primera etapa del proceso consistió en montar Google Drive como sistema de almacenamiento, desde donde se accedió tanto al conjunto de imágenes como al archivo de anotaciones JSON personalizado. Dicho archivo incluía, para cada imagen, una consulta (*prompt*) y una etiqueta.

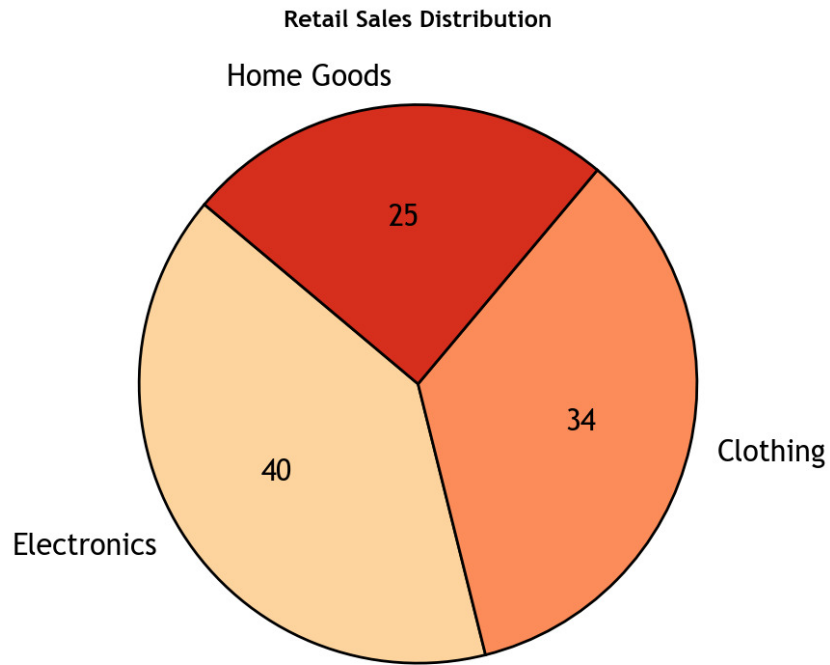
A continuación, se procedió a la configuración del entorno con las versiones compatibles de las librerías necesarias: *transformers*, *torch*, *accelerate*, *peft*, *bitsandbytes*, *pillow*, *sentencepiece*, *decord* y *torchvision*. Se utilizó el modelo base *openbmb/MiniCPM-V-2_6*, cargado desde Hugging Face, y se adaptó para entrenamiento eficiente en GPU mediante técnicas de cuantización en 4 bits y entrenamiento eficiente por parámetros *PEFT* mediante *LoRA*.

Para ello, se aplicó una configuración de *LoRA* con parámetros específicos y se indicaron los módulos objetivo *q_proj* y *v_proj*. Se preparó el modelo con *prepare_model_for_kbit_training* y luego se aplicó *get_peft_model* para encapsularlo en el esquema *LoRA*.

Durante el entrenamiento, se procesó cada imagen asociada a su respectivo *prompt*. Se convirtió a tensor la información de texto y se utilizó un bucle de entrenamiento personalizado. A cada paso, el modelo recibía como entrada una imagen y una consulta textual. Esta implementación permite que el modelo aprenda a asociar características visuales con posibles manipulaciones gráficas descritas en los *prompts*, desarrollando así una comprensión más ajustada al contexto de desinformación visual.

3.5 Despliegue y evaluación del modelo

Se evidenciarán algunas de las pruebas realizadas a imágenes tanto engañosas como no. Esto se hace para demostrar que el modelo funciona de manera adecuada. Tras un análisis comparativo de diversos *prompts*, se seleccionaron tres como los más efectivos. Dos de ellos fueron elaborados tomando como referencia las estrategias propuestas por los autores G. Zheng *et al.* [19] y B. Chen *et al.* [20], cuyas metodologías sirvieron de guía para su formulación. El tercer *prompt* fue diseñado de manera original, basado en una estructura de pregunta comúnmente utilizada por los usuarios en contextos reales. Las pruebas realizadas demostraron que el uso de estos tres *prompts* contribuye a una mejora significativa en la calidad de las respuestas generadas por el modelo. A continuación se mostrarán 2 ejemplos, primero una imagen engañosa y luego la otra imagen sin alteraciones, esto quiere decir que no es engañosa:



Source: National Retail Federation

Figura 4: Visualización de datos 1
[18]

En la Tabla 1 se indica los *prompts* utilizado para analizar la Figura 4, así como la respuesta generada por el modelo y si la interpretación fue considerada correcta o no.

<i>PROMPTS</i>	RESPUESTA DEL MODELO	¿INTERPRETACIÓN CORRECTA?
----------------	----------------------	---------------------------

<p>Is that graph misleading?</p>	<p>The graph is misleading; it represents the distribution of retail sales among three categories: Home Goods, Electronics, and Clothing. The segments are clearly labeled with their respective percentages, making it easy for viewers to understand how each category contributes to the total retail sales.</p>	<p>El modelo confirmó que es un gráfico engañoso</p>
<p>I'm going to show you a picture of a graph. First, describe each visual element (axes, labels, colors, graph type). Then, think step by step about possible visual problems: Does the Y-axis start at zero? Are the proportions consistent? Could the colors be misleading? Finally, answer the question: Is this graph misleading? And explain why.</p>	<p>The pie chart displays the distribution of retail sales, with each segment representing a different category: Home Goods (25%), Electronics (40%), and Clothing (34%). The colors used are red for Home Goods, beige for Electronics, and orange for Clothing. There is no Y-axis present in this type of graph; instead, the proportions are indicated by the size of the segments relative to the whole circle. The Y-axis does not start at zero; it begins at 0% which is appropriate for a percentage-based representation. The proportions add up to approximately 99%, which suggests that there might be a small rounding error or an incomplete data point missing from the chart. Considering these points, there is a slight discrepancy when the total does not reach 100%, Therefore, the graph is misleading.</p>	<p>El modelo detalló mejor al gráfico y su tendencia es indicar que el gráfico puede ser engañoso</p>

<p>First, describe what you see: axes, ranges, labels, colors, scale. Second, critically analyze each element: Are there truncated axes? Misleading colors? Is there a logarithmic scale without warning? Finally, based on the above, is the graph misleading? Detail which elements are misleading.</p>	<p>First: The graph is a pie chart representing the distribution of retail sales among three categories: Home Goods, Electronics, and Clothing. The data source is cited as the National Retail Federation. Second, since the graph uses percentages and there is no indication of a logarithmic scale, there is no warning needed. There are no visible axes; thus, it's impossible to determine if they are truncated. The colors used do not inherently mislead; they clearly differentiate between the categories. In conclusion is misleading</p>	<p>Detalló de manera acertada el gráfico y afirmó que el gráfico es engañoso</p>
---	--	--

Tabla 1: Análisis de visualización de datos 1

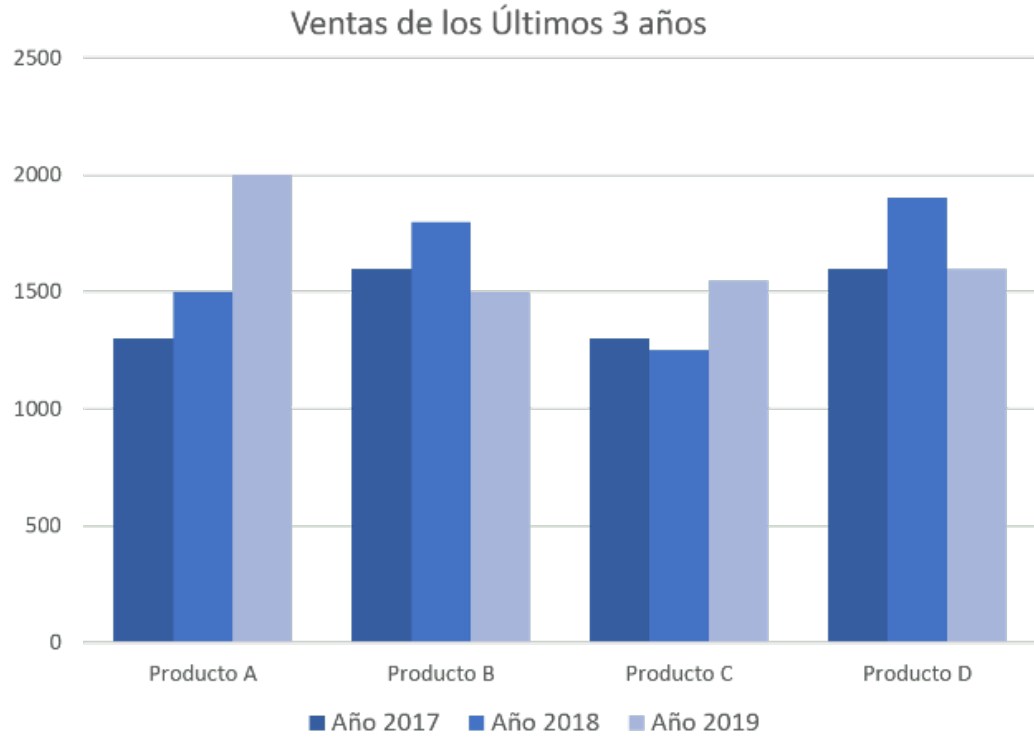


Figura 5: Visualización de datos 7
[13]

En la Tabla 2 se indica los *prompts* utilizado para analizar la Figura 5, así como la respuesta generada por el modelo y si la interpretación fue considerada correcta o no.

PROMPTS	RESPUESTA DEL MODELO	¿INTERPRETACIÓN CORRECTA?
Is that graph misleading?	Based on the analysis, the chart may not be misleading due to clear axis labels; it's missing some features, such as the legend.	El modelo dio una respuesta correcta

<p>I'm going to show you a picture of a graph. First, describe each visual element (axes, labels, colors, graph type). Then, think step by step about possible visual problems: Does the Y-axis start at zero? Are the proportions consistent? Could the colors be misleading? Finally, answer the question: Is this graph misleading? And explain why.</p>	<p>The graph is a bar chart with the Y-axis labeled in increments of 100, starting at zero. Each product category (Product A, Product B, Product C) has three bars representing different years: 2017 (light blue), 2018 (dark blue), and 2019 (white). The X-axis labels these categories sequentially. The Y-axis starts at zero, which is standard for measuring quantities that can be negative or have no lower limit. The colors used light blue, dark blue, and white are distinct enough to differentiate between the years, which helps in reading the graph without confusion. Given these observations, the graph appears to be well-constructed and not inherently misleading if the proportions are consistent over time.</p>	<p>El modelo detalló mejor al gráfico</p>
<p>First, describe what you see: axes, ranges, labels, colors, scale. Second, critically analyze each element: Are there truncated axes? Misleading colors? Is there a logarithmic scale without warning? Finally, based on the above, is the graph misleading? Detail which elements are misleading.</p>	<p>First: El gráfico parece un gráfico de barras que muestra los valores de cuatro categorías diferentes a lo largo de tres años. The x-axis represents the products (Product 1 to Product 4). The y-axis represents the values (in arbitrary units), with labels at intervals of 250. Second, each category has a label indicating the year (Fly 2027, Fly 2028, and Fly 2029). The scale seems consistent across the axes, but there's no indication if it's logarithmic or linear. In conclusion, No, the graph is not misleading.</p>	<p>Detalló de manera acertada el gráfico y afirmó que el gráfico no es engañoso</p>

Tabla 2: Análisis de visualización de datos 7

4. DISCUSIÓN

Los resultados obtenidos en este trabajo reflejan de forma clara el potencial de los modelos de lenguaje de gran tamaño (LLMs), específicamente aquellos con arquitectura multimodal, para abordar la detección de visualizaciones de datos engañosas desde una perspectiva centrada en sus elementos primitivos. A lo largo del proceso experimental, se evidenció cómo el ajuste fino del modelo seleccionado permitió mejorar sustancialmente la calidad de las interpretaciones generadas, tanto en precisión como en argumentación.

La evolución observada en las respuestas del modelo tras la integración de estrategias de *prompting* más elaboradas, ha demostrado que no basta con simplemente preguntarle a un modelo si un gráfico es engañoso. La clave radica en cómo se le formula la consulta, ya que se puede producir un posible sesgo por la formulación de un mal *prompt*, dado que los modelos de lenguaje de gran tamaño son altamente sensibles al enunciado de la instrucción, el tipo de pregunta planteada puede condicionar significativamente la respuesta. Por ejemplo, un *prompt* que contiene términos como “engañoso” o “identifica errores” puede predisponer al modelo a asumir que existe algún problema en la visualización, incluso si no lo hay. Por el contrario, un *prompt* excesivamente neutro puede conducir a respuestas superficiales o no comprometidas [21]. En el presente trabajo, se ha observado que los *prompts* más estructurados tienden a guiar mejor al modelo, existe un equilibrio delicado entre estructurar el razonamiento del modelo y no sesgar su juicio. Gracias a estos *prompts* estructurados, condujeron al modelo a examinar individualmente los componentes del gráfico (ejes, escala, colores, leyendas), permitiendo obtener razonamientos más completos, coherentes y fundamentados. Esta mejora fue especialmente notoria en gráficas con distorsiones sutiles.

Asimismo, se corroboró que los LLMs, aunque poderosos, pueden ser fácilmente confundidos por gráficos manipulados si no se les entrena con ejemplos adecuados y diversos. La incorporación de datos sintéticos y gráficos reales permitió al modelo enfrentarse a una gama variada de engaños, desde truncamiento de ejes hasta uso indebido del color o escalas no uniformes. Esto refuerza la necesidad de contar con *datasets* más robustos y representativos que reflejen los desafíos reales de la desinformación gráfica.

Por otra parte, se hizo evidente que no todos los modelos disponibles en plataformas como Hugging Face están preparados para este tipo de tareas. El descarte de alternativas como *microsoft/git-base* y *MiniCPM-V-2_6-int4* en las primeras fases del proyecto pone de manifiesto la importancia de una selección cuidadosa basada no solo en la disponibilidad del modelo, sino también en sus capacidades multimodales reales, su compatibilidad con ajuste fino y su rendimiento frente a *prompts* complejos.

En términos generales, el modelo ajustado mostró una mejora grande en la interpretación y detección de gráficos engañosos, no solo señalando el posible engaño, sino también proporcionando justificaciones razonadas. Esta capacidad argumentativa resulta fundamental si se aspira a integrar estos modelos en contextos donde la transparencia y la explicabilidad son esenciales, como en el periodismo, la docencia o la investigación científica.

5. Conclusiones

En este trabajo se ha abordado una problemática actual y relevante que es la detección de visualizaciones de datos engañosas en contextos donde la información visual es cada vez más utilizada para influir en la opinión pública, la toma de decisiones y la interpretación científica. Frente a este desafío, se propuso el ajuste fino de modelos de lenguaje de gran tamaño (LLMs) con capacidades multimodales, centrándose en el análisis de los elementos gráficos primitivos como ejes, proporciones, formas y colores.

A través de una revisión sistemática de literatura, se identificaron los enfoques más relevantes en cuanto a modelos, *datasets*, técnicas de *fine-tuning* e ingeniería de *prompts* aplicados a tareas de interpretación visual. Posteriormente, se diseñó e implementó una arquitectura experimental basada en el modelo *openbmb/MiniCPM-V-2_6*, adaptada mediante técnicas de ajuste eficiente (LoRA, PEFT) y evaluada con un conjunto de gráficos tanto reales como sintéticos, etiquetados según su grado de veracidad.

Los resultados obtenidos permiten concluir que los LLMs, cuando son correctamente ajustados y guiados mediante *prompts* estructurados, pueden identificar patrones gráficos engañosos con razonamientos detallados, aportando explicaciones que van más allá de una simple clasificación binaria. Este enfoque no solo mejora la precisión del modelo frente a manipulaciones visuales, sino que también promueve una interpretación más crítica y transparente por parte del usuario final.

Asimismo, el trabajo ha evidenciado las principales limitaciones técnicas y operativas de los entornos de desarrollo actuales, como Google Colab Pro, así como la escasez de *datasets* públicos suficientemente variados y explicativos. Estas restricciones constituyen oportunidades para futuras investigaciones.

En definitiva, este estudio ha demostrado que es viable y muy necesario, avanzar hacia sistemas basados en IA capaces de detectar visualizaciones engañosas de manera explicable, robusta y contextualizada. Los resultados obtenidos sientan las bases para el desarrollo de herramientas automatizadas que fortalezcan la alfabetización visual, apoyen la verificación de información y contribuyan a combatir la desinformación gráfica en entornos digitales, educativos y profesionales.

6. Futuras líneas de investigación

Los avances logrados en este trabajo abren múltiples posibilidades para continuar profundizando en la intersección entre modelos de lenguaje multimodales y la interpretación crítica de visualizaciones de datos. Algunas líneas de investigación que podrían explorarse en futuros desarrollos incluyen:

- **Ampliación y estandarización de *datasets*:** La creación de un corpus amplio y balanceado de visualizaciones engañosas y no engañosas, etiquetado con metadatos detallados, permitiría mejorar la generalización de los modelos y facilitar comparaciones entre investigaciones.
- **Modelos más ligeros y eficientes:** Dado que los recursos computacionales siguen siendo una limitación, resulta pertinente explorar arquitecturas más livianas, basadas en técnicas como *knowledge distillation* o *sparsity*, que mantengan un rendimiento aceptable en entornos con recursos limitados. En particular, el *knowledge distillation* consiste en transferir el conocimiento de un modelo grande (denominado teacher) a uno más pequeño (student), entrenando a este último para que imite las salidas del primero. Este proceso permite reducir la complejidad del modelo sin comprometer significativamente su precisión. Por otro lado, la técnica de *sparsity* busca eliminar o reducir ciertos parámetros del modelo, por ejemplo, estableciendo pesos cercanos o iguales a cero, con el fin de hacerlo más eficiente en términos de memoria y cómputo. Esto se logra mediante estrategias como *pruning*, cuantización o regularización, que permiten obtener representaciones más compactas conservando una buena capacidad de generalización.
- **Aplicación a dominios específicos:** Explorar el uso de estos modelos en contextos especializados como periodismo de datos, educación secundaria o revisión científica permitiría adaptar el sistema a las necesidades concretas de cada disciplina.

Estas propuestas contribuirían a consolidar un marco tecnológico que no solo detecte visualizaciones engañosas, sino que también fomente una cultura visual crítica y responsable en la sociedad.

Referencias

1. L. Y.-H. Lo and H. Qu, “How good (or bad) are llms at detecting misleading visualizations?,” *arXiv*, jul 2024.
2. G. Li, W. Liu, Y. Wu, P. Wang, S. Huang, H. Xu, and J. Wu, “From misleading queries to accurate answers: A three-stage fine-tuning method for llms,” *arXiv*, apr 2025.
3. Z. Li, W. Zhang, and J. Tao, “Flowllm: Large language model driven flow visualization,” *Visual Informatics*, may 2025. available May 2.2025.

4. L. Wilkinson, "The grammar of graphics," in *Handbook of Computational Statistics: Concepts and Methods*, pp. 375–414, Berlin, Heidelberg: Springer Berlin Heidelberg, 2015.
5. D. A. Szafr, "Five ways visualizations can mislead (and how to fix them)," *Interactions*, vol. 25, no. 4, pp. 26–33, 2018.
6. B. Kitchenham, "Procedures for performing systematic reviews," *Dept. Comput. Sci., Keele Univ., Keele, U.K., Tech. Rep.0400011T.1.*, vol. 33, pp. 1–26, 2004.
7. B. Kitchenham and S. Charters, "Guidelines for performing systematic literature reviews in software engineering," *School Comput. Sci. Math., Keele Univ., Keele, U.K., Tech. Rep. EBSE-2007-01.*, 2007.
8. M. Dixon-Woods *et al.*, "How can systematic reviews incorporate qualitative research? a critical perspective," *Qualitative Research*, vol. 6, no. 1, pp. 27–44, 2006.
9. M. Lisnic, C. Polychronis, A. Lex, and M. Kogan, "Misleading beyond visual tricks: How people actually lie with charts," in *Proc. CHI '23*, (New York, NY, USA), pp. 1–21, ACM, apr 2023.
10. J. Wang, Z. Yang, X. Hu, L. Li, K. Lin, Z. Gan, Z. Liu, and C. Liu, "Git: A generative image-to-text transformer for vision and language," tech. rep., Microsoft Research, 2022.
11. M. J. Nelson and A. K. Hoover, "Notes on using google colabatory in ai education," in *Proc. ACM Conf. Innov. Technol. Comput. Sci. Educ. (ITICSE)*, pp. 533–534, Assoc Computing Machinery, 2020.
12. T. Wolf *et al.*, "Transformers: State-of-the-art natural language processing," in *Proc. 2020 Conf. Empirical Methods Natural Language Processing: System Demonstrations*, pp. 38–45, 2020.
13. T. Kudo and J. Richardson, "Sentencepiece: A simple and language independent subword tokenizer and detokenizer for neural text processing," in *Proc. 2018 Conf. Empirical Methods in Natural Language Processing: System Demonstrations*, (Brussels, Belgium), pp. 66–71, 2018.
14. R. Patil, P. Khot, and V. Gudivada, "Analyzing llama3 performance on classification task using lora and qlora techniques," *Applied Sciences*, vol. 15, no. 6, p. 3087, 2025.
15. openbmb, "MiniCPM-V-2_6 · Hugging Face." url: https://huggingface.co/openbmb/MiniCPM-V-2_6. Accessed: 2025-06-23.
16. A. Ait, J. L. Canovas Izquierdo, and J. Cabot, "Hfcommunity: A tool to analyze the hugging face hub community," in *2023 IEEE Int. Conf. Softw. Anal. Evol. Reengineering (SANER)*, (Los Alamitos), pp. 728–732, IEEE Computer Soc, 2023.
17. open bmb, "open-bmb." url: <https://www.openbmb.cn/en/about-us>. Accessed: 2025-06-23.
18. A. Masry *et al.*, "Chartqa: A benchmark for question answering about charts with visual and logical reasoning," *arXiv*, 2022. arXiv:2203.10244.
19. G. Zheng, B. Yang, J. Tang, H.-Y. Zhou, and S. Yang, "Ddcot: Duty-distinct chain-of-thought prompting for multimodal reasoning in language models," *arXiv*, oct 2023. arXiv:2310.16436.
20. B. Chen, Z. Zhang, N. Langrené, and S. Zhu, "Unleashing the potential of prompt engineering for large language models," *Patterns*, vol. 6, p. 101260, may 2025.

21. X. Yang, R. Zhan, D. F. Wong, S. Yang, J. Wu, and L. S. Chao, "Rethinking prompt-based debiasing in large language models," 2025. arXiv:2503.09219.

Desambiguación de nombres de autor en bases de datos bibliográficas con pocos metadatos

Diego Uribe Esquivel, Ángel Francisco Zazo Rodríguez

Universidad de Salamanca
Departamento de Informática y Automática
Facultad de Ciencias.
Plaza de los Caídos s/n
37008 Salamanca, España
{idu061413, angelzazo}@usal.es

Resumen Este trabajo explora diversas técnicas para implementar la desambiguación de nombres de autor (AND) en bases de datos bibliográficas antiguas con metadatos limitados. Se investigaron distintos métodos que hayan sido usados en este campo, como por ejemplo métodos de similitud sintáctica, como Levenshtein y Jaro-Winkler, técnicas de minería de texto; por ontologías, similitud sintáctica como TF-IDF, Word Embeddings como Word2Vec, BERT y embeddings con LLMs), enfoques de minería de datos abarcando algoritmos de Machine Learning supervisados, no supervisados y combinaciones (semi-supervisados). Así mismo, se abarcaron algunas técnicas de Deep Learning e incrustaciones de redes (Network Embeddings). La implementación del proyecto combina 3 métodos principales, la similitud de texto para nombres de autores, el Word Embedding de títulos, usando el LLM "nomicz" la comparación por medio de la similitud del coseno, junto con redes de coautoría. Así mismo, se emplean herramientas como Apache Solr y ChromaDB para una gestión y recuperación eficiente de los datos.

Keywords: Desambiguación de Nombres de Autor (AND), metadatos limitados, Machine Learning, supervisados y no supervisados, Word Embeddings, Network Embeddings, LLMs, cosine similarity

1. Introducción

El aspecto de la ambigüedad en los nombres de los autores es un tema controversial, que ha existido durante mucho tiempo atrás, pero que se ha hecho consciente en épocas modernas gracias al gran acceso a la información que se posee hoy en día. Dicho esto, se vuelve necesario identificar a los autores y diferenciarlos correctamente de otros autores, para así conocer los autores reales de las obras.

Para cualquier trabajo de literatura, ya sea una obra histórica, una novela o incluso un trabajo de investigación, un problema fundamental es el de identificar el o los individuos quienes redactaron. Este puede representar un problema grande por atacar, por más simple que pueda parecer.

Un autor podría tener el mismo nombre a través de toda su trayectoria en los libros, lo que facilitaría la desambiguación, aunque esto no sería la única condición determinante, ya que existen muchos otros parámetros, además del nombre, que pueden ser diferentes a la persona que se tiene en mente.

Una de las principales problemáticas de la desambiguación consiste en la identificación de autores por su nombre. Hay dos padecimientos que pueden dificultar este proceso [2, 9, 18], y son:

- Las variaciones en el nombre de un mismo autor a lo largo de sus publicaciones (o mejor conocido como **Sinonimia**).
- La aparición de múltiples autores diferentes compartiendo el mismo nombre (mejor conocido en este ámbito como **Homonimia**, o **Polisemia**, en algunos artículos), o similar.

Con el paso del tiempo y a la edad del internet, se han creado bases de datos en donde se guardan bibliografías grandes, como artículos científicos, periódicos, patentes, novelas, y una infinidad de obras que involucran autoría. Estas bases de datos se crearon con el propósito de tener toda la información centralizada dentro de un solo repositorio, para así ayudar a las búsquedas de personas dedicadas a la investigación. [9]

En este caso, los registros disponibles contienen información antigua, que proviene de épocas en las que probablemente no existían formas estandarizadas de identificar a una persona como autor. Por lo tanto, es difícil distinguir al autor del resto de autores, debido a la falta de metadatos que esclarezcan la desambiguación.

Para un humano es más fácil desambiguar un autor, pues además de revisar el nombre y buscar autores con uno similar, revisa la bibliografía, el texto contenido y el significado semántico, y busca autores con similitud en el perfil. Sin embargo, este mismo procedimiento es inviable cuando se tratan con miles de obras a desambiguar, lo que es una limitante al lidiar con las bases de datos masivas. [19]

Es por eso que se deben contar con métodos de desambiguación efectiva, ya que se tiene un número elevado de autores en ellas, y que además, debido a los avances tecnológicos, esta curva crece exponencialmente con el paso del tiempo. [9, 10, 23]

2. Objetivos de este trabajo de investigación

Como objetivo principal de este Trabajo de Fin de Máster, se hizo una investigación sistemática sobre los diferentes métodos para

la desambiguación de nombres de autores (AND), y se elaboró una estrategia seleccionando los métodos más convenientes para la AND al disponer de metadatos escasos. Se elaboró un programa en lenguaje Python que sirvió como herramienta para emplear las estrategias y evaluar su uso.

2.1 Bibliotecas digitales

Las bases de datos de esta investigación fueron bibliotecas digitales en las que se resguardan copias de las obras en formato digital. Entre ellas están implicadas acervos generales, bibliotecas de universidades, y proyectos de gran escala de reserva de títulos. En este estudio, se analizarán casos de estudio sobre las siguientes bases de datos:

- La Biblioteca General Histórica de la Universidad de Salamanca
- Fichero de Autoridades Virtual Internacional (VIAF, por sus siglas en inglés, que se traduce como Virtual International Authority File)
- La Biblioteca Nacional de España (o BNE)

3. Métodos para AND usados en otros estudios

Han habido varios estudios y esfuerzos de diversos autores por ayudar a desambiguar las bases de datos masivas. Para hacerlo, existen dos paradigmas principales [18]:

- **Agrupamiento de Autores (Clustering):** Este enfoque busca agrupar referencias que pertenezcan al mismo autor mediante el análisis de sus atributos y el uso de funciones de similitud o técnicas de agrupamiento.
- **Asignación de Autores (Clasificación):** Los métodos de asignación construyen un modelo que representa a un autor y atribuyen directamente los registros de autoría a sus respectivos autores, a menudo empleando técnicas de clasificación o agrupamiento.

3.1 Técnicas de minería de datos

Actualmente se han estado usando métodos más modernos, incluso con el auge de la Inteligencia Artificial, que implican el uso de:

- Técnicas de **Machine Learning** y **aprendizaje automático**.
- Redes neuronales de diferentes tipos; **Deep learning**

Los modelos modernos empleados para AND sirven para llevar a cabo la desambiguación de forma eficiente, y en algunos casos, automáticamente, de tal manera que ahorra mucho tiempo y trabajo al ser humano evitando hacer una desambiguación manual. [3, 6, 9, 10]

3.1.1 Machine Learning: Métodos supervisados Estos métodos utilizan datos de entrenamiento previamente etiquetados¹, los cuales consisten en dos pares de perfiles², que se marcan como coincidentes o no coincidentes con el mismo autor [13, 14, 16]. Estos datos se usan generalmente para entrenar a un clasificador, los cuales tienen un mejor rendimiento que los enfoques no supervisados [14]. Sin embargo, se requiere de un gran esfuerzo para tener estos conjuntos de datos correctamente etiquetados, si se preparan de forma manual, y además deben ser de alta calidad. [7]

Algunos enfoques comunes enlistados en análisis de las investigaciones consultadas, pueden ser:

- **Máquinas de Soporte Vectorial (SVM, por sus siglas en inglés)** [3, 4, 7–9, 22]
- **Algoritmo Naïve-Bayes** [3, 4, 8, 9, 22]
- **Decision trees** [7, 13–15]
- **Random Forests** [14–16, 20, 23]
- **Regresión logística** [9, 13, 15, 18, 20]

Previo a la clasificación, para tener un mejor aprovechamiento de estos conjuntos de datos, se pasan por una etapa de preprocesamiento [6, 7, 12, 13, 22], que consiste en eliminar impurezas que puedan causarles ruido a los modelos. Estas técnicas implican la normalización del texto a minúsculas, eliminación de acentos, signos de puntuación y caracteres especiales, tokenización de palabras, y, la técnica de “bloqueo” (o blocking en inglés), que consiste en agrupar todos los autores con el mismo apellido e inicial de su primer nombre (técnica llamada Last Name – First Initial, LN-FI [8, 14]), lo que reduce drásticamente la complejidad computacional, pero muchas veces podría generar falsos positivos.

3.1.2 Machine Learning: Métodos no supervisados Los algoritmos no supervisados, por el contrario, no requieren de datos etiquetados previamente para su entrenamiento, por lo que son útiles cuando no se disponen de etiquetas, por ejemplo:

1. **Algoritmo de agrupación K-means** [7, 13, 17, 19]
2. **Agrupación Espacial Basada en la Densidad de Aplicaciones con Ruido (Density Based Spatial Clustering of Applications with Noise, DBSCAN)** [13–15, 22]
3. **Clustering jerárquico aglomerativo (o Hierarchical Agglomerative Clustering en inglés, HAC)** [3, 7, 17, 22]
4. **Clustering espectral** [7, 9, 13, 19]

¹ Contienen las variables y el resultado que debería ser predicho

² O artículos, dependiendo de lo que se quiera comparar

3.1.3 Machine Learning: Métodos semi-supervisados Existen también **enfoques de aprendizaje semisupervisado**, los cuales son una estrategia híbrida de llevar a cabo las tareas de desambiguación. Éstas combinan ambos métodos en sí para llevar a cabo el proceso de AND [7]. Requieren tomar una pequeña cantidad de datos etiquetados con una mayor cantidad de datos no etiquetados para su entrenamiento, lo que resulta en un método más sencillo de entrenar que los modelos supervisados [6,14], y el uso de lo mejor de dos mundos.

3.2 Minería de textos

Estos métodos tienen un enfoque un poco diferente a los anteriores. Implican la transformación de texto no estructurado, como el texto contenido dentro de artículos digitales, revistas, sitios web, etc. en un formato estructurado, con la finalidad de identificar patrones significativos y tendencias textuales dentro de ellos. Involucran en sí el uso de técnicas de Procesamiento de Lenguaje Natural (PLN, o NLP por Natural Language Processing en inglés). [12]

Para estos métodos, se encontraron dos diferentes tipos de técnicas existentes en la bibliografía [7, 12, 14]. La primera consiste en identificación de términos y temas en común:

1. **TF-IDF** [1, 7–9, 14, 19]
2. **Modelos de Tópicos latentes**, como **Probabilistic Latent Semantic Analysis (PLSA)** o **Latent Dirichlet Allocation (LDA)**. Estos modelos asignan vectores probabilísticos a los tópicos de cada cita, finalmente utilizando estas distribuciones para medir la similitud entre artículos. [7, 19]
3. **Reconocimiento de Entidades Nombradas** o **Named Entity Recognition, NER**, en el cual se identifican categorías predefinidas de objetos en un cuerpo de texto. [19]

Los algoritmos de la segunda técnica se basan en el **Word Embedding**. Esta técnica implica el uso de redes neuronales, que codifican una frase a un formato numérico con el uso de vectores de alta dimensionalidad para su mejor entendimiento en las máquinas.

1. **Char2Vec** es el primer método sencillo encontrado en [3]. Esta estrategia se basa en la codificación de nombres u oraciones de tamaño muy corto. Aprende la representación de los textos letra por letra.
2. **Word2Vec** [12], el cual aprende las palabras dadas de los textos. Se tiene que entrenar previamente con un conjunto de datos. Hay dos estrategias principales por las que pueden ser entrenados estos modelos:
 - a) **Skip-gram**: El modelo es entrenado para predecir las palabras alrededor de una misma palabra dada.

- b) **Continuous Bag-of-Words**: El modelo es entrenado para predecir las palabras alrededor de una misma palabra dada.
3. **Enfoque GloVe** [3]: Este enfoque se basa en el mismo principio que el método Word2Vec. La diferencia es que este puede captar relaciones más profundas en las palabras, lo que puede ayudar a asociar palabras raras pero relevantes que podrían salirse del vocabulario que usa un ser humano de manera normal.
 4. **Doc2Vec** [17]: Éste tiene el mismo funcionamiento que el método anterior Word2Vec, sin embargo, este método funciona para frases de texto largas.
 5. **BERT**. Este es un modelo que consiste en redes de codificación y decodificación, y usa el mecanismo de autoatención. Este modelo aprende a representar texto como una secuencia de vectores utilizando aprendizaje auto-supervisado. Se puede ver el ejemplo *SciBert* en [11], o su empleo en las técnicas del modelo *WhoIs?* [4].
 6. **Modelos de Lenguaje Grandes (LLMs) de embedding**: Estos modelos son LLMs que han sido entrenados y programados para fungir únicamente como un codificador, por lo que no devolverán a la salida ningún texto, sino el vector de la incrustación con múltiples dimensiones. Fue una línea futura del trabajo en [22].

3.3 Deep Learning

En la bibliografía se menciona también el uso de técnicas de aprendizaje profundo, y dichas técnicas son cada vez más populares en el campo del AND.

3.3.1 Redes Neuronales (RNs) El uso de redes neuronales se ha vuelto cada vez más popular en muchos campos en los que es implicado el aprendizaje automático de características, ya que tienen la capacidad de aprender características complejas y representaciones abstractas de los datos. [11, 21]

Se tienen varios trabajos basados en el uso de redes neuronales, como por ejemplo, los estudios encontrados en [18], en donde... O también, para uso en el aspecto de la desambiguación para autores con nombres Vietnamitas en [21].

Las RNs más usadas pueden ser:

- **Redes Neuronales Profundas (Deep Neural Networks, DNN)** [21]: este tipo de RNs suele usar más de una capa ocultas en el procesamiento, lo que les permite aprender características complejas.
- **Redes Neuronales Recurrentes (Recurrent Neural Network, RNN)** [10]: Estas redes suelen ser usadas para procesar datos secuenciales.

Las desventajas [6, 19] a la hora de utilizar estos métodos, al igual que los métodos supervisados, son que:

1. Se necesita de una cantidad grande de datos previamente etiquetados de buena calidad para el entrenamiento, así como un buen hardware para el entrenamiento (como una buena tarjeta de video) que requiere menos tiempo de procesamiento, y mucho tiempo para entrenarlas.
2. Otra desventaja de éstas identificada, es que, al enfocarse tanto en las semánticas y las relaciones entre conceptos de los textos, tienden a ignorar campos importantes de la bibliografía, los autores y lugares (metadatos).

3.3.2 Incrustación de Redes : Los modelos de redes se han vuelto un enfoque cada vez más popular en el aspecto del AND. Estos métodos transforman los autores y/o publicaciones en nodos, dentro de un espacio vectorial de baja dimensionalidad, en donde se capturan similitudes semánticas y relaciones en la red. [3]

Existen las **Redes Neuronales Gráficas (Graph Neural Networks, o GNNs)**, por ejemplo HetGNN y su mejora HetGNN with SciBERT (HGNN-S), mencionadas en [11], en donde se incorporan redes neuronales y mecanismos de atención para permitir interacciones profundas entre las incrustaciones iniciales. Pueden existir algunos tipos más específicos de éstas [6]:

1. **Graph Attention Networks (GAT)**: En este tipo de GNNs se aplica un mecanismo de atención multicabezal para modelar la importancia de los nodos vecinos cuando se le agrega información a un nodo objetivo, y las incrustaciones resultantes se fusionan mediante una capa completamente conectada.
2. **Graph Auto Encoders (o Autoencoders Gráficos)**: Diseñadas para aprender representaciones comprimidas y significativas de datos estructurados por medio de grafos. Éstas han sido usadas para refinar representaciones de documentos basadas en atributos de los mismos y las relaciones de autores.

En [21] se desarrolló un sistema de relaciones de entidades a base de métodos gráficos para una base de conocimientos académicos. En [10, 17] se describen varias investigaciones en las que se usan Redes Neuronales Profundas (DNN), dos proyectos llamados DeepWalk y Metapath2Vec.

3.4 Métodos de normalización y preprocesamiento de los datos

Se tiene que tener en cuenta el preprocesamiento de los datos, pues es un proceso fundamental por el que deben pasar los conjuntos para el entrenamiento de métodos, tanto de Machine Learning como Deep Learning.

En la bibliografía se mencionan varios datasets que han sido creados con la intención de funcionar como base para la creación de métodos de desambiguación futuros, como S2AND [20] y AggAND [23].

4. Implementación del proyecto

En la parte práctica de este proyecto se evaluará el uso de distintas técnicas de minería de datos y aprendizaje máquina cuando se dispone de pocos atributos en los registros bibliográficos.

4.1 Análisis previo; datasets usados

Para este trabajo, se usaron tres conjuntos de datos tomados de las fuentes oficiales enlistadas en la sección 2.3. Estos conjuntos de datos contienen información de los registros de autoría.

4.1.1 BNE (dataset bne.pickle) Los campos contenidos dentro de este dataset (no todos los registros tienen estos campos llenos) son:

- Label
- Othernames
- Titles
- Gender
- Coauthors
- URLs
- Dates (tiene que ver con bdate “birth date” y ddate “death date”)

Como a algunos registros les faltan estos campos, se decidió elegir los siguientes campos:

- Label
- Dates
- Titles
- Othernames
- Gender
- y Coauthors

Se descartó el uso del campo URLs, ya que no contiene relevancia para obtener información para probar los métodos de desambiguación.

Este dataset contiene un total de 46750 registros. Los porcentajes de disponibilidad de estos campos pueden verse con respecto a la siguiente imagen:

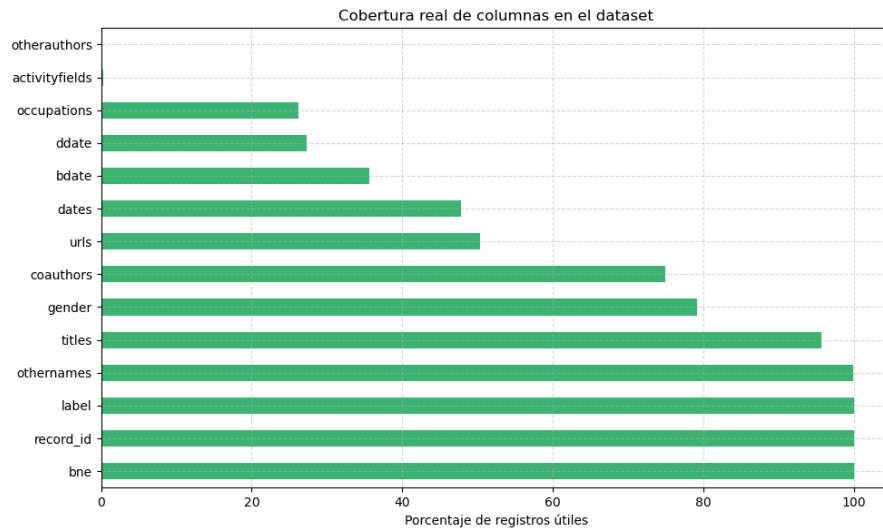


Figura 1: Porcentaje de registros útiles en dataset BNE

4.1.2 Biblioteca General Histórica de la USAL (dataset `nombrs.pickle`)

Este dataset contiene los siguientes campos:

- Clave: “apellido(s), nombre, fnac-fmor”
- Name
- Dates (fecha de nacimiento y fecha de muerte). Estos campos pueden contener “circa” (alrededor de), y floruit (activo en esas fechas)
- Titles
- Coauthors

Este dataset contiene un total de 49614 registros de autores, con sus respectivos títulos y coautores. En la siguiente imagen se pueden apreciar la usabilidad de los campos:

4.1.3 VIAF (dataset `viaf.pickle`)

Este dataset tiene los siguientes campos:

- record_id (que en este caso es el VIAF ID)
- Titles
- Names (este dataset incluye variaciones del nombre mediante el cual se encuentra el autor en los registros)
- Vid (el ID que VIAF le ha asignado al autor)
- Wikipedias (Los enlaces a los artículos de Wikipedia que pertenecen a las obras de este autor)
- Occupations
- Dates
- Coauthors

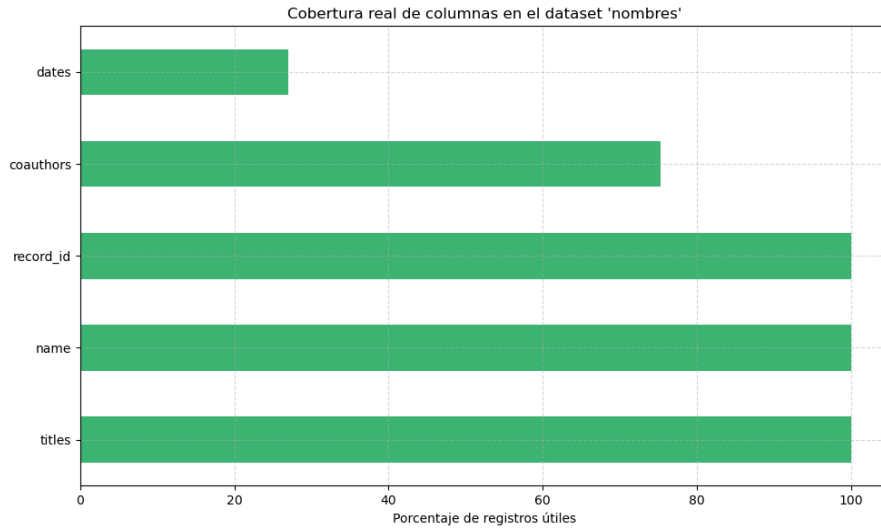


Figura 2: Porcentaje de registros útiles en dataset nombres.pickle

■ FieldOfActivity

Este dataset contiene un total de 407229 registros de autores diferentes, cada uno con su propia lista de títulos y coautores. En la siguiente gráfica se puede ver la usabilidad de cada campo entre todos los registros:

Se puede observar que los principales campos son VID (y su equivalente, Record_ID), names, gender, dates, titles y coauthors.

Dataset	Proveniencia	Total de registros
names.pickle	Biblioteca General Histórica de la Universidad de Salamanca	49614
viaf.pickle	Virtual International Authority File (VIAF)	407229
bne.pickle	Biblioteca Nacional de España	46750

Tabla 1: Descripciones de los datasets

En cada uno de estos datasets, se tienen títulos de obras que vienen en diferentes idiomas. Se encontraron títulos también en alemán, inglés, francés, italiano, griego, ruso, chino, latín, entre otros tantos. Esto indica que contamos con una variedad políglota de registros.

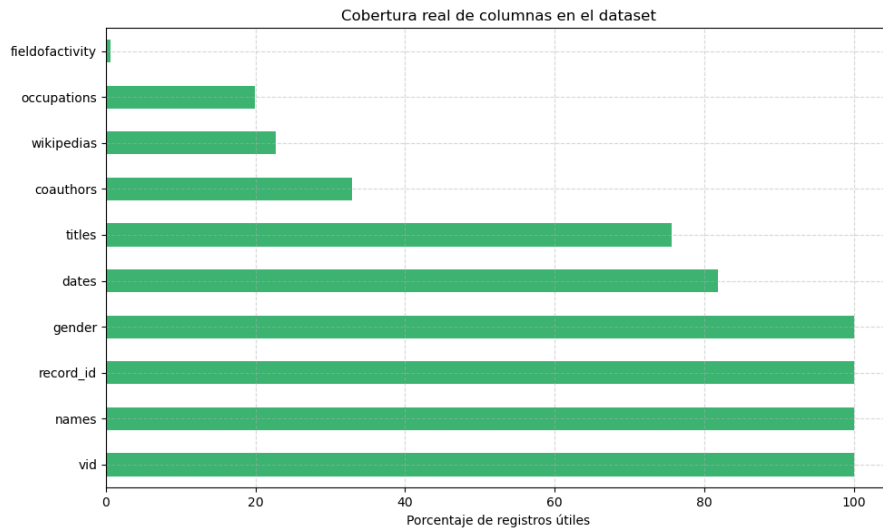


Figura 3: Porcentaje de registros útiles en dataset VIAF

Estos tres datasets también fueron convertidos en formato JSON (Javascript Object Notation). En este formato, se pueden insertar en un sistema de bases de datos de documentos, en el cual se puedan hacer búsquedas por medio de lo que serían “diccionarios” en lenguaje Python, que es, por cada registro, se puede buscar en cada campo. Un ejemplo, para clarificación, puede ser el siguiente registro encontrado en la VIAF:

```
{
  "id": "66049736",
  "source": "VIAF",
  "names": ["Ercolani, Mario", "Mario Ercolani"],
  "gender": "male",
  "dates": "1913:2008",
  "titles": ["I conti finanziari della Banca d'Italia.",
    "Geografia della delinquenza in Italia : nuovi metodi di
    rappresentazione grafica", "Una storia lunga un secolo :
    Lega braccianti CGIL di Mezzano, 1901-2001", "Vita della B.
    Margarita di Savoia, marchesa di Monferrato, detta la
    Grande"],
  "wikipedias": ["https://it.wikipedia.org/wiki/Mario_Erco
    lani"],
  "_version_": 1836613405213982721,
  "_root_": "66049736"
}
```

5. Aplicación

El resultado del análisis tanto de los métodos como de los metadatos que se disponen dió como resultado la idea de juntar 3 métodos de conveniencia para AND de bases de datos con pocos metadatos. Estos 3 métodos tienen enfoques de información distintas, y son los siguientes:

1. Desambiguación de nombres de autor por **puntajes de similitud de cadenas**, como la distancia de Levenshtein, Jaro-Winkler, presencia de subcadenas (Partial Ratio de RapidFuzz), Sequence-Matcher, y Jaccard.
2. **Embeddings en los títulos** de los textos por medio de embeddings con el LLM nomic, y comparación de los vectores resultantes mediante la similitud del coseno.
3. **Redes de autores y coautorías**, para obtener métricas de similitud de coautores.
4. Adicionalmente, también se agregó la **comparación entre los años de vida** de un autor.

A su vez, ya que se tiene una cantidad considerablemente grande de datos, se planteó el uso de dos sistemas de almacenamiento y recuperación de la información. El primero se llama **Apache Solr**. Este se basa en el uso de documentos, con los cuales se puede recuperar la información en diferentes formatos, como JSON, XML, CSVs, entre otros.

El segundo sistema de almacenamiento y recuperación de la información es **ChromaDB**. Este sistema permite el almacenamiento y recuperación rápidos de vectores, con lo cual facilita ampliamente la búsqueda de textos con semánticas similares. La ventaja de este sistema es que devuelve la similitud del coseno de los textos buscados al dar los vectores que menor distancia del coseno tengan³.

5.1 Esquema de funcionamiento

Todo el código para la aplicación se desarrolló en el lenguaje de programación Python. Esta aplicación aglomera el uso de estos métodos descritos, conectándose en un funcionamiento en conjunto con los métodos más adecuados para este estudio. Este programa, combinado con técnicas de análisis rápido por medio de la plataforma Solr proporcionan una forma eficaz de desambiguar autores, que implica el uso de técnicas avanzadas de machine learning y embedded networks.

Un esquema general del funcionamiento de este sistema sería el siguiente:

1. El funcionamiento para importar autores al sistema (Fig. 4)
2. El funcionamiento para las tareas de AND (Fig. 5)

³ Es decir, los vectores de los títulos que más se parezcan.

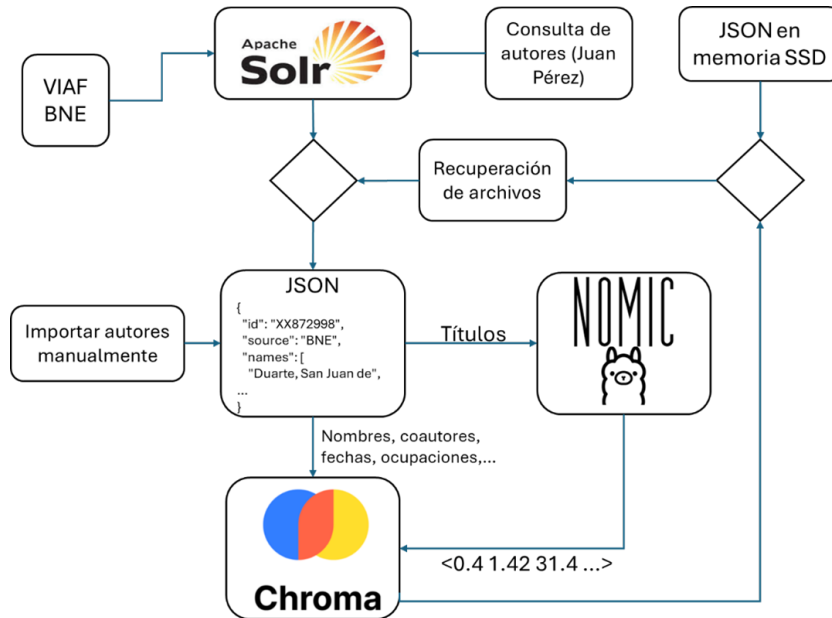


Figura 4: Funcionamiento del programa para importar autores

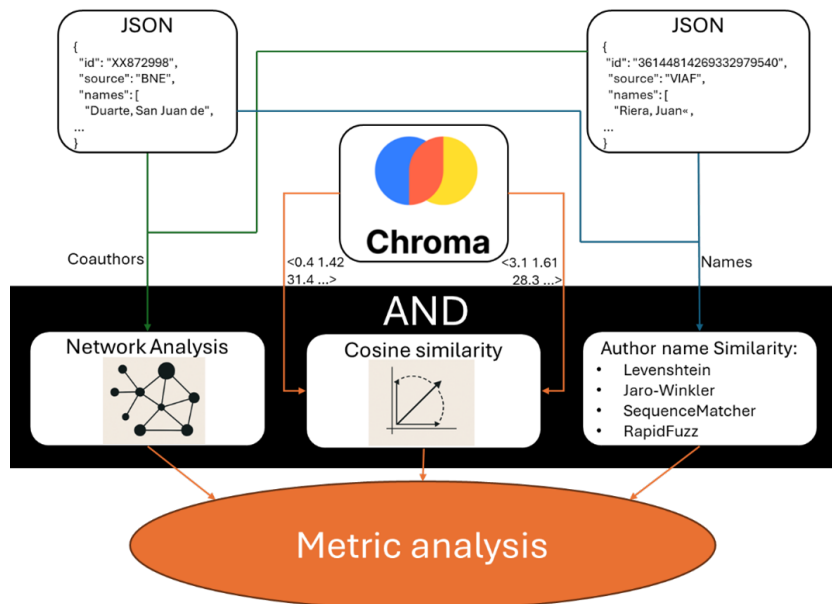


Figura 5: Algoritmo para ejecutar AND

5.2 Estrategia de pruebas con datos reales

Se probaron las herramientas para intentar desambiguar nombres de autores de bases de datos cruzadas, tanto de la VIAF como de la BNE o nombres manuales, y pueden tener una probabilidad alta, media o baja de coincidencia.

Se tomaron ejemplos del dataset `similitud.ods`. Este dataset contiene varios ejemplos en los que se probaron anteriormente métodos de similitudes de cadenas sobre pares de nombres de autores y pares de títulos. Estos ejemplos fueron comparados inicialmente con métodos heurísticos.

La metodología tiene el mismo procedimiento que el algoritmo de este programa, y fue la siguiente:

1. Se buscaron pares de registros en el dataset `similitud.ods`, de cuatro tipos:
 - a) Que se pudieran desambiguar mapeando sus registros entre las bases de datos VIAF o BNE (pudiendo pertenecer al mismo autor en VIAF o BNE dos veces, o que esté combinado entre VIAF y BNE)
 - b) Que se quedaran ambiguos, obteniendo así resultados de métricas que pudieran sugerir similitud, pero sin ser cierta, y
 - c) Que denotaran que no hay parecido entre dos perfiles de autores, llevando a cabo una desambiguación correcta para probar que se trata de dos autores diferentes.
 - d) También se lidió con casos en los que se tuvo que volver a comparar pares de perfiles de autores que se encontraban en el dataset `similitud.ods`, para así probar el AND con las herramientas proporcionadas. Se insertó uno de los dos perfiles siempre, obteniendo el otro de los documentos en Solr.
2. Se mapeaban estos registros en los documentos de Solr por medio del ID o el nombre del autor, para así obtener los perfiles a importar.
3. Se importaba el documento que contenía al perfil del autor al sistema por medio del ID (si existía en las bases de datos), o se insertaba el perfil manualmente con los campos disponibles ⁴.
4. Se analizaban y comparaban los perfiles por similitud de cadenas en los nombres, por semánticas en los títulos y los coautores en común
5. Estos datos se almacenaron en una hoja de datos de Excel, para su análisis manual. Esto con el fin de probar si la herramienta está siendo efectiva o no, comprobar cualquier fallo y plantear áreas de mejora.

El conjunto de pruebas consiste en los siguientes nombres, tomados tanto de BNE como de VIAF y del archivo `similitud.ods`:

Se puede apreciar cómo hay autores que se pueden desambiguar por el hombre de manera rápida al darle un vistazo a un par, y también hay autores que no se pueden desambiguar debido a la falta de datos.

⁴ nombres, títulos, fechas, género, coautores

ID	Nombre	Clase	Clase	Clase	Clase	Clase	Clase
XX142452	Lemus Juan Bautista	BNE	S	9551175	VIAF		
XX079767	Ortega Rubio, Juan	BNE	S	11547645	VIAF		
XX032263	"Rossi, Francesco", "Roberto, Roberto", Bañales García, Misael	BNE	S	3007703	VIAF		
XX042031	"Rico, Francisco", "Río de Guayana, Francisco", "Río, Francisco"	BNE	S	13952701	VIAF		
XX041408	"Rico, Francisco", "Río de Guayana, Francisco", "Río, Francisco"	BNE	S	8827547	VIAF		
XX028760	"Rico, Francisco", "Río de Guayana, Francisco", "Río, Francisco"	BNE	S	8827547	VIAF		
7627670	Rosendo, Wilfredo	VIAF	S	20491597	VIAF		
XX027813	"Rosendo y Rodríguez de la Peña, Rodríguez, Rosendo, José"	BNE	S	11373659	VIAF		
5602456	"Quintero y Villegas, F. G. de 1580-1645 Francisco Gómez de "Manojo o Benavente, G. en 1580-1645 Boparcoscaze"	VIAF	S	XX039651	BNE		
70135212	"Qui, Alexander, 1936-1989"	VIAF	S	XX037703	BNE		
XX022523	"Qui, Alexander, 1936-1989"	VIAF	S	XX037703	BNE		
XX048620	"Quijano Riederer, J. Esteban, Desiderio" "Quijano de Riederer"	BNE	S	8767386	VIAF		
92340378	"Quijano Riederer, Esteban" "Quijano Riederer, Desiderio"	BNE	S	XX037703	BNE		
XX0109178	Quijano López, Ángel	BNE	S	8633823	VIAF		
836770	Quijano, Teresa 1715-1764	VIAF	S	manusc:24863-3464 486-a-3b-a-3b-a-3b-a-746	manusc		
19783335	"María del Pilar de Torres Luna" "Torres Luna, María Pilar de 1937" "	VIAF	S	XX038447	BNE		
XX024500	"Quilón Zamora, Pedro del" "Miguel Zamora Estroza, Pedro"	VIAF	S	10238695	VIAF		
9093223	"Quirós, Jaime, 1907-1991", "Quirós, Jaime"	VIAF	S	XX032923	BNE		
51848459	"Quirós, Francisco María, 1834-1778" "Quirós, Francisco"	VIAF	S	3692546	VIAF		
XX0311518	"Quirós Sánchez, Juan Francisco" "Quirós"	BNE	S	8790989	VIAF		
XX072915	"Quirós, Nicolás"	BNE	S	8603784	VIAF		
18067443	"Quirós, Blas" "Fabricius, Blasius"	BNE	Posible	00046282	BNE		

Figura 6: Conjunto de autores que pasaron por el proceso de desambiguación en el programa desarrollado, junto con el ID del autor comparado

5.3 Experimento

Se tomó una muestra de 50 pares de nombres a desambiguar; algunos estaban presentes en VIAF, otros en BNE y hubieron unos cuantos que se agregaron manualmente.

En la Fig. 8 y 9 se pueden ver algunos ejemplos de las pruebas realizadas con la herramienta; se puede ver en él cómo se lograron identificar 15 autores que tenían otro registro asociado en la base de datos opuesta, como se puede ver en las búsquedas 1-3, 7, 11, 13-18, 20-33.

Los métodos de similitud de cadenas probaron tener un buen grado de efectividad, pues daban un primer indicio de similitud. Habían nombres que estaban exactamente igual en ambos registros, por lo que solo se tomaba en cuenta que eran registros coincidentes, y en la hoja de datos se ponía el siguiente par de nombres con un puntaje que no fuera similitud 1, de esta manera se pueden apreciar las métricas. Como por ejemplo, entre los IDs XX5557283 y 5007703:

Rossi, Francesco = Rossi, Franco
 Razón: Partial Ratio alto (>0.8)
 Levenshtein: 3
 Jaro-Winkler: 0.960
 Ratio: 0.897
 Partial Ratio: 0.957
 SequenceMatcher: 0.889
 Jaccard: 0.333

Se logró hacer que esta estrategia fuera un poco más robusta al tener varias métricas de similitud, para así poder ver comparativas entre pares de nombres como la siguiente, en la cual el Partial Ratio y la Jaro-Winkler dan un puntaje alto:

21384621	"Rosendo Gali et sa familia, 1920", "Rosendo Gali, Rosendo", 1884-1884"	VIAF	Posible	0087959	BNE		
36925746	"Rosendo, 1884-1778", "Rosendo 1884-1778", "Rosendo F. M. A. de 1884-1778", "Rosendo, Francisco de", "Rosendo, Francisco M. de 1884-1778", "Rosendo López, M. de Carme"	VIAF	Posible	2374678038903772095	VIAF		
57841910	"Rosendo López, M. de Carme"	VIAF	Posible	XX037746	BNE		
7130160676	"Rosendo, Gerónimo de"	VIAF	Posible	7130160676	VIAF		
4303560306	"Rosendo, Gerónimo de"	VIAF	Posible	4303560306	VIAF		
4303560306	"Rosendo, Gerónimo de"	VIAF	Posible	4303560306	VIAF		
70217784	Rosendo Pallas, José María	VIAF	Posible	46102279	VIAF		
XX028415	Rosendo, Marcos Felipe	BNE	Posible	46102279	VIAF		
XX0482375	Rosendo, Marcos Felipe	BNE	Posible	23746777	VIAF		
6501994746	"Rosendo, Marcos Felipe"	VIAF	Posible	4303560306	VIAF		
5271678013	"Rosendo, Marcos Felipe Anselmi de, 1884-1778"	VIAF	Posible	51584369	VIAF		
86177269	Rosendo, Juan, de la familia XV	VIAF	Posible	90018144	VIAF		
521184400	"Rosendo, Juan, de la familia XV"	VIAF	Posible	8827547	VIAF		
3478410000	"Rosendo, Juan, de la familia XV"	VIAF	Posible	00844468	BNE		
21146939	"Rosendo, Juan, de la familia XV"	VIAF	Posible	23746777	VIAF		
80277547	"Rosendo, Juan, de la familia XV"	VIAF	Posible	00844468	BNE		
405883	Artaño, Francisco de, active 18th century	VIAF	No	23925258	BNE		
28609158	Artaño, Francisco de, active 18th century	VIAF	No	XX049888	VIAF		
115819998	"Artaño, Francisco de, active 18th century"	VIAF	No	23925258	BNE		
14731480	"Artaño, Francisco de, active 18th century"	VIAF	No	115819998	VIAF		
1512493328	Artaño, Francisco de, active 18th century	VIAF	No	4878-3868-000004005	manusc		
8851301374	Artaño, Francisco de, active 18th century	VIAF	No	-	-		
5128170836	"Artaño, Francisco de, active 18th century"	VIAF	No	-	-		
4801840000	"Artaño, Francisco de, active 18th century"	VIAF	No	-	-		
521184400	Artaño, Francisco de, active 18th century	VIAF	No	-	-		
3478410000	Artaño, Francisco de, active 18th century	VIAF	No	-	-		
1512493328	Artaño, Francisco de, active 18th century	VIAF	No	-	-		
8851301374	Artaño, Francisco de, active 18th century	VIAF	No	-	-		
115819998	Artaño, Francisco de, active 18th century	VIAF	No	-	-		
86503729	Artaño, Francisco de, active 18th century	VIAF	No	-	-		

Figura 7: Continuación de 6

Rossi, Francesco = Rossi, Francesco de', 1503-environ 1574
 Razón: Partial Ratio alto (>0.8)
 Levenshtein: 20
 Jaro-Winkler: 0.886
 Ratio: 0.582
 Partial Ratio: 1.000
 SequenceMatcher: 0.600
 Jaccard: 0.400

No se disponían de reglas con las que se pudiera distinguir automáticamente entre una coincidencia alta o baja (que era el propósito de agregar una razón por la cual mostrar el resultado); esto lo hacía el propio usuario con su criterio.

En las pruebas, se puede ver cómo el método de comparación por embeddings tiene un muy buen desempeño, pues hay títulos que son idénticos, y éste da un score =1. Incluso hay algunos que difieren por algún carácter, pero dan una coincidencia por arriba de 0.9, como en el siguiente ejemplo de comparativa entre XX1148499 y 86881785:

Discvrso apologetico y alegación en derecho en favor de los Collegiales del mayor del Arçobispado = Discvrso apologetico y alegación en derecho en favor de los Collegiales del mayor del Arçobispo (0.988)

Y hay algunos títulos que denotan parecido en el contexto, como este otro ejemplo:

Secunda pars. Tractatus in materia equorum = Tertia pars. Tractatus in materia equorum (0.932)

E incluso si alguna cadena tiene más texto, las predicciones de similitud se acercan mucho:

Historia de Principi di Este = Historia de principi di Este... primo volume nel quale si contengono congiuntamente le cose principali dalla rivolutione del Romano Imp. in fino al MCCCCLXXVI. (0.809)

Al evaluar el funcionamiento de la red de coautores se pudo apreciar que el formato en los nombres de los datasets no estaba estandarizado. Por ejemplo, en el autor Fabricius, Blasius [ID de VIAF 196097443], el autor "Łaski, Jan (1499-1560)" se encontraba en la BNE como "Łaski, Jan"; con este ejemplo se aprecia cómo los nombres pueden variar ligeramente tanto en caracteres (Ł vs L), como en la inclusión de las fechas (1499-1560), o también en el orden [apellido, nombre] o [nombre, apellido]. Por estas diferencias, no se pudo hacer la asociación de coautores en algunos casos, y por lo tanto el método fallaba.

Con algunos autores se encontró que las fechas de nacimiento / muerte estaban en un formato que no coincidía; por ejemplo, en el

#	ID1	Nombre autor 1	DS	Nombre autor 2	ID2	DS2	mbres más similitud	Similitud	Fecha 1	Fecha 2	Coincidencia	Títulos y similitud	Similares	MATCH / NO MATCH	
1	XX1149499	Larraz, Juan Bautista	BNE	"Larraz, Juan Bautista, 1649-1649"; "Larraz, Juan Bautista, m. 1649"; "Juan Bautista Larraz"; "Larraz, Juan Bautista, Spania 1649-1649"; "Larraz, Juan Bautista, c. 1649"	86981785	VIAF	Larraz, Juan Bautista; Larraz, Juan Bautista, 1645	Reason: Partial Ratio alto (0.8) Levenshtein: 3 Jaro-Winkler: 0.905 Part Ratio: 0.792 Part Ratio: 0.500 SequenceMatch: 0.289 Jaccard: 0.300	m. 1645	1589-1645	COINCIDENCIA ALTA - Los análisis de fechas solapan	Títulos y similitud Top 5 pares con similitud mayor a 1.000: 1. Rodríguez de la Peña, H. (1886-1892) 2. Rodríguez de la Peña, H. (1886-1892) 3. Rodríguez de la Peña, H. (1886-1892) 4. Rodríguez de la Peña, H. (1886-1892) 5. Rodríguez de la Peña, H. (1886-1892)	(Vista de Juan Bautista Larraz)	MATCH	
2	XX0878787	"Ortega Rubio, Juan"; "Ortega y Rubio, Juan"	BNE	"Ortega Rubio, Juan, 1845-1921"; "Ortega Rubio, Juan, Rubio Historien español"; "Ortega Rubio, Juan"; "Ortega Rubio, Juan, 1845"	115479645	VIAF	Ortega Rubio, Juan; Ortega Rubio, Juan	Levenshtein: 0 Jaro-Winkler: 1.000 Part Ratio: 1.000 Part Ratio: 1.000 SequenceMatch: 1.000	1845, 1921	1845, 1921	COINCIDENCIA EXACTA - Las fechas son idénticas	Títulos y similitud Top 5 pares con similitud mayor a 1.000: 1. Rodríguez de la Peña, H. (1886-1892) 2. Rodríguez de la Peña, H. (1886-1892) 3. Rodríguez de la Peña, H. (1886-1892) 4. Rodríguez de la Peña, H. (1886-1892) 5. Rodríguez de la Peña, H. (1886-1892)	(Rodríguez de la Peña, H.)	MATCH	
3	XX5557383	"Rossi, Francisco"; "Rubus, Roscius"	BNE	"Rossi, Francisco del 1503-aproximada 1574"; "Rossi, Francisco, 1503-1574"; "Rossi, Francisco"; "Rossi, Francisco, 1524-1574"; "Francisco Rossi"; "Rossi, Francisco, 1503-1574"; "Rossi, Francisco de (1503-c. 1574)"; "Rossi, Francisco del 1503-circa 1574"	SE+06	VIAF	Rossi, Francisco; Rossi, Francisco	Levenshtein: 3 Jaro-Winkler: 0.960 Part Ratio: 0.897 Part Ratio: 0.957 SequenceMatch: 0.889	1	1521-1574	1503-1574	COINCIDENCIA ALTA - Los rangos de fechas se solapan	Títulos y similitud Top 5 pares con similitud mayor a 1.000: 1. Rodríguez de la Peña, H. (1886-1892) 2. Rodríguez de la Peña, H. (1886-1892) 3. Rodríguez de la Peña, H. (1886-1892) 4. Rodríguez de la Peña, H. (1886-1892) 5. Rodríguez de la Peña, H. (1886-1892)	(Francisco Rossi)	MATCH
4	333116804	"Barber, Jean, activo c. XV"	VIAF	"Barber, Jean, c. 1496-1515"; "Jean Barber francouzais y videtur, c. 1496-1515"; "1496 y Londyn, v letech 1498-1499 v Westmínsku a letech 1502-1515 v sístě"	9061814	VIAF	Barber, Jean, activo c. XV; Barber, Jean, c. 1496-1515	Levenshtein: 13 Jaro-Winkler: 0.837 Part Ratio: 0.576 Part Ratio: 0.214 SequenceMatch: 0.556	S.VX	1496-1515	Fechas parciales con campos probabilidad baja	No hay títulos	0	NO MATCH	
5.1	231949335	"Riz, Francisco (Spanish painter and scenographer, 1614-1685)"; "Riz, Francisco (1608-1685)"; "Francisco Riz"	VIAF	"Riz, Francisco"; "Riz de Quevedo, Francisco"; "Riz, Francisco"	XX941408	BNE	Riz, Francisco; Riz, Francisco	Levenshtein: 9 Jaro-Winkler: 0.922 Part Ratio: 0.790 Part Ratio: 1.000 SequenceMatch: 0.757	[1608, 1685]	[1614-1685]	COINCIDENCIA ALTA - Los rangos de fechas se solapan	No hay títulos	0	NO MATCH	
5.2	231949335	"Riz, Francisco (Spanish painter and scenographer, 1614-1685)"; "Riz, Francisco (1608-1685)"; "Francisco Riz"	VIAF	"Riz, Francisco, 1614-1685"; "Francisco Riz pintor barroco español"	88277547	VIAF	Riz, Francisco; Riz, Francisco	Levenshtein: 2 Jaro-Winkler: 0.956 Part Ratio: 0.902 Part Ratio: 0.913 SequenceMatch: 0.913	[1608, 1685]	[1614-1685]	COINCIDENCIA ALTA - Los rangos de fechas se solapan	No hay títulos	0	NO MATCH	
5.3	XX941408	"Riz, Francisco"; "Riz de Quevedo, Francisco"; "Riz, Francisco"	BNE	"Riz, Francisco, 1614-1685"; "Francisco Riz pintor barroco español"	88277547	VIAF	Riz, Francisco; Riz, Francisco	Reason: Partial Ratio alto (0.8) Levenshtein: 2 Jaro-Winkler: 0.956 Part Ratio: 0.902 Part Ratio: 0.913 SequenceMatch: 0.913	1614, 1685	1614, 1685	MATCH EXACTO	No hay títulos	0	MATCH	
5.4	XX941408	"Riz, Francisco"; "Riz de Quevedo, Francisco"; "Riz, Francisco"	BNE	"Riz, Francisco"	XX5287650	BNE	Riz, Francisco; Riz, Francisco	Reason: Partial Ratio alto (0.8) Levenshtein: 0 Jaro-Winkler: 1.000 Part Ratio: 1.000 Part Ratio: 1.000 SequenceMatch: 1.000	1614, 1685	1614, 1685	MATCH EXACTO	No hay títulos	0	NO MATCH	
5.5	231949335	"Riz, Francisco (Spanish painter and scenographer, 1614-1685)"; "Riz, Francisco (1608-1685)"; "Francisco Riz"	VIAF	"Riz, Francisco"	XX5287650	BNE	Riz, Francisco; Riz, Francisco	Reason: Partial Ratio alto (0.8) Levenshtein: 9 Jaro-Winkler: 0.870 Part Ratio: 0.792 Part Ratio: 1.000 SequenceMatch: 0.767 Jaccard: 0.260	[1608, 1685]	[1614-1685]	COINCIDENCIA ALTA - Los rangos de fechas se solapan	No hay títulos	0	NO MATCH	
5.6	88277547	"Riz, Francisco, 1614-1685"; "Francisco Riz pintor barroco español"	VIAF	"Riz, Francisco"	XX5287650	BNE	Riz, Francisco; Riz, Francisco	Reason: Partial Ratio alto (0.8) Levenshtein: 2 Jaro-Winkler: 0.956 Part Ratio: 0.902 Part Ratio: 0.913 SequenceMatch: 0.913	1614, 1685	1614, 1685	MATCH EXACTO	No hay títulos	0	MATCH	
6	591989471	"Hoffmann, Johannes, 1369-1451"	VIAF	"Johann v Biskup Michl: 1375-1451"; "Johann v Biskup Michl, Bischof von Prag 1375-1451"; "Hoffmann, Johannes, 1451-1451"; "Hoffmann, Johannes, v. c. d. Meissen, m. 1451"; "Hoffmann, Johannes v. c. d. Meissen, m. 1451"	47699038	VIAF	Hoffmann, Johannes, 1369-1451; Hoffmann, Johannes, v. c. d. Meissen, m. 1451	Reason: Partial Ratio alto (0.8) Levenshtein: 18 Jaro-Winkler: 0.870 Part Ratio: 0.649 Part Ratio: 0.818 SequenceMatch: 0.667 Jaccard: 0.250	1375, 1451	1369, 1451	COINCIDENCIA ALTA - Los rangos de fechas se solapan	No hay títulos	0	NO MATCH	
7	XX048620	"Tramada de Reitorado"; "Tramada de Reitorado"; "Tramada de Reitorado"; "Tramada de Reitorado"; "Tramada de Reitorado"; "Tramada de Reitorado"; "Tramada de Reitorado"; "Tramada de Reitorado"; "Tramada de Reitorado"; "Tramada de Reitorado"	BNE	"Tramada de Reitorado"; "Tramada de Reitorado"; "Tramada de Reitorado"; "Tramada de Reitorado"; "Tramada de Reitorado"; "Tramada de Reitorado"; "Tramada de Reitorado"; "Tramada de Reitorado"; "Tramada de Reitorado"; "Tramada de Reitorado"	6767396	VIAF	Tramada de Reitorado; Tramada de Reitorado	Reason: Partial Ratio alto (0.8) Levenshtein: 0 Jaro-Winkler: 1.000 Part Ratio: 1.000 Part Ratio: 1.000 SequenceMatch: 1.000	1467, 1530	1469-1530	ALTA	Títulos con similitud 1.000 (Total 21) Top 5 pares con similitud mayor a 1.000: 1. Rodríguez de la Peña, H. (1886-1892) 2. Rodríguez de la Peña, H. (1886-1892) 3. Rodríguez de la Peña, H. (1886-1892) 4. Rodríguez de la Peña, H. (1886-1892) 5. Rodríguez de la Peña, H. (1886-1892)	(Tramada de Reitorado)	MATCH	
8	XX045201	"Bañuelos García, Misael"; "Bañuelos, Misael"	BNE	"Bañuelos, Misael, 1887-1954"; "Misael Bañuelos"; "Misael Bañuelos García"; "Bañuelos, M."	13952701	VIAF	Bañuelos, Misael; Bañuelos, Misael	Reason: Partial Ratio alto (0.8) Levenshtein: 0 Jaro-Winkler: 0.933 Part Ratio: 0.786 Part Ratio: 1.000	1887, 1954	1887, 1954	MATCH EXACTO	Títulos con similitud 1.000 (Total 10) Top 5 pares con similitud mayor a 1.000: 1. Rodríguez de la Peña, H. (1886-1892) 2. Rodríguez de la Peña, H. (1886-1892) 3. Rodríguez de la Peña, H. (1886-1892) 4. Rodríguez de la Peña, H. (1886-1892) 5. Rodríguez de la Peña, H. (1886-1892)	(Misael Bañuelos)	MATCH	
9	78676373	"Henrich, Wilhelm, 1887-1972"; "Henrich, Wilhelm, 1887-1972"; "Henrich, Wilhelm"; "Henrich, Wilhelm"; "Henrich, Wilhelm (Gothmann)"	VIAF	"Henrich, Wilhelm, 1887-1972"; "Henrich, Wilhelm, 1887-1972"; "Henrich, Wilhelm"; "Henrich, Wilhelm"; "Henrich, Wilhelm (Gothmann)"	204911897	VIAF	Henrich, Wilhelm; Henrich, Wilhelm	Reason: Partial Ratio alto (0.8) Levenshtein: 0 Jaro-Winkler: 1.000 Part Ratio: 0.971 Part Ratio: 1.000 SequenceMatch: 1.000 Jaccard: 0.333	1887-1972	1887-1972	INFORMACION INSUFICIENTE - Solo uno de los perfiles tiene fechas disponibles	Títulos con similitud 1.000 (Total 10) Top 5 pares con similitud mayor a 1.000: 1. Rodríguez de la Peña, H. (1886-1892) 2. Rodríguez de la Peña, H. (1886-1892) 3. Rodríguez de la Peña, H. (1886-1892) 4. Rodríguez de la Peña, H. (1886-1892) 5. Rodríguez de la Peña, H. (1886-1892)	(Henrich, Wilhelm)	MATCH	
10	XX027813	"Rodríguez de la Peña, Hipólito"; "Rodríguez de la Peña, H."; "Rodríguez de la Peña, H."; "Rodríguez de la Peña, H."	BNE	"Rodríguez de la Peña, H., 1886-1892"; "Rodríguez de la Peña, H., 1886-1892"; "Rodríguez de la Peña, H., 1886-1892"; "Rodríguez de la Peña, H., 1886-1892"	113776359	VIAF	Rodríguez de la Peña, H.; Rodríguez de la Peña, H.	Reason: Partial Ratio alto (0.8) Levenshtein: 0 Jaro-Winkler: 0.963 Part Ratio: 0.892 Part Ratio: 1.000	[1886, 1952]	[1886, 1952]	COINCIDENCIA EXACTA - Las fechas son idénticas	Títulos con similitud 1.000 (Total 10) Top 5 pares con similitud mayor a 1.000: 1. Rodríguez de la Peña, H. (1886-1892) 2. Rodríguez de la Peña, H. (1886-1892) 3. Rodríguez de la Peña, H. (1886-1892) 4. Rodríguez de la Peña, H. (1886-1892) 5. Rodríguez de la Peña, H. (1886-1892)	(Rodríguez de la Peña, H.)	MATCH	
11	56604566	"Queredo y Villacorta, Francisco Gómez de 1580-1645"; "Queredo y Villacorta, Francisco Gómez de 1580-1645"; "Queredo y Villacorta, Francisco Gómez de 1580-1645"; "Queredo y Villacorta, Francisco Gómez de 1580-1645"	VIAF	"Queredo y Villacorta, Francisco Gómez de 1580-1645"; "Queredo y Villacorta, Francisco Gómez de 1580-1645"; "Queredo y Villacorta, Francisco Gómez de 1580-1645"; "Queredo y Villacorta, Francisco Gómez de 1580-1645"	XX1066651	BNE	Queredo y Villacorta, Francisco Gómez de 1580-1645; Queredo y Villacorta, Francisco Gómez de 1580-1645	Reason: Partial Ratio alto (0.8) Levenshtein: 9 Jaro-Winkler: 0.938 Part Ratio: 0.892 Part Ratio: 1.000 SequenceMatch: 0.938	1580, 1645	1580, 1645	COINCIDENCIA EXACTA - Las fechas son idénticas	Títulos con similitud 1.000 (Total 10) Top 5 pares con similitud mayor a 1.000: 1. Rodríguez de la Peña, H. (1886-1892) 2. Rodríguez de la Peña, H. (1886-1892) 3. Rodríguez de la Peña, H. (1886-1892) 4. Rodríguez de la Peña, H. (1886-1892) 5. Rodríguez de la Peña, H. (1886-1892)	(Queredo y Villacorta, Francisco Gómez de 1580-1645)	MATCH	

Figura 8: Parte de las pruebas realizadas a los autores en la fig 9-10

autor Fabricius, Blaise, s. XVI, y su registro en la VIAF que tenían formatos diversos:

- Fabricius, Blasius, -1577
- Fabricius, Blaise, 15.-15..., imprimeur
- Fabricius, Blasius (15.-1577).

Esto hacía difícil el reconocimiento de coincidencias de fechas, a pesar de que ambos perfiles claramente se sitúan en la misma época.

Hubieron algunos casos en los que las fechas eran inexactas y se tenían muchos rangos, como por ejemplo al comparar los dos registros XX5557283 de BNE y 5007703 de VIAF, en donde en BNE se tiene fl. 1521:1574, y en VIAF 1503:1574 pero los nombres contienen más fechas posibles del autor. Esto demuestra que hay múltiples fechas para los autores implícitas en el dataset dentro de los nombres.

Existieron casos en los que no se podía asegurar la época del autor debido a ambigüedades y contradicciones en las fechas que demostraban los registros en el campo “dates” y las fechas contenidas en los nombres, como por ejemplo en el registro 313230543, en donde en el campo “dates se tenía “1942:”, pero en una variante del nombre de autor se tenía “*Barbier, Jean active 15th century*”:

```
{
  "id": "313230543",
  "source": "VIAF",
  "names": ["Barbier, Jean 1942-....", "Barbier, Jean active 15th
  century", "Barbier, Jean, 1942-...., géologue"],
  "gender": "male",
  "dates": "1942:",
  "titles": ["Atlas départemental", "Conséquences géochimiques
  ... Lauriol, Isabelle.", "Boubounelle,
  Nicole."],
  "_version_": 1836613417880780823,
  "_root_": "313230543"
}
```

Se pudo probar la efectividad del LLM *nomie* a la hora de hacer la incrustación de texto en lenguajes muy diversos, como en el registro 115819998 que contiene un nombre japonés *Ezaki, Josefina Keiko*, como se muestra en la Fig. 10. Se pretendió desambiguar este nombre al buscar nombres similares con los mismos caracteres en japonés, y se importaron otros 3 registros, los cuales contienen títulos en alfabeto japonés: 253827815, 1636913 y 115915859⁵

Para el mismo ejemplo anterior, si se compara entre pares de nombres con caracteres en japonés y caracteres en español, se puede ver cómo se tienen métricas de distancia. Buscando la traducción de los nombres de caracteres en japonés en Google Translate, en Fig. 11.

⁵ Todos estos registros vienen de VIAF.

```

{
  "id": "115915859",
  "source": "VIAF",
  "names": ["Minami, Keiko, 1911-2004", "南, 桂子, 1911-2004", "南桂子"],
  "gender": "female",
  "dates": "1911:2004",
  "titles": ["arbre de paix carte de vœux pour l'UNICEF, ...", "Yozo
Hamaguchi et Keiko Minami", "南桂子の世界 : 空・鳥・水...", "南桂子全版画作品集", "
南桂子展 = Keiko Minami retrospective exhibition : 生誕100年", "船の旅 : 詩と
童話と銅版画 : 南桂子の世界", "銅版画家南桂子 : スルヘンの小さな王国へ"],
  "coauthors": ["Galerie Berggruen Paris", "Atelier Georges
Leblanc", "Imperial hotel", "Hamaguchi, Yōzō, 1909-2000.", "浜口, 陽三 1909-
2000", "高岡市美術館", "美津島, 徳蔵", "中山, 公男 1927-2008", "ミュージアム三ツツヤサロ
ン", "練馬区立美術館.", "小磯記念美術館 (神戸市立)", "Galerie Gerda
Bassenge"],
  "wikipedias": ["https://ar.wikipedia.org/wiki/مينامى_كيكو", "https://en.wik
ipedia.org/wiki/Keiko_Minami", "https://ja.wikipedia.org/wiki/南桂子"],
  ...
  {
    "id": "253827815",
    "source": "VIAF",
    "names": ["江崎, 清"],
    "titles": ["Shajō jinbutsu satsuei nyūmon", "寫場人物撮影入門"],
    "_version_": "1836613406586568719",
    "_root_": "253827815"
  }
  {
    "id": "1636913",
    "source": "VIAF",
    "names": ["Hirohashi, Keiko", "広橋, 桂子, 1931-", "広橋, 桂子"],
    "dates": "1931:",
    "titles": ["Ki no hako.", "Wood package, c1987", "木の箱"],
    "occupations": ["designers"],
    "_version_": "1836613404775677962",
    "_root_": "1636913"
  }
}

```

Figure 10: JSON con registro que contiene caracteres en japonés

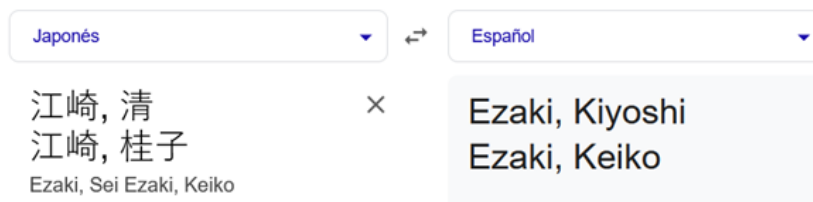


Figure 11: Muestra de la traducción en Google Translate y la transliteración de japonés al alfabeto español

6. Conclusión

La desambiguación de nombres de autor es un campo de investigación muy amplio que se ha estado desarrollando de manera muy rápida con el paso del tiempo y la aparición de nuevas tecnologías disponibles. Se pueden lograr desarrollar nuevas herramientas que sean cada vez más *inteligentes*, y que puedan facilitar el trabajo de desambiguación, hasta incluso hacerla automática.

Las cuatro estrategias empleadas en la desambiguación probaron ser efectivas; se pudieron obtener algunos autores desambiguados entre dos bases de datos, e incluso se pudo identificar que habían autores duplicados en la base de datos de VIAF.

La herramienta es muy útil en cuanto a búsquedas de asignación de autores, pero para agrupamiento se necesita extraer conocimiento previo de los autores, y ésta no emplea técnicas de agrupamiento masivos, por lo que solo es recomendable usarla si se pretende ya tener un autor en mente a desambiguar.

Se recomienda mucho el uso de herramientas políglotas, para poder tener tolerancia ante la transliteración de textos.

7. Líneas futuras de investigación y desarrollo

Existen áreas de mejora para este proyecto, puesto a que había una fecha límite para entregar el trabajo, no se pudieron experimentar lo suficiente y probar con varios métodos para elegir cuál es el mejor.

Para la siguiente etapa de funcionamiento del proyecto en este trabajo, se podría proponer mejorar el algoritmo entrenando a un clasificador una vez que se disponga de varios resultados de las pruebas hechas, para automatizar todo el flujo de trabajo.

Se puede plantear hacer más robusto el módulo de la red de coautores, como por ejemplo tolerar nombres similares estableciendo un umbral de similitud de cadenas (J. Carlos, Juan Carlos)⁶.

Se pueden crear *embedded networks* que contengan tanto los coautores como las semánticas de los títulos, de esta manera se pueden comparar los coautores y enlazar por otro método que no sea solo el nombre, de esta manera construir una red que tenga la capacidad *borrosa*.

También es conveniente depurar los datasets usados para asegurar la efectiva separación de todos los campos, o implementar técnicas de extracción de información en, por ejemplo los nombres con las fechas.

Hay escenarios en donde se tienen muchos títulos para comparar en un par de registros. En estos casos, para optimizar el rendimiento del sistema, sería conveniente plantear que al encontrar un título que

⁶ Nombre hipotético

coincida completamente con otro⁷, entonces que ese título ya no sea objeto de comparación con los demás.

Referencias

1. Arif, T.: Exploring The Use Of Hybrid Similarity Measure For Author Name Disambiguation. *INTERNATIONAL JOURNAL OF SCIENTIFIC & TECHNOLOGY RESEARCH* **4** (2015). https://www.researchgate.net/publication/287473674_Exploring_The_Use_Of_Hybrid_Similarity_Measure_For_Author_Name_Disambiguation
2. Beall, J.: Metadata for Name Disambiguation and Collocation. *Future Internet* (2010). <https://doi.org/10.3390/fi2010001>
3. Boukhers, Z., Bahubali, N.A.: Deep author name disambiguation using DBLP data. *Int. J. Digit. Libr.* (2023). <https://doi.org/10.1007/s00799-023-00361-6>
4. Boukhers, Z., Bahubali, N.A.: Whois? Deep Author Name Disambiguation using Bibliographic Data. In: 26th Int. Conf. on Theory and Practice of Digital Libraries, Koblenz, Germany (2022). <https://doi.org/10.48550/arXiv.2207.04772>
5. Boukhers, Z., Bahubali, N., Chandrasekaran, A.T., Anand, A., Prasad, S.M.G., Aralappa, S.: Bib2Auth: Deep Learning Approach for Author Disambiguation using Bibliographic Data. (2021). <https://doi.org/10.48550/arXiv.2107.04382>
6. Capelli, F., Colavizza, G., Peroni, S.: Recent Developments in Deep Learning-based Author Name Disambiguation. In: 21st Conf. on Information and Research Science Connecting to Digital and Library Science, Udine, Italy (2024)
7. Daud, A., Amjad, T., Shoaib, M.: Author Name Disambiguation in Bibliographic Databases: A Survey. **2**(9), 87–110 (2021). <https://doi.org/10.1111/RpJC.2020.DOINumber>
8. De Bonis, M., Falchi, F., Manghi, P.: Graph-based methods for Author Name Disambiguation: a survey. *PeerJ Comput. Sci.* **9** (2023). <https://doi.org/10.7717/peerj-cs.1536>
9. Ferreira, A.A., Gonçalves, M.A., Laender, A.H.: A brief survey of automatic Methods for Author Name Disambiguation. *SIGMOD Record* (2012). <https://doi.org/10.1145/2350036.2350040>
10. Wang, G., Hu, W., Silva, F.N.: Author name disambiguation based on heterogeneous graph neural network. *Indiana University* (2025). <https://doi.org/10.1371/journal.pone.0310992>
11. Huang, Z., Zhang, H., Hao, C., Yang, H., Wu, H.: A cross-domain transfer learning model for author name disambiguation on heterogeneous graph with pretrained language model. *Knowl.-Based Syst.* **305** (2024). <https://doi.org/10.1016/j.knosys.2024.112624>
12. Jatnika, D., Bijaksana, M.A., Suryani, A.A.: Word2Vec Model Analysis for Semantic Similarities in English Words. Elsevier (2019) 160–167
13. Kim, J., Kim, J.: ANDez: An open-source tool for author name disambiguation using machine learning. *SoftwareX* **26** (2024). <https://doi.org/10.1016/j.softx.2024.101719>

⁷ Es decir, con un score 1 o muy cercano

14. Sanyal, D.K., Bhowmick, P.K., Das, P.P.: A review of author name disambiguation techniques for the PubMed bibliographic database. *J. Inf. Sci.* (2019) 1–28. <https://doi.org/10.1177/0165551519888605>
15. Menon, A.: Large Scale Author Name Disambiguation in Scholarly Databases. Thesis, The Pennsylvania State University (2021)
16. Müller, M.-C.: Semantic Author Name Disambiguation with Word Embeddings. In: *Research and Advanced Technology for Digital Libraries* (2017). <https://doi.org/10.1007/978-3-319-67008-9>
17. Qiao, Z., Du, Y., Fu, Y., Wang, P., Zhou, Y.: Unsupervised Author Disambiguation using Heterogeneous Graph Convolutional Network Embedding. In: *IEEE Int. Conf. on Big Data* (2019) 910–919. <https://doi.org/10.1109/BigData47090.2019.9005458>
18. Rodrigues, N.S., Mariano, A.M., Ralha, C.G.: Author name disambiguation literature review with consolidated meta-analytic approach. *Int. J. Digit. Libr.* **24** (2024) 765–785. <https://doi.org/10.1007/s00799-024-00398-1>
19. Smalheiser, N.R., Torvik, V.I.: Author Name Disambiguation. *Ann. Rev. Inf. Sci. Technol.* **43** (2009)
20. Subramanian, S., King, D., Downey, D., Feldman, S.: S2AND: A Benchmark and Evaluation System for Author Name Disambiguation. In: *ACM/IEEE Joint Conf. on Digital Libraries* (2022). <https://doi.org/10.1109/JCDL52503.2021.00029>
21. Tran, H.N., Huynh, T., Do, T.: Author Name Disambiguation by Using Deep Neural Network. University of Information Technology, Vietnam (2017)
22. Ye, F., Xia, Z., Ling, Z., Wu, L.: Multi-view contrastive and cluster-guided learning for author name disambiguation. *Expert Syst. Appl.* **289** (2025) 128324. <https://doi.org/10.1016/j.eswa.2025.128324>
23. Zhang, L., Huang, Y., Yang, J., Lu, W.: Aggregating large-scale databases for PubMed author name disambiguation. *J. Am. Med. Inform. Assoc.* **28**(9) (2021) 1919–1927. <https://doi.org/10.1093/jamia/ocab095>

Espacios de datos y su aplicación en el sector Agrotech

Giannina Milagros Miguel Jurado, Dra. Belén Pérez Lancho, and Dra. Ana Belén Gil González

Universidad de Salamanca
Departamento de Informática y Automática
Facultad de Ciencias.
Plaza de los Caídos s/n
37008 Salamanca, España
{idu071268,lancho,abg}@usal.es

Resumen Este Trabajo de Fin de Máster propone el diseño de un espacio de datos agrícola basado en microservicios e integrado con fuentes públicas (SIAR) y privadas, alineado con principios europeos de soberanía e interoperabilidad. Sobre esta arquitectura se desarrolló un caso de uso para la gestión inteligente del riego, estimando la Evapotranspiración de Referencia (ET_o) mediante modelos de Aprendizaje Automático. Se compararon Random Forest y Perceptrón Multicapa con datos de estaciones agroclimáticas de Soria, siendo RF el más preciso (R^2 y NSE hasta 0,986). Los resultados evidencian que es posible estimar ET_o con menos variables que el método tradicional de Penman-Monteith, facilitando su aplicación operativa. El estudio concluye que la combinación de espacios de datos e IA mejora la eficiencia en la toma de decisiones hídricas y permite soluciones Agrotech escalables.

Palabras clave: espacio de datos, evapotranspiración de referencia, aprendizaje automático

1. Introducción

La Agricultura 4.0 impulsa una transformación en el sector agrícola mediante el uso de tecnologías avanzadas como el Internet de las Cosas (IoT), Big Data e Inteligencia Artificial (IA), lo que permite optimizar la producción, mejorar la sostenibilidad y fortalecer la seguridad alimentaria global [1, 2]. En este contexto, los espacios de datos agroalimentarios se consolidan como herramientas claves para integrar y compartir datos de forma interoperable, segura y reutilizable entre múltiples actores [3], para optimizar prácticas agronómicas y mejorar la toma de decisiones [4].

Entre los datos esenciales para la toma de decisiones agrónomas, destaca la Evapotranspiración de Referencia (ET_o), un parámetro

crítico para estimar las necesidades hídricas de los cultivos. Sin embargo, uno de los modelos estándar FAO56-PM [5] más utilizados requiere una amplia cantidad de variables meteorológicas, lo cual representa una barrera en regiones con infraestructura limitada o con escasez de datos [6].

Dado este desafío, diversos estudios han planteado métodos alternativos para estimar la ETo con menos variables [5, 7]. En este contexto, la presente investigación evalúa algoritmos de inteligencia artificial (IA) por su capacidad para modelar relaciones no lineales y aprovechar datos históricos, buscando mejorar la estimación de ETo.

Además, se propone integrar estos modelos IA dentro de un ecosistema de datos agrícolas, permitiendo a agricultores y asesores acceder a estimaciones confiables y calcular con mayor precisión el déficit hídrico. Esto contribuye a una gestión del riego más eficiente y sostenible [4].

1.1 Trabajos relacionados

Diversos estudios han explorado la estimación de la Evapotranspiración de Referencia (ETo) mediante enfoques empíricos y técnicas de inteligencia artificial (IA), destacando el uso de algoritmos de aprendizaje automático (ML) y aprendizaje profundo (DL) [6, 8, 9]. Estas investigaciones comparan modelos tradicionales como FAO-56 Penman-Monteith (PM) con métodos basados en IA, buscando alternativas eficientes que requieran menos variables meteorológicas.

Por ejemplo, en [8], se evaluó la precisión de tres modelos empíricos (HS, PT, TU) y tres de Aprendizaje Automático (LR, RF, NN), concluyendo que RF y NN superan en rendimiento a los métodos empíricos no calibrados. De forma similar, [6] analizó diversos algoritmos (MLP, GRNN y XGBoost), usando únicamente variables de temperatura, y demostró su eficacia en contextos con datos limitados.

En [9], se aplicó un análisis de componentes principales para reducir las variables climáticas y se entrenó una red neuronal profunda con solo tres variables (Temperatura máxima, Temperatura mínima y radiación), obteniendo resultados comparables al modelo Penman-Monteith. Asimismo, el estudio de [10], desarrollado en Marruecos, evaluó ocho modelos empíricos, cuatro de ML y combinaciones híbridas con distintas configuraciones de entrada, confirmando que los modelos de ML superan significativamente a los métodos empíricos en precisión, evaluados con métricas como KGE, R^2 , RMSE y RRSE.

Finalmente, investigaciones recientes también destacan el uso de modelos híbridos que combinan algoritmos de ML y DL para mejorar aún más la estimación de ETo [10, 11]. En conjunto, estos estudios reflejan el creciente potencial de la IA para abordar los desafíos de estimación de ETo, especialmente en escenarios con datos meteorológicos incompletos o necesidad de predicciones en tiempo real.

1.2 Problema Planteado

Una gestión eficiente del riego requiere estimar con precisión la Evapotranspiración de Referencia (ET_O), ya que esta variable es fundamental para calcular la demanda hídrica de los cultivos y evaluar el déficit hídrico. Sin embargo, el método estándar FAO56-PM más aceptado para su cálculo exige una amplia cantidad de datos meteorológicos que no siempre están disponibles, especialmente en zonas rurales con baja cobertura instrumental.

Esta dependencia de variables agroclimáticas de difícil acceso limita la aplicación operativa del cálculo de ET_O , lo que repercute negativamente en la planificación hídrica, el rendimiento agrícola y la sostenibilidad del recurso agua.

1.3 Objetivo

El objetivo de este trabajo es estudiar los espacios de datos como marco de referencia para la integración de fuentes de datos climáticos, mediante la propuesta de un caso de estudio que incorpore modelos de estimación de la evapotranspiración de referencia (ET_O) basados en algoritmos de aprendizaje automático y analizar su potencial para fortalecer la gestión agrícola.

Más concretamente podemos destacar dos objetivos específicos:

- Desarrollar y evaluar modelos de estimación de la evapotranspiración de referencia (ET_O) utilizando algoritmos de aprendizaje automático a partir de variables agroclimáticas.
- Proponer el diseño de la arquitectura técnica de una plataforma integrada a un ecosistema de datos agrícolas, centrada en la gestión inteligente del riego, como caso de estudio.

1.4 Espacio de Datos

En la economía digital, los espacios de datos (*Data Spaces*) se consolidan como una arquitectura clave para promover el intercambio soberano, seguro e interoperable de datos entre múltiples actores. Estos entornos federados permiten que los datos se compartan con garantías técnicas, jurídicas y organizativas, respetando las condiciones impuestas por los propietarios del dato. Iniciativas como Gaia-X y la *International Data Spaces Association (IDSA)*, están liderando el diseño de infraestructuras confiables para estos fines [12, 13].

Según el glosario [14], un espacio de datos es una estructura distribuida, que se define por un marco de gobernanza, que permite transacciones de datos confiables entre los diversos participantes y, a su vez, otorga un respaldo a la confianza y soberanía de los datos. Un espacio de datos se implementa mediante una o más infraestructuras y admite uno o más casos de uso.

1.5 Riego de cultivos

El riego es una práctica agrícola que aporta agua de forma artificial para cubrir las necesidades hídricas de los cultivos y asegurar su buen desarrollo. Su método de aplicación depende de factores como el tipo de cultivo, suelo, topografía, disponibilidad de agua, costos y su impacto ambiental.

La programación de riegos implica determinar cuándo se debe regar y cuánta agua se debe aplicar. Para esto, es imprescindible conocer las características del cultivo, las características físicas del suelo y las condiciones climáticas de la zona.

Para llevar a cabo una programación de riego eficiente, es necesario estimar con precisión las necesidades hídricas del cultivo. En este sentido, uno de los métodos más utilizados es el balance de agua, el cual tiene en cuenta tanto los aportes (como la lluvia y el riego) como las pérdidas de agua (principalmente por evapotranspiración) en el sistema suelo-planta. La diferencia entre estas entradas y salidas determina los cambios en la humedad del suelo [15].

1.5.1 Evapotranspiración de referencia (ET_o) La evapotranspiración de referencia (ET_o), es la evapotranspiración que se produciría en un campo de gramíneas de 12 cm de altura, sin falta de agua y con determinadas características aerodinámicas y de albedo. En el concepto se hace referencia a características aerodinámicas, ya que la resistencia al flujo de aire afecta la transferencia de vapor de agua entre las hojas de las plantas y la atmósfera, y al albedo como la capacidad de la superficie vegetal para reflejar la radiación solar [6].

Además, de acuerdo con [16], ET_o es un parámetro climático calculado a partir de datos meteorológicos que mide la capacidad de evaporación de la atmósfera en un lugar y tiempo específicos, sin considerar características del cultivo ni del suelo.

En ocasiones la ingeniería de cultivos, especialmente en el campo de la Hidrología, se utilizan estos conceptos de Evapotranspiración de referencia (ET_o) y Evapotranspiración potencial (ETP) de manera intercambiable al analizar el balance hídrico dentro de una cuenca y se emplean fórmulas diseñadas para calcular tanto ETP como ET_o sin hacer distinción, asumiendo que ambos son similares [17].

1.5.2 Déficit Hídrico El déficit hídrico se define como la diferencia entre la demanda hídrica del cultivo (expresada mediante la evapotranspiración potencial o de referencia) y la cantidad de agua efectivamente disponible para las plantas. Este déficit puede ocurrir por insuficiente precipitación o riego, lo que conlleva a una reducción en la tasa de transpiración y, en consecuencia, en el desarrollo y rendimiento del cultivo [18]. Uno de los métodos del cálculo del déficit hídrico, de acuerdo al autor [15], se expresa de la siguiente manera:

$$DH = K_c \cdot ET_o - P_e \quad (1)$$

1.6 Aprendizaje Automático

El aprendizaje automático (ML o *Machine Learning*) es un proceso mediante el cual los ordenadores son capaces de aprender por sí mismos a partir de datos. En lugar de ser programados explícitamente para resolver un problema, los sistemas de aprendizaje automático adquieren conocimientos mediante la experiencia, aprovechando los datos y ajustando sus comportamientos a medida que reciben más información. Este proceso permite que los sistemas resuelvan problemas que no han sido explícitamente programados para manejar, utilizando un modelo adaptativo basado en patrones aprendidos de los datos [19]. Para que el proceso de aprendizaje automático pueda ser efectivo, es crucial realizar un preprocesamiento de los datos antes de alimentar un modelo. El preprocesamiento de datos implica limpiar, transformar y seleccionar las variables más relevantes.

Existen diferentes técnicas para obtener modelos de predicción basados en aprendizaje automático y los utilizados en este trabajo son dos:

1.6.1 Random Forest (RF) Es un algoritmo de aprendizaje supervisado que utiliza el enfoque de *bagging* para construir múltiples árboles de decisión. Cada árbol se entrena con diferentes subconjuntos de datos y seleccionando aleatoriamente un conjunto de características en cada división. El objetivo es disminuir la varianza de los árboles individuales y mejorar la precisión de las predicciones, manteniendo el sesgo bajo. Para hacer una predicción final, el algoritmo combina los resultados de todos los árboles, lo que lo hace robusto y efectivo tanto para clasificación como para regresión [20].

1.6.2 Redes Neuronales (NN) Son modelos matemáticos inspirados en la forma en que funcionan las neuronas en el cerebro. Están compuestas por capas de nodos (o "neuronas"), donde cada nodo realiza un cálculo y transmite el resultado a otros nodos de la siguiente capa. Su objetivo es mapear un conjunto de variables de entrada a un conjunto de variables de salida, utilizando parámetros adaptables, que se ajustan durante el proceso de entrenamiento [21]. Pueden utilizarse para tareas como el reconocimiento de patrones, compresión de información, reducción de la dimensionalidad, agrupamiento, clasificación y visualización [22].

El Perceptrón Multicapa (MLP o *Mult Layer Perceptron*) es un tipo de red neuronal usada para resolver problemas de aproximación funcional y clasificación cuando los datos no son linealmente separables. Este modelo tiene una estructura en cascada, con una o más capas ocultas entre la entrada y la salida [22].

2. Metodología

Los objetivos de esta investigación se desarrollan en dos etapas, como se muestra en la Figura 1. El proceso inicia con una revisión de bibliográfica clave y marcos de referencia, lo que permitió establecer los componentes técnicos necesarios para integrar el caso de estudio en un espacio de datos agrícola.

A partir de esta base, se abordan dos etapas principales:

Primera etapa: Enfocada en la gestión del riego, incluye la identificación de roles dentro del ecosistema de datos, definición de requerimientos y casos de uso, y culmina con el diseño técnico de la arquitectura que alojará los modelos de estimación de ETo.

Segunda etapa: Centrada en la propuesta de valor, consiste en el entrenamiento de modelos de aprendizaje automático para estimar la ETo, los cuales se integran como un servicio analítico dentro del espacio de datos, accesible para usuarios autorizados.

Como aplicación práctica, se plantea el uso de estos modelos para el cálculo del déficit hídrico.

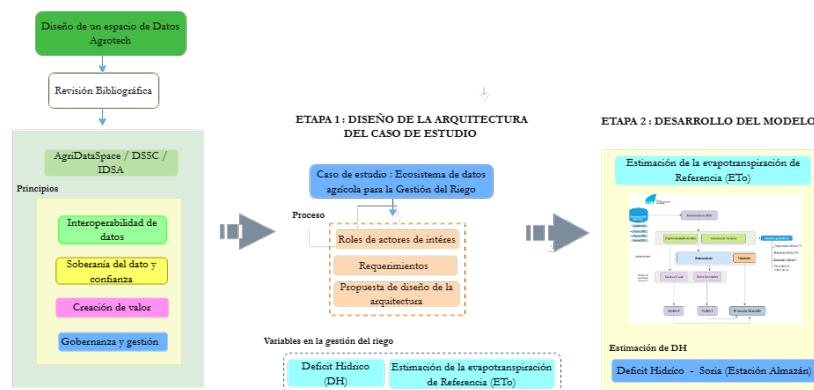


Figura 1: Metodología General

2.1 Descripción del espacio de gestión de riego agrícola

2.1.1 Caso de uso: Se plantea el desarrollo de una solución digital basada en los principios de los espacios de datos agrícolas, cuyo objetivo es optimizar la gestión del riego mediante la estimación automatizada y localizada de la evapotranspiración de referencia (ETo).

2.1.2 Actores del Ecosistema de Datos Agrícola Los proveedores de datos, como el Sistema de Información Agroclimática para el

Regadío (SIAR) y los propios agricultores, aportan información agroclimática clave. Los proveedores de servicios operan plataformas que permiten visualizar datos y estimar ETo y déficit hídrico. Los proveedores tecnológicos desarrollan los modelos de ML e implementan la infraestructura digital. Los consumidores, como agricultores, asesores y cooperativas, utilizan esta información para optimizar la gestión del riego. Finalmente, los intermediarios de datos integran, transforman y redistribuyen datos provenientes de diversas fuentes para su uso contextualizado.

2.2 Propuesta de valor: Estimación de la evapotranspiración de referencia (ETo)

El área de estudio corresponde a la provincia de Soria, ubicada en Castilla y León, al noreste de España que comparte similitudes climáticas con regiones altoandinas del Perú, en ambas zonas se practica la agricultura pero, a diferencia de Soria, en dichas regiones existe una carencia de estaciones agroclimáticas lo cual dificulta la planificación agronómica, lo que refuerza la importancia de contar con modelos predictivos que funcionen con un número reducido de variables.

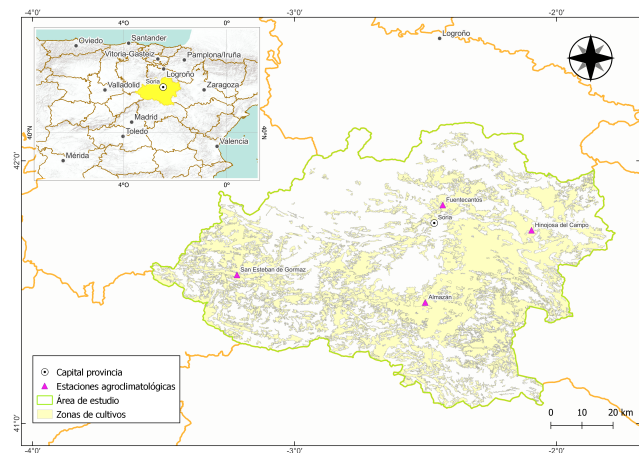


Figura 2: Ubicación geográfica del área de estudio y distribución de estaciones agrometeorológicas.

La base de datos utilizada en este estudio se generó a partir de información recopilada por cuatro estaciones agroclimáticas del SIAR ubicadas en la provincia de Soria: SO01 (Almazán), SO02 (San Esteban de Gormaz), SO03 (Fuentecantos) y SO101 (Hinojosa del Campo).

Dado que la estimación de la evapotranspiración de referencia (ETo) es un componente fundamental para una gestión eficiente del

riego, se desarrolló un flujo metodológico específico para su predicción, el cual se presenta en la Figura 3.

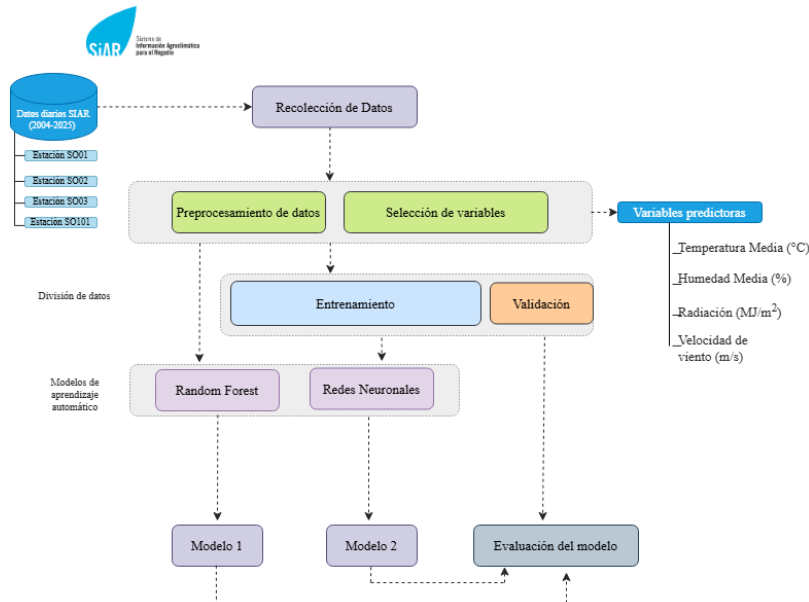


Figura 3: Flujograma del modelo de predicción de ETo.

2.3 Preprocesamiento de Datos

Del total de datos, se seleccionaron las variables climáticas utilizadas en las fórmulas de cálculo tradicional: Temperatura Media (°C), Temperatura Máxima (°C), Temperatura Mínima (°C), Humedad Media (%), Humedad Máxima (%), Humedad Mínima (%), Velocidad de viento (m/s), Radiación (MJ/m²) y Evapotranspiración de referencia (ETo).

Dado que las estaciones del SIAR no operan de forma homogénea en el tiempo, se aplicó un filtro de consistencia temporal para seleccionar solo aquellas con registros completos, tomando como referencia la estación SO101 (Hinojosa del Campo), activa desde 2004. Esto permitió establecer un periodo común de análisis (2004-05-12 a 2025-04-09) y garantizar comparaciones homogéneas entre estaciones.

Al analizar los registros, se identificaron datos faltantes en todas las estaciones, siendo Hinojosa del Campo la más afectada, especialmente en los meses de noviembre a enero durante 2012–2013. Para asegurar la continuidad temporal, se completaron las fechas faltantes

en todos los conjuntos de datos, asignando valores NaN a los días sin datos.

En las estaciones con pocos valores ausentes (S001, S002 y S003), se aplicó regresión lineal simple usando datos vecinos. En la estación SO101, que presentaba más del 15 % de datos faltantes en 2012 y un 3,8 % en 2013, se evaluaron métodos adicionales. Aunque no se halló autocorrelación espacial significativa por el número reducido de estaciones, las altas correlaciones temporales permitieron imputar temperatura, radiación y ETo con el promedio de las estaciones más correlacionadas (S001 y S003). Para humedad y viento, se utilizó la media histórica mensual de la propia estación.

Para reducir la dimensionalidad del conjunto de datos y conservar las variables más relevantes para la predicción de la ETo, se aplicó un modelo de Random Forest Regressor a los datos de la estación Almazán (S001). El análisis mostró que las variables con mayor importancia relativa fueron Radiación (0.82) y Temperatura Media (0.132). También se evaluó la matriz de correlación, identificando redundancias entre variables similares (por ejemplo, las temperaturas y las humedades).

Como resultado, se seleccionaron cuatro variables clave para el entrenamiento de modelos: Radiación, Temperatura Media, Velocidad del Viento y Humedad Media, por su relevancia y menor redundancia.

Los datos se dividieron en un 80 % para entrenamiento y 20 % para prueba. Se entrenaron y compararon dos modelos: un Random Forest Regressor con 80 árboles y una red neuronal multicapa (MLP). La red fue implementada con NumPy, usando una arquitectura simple: una capa de entrada (una variable por nodo), una capa oculta con tres neuronas y una capa de salida para predecir la ETo.

3. Resultados

Previo al entrenamiento, los datos fueron escalados con *Min-Max Scaling*, y el modelo MLP se entrenó durante 100 épocas con *learning rate* de 0.1, usando retropropagación. Ambos modelos se evaluaron en distintos experimentos, manteniendo fija la arquitectura, con el objetivo de identificar el conjunto mínimo de variables necesario para predecir ETo y, en consecuencia, calcular el déficit hídrico.

En línea con los objetivos del estudio, se definieron dos metas principales: (1) diseñar la arquitectura del caso de estudio para una gestión eficiente del riego y (2) entrenar modelos de aprendizaje automático para estimar la evapotranspiración de referencia (ETo). A continuación, se presentan los resultados obtenidos en ambas etapas, abarcando tanto el diseño arquitectónico como el desempeño de los modelos.

3.1 Propuesta del Diseño arquitectónico del caso de estudio

La Figura 4 muestra el diseño arquitectónico de una plataforma para la gestión del riego agrícola, basada en un patrón de microservicios.

Esta propuesta promueve la modularidad, escalabilidad y mantenibilidad, en línea con los principios de interoperabilidad y soberanía de los espacios de datos modernos. Entre sus componentes clave, destaca el Servicio Modelo ETo, encargado de ejecutar los modelos entrenados (Random Forest y MLP) para estimar la evapotranspiración de referencia (ETo), activado por eventos publicados a través del *Message Broker*. A partir de esta estimación, el módulo de Cálculo Hídrico permite obtener los valores de la Evapotranspiración del cultivo en condiciones estándar (ETc) y déficit hídrico. Ambos servicios se integran dentro de una arquitectura interoperable, facilitando el análisis y visualización de resultados para una gestión hídrica más eficiente.

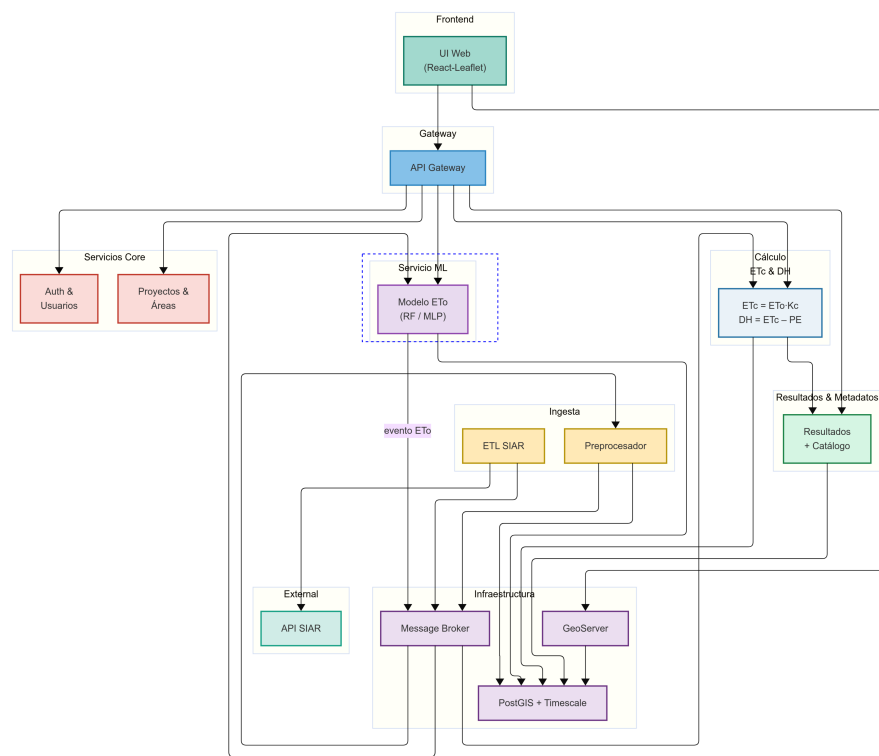


Figura 4: Arquitectura de microservicios del caso de estudio (ETo)

3.2 Estimación de la variable de evapotranspiración de referencia (ETo)

Este análisis aplica modelos de aprendizaje automático (Random Forest y MLP) para estimar la evapotranspiración de referencia

(ET_o), utilizando datos agroclimáticos del SIAR provenientes de estaciones en la provincia de Soria.

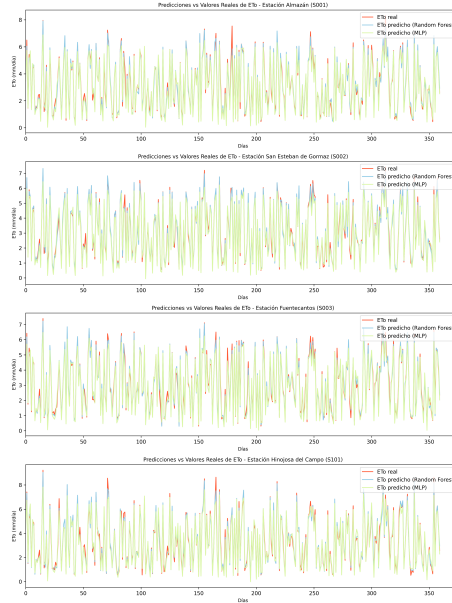


Figura 5: Serie de tiempo de los datos predichos vs reales de ET_o, usando cuatro variables de entrada.

En el experimento 1, se entrenaron los modelos Random Forest y MLP utilizando cuatro variables: Temperatura Media, Humedad Media, Radiación y Velocidad del Viento. Ambos modelos mostraron una buena correspondencia con los valores reales de ET_o, indicando una capacidad predictiva adecuada. Sin embargo, Random Forest presentó un mejor ajuste general en las cuatro estaciones analizadas.

La Figura 5 muestra que ambos modelos, Random Forest y MLP, capturan adecuadamente la dinámica temporal de la ET_o a lo largo del año. No obstante, Random Forest se ajusta mejor a las variaciones abruptas, especialmente en Almazán y Hinojosa del Campo, mientras que MLP tiende a suavizar los extremos y subestimar los valores altos, especialmente en San Esteban de Gormaz y Fuentescantos.

La Tabla 1 muestra que el modelo Random Forest superó consistentemente al MLP en todas las estaciones, con menores errores (MAE, MSE, RMSE) y mayores valores de R² y NSE. El mejor desempeño se obtuvo en San Esteban de Gormaz (S002), donde RF alcanzó un R² de 0.9852 y un NSE de 0.9861. En el resto de estaciones, RF también mostró mejor capacidad explicativa, especialmente en la predicción de valores extremos de ET_o.

Tabla 1: Métricas de evaluación - Experimento 1

Estación	Modelo	MAE	MSE	RMSE	R ²	NSE
2*Almazán (S001)	MLP	0.2669	0.1294	0.3598	0.9669	0.9644
	RF	0.1784	0.0579	0.2407	0.9852	0.9848
2*San Esteban de Gormaz (S002)	MLP	0.2662	0.1233	0.3512	0.9662	0.9631
	RF	0.1654	0.0495	0.2224	0.9864	0.9861
2*Fuentecantos (S003)	MLP	0.2645	0.1238	0.3518	0.9649	0.9626
	RF	0.1832	0.0607	0.2463	0.9828	0.9824
2*Hinojosa del Campo (S101)	MLP	0.2887	0.1685	0.4105	0.9641	0.9613
	RF	0.2230	0.0939	0.3065	0.9800	0.9793

En el segundo experimento se entrenaron los modelos Random Forest y MLP usando tres variables: Temperatura Media, Radiación y Velocidad del Viento. A pesar de reducir la cantidad de variables respecto al experimento anterior, ambos modelos mantuvieron un buen desempeño predictivo. En la Figura 6 muestra que las predicciones siguen de forma ajustada la evolución diaria de ETo a lo largo del año.

Random Forest destaca por su mayor precisión al capturar los picos y valles, incluso en condiciones variables como en Hinojosa del Campo (SO101). Por su parte, el modelo MLP tiende a suavizar la serie, mostrando ligeros desfases en los valores extremos, especialmente en Fuentecantos (S003) y Almazán (S001), aunque reproduce correctamente la tendencia estacional.

La Tabla 2 muestra que Random Forest superó consistentemente al MLP en todas las estaciones y métricas evaluadas, con menores errores (MAE, MSE, RMSE) y mayores valores de R² y NSE. El mejor rendimiento se observó en San Esteban de Gormaz (S002), con un R² de 0.9824. Las diferencias entre modelos fueron más marcadas en S002 y Almazán (S001). Aunque Fuentecantos (S003) mostró buenos resultados en ambos casos, RF mantuvo una ligera ventaja. En Hinojosa del Campo (S101), ambos modelos tuvieron un desempeño algo menor, pero RF siguió siendo más preciso, especialmente en la reducción del error absoluto.

En el último experimento se entrenaron los modelos utilizando solo dos variables: Temperatura Media y Radiación. En la evaluación temporal (Figura 7), MLP siguió la serie de forma más suavizada, mientras que Random Forest mostró mayor capacidad para capturar cambios abruptos y valores extremos.

La disminución de variables afectó la precisión de ambos modelos, pero Random Forest mantuvo una mayor robustez, evidenciando mejor adaptación en contextos con alta variabilidad temporal, donde el MLP mostró limitaciones más marcadas.

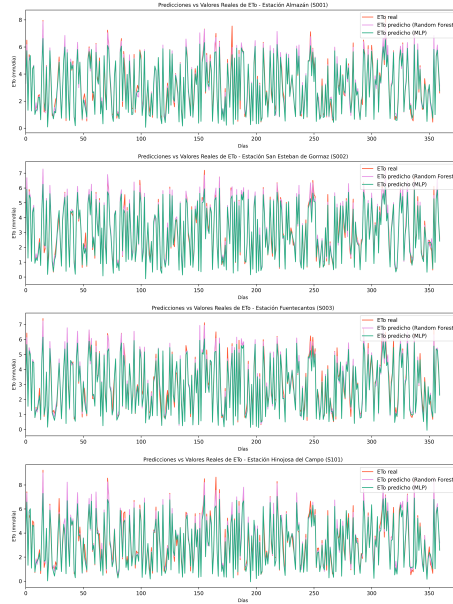


Figura 6: Serie de tiempo de los datos predichos vs reales de ETo, usando tres variables de entrada.

Tabla 2: Métricas de evaluación - Experimento 2

Estación	Modelo	MAE	MSE	RMSE	R ²	NSE
Almazán (S001)	MLP	0.2805	0.1457	0.3817	0.9628	0.9593
	RF	0.2125	0.0827	0.2875	0.9789	0.9783
San Esteban de Gormaz (S002)	MLP	0.2717	0.1306	0.3614	0.9642	0.9609
	RF	0.1904	0.0640	0.2531	0.9824	0.9819
Fuentecantos (S003)	MLP	0.2910	0.1534	0.3916	0.9566	0.9526
	RF	0.2184	0.0861	0.2935	0.9756	0.9748
2*Hinojosa del Campo (S101)	MLP	0.3896	0.2880	0.5367	0.9387	0.9305
	RF	0.2964	0.1582	0.3978	0.9663	0.9650

La Tabla 3 confirma que Random Forest supera al MLP en todas las estaciones, con menores errores y mayores valores de R² y NSE. Aunque solo se usaron dos variables, ambos modelos mantuvieron una buena capacidad explicativa, aunque con menor precisión que en los experimentos anteriores. Ambos presentan limitaciones para capturar picos extremos de ETo, pero logran reproducir adecuadamente la tendencia general de la serie temporal.

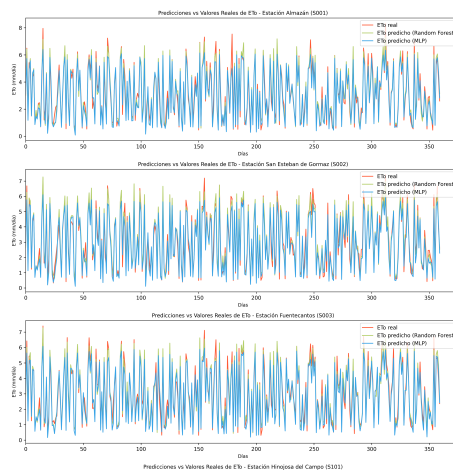


Tabla 3: Métricas de evaluación - Experimento 3

Estación	Modelo	MAE	MSE	RMSE	R ²	NSE
Almazán (SO01)	MLP	0.3325	0.2020	0.4494	0.9484	0.9432
	RF	0.3228	0.1824	0.4270	0.9534	0.9517
San Esteban de Gormaz (SO02)	MLP	0.3617	0.2278	0.4773	0.9375	0.9301
	RF	0.3206	0.1759	0.4194	0.9517	0.9500
Fuentecantos (SO03)	MLP	0.3373	0.2064	0.4543	0.9416	0.9358
	RF	0.3208	0.1758	0.4193	0.9502	0.9485
Hinojosa del Campo (SO101)	MLP	0.4242	0.3494	0.5911	0.9256	0.9153
	RF	0.4057	0.3035	0.5509	0.9354	0.9322

Como se observa en la Figura 8, el modelo Random Forest (RF) demostró una mayor solidez que el MLP, manteniendo valores de RMSE más bajos en todas las estaciones. No obstante, al reducir el número de variables predictoras, ambos modelos registraron un incremento progresivo en el error, evidenciando una pérdida de precisión en la estimación de ETo. Esta afectación fue más marcada en el MLP, mientras que RF mostró mayor resistencia a la disminución de información.

Cabe señalar que la estación Hinojosa del Campo fue la más sensible a esta reducción, presentando una caída significativa en el rendimiento. En cambio, San Esteban de Gormaz experimentó la menor variación, lo que indica una mayor consistencia del modelo en dicha estación. En general, la disminución en el ajuste y el aumento de errores reflejan que contar con más variables climáticas mejora la estimación de ETo. No obstante, los modelos siguen siendo útiles en escenarios con datos limitados. Random Forest destaca por mantener un buen equilibrio entre precisión y simplicidad, mientras que el MLP, aunque competitivo, muestra mayor sensibilidad ante la reducción de información.

3.3 Aplicación de los valores estimados de ETo

Para aplicar los modelos de Random Forest y MLP en un ecosistema de gestión del riego, se siguieron tres pasos: (a) obtención de parámetros clave (K_c , ET_c y Pe), (b) estimación de ETo con los modelos entrenados y (c) cálculo del déficit hídrico.

Como referencia, se seleccionó un punto de muestreo cercano a la estación Almazán (SO01) mediante el visor del SIAR (<https://espaciosiar.es/visor>), correspondiente a noviembre de 2023. En este punto, se obtuvo un valor de ETo de 33 mm/mes y un ET_c de 38 mm/mes. La precipitación efectiva (Pe) se obtuvo de forma complementaria desde la plataforma del SIAR (<https://servicio.mapa.gob.es/websiar/>), junto con las variables climáticas empleadas como entrada en los modelos (temperatura media, radiación, velocidad del viento y ETo).

La constante de cultivo (K_c) se calculó a partir de los valores de evapotranspiración del cultivo (ET_c) y la evapotranspiración de referencia (ET_o), usando la relación mostrada en la Ecuación 1. Para obtener

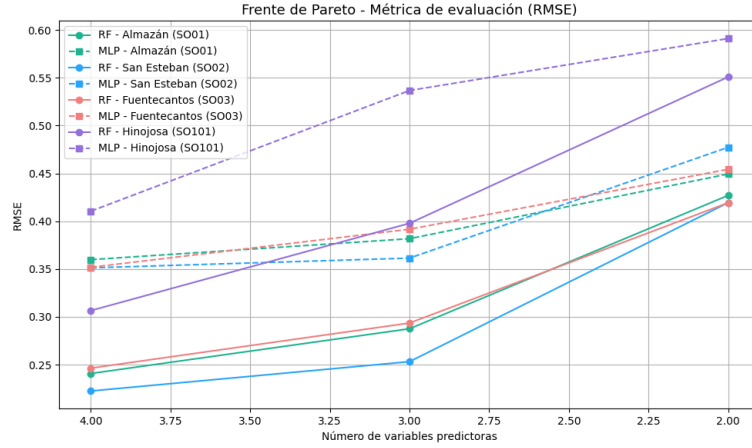


Figura 8: Fronte de Pareto de la métrica de evaluación (RMSE), respecto a la cantidad de variables predictoras.

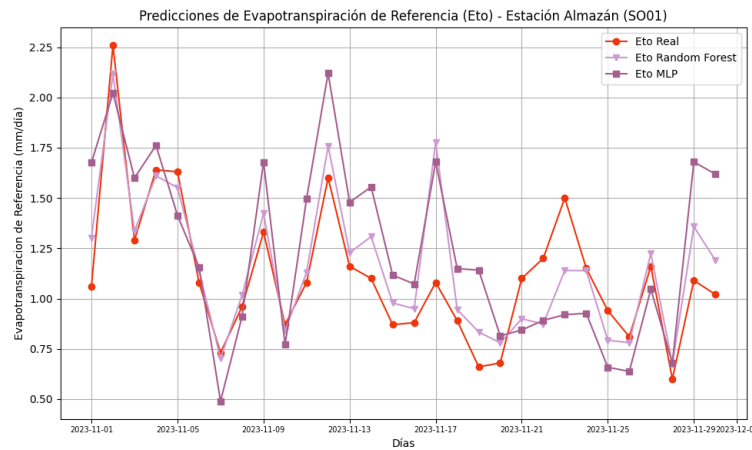


Figura 9: Valores reales vs calculados de ETo en la estación Almazán (SO01)

ET_o , se utilizaron los valores estimados por los modelos desarrollados en esta investigación.

La Figura 9 muestra que ambos modelos, Random Forest y MLP, reproducen adecuadamente la tendencia diaria de la ETo en la estación Almazán (SO01). Aunque el MLP presenta mayor variabilidad y errores en los picos y valles, especialmente en comparación con RF, el desempeño general de ambos modelos es satisfactorio para el periodo analizado.

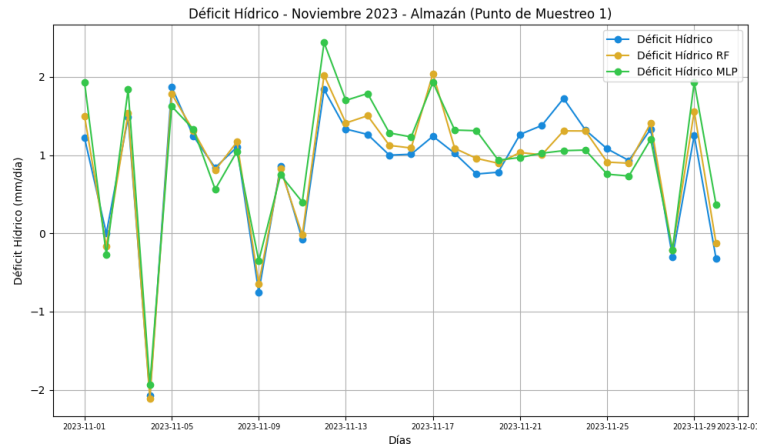


Figura 10: Déficit Hídrico calculado

La Figura 10 ilustra la evolución diaria del déficit hídrico. Ambos modelos capturan adecuadamente la tendencia general, incluyendo los días con déficit máximo y mínimo, aunque con ligeras subestimaciones en los extremos. En algunos casos, como el 17 de noviembre, los modelos estimaron déficits máximos en fechas distintas a la observada real, lo que evidencia pequeñas desviaciones.

En conjunto, los resultados confirman la utilidad de ambos modelos para predecir el déficit hídrico, ofreciendo una herramienta valiosa para la planificación del riego y la gestión agronómica.

4. Conclusiones

La investigación demostró que los modelos de inteligencia artificial, especialmente Random Forest (RF), son altamente eficaces para estimar la evapotranspiración de referencia (ET_o), variable clave en el cálculo del déficit hídrico y la gestión del riego. En todos los experimentos y estaciones, RF obtuvo consistentemente mayores valores de R² y NSE, y menores errores que MLP, confirmando su idoneidad para problemas de regresión.

El análisis de Pareto evidenció que el mejor rendimiento se obtiene al emplear las cuatro variables predictoras, logrando menor RMSE y mayores valores de R² y NSE. No obstante, al reducir a solo dos variables (temperatura media y radiación), el modelo RF sigue ofreciendo un rendimiento robusto, con R² superiores al 94% y NSE por encima del 91%. Esto lo convierte en una alternativa eficiente y práctica en escenarios con datos climáticos limitados.

La integración de estos modelos en un sistema permite generar series diarias de déficit hídrico, facilitando la detección de eventos de

estrés o exceso hídrico y mejorando la toma de decisiones. Además, se evidencia que los modelos desarrollados pueden sustituir eficazmente métodos tradicionales como Penman-Monteith, con menor demanda de variables.

Finalmente, se propuso una arquitectura basada en microservicios compuesta por seis bloques funcionales, que permite una gestión modular del sistema, facilita la integración de datos heterogéneos y responde a las necesidades del sector agrícola, en línea con los principios de interoperabilidad y soberanía que caracterizan a los espacios de datos.

4.1 Futuras líneas de trabajo

Dado que el modelo entrenado no presentó inconvenientes al generar resultados precisos, una línea futura de investigación podría centrarse en el uso de información climática derivada de imágenes satelitales, las cuales ofrecen mayor cobertura espacial. Esto permitiría estimar la ETo a nivel de píxel, ampliando así el alcance geográfico de los análisis y su aplicabilidad en distintas regiones agrícolas.

Otra posible línea de ampliación de este trabajo podría ser, diseñar un módulo avanzado de recomendación de riego que combine los pronósticos de ETo y ETc con factores económicos como el costo del agua y de la energía, con el objetivo de optimizar la planificación del riego desde una perspectiva orientada a la eficiencia económica y al uso racional de los recursos.

5. Agradecimientos

Este trabajo se enmarca en el proyecto Cátedra Internacional en IA Fiable y Reto Demográfico (DemIA) financiado por la Secretaría de Estado de Digitalización e Inteligencia Artificial, Ministerio para la Transformación Digital y de la Función Pública (Referencia: TSI-100933-2023-0001).

Referencias

1. S. Wolfert, L. Ge, C. Verdouw, and M.-J. Bogaardt, "Big data in smart farming – a review," *Agricultural Systems*, vol. 153, pp. 69–80, 2017.
2. A. Cravero, S. Pardo, P. Galeas, J. López Fenner, and M. Caniupán, "Data type and data sources for agricultural big data and machine learning," *Sustainability*, vol. 14, no. 23, p. 16131, 2022.
3. G. Brunori, M. Bacco, C. Puerta-Piñero, M. T. Borzacchiello, and E. Stormer, "Agri-food data spaces: Highlighting the need for a farm-centered strategy," *Data in Brief*, vol. 59, p. 111388, 2025.
4. R. Giaffreda, R. García, R. Palma, and et al., "Building a european framework for the secure and trusted data space for agriculture," Deliverable Report Grant Agreement No. 101083401, FONDAZIONE BRUNO

- KESSLER (FBK), Mar. 2023. Work Package WP3: Data Space technological landscape. Submission Date: 31 March 2023. Call: DIGITAL-2021-PREFACTS-DS-01.
5. R. G. Allen, L. S. Pereira, D. Raes, and M. Smith, *Crop evapotranspiration - Guidelines for computing crop water requirements*, vol. 56 of *FAO Irrigation and drainage paper*. Food and Agriculture Organization of the United Nations, 1998.
 6. J. A. Bellido-Jiménez, J. Estévez, and A. P. García-Marín, “New machine learning approaches to improve reference evapotranspiration estimates using intra-daily temperature-based variables in a semi-arid region of Spain,” *Agricultural Water Management*, vol. 245, p. 106558, 2021.
 7. J. Doorenbos and W. O. Pruitt, *Las necesidades de agua de los cultivos*. Estudio FAO Riego y Drenaje No. 24, Roma, Italia: FAO - Organización de las Naciones Unidas para la Alimentación y la Agricultura, 1986. (FAO77).
 8. N. S. Hailegnaw, H. K. Bayabil, M. L. Berihun, F. T. Teshome, V. Shelia, and F. Getachew, “Integrating machine learning and empirical evapotranspiration modeling with dssat: Implications for agricultural water management,” *Science of the Total Environment*, vol. 912, p. 169403, 2024.
 9. M. Nagappan, V. Gopalakrishnan, and M. Alagappan, “Prediction of reference evapotranspiration for irrigation scheduling using machine learning,” *Hydrological Sciences Journal*, vol. 65, no. 16, pp. 2669–2677, 2020.
 10. S. Acharki, A. Raza, D. K. Vishwakarma, and et al., “Comparative assessment of empirical and hybrid machine learning models for estimating daily reference evapotranspiration in sub-humid and semi-arid climates,” *Scientific Reports*, vol. 15, p. 2542, 2025.
 11. M. Valipour, M. A. Gholami Sefidkouhi, M. Raeini-Sarjaz, and S. M. Guzman, “A hybrid data-driven machine learning technique for evapotranspiration modeling in various climates,” *Atmosphere*, vol. 10, no. 311, 2019.
 12. B. Otto, S. Steinbuß, and A. e. a. Teuscher, “Ids reference architecture model 3.0,” tech. rep., International Data Spaces Association, 2019.
 13. Gaia-X Data Space Business Committee, “Data space business committee – position paper,” 2021. Position paper on sector-specific data spaces including agriculture.
 14. Data Spaces Support Centre (DSSC), “1st collaborative discussion paper Why data spaces A business and user’s perspective,” tech. rep., nov 2022.
 15. R. F. Gómez, R. Ávila Alabarces, M. L. Rodríguez, P. G. Zafra, and N. A. O. Gutiérrez, *Manual de Riego para Agricultores. Módulo 1: Fundamentos del Riego*. Agricultura: formación, Sevilla: Junta de Andalucía. Instituto de Investigación y Formación Agraria y Pesquera. Consejería de Agricultura y Pesca, 2010. Depósito Legal: SE-1942-2010.
 16. FAO, *Guías para la determinación de los requerimientos de agua de los cultivos*. Serie de Riego y Drenaje No. 56, Roma: FAO, 1990. Disponible en: FAO - Food and Agriculture Organization of the United Nations.
 17. M. I. Sánchez, *Métodos para el estudio de la evaporación y evapotranspiración*, vol. 3. Cuadernos Técnicos Sociedad Española de Geomorfología, 1992.
 18. M. Villodre Carrilero, *Evaluación de estrategias de riego deficitario controlado en almendro en un contexto de cambio climático*. Tesis doctoral, Universidad Politécnica de Cartagena, 2022.

19. F. Berzal, *Redes Neuronales & Deep Learning*. Granada: Edición independiente, 2018. Edición a todo color.
20. T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. New York: Springer, 2009.
21. C. M. Bishop, *Pattern Recognition and Machine Learning*. New York: Springer, 2006.
22. J. H. Orallo, M. J. R. Quintana, and C. F. Ramírez, *Introducción a la Minería de Datos*. Madrid, España: Pearson Educación, S.A., 2004. Última reimpresión 2005.

Clasificación Automática de Textos de Hotelería Basada en Transformadores

Joel Jimenez Valido, María Angélica González Arrieta, and Guillermo
Hernández

Universidad de Salamanca
Departamento de Informática y Automática
Facultad de Ciencias.
Plaza de los Caídos s/n
37008 Salamanca, España
{idu064265, ange1a, gillehg}@usal.es

Resumen La heterogeneidad en las descripciones de los tipos de habitación de hotel presenta un desafío significativo para la distribución en línea, impidiendo la consolidación eficiente de tarifas y disponibilidad. Este artículo aborda este problema implementando y evaluando dos paradigmas de Procesamiento del Lenguaje Natural (PLN) para la clasificación automática de estas descripciones en 15 categorías predefinidas. Comparamos un modelo clásico (TF-IDF con Regresión Logística) frente a un modelo conexionista de última generación (fine-tuning de XLM-RoBERTa-base). Los resultados demuestran la abrumadora superioridad del enfoque basado en Transformers, que alcanzó una precisión de 0.9998 frente al 0.9915 del modelo clásico, resolviendo eficazmente el problema de negocio. Aunque el modelo Transformer es computacionalmente más intensivo, su precisión justifica su uso para una solución robusta y escalable a nivel mundial.

Keywords: Procesamiento del Lenguaje Natural, Clasificación de Texto, Modelos Transformer, XLM-RoBERTa, Distribución Hotelera, TF-IDF

1. Introducción

El ecosistema de la distribución hotelera en línea es masivo, fragmentado y heterogéneo. Un problema central es la falta de una nomenclatura estandarizada para productos idénticos; los proveedores utilizan sus propios términos, lo que lleva a un “efecto de teléfono roto” en la cadena de suministro de datos. Esto resulta en la duplicación de inventarios, la incapacidad de consolidar ofertas para ofrecer el mejor precio y una mala experiencia de usuario. Este estudio enmarca el problema como un problema de clasificación de texto multiclase: mapear descripciones de texto libre a uno de los 15 códigos de habitación estandarizados y predefinidos (por ejemplo, Single, Double, Suite, Twin).

1.1 Limitaciones de los Enfoques Tradicionales

Históricamente, este mapeo se ha basado en expresiones regulares (regex), un método plagado de limitaciones para tareas complejas de PLN. Los sistemas basados en regex son frágiles, costosos de mantener e incapaces de capturar matices semánticos (p. ej., no pueden entender inherentemente que "Habitación para dos.^{es} equivalente a "Habitación Doble"). También dependen del idioma y pueden introducir vulnerabilidades de seguridad como ReDoS [5].

1.2 Objetivos de la Investigación

Los objetivos principales de esta investigación fueron:

- Proporcionar una solución robusta para la clasificación automática de las descripciones de tipos de habitación de hotel.
- Realizar una evaluación comparativa de un enfoque clásico de PLN (TF-IDF con Regresión Logística) y un enfoque conexionista moderno (fine-tuning de XLM-RoBERTa).
- Analizar la naturaleza de los errores de cada modelo para comprender sus fortalezas y debilidades cualitativas.
- Medir el coste computacional para determinar la viabilidad en un entorno de producción.

2. Metodología

El estudio utilizó un conjunto de datos de 529,476 descripciones de habitaciones de hotel de un entorno de producción de una OTA. Este se dividió en un conjunto de entrenamiento (423,580) y uno de prueba (105,896) mediante un muestreo estratificado para preservar la distribución de las 15 clases.

2.1 Infraestructura y Herramientas

El entrenamiento se realizó en un portátil Dell XPS 13 conectado a una eGPU MSI NVIDIA GeForce RTX 3080. Las bibliotecas de software clave incluyeron Python, Scikit-learn, PyTorch y Hugging Face Transformers.

2.2 Modelo Clásico: TF-IDF + Regresión Logística

El pipeline clásico implicó varios pasos:

1. **Preprocesamiento:** Se limpiaron los datos y se codificaron numéricamente las etiquetas.

2. **Extracción de Características:** El texto se convirtió en una matriz numérica utilizando TF-IDF, configurado para considerar unigramas y bigramas, eliminar palabras vacías (stop words) y limitar el vocabulario a 5,000 características para evitar el sobreajuste.
3. **Clasificación:** Se entrenó un clasificador de Regresión Logística utilizando una estrategia One-vs-Rest (OvR), integrado en un pipeline para evitar la fuga de datos [2].

2.3 Modelo Conexionista: Fine-tuning de XLM-RoBERTa

El enfoque moderno aprovechó un modelo Transformer [6] multilingüe preentrenado:

1. **Arquitectura:** Utilizamos ‘xlm-roberta-base’, preentrenado en 100 idiomas, lo que lo hace ideal para el dominio hotelero multilingüe [1]. Los hiperparámetros clave incluyeron una longitud máxima de secuencia de 128, un tamaño de lote de 16/32, 3 épocas y una tasa de aprendizaje baja de $2e-5$.
2. **Preparación de Datos:** El texto fue tokenizado utilizando el tokenizador específico del modelo, y los datos se estructuraron usando las clases ‘Dataset’ y ‘DataLoader’ de PyTorch.
3. **Fine-Tuning:** El modelo preentrenado se cargó con una nueva cabecera de clasificación. Usamos el optimizador AdamW con un planificador de calentamiento lineal (linear warmup) e implementamos la detención temprana (early stopping) para guardar el modelo con mejor rendimiento basado en la precisión de validación.

2.4 Evaluación

El rendimiento del modelo se evaluó utilizando Accuracy, Precision, Recall y F1-Score con promedio ponderado, complementado con matrices de confusión y análisis ROC-AUC para comprender los patrones de error y el poder discriminativo [3, 4].

3. Resultados y Discusión

3.1 Rendimiento Cuantitativo

El conjunto de datos estaba significativamente desequilibrado, con clases como "Double"(161,063 muestras) y "Suite"(99,156) dominando sobre clases minoritarias como "Quintuple Room"(240). Esto resalta la necesidad de métricas ponderadas. Como se muestra en la Tabla 1, el modelo XLM-RoBERTa demostró superioridad, logrando puntuaciones casi perfectas en todas las métricas.

Aunque el modelo TF-IDF tuvo un buen rendimiento ($>99.1\%$), su recall disminuyó para las clases minoritarias como "Shared Room"(0.91) y "Quintuple Room"(0.85). En contraste, XLM-RoBERTa mantuvo un rendimiento casi perfecto incluso para estas clases subrepresentadas, demostrando el poder del aprendizaje por transferencia.

Tabla 1: Comparación del Rendimiento General de los Modelos.

Métrica	TF-IDF + Regresión Logística	XLM-RoBERTa (Ajuste Fino)
Accuracy	0.9915	0.9998
Precisión (Ponderada)	0.9914	1.00
Recall (Ponderado)	0.9915	1.00
F1-Score (Ponderado)	0.9914	1.00

3.2 Análisis Cualitativo de Errores

Las matrices de confusión revelaron la naturaleza de los errores. Los errores del modelo TF-IDF fueron sistémicos, a menudo confundiendo categorías semánticamente similares. Por ejemplo, clasificó erróneamente "Shared Room" como "Double" 205 veces y tuvo dificultades para desambiguar ".Apartment" de otros tipos. La matriz del modelo Transformer fue casi perfecta, reduciendo estos puntos de confusión principales a solo uno o dos errores.

Un análisis más profundo de los 14 errores del Transformer mostró que no se debían a limitaciones del modelo, sino a problemas de calidad de los datos. Los errores surgieron de:

- **Ruido en los Datos:** Un error tipográfico como ".Aquad.en Suite - Executive - Aquadçausó una clasificación errónea porque el modelo priorizó correctamente el token claro "Suite".
- **Ambigüedad Inherente:** "Family Jr. Suite" se clasificó como "Family Room", una inferencia lógica dados los términos en competencia.
- **Errores de Etiquetado:** "Bungalow (1 Bedroom) with living room" se clasificó como "Bungalow", lo que coincide perfectamente con el texto, sugiriendo que la etiqueta original de ".Apartment.er" era incorrecta.

Esto indica que el modelo razona lógicamente "sobre datos ambiguos o ruidosos, y las mejoras futuras dependen de la curación de datos, no de la complejidad del modelo.

3.3 Eficiencia Computacional

Existe un compromiso significativo entre la precisión y el coste computacional (Tabla 2). El modelo TF-IDF es 98 veces más rápido de entrenar y mucho más rápido en la inferencia. Esto se debe a que TF-IDF es un cálculo estadístico simple, mientras que el fine-tuning de un Transformer implica la retropropagación a través de cientos de millones de parámetros. Si bien la precisión del Transformer es superior, TF-IDF podría ser una solución "suficientemente buena" si el re-entrenamiento frecuente y los recursos computacionales limitados son restricciones importantes.

Tabla 2: Comparación de la Eficiencia Computacional.

Métrica	TF-IDF + Re- gresión Logís- tica	XLM- RoBERTa (Ajuste Fino)
Tiempo de Entrenamiento	2.3 min	3.8 horas
Tiempo de Inferencia (105,896 muestras)	4.08 segundos	1 hora
Inferencias por Muestra	Casi instantáneo	0.04 seg/mues- tra

4. Conclusión

Este trabajo ha demostrado de manera concluyente que el fine-tuning de un modelo Transformer preentrenado como XLM-RoBERTa es superior a los métodos clásicos para clasificar las descripciones de las habitaciones de hotel. Con una precisión de 0.9998, el modelo resuelve eficazmente un problema de negocio crítico para la industria de viajes en línea, permitiendo una agregación de inventario más limpia, precios más competitivos y una mejor experiencia para el cliente. De manera crucial, el análisis reveló que el límite de rendimiento del modelo no está definido por sus propias capacidades, sino por la calidad de los datos de entrada. Este estudio subraya el poder de los grandes modelos de lenguaje y el aprendizaje por transferencia para resolver problemas específicos de PLN del mundo real.

5. Trabajo Futuro

El trabajo futuro procederá a lo largo de dos vías principales:

1. **Aplicación a Nuevos Dominios:** La misma metodología se puede aplicar directamente para clasificar los regímenes alimenticios, probablemente con éxito. Un desafío más complejo es la resolución de entidades para el mapeo de hoteles, que requerirá técnicas avanzadas como el aprendizaje de similitud (similarity learning) o modelos basados en grafos para manejar las variaciones en nombres y direcciones.

Referencias

1. Conneau, A., Khandelwal, K., Goyal, N., Chaudhary, V., Wenzek, G., Guzmán, F., Grave, E., Ott, M., Zettlemoyer, L., & Stoyanov, V. (2019). Unsupervised Cross-lingual Representation Learning at Scale. *ArXiv, abs/1911.02116*. <https://api.semanticscholar.org/CorpusID:207880568>.
2. Cox, D. R. (1958). The Regression Analysis of Binary Sequences. *Journal of the royal statistical society series b-methodological*, 20, 215–232. <https://api.semanticscholar.org/CorpusID:125694386>.

3. Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognit. Lett.*, 27, 861–874. <https://api.semanticscholar.org/CorpusID:2027090>.
4. Powers, D. M. W. (2011). Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. *ArXiv, abs/2010.16061*. <https://api.semanticscholar.org/CorpusID:3770261>.
5. Servant, F. (n.d.). Regexes are Hard: Decision-making, Difficulties, and Risks in Programming Regular Expressions. <https://fservant.github.io/papers>.
6. Vaswani, A., Shazeer, N. M., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention is All you Need. *Neural Information Processing Systems*. <https://api.semanticscholar.org/CorpusID:13756489>.

Aplicación de técnicas de lingüística computacional y visualización para identificar patrones en la creación de historias

Pablo Díez Gastón, Roberto Therón Sánchez, and Andrea Vázquez Ingelmo

Universidad de Salamanca
Departamento de Informática y Automática
Facultad de Ciencias.
Plaza de los Caídos s/n
37008 Salamanca, España
{pablodiegastón, theron, andreavázquez}@usal.es

Resumen El presente artículo tiene como objetivo principal desarrollar una herramienta de análisis narrativo que permita representar visualmente la estructura de un guion cinematográfico y compararla con modelos narrativos clásicos como el Viaje del Héroe y la Pirámide de Freytag. Para ello, se ha empleado un enfoque basado en técnicas de procesamiento de lenguaje natural y clasificación automática, combinando la detección de puntos clave narrativos con modelos de clasificación *zero-shot*. Esta aproximación ha permitido generar visualizaciones que reflejan la evolución dramática del guion y su grado de alineación con estructuras narrativas canónicas. El trabajo no solo contribuye a reducir la brecha entre la teoría de la narrativa y las aplicaciones computacionales, sino que también proporciona una herramienta práctica para el análisis sistemático de guiones. Además, sienta las bases para futuras aplicaciones en escritura asistida, análisis comparativo entre géneros y desarrollo de sistemas narrativos inteligentes.

Keywords: Narrativa, Visualización, Zero-Shot, Viaje del Héroe, Pirámide de Freytag

1. Introducción

La narrativa ha sido desde siempre el vehículo primordial para transmitir experiencias, valores y emociones. Sin embargo, la forma en que estructuramos una historia —la disposición de sus eventos, giros y clímax— no es arbitraria [1], sino que obedece a patrones profundos, reconocibles tanto por el lector como por el guionista. En el ámbito

cinematográfico, estos patrones se convierten en herramientas de diseño de guion [2] y en objetos de estudio para entender por qué algunas películas resuenan más que otras tanto para la crítica como para el espectador.

En paralelo al auge de la digitalización de textos y al crecimiento de la capacidad de cómputo, ha emergido una incipiente línea de investigación que combina la lingüística computacional con la visualización de datos [3] para obtener información acerca de la estructura narrativa a gran escala. Gracias a técnicas de procesamiento de lenguaje natural y a la disponibilidad de corpus extensos, como colecciones de guiones o novelas digitalizadas, es posible cuantificar trayectorias emocionales, detectar recursos narrativos y mapear ritmos dramáticos de manera automática.

Aun así, el aprovechamiento de estos hallazgos en el análisis de guiones cinematográficos presenta retos específicos. Por un lado, los guiones combinan texto con indicaciones temporales y visuales, lo cual exige adaptar los algoritmos de PLN (Procesamiento del lenguaje natural) para reconocer no solo la carga semántica de los diálogos, sino también momentos clave en la puesta en escena. Por otro, la transición de la dimensión textual a la visual plantea preguntas de diseño: ¿Cómo representar de modo intuitivo el encadenamiento de etapas narrativas? ¿Qué métricas cuantitativas pueden indicar el grado de apego de un guion a estructuras clásicas como el “Viaje del Héroe” o el esquema de tres actos?

1.1 Objetivos

El objetivo principal que persigue este trabajo es ser capaz de extraer y representar un patrón visual de un guion de cine que nos permita compararlo con los patrones visuales previamente obtenidos de estructuras narrativas comunes como pueden ser la pirámide de Freytag, entre otros.

Además, durante la realización del presente proyecto se abordarán los siguientes objetivos secundarios:

- Realizar una revisión de los estudios previos realizados en el ámbito para analizar sus capacidades, alcance y potencial aplicación a las tareas que atañen este estudio así como analizar posibles líneas de investigación inexploradas o poco abordadas.
- Seleccionar un conjunto de textos adecuado, que represente varios géneros para poder demostrar la aplicabilidad del proyecto en multitud de frentes.
- Investigar y analizar diferentes técnicas empleadas para el procesamiento del lenguaje natural y aplicarlas a los casos de estudio de entre el conjunto de datos seleccionado.

1.2 Aplicabilidad e impacto esperado

Basándose en las limitaciones y oportunidades identificadas en la literatura, este trabajo contribuirá al estado del arte mediante:

- Integración teórico-práctica: Combinar fundamentos narrativos con técnicas computacionales avanzadas aportando un punto de vista que combina varias técnicas de clasificación y visualización.
- Análisis comparativo sistemático: Establecer metodologías para comparar estructuras narrativas entre diferentes géneros y estilos.
- Visualización interactiva: Crear representaciones visuales que faciliten la identificación de patrones narrativos en los textos y que sean interactivas para aportar flexibilidad en la interpretación de resultados.

2. Estado del arte

Debido a la creciente disponibilidad de guiones digitalizados y al avance de las técnicas de PLN, el estudio de la estructura narrativa cinematográfica ha evolucionado de la mera lectura cualitativa a un análisis cuantitativo y visual. En este contexto, la visualización de datos narrativos permite explorar cómo se distribuyen los eventos clave, los giros de la trama y las transformaciones de los personajes a lo largo de la obra, facilitando la detección de patrones clásicos como el “Viaje del Héroe” [4] o la pirámide de Freytag [5], como se ve en la figura 1.

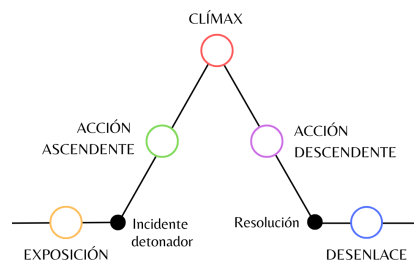


Figura 1: Pirámide de Freytag [6]

A continuación, se resumen las principales líneas de investigación relevantes para este trabajo:

1. **Identificación automática de turning points.** Papalampidi et al. [2] definen los *turning points* como momentos decisivos en la estructura narrativa que marcan la transición entre actos. Mediante anotaciones en sinopsis y redes neuronales, se han desarrollado modelos capaces de identificar automáticamente estos puntos clave en guiones.
2. **Resumen automático mediante grafos dispersos.** En trabajos posteriores [7], los mismos autores proponen representar el guion como un grafo donde escenas relevantes (TPs) actúan como nodos. Una GCN clasifica escenas clave que permiten generar resúmenes estructurados, mejorando métricas como *Total Agreement* y *Partial Agreement* (ver figura 2).

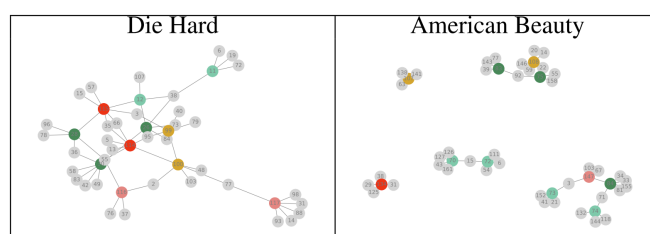


Figura 2: Grafo narrativo de escenas clave basado en [7].

3. **Preentrenamiento narrativo y comprensión estructural.** Zhao et al. [8] introducen la técnica *Learning by Narrating*, adaptando BART para mejorar la clasificación zero-shot de etapas narrativas. Se entrena con pares diálogo-sinopsis para captar progresiones dramáticas.
4. **Tipologías narrativas y arcos emocionales.** Reagan et al. [9] identifican seis arcos emocionales comunes en la narrativa literaria, y otros estudios los han extrapolado al cine. Las curvas sentimentales obtenidas se han correlacionado con métricas de éxito comercial [10].
5. **Herramientas IA para guionistas.** Han surgido sistemas interactivos como:
 - *ScriptViz* [11], que genera imágenes de referencia a partir de texto usando modelos CLIP.
 - *Portrayal* [12], que analiza la evolución de personajes mediante técnicas de *clustering* y visualización temporal.
 Estas herramientas enriquecen la escritura mediante análisis estructural y visualización narrativa.
6. **Generación automática de descripciones.** AutoAD [13] integra información visual y textual para producir descripciones coherentes de escenas. Emplea modelos como CLIP y GPT para mejorar la comprensión contextual en narrativa audiovisual.

7. **Análisis del sentimiento en guiones.** Estudios como [10] muestran cómo las curvas de sentimiento a lo largo del guion se relacionan con el éxito de las películas. Se han aplicado modelos supervisados (SVM, Transformers) y técnicas de aprendizaje transferido para analizar emociones en diálogos.
8. **Lingüística de corpus y anotación estructurada.** El trabajo de Oostdijk [14] ofrece un marco formal para la construcción y evaluación de corpus anotados. Su enfoque sobre precisión, cobertura y coherencia resulta clave para validar herramientas automáticas en PLN y análisis narrativo.

3. Técnicas de clasificación

Las herramientas de clasificación constituyen un componente fundamental en el análisis computacional de textos narrativos. Estas técnicas permiten asignar automáticamente etiquetas o categorías a fragmentos de texto, facilitando la identificación de elementos narrativos como escenas clave, emociones, o etapas específicas dentro de una estructura narrativa. En el contexto del análisis de guiones, las herramientas de clasificación posibilitan detectar patrones y momentos significativos sin intervención manual, agilizando así el proceso de estudio y comparación de múltiples obras.

3.1 Modelos supervisados

Los modelos supervisados son técnicas de aprendizaje automático que requieren un conjunto de datos previamente etiquetado para aprender a clasificar nuevos ejemplos. En el análisis narrativo, estos modelos se entrenan con fragmentos de texto donde se han identificado manualmente elementos narrativos específicos, como escenas, emociones o eventos clave. Una vez entrenados, pueden predecir con precisión estas categorías en guiones nuevos, siempre que sean similares a los ejemplos de entrenamiento. Aunque su desempeño suele ser alto cuando se dispone de datos adecuados, su principal limitación radica en la necesidad de contar con grandes volúmenes de datos anotados, lo que puede ser costoso y laborioso de obtener.

3.2 Clasificación *Zero-Shot*

La clasificación *zero-shot* (o *zero-shot classification*) es una técnica de aprendizaje automático que permite a un modelo asignar etiquetas a ejemplos sin haber sido entrenado explícitamente con ejemplos de esas etiquetas durante el entrenamiento. En otras palabras, el modelo puede generalizar a nuevas clases no vistas previamente, utilizando únicamente descripciones textuales o semánticas de las etiquetas. Una

aplicación conocida de esta técnica es el traslado de conocimiento obtenido en un idioma a otro del que generalmente se tiene menor cantidad de información [15].

Este enfoque se basa en modelos de lenguaje preentrenados a gran escala, como los modelos basados en arquitecturas transformer (por ejemplo, BART [16], RoBERTa [17] o GPT [18]), que han sido entrenados con grandes cantidades de datos no estructurados. Gracias a esto, estos modelos adquieren una comprensión contextual profunda del lenguaje que les permite comparar el significado de una entrada con una lista arbitraria de etiquetas.

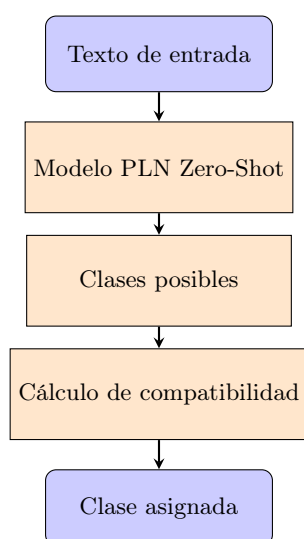


Figura 3: Funcionamiento del modelo de clasificación *Zero-Shot*

En la clasificación *zero-shot*, el modelo convierte tanto el texto de entrada como las etiquetas candidatas en representaciones vectoriales (*embeddings*) en un mismo espacio semántico. Luego, evalúa la similitud entre ellos para determinar cuál etiqueta es la más adecuada. Por ejemplo, se puede aplicar a tareas como:

- Clasificar reseñas en categorías como positiva, negativa, neutral, sin haber sido entrenado específicamente con esos términos.
- Identificar el tipo de evento en una escena narrativa (ej. conflicto, revelación, clímax) con solo proporcionar estas etiquetas como opciones.
- Clasificar elementos narrativos dentro de partes de una estructura narrativa.

Para identificar escenas que corresponden a un *change of plans*, se proporciona al modelo una breve definición de ese concepto (por

ejemplo, “La escena en la que el personaje principal cambia de objetivo o descubre un nuevo obstáculo que le obliga a tomar otro rumbo”) y se compara esta descripción con el contenido de cada escena. El modelo asigna a cada escena la etiqueta más compatible según su comprensión del significado, sin necesidad de entrenamiento adicional.

Sin embargo, esta metodología también presenta desafíos importantes. Uno de los principales es la sensibilidad a la formulación de las etiquetas: cambios mínimos en la redacción pueden alterar los resultados de clasificación. Además, el modelo puede tener dificultades para captar relaciones narrativas implícitas o para interpretar correctamente escenas que dependen de un contexto más amplio, lo que puede conducir a errores de asignación. Aun así, los resultados muestran que esta técnica es una herramienta viable para iniciar el análisis automático de estructuras narrativas en guiones.

4. Técnicas y herramientas de visualización

Uno de los objetivos principales de este trabajo es no solo detectar patrones narrativos a partir de guiones cinematográficos, sino también representarlos de forma visual para facilitar su análisis por parte de guionistas e investigadores. Para ello, se han empleado diversas técnicas de visualización de datos interactivas que permiten explorar tanto la estructura dramática como los momentos clave de cada guion.

Todas las representaciones gráficas han sido desarrolladas empleando la biblioteca `Plotly` en su versión para Python, lo que ha permitido generar visualizaciones dinámicas e interactivas fácilmente integrables en entornos web.

1. **Interfaz interactiva para el análisis narrativo.** Se ha desarrollado una herramienta basada en `Dash` y `Plotly` que permite navegar por las estructuras detectadas en los guiones, filtrando por fases del *Viaje del Héroe* o tipos de *turning points*. Incluye un gráfico de anillo (*donut chart*) generado con `plotly.express.pie(hole=0.5)` que resume la proporción de escenas asignadas a cada fase y permite interacciones para centrar el análisis.
2. **Línea temporal de momentos clave.** Se ha implementado una visualización que representa los *turning points* relevantes sobre una escala temporal normalizada (de 0 a 1). Cada tipo de momento narrativo se identifica con un color único y una intensidad proporcional a la confianza (*score*) del modelo.
3. **Distribuciones y estadísticas del corpus.** Para contextualizar los datos, se han generado histogramas y gráficos de barras que muestran aspectos como la distribución de películas por década, género y longitud de guion, lo cual ayuda a identificar patrones globales en el conjunto analizado.
 - La distribución de películas por década.

- La proporción de guiones por género cinematográfico.
- La evolución de la longitud media de los diálogos por década.

Estas visualizaciones permiten observar tendencias temporales y diferencias estructurales entre géneros, sirviendo de contexto a los análisis más específicos de las estructuras narrativas.

4.1 Representación del Viaje del Héroe por escena

Finalmente, se ha desarrollado una visualización que recorre escena por escena la asignación de fases del *Viaje del Héroe*, permitiendo observar transiciones, ausencias o agrupamientos de fases. Esta visualización se construye a partir de una matriz de calor donde el eje horizontal representa la progresión del guion (por escena o por posición normalizada) y el eje vertical las diferentes fases del viaje, coloreando los puntos donde se detecta correspondencia.

Análisis narrativo: Viaje del Héroe

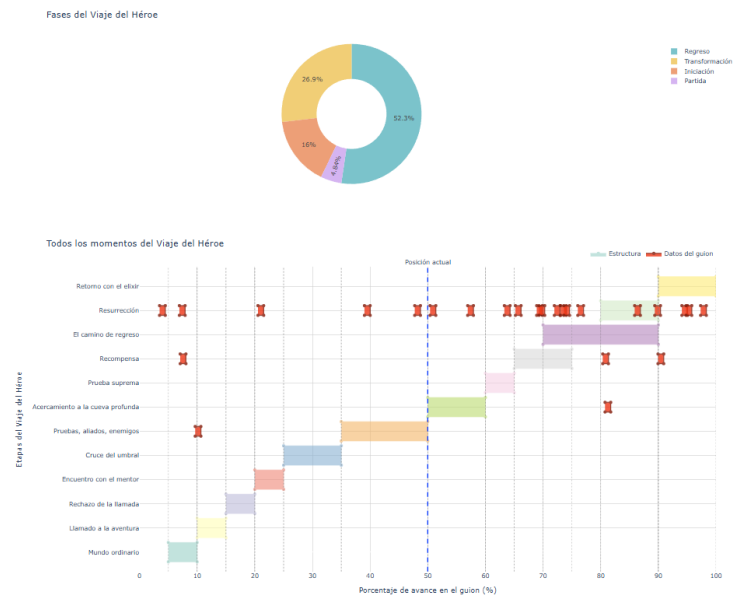


Figura 4: Herramienta de visualización desarrollada

4.2 Representación visual de la adecuación al patrón narrativo

En el contexto de este trabajo, se ha empleado un enfoque *zero-shot* para asignar fases narrativas a las escenas de los guiones sin

requerir entrenamiento supervisado específico. Dado que no se dispone de una verdad de referencia completa para cada guion, las métricas tradicionales de evaluación (como *precision*, *recall* o *F1-score*) no resultan adecuadas. En su lugar, se han definido métricas específicas para evaluar el grado de alineación con la estructura narrativa teórica.

- **Exactitud directa (*Accuracy*):** proporción de escenas clasificadas con la fase más probable que coincide con la esperada en un patrón ideal.
- **Distancia narrativa media:** medida basada en la posición relativa de cada fase dentro de la secuencia teórica. Se calcula la distancia entre la fase predicha y la fase esperada dentro del patrón narrativo (por ejemplo, del *Viaje del Héroe*), penalizando más aquellas asignaciones fuera de orden.
- **Cobertura de fases (*Coverage*):** porcentaje de fases del patrón que están representadas al menos una vez en la predicción para el guion.

Estas métricas se visualizan mediante un gráfico de barras horizontales donde cada barra representa una métrica para un guion determinado. La exactitud se representa con una barra sólida, la distancia narrativa media se visualiza de forma invertida (a menor valor, mayor ajuste), y la cobertura se indica mediante un marcador sobre la barra. Esta representación facilita identificar rápidamente guiones que muestran una buena adecuación global al patrón frente a aquellos que presentan desviaciones significativas.

4.3 Gráficas para explorar el dataset

El tipo de gráficas empleadas para explorar el conjunto de datos será principalmente el histograma, utilizado para representar tanto el número de guiones por década como la distribución de películas por año y por género. Este tipo de visualización resulta especialmente adecuado para variables categóricas o discretas, ya que permite observar de forma clara la frecuencia de aparición de cada categoría, facilitando la detección de tendencias generales o desequilibrios en la muestra.

Por otro lado, para analizar la evolución de la longitud promedio de los diálogos a lo largo del tiempo, se opta por una gráfica lineal. Esta elección permite visualizar con claridad la progresión temporal de una variable continua, resaltando posibles patrones de crecimiento, estancamiento o disminución en la extensión de los diálogos a lo largo del corpus.

5. Metodología

La metodología empleada para la realización del presente trabajo ha consistido en la investigación de diferentes artículos en diferentes

repositorios y el posterior desarrollo de una herramienta de extracción de información a partir del conjunto de datos, comúnmente denominado *dataset*, de guiones de cine que se explorará en siguiente apartado.

5.1 Justificación de la relevancia del estudio en el campo de investigación

A partir de la revisión de la literatura, se identifican varios *gaps* importantes en la investigación actual:

- Desconexión entre teoría narratológica y aplicaciones computacionales: Como señalan Piper et al. (2021) [19], existe una notable separación entre el trabajo teórico en narratología y las aplicaciones prácticas en PLN.
- Limitaciones en la comprensión cognitiva: Zhu et al. (2023) [20] destacan que los modelos actuales, aunque técnicamente sofisticados, no capturan completamente los procesos cognitivos subyacentes en la creación narrativa.
- Falta de herramientas integradas para guionistas: Mientras que trabajos como el de Papalampidi et al. (2019) [21] desarrollan técnicas para identificar elementos estructurales, no existe una integración de estas técnicas en herramientas prácticas para creadores de contenido.
- Ausencia de análisis comparativo sistemático: La literatura carece de estudios que permitan comparar sistemáticamente diferentes estructuras narrativas y sus patrones.

El presente trabajo busca abordar estos *gaps* mediante el empleo de técnicas y de herramientas de visualización que permitan visualizar la información extraída mediante lingüística computacional y análisis de datos, proporcionando a los guionistas una herramienta visual para analizar y comparar estructuras narrativas de manera sistemática.

Algunos trabajos recientes han aplicado técnicas de clasificación *zero-shot* para asignar etiquetas narrativas de forma automática a segmentos de texto [22], mientras que otros se centran en la detección manual o automática de *puntos clave* narrativos en función de modelos estructurales [7]. Sin embargo, apenas se han explorado propuestas que combinen ambos enfoques de forma coordinada para analizar la alineación de los guiones con estructuras narrativas clásicas.

5.2 Conjunto de datos empleado

Para llevar a cabo el análisis de los patrones narrativos en guiones cinematográficos, se ha empleado el corpus **Movie Scripts Corpus** [23], disponible en la plataforma **Kaggle**. Este conjunto de datos ha sido recopilado con el objetivo de facilitar el análisis de guiones mediante técnicas de aprendizaje automático y procesamiento del lenguaje natural.

El corpus incluye 2858 guiones completos en formato texto, tanto en su versión original como lematizada. Además, contiene anotaciones estructurales realizadas de manera manual (en 33 documentos) y automática, empleando modelos como *BERT* [24] y el anotador basado en reglas ScreenPy [25]. Estas anotaciones identifican componentes clave de la estructura narrativa, como escenas, descripciones, diálogos o transiciones que podrán ser utilizadas para extraer información relevante para el estudio.

Es importante señalar algunas limitaciones inherentes al propio conjunto de datos. En primer lugar, aunque el corpus tiene un tamaño considerable, su distribución por géneros, décadas y países no es uniforme, lo cual puede introducir sesgos en los análisis realizados. La predominancia de producciones estadounidenses y de ciertos géneros mayoritarios como el drama o la comedia puede condicionar la representación de otras formas narrativas menos convencionales o pertenecientes a otras culturas cinematográficas.

Asimismo, la anotación automática de las estructuras narrativas, aunque útil para facilitar el análisis a gran escala, puede no capturar adecuadamente los matices de guiones con estructuras no lineales, múltiples tramas o experimentación formal. El enfoque utilizado en este trabajo, centrado en el análisis textual y estructural de la información contenida en los guiones, no incorpora aspectos audiovisuales esenciales para la construcción narrativa, como la música, la puesta en escena o la interpretación de los actores.

Estas limitaciones no desmerecen el valor del corpus, pero sí invitan a interpretar los resultados con cautela y a considerar posibles extensiones futuras que incorporen una mayor diversidad de fuentes y una aproximación más multimodal al análisis narrativo.

5.3 Patrón narrativo empleado

El análisis narrativo llevado a cabo en este trabajo se basa en la detección automática de momentos clave dentro de guiones cinematográficos. Para evaluar la coherencia estructural de los mismos y estudiar su alineación con formas narrativas consolidadas, se han seleccionado dos modelos ampliamente aceptados en el ámbito de la teoría narrativa: el *Viaje del Héroe* y la *Pirámide de Freytag*.

La elección de estos dos marcos responde a su complementariedad y a su relevancia tanto teórica como práctica:

- El **Viaje del Héroe**, propuesto por Joseph Campbell y adaptado por autores como Christopher Vogler, es un modelo arquetípico que describe la evolución del protagonista a lo largo de una historia de transformación personal. Su enfoque centrado en el personaje resulta especialmente útil para analizar el desarrollo emocional y simbólico del héroe en narrativas orientadas al viaje interior o al

crecimiento individual, elementos frecuentes en el cine narrativo contemporáneo.

- Por otro lado, la **Pirámide de Freytag** proporciona una visión estructural más general de la historia, basada en la progresión de la tensión dramática. Este modelo permite observar cómo se construyen y resuelven los conflictos narrativos a nivel global, independientemente del arco específico de un personaje. Su aplicación es especialmente pertinente en guiones de géneros como el drama, el thriller o la tragedia clásica, donde la estructura en cinco actos continúa vigente.

El uso combinado de ambos modelos permite obtener una representación más rica y multi-dimensional de la narrativa: por un lado, uno de ellos permite evaluar la evolución del protagonista como motor del relato; por otro, se analiza la dinámica global de la historia en términos de tensión y resolución. Esta doble perspectiva contribuye a una caracterización más precisa de la estructura narrativa de los guiones y permite detectar patrones recurrentes o desviaciones significativas respecto a las formas canónicas.

5.4 Análisis exploratorio de los datos

En un primer lugar, como se ve en la figura 5, el conjunto de películas del dataset puede considerarse variado en términos de años de estreno de las películas, con una notable abundancia de ejemplos de películas entre los años 2000 y la década de 2010 como puede verse en la figura 6. Por otro lado, clasificándolo por género de películas contiene predominantemente dramas y comedias como se ve en la figura 7.

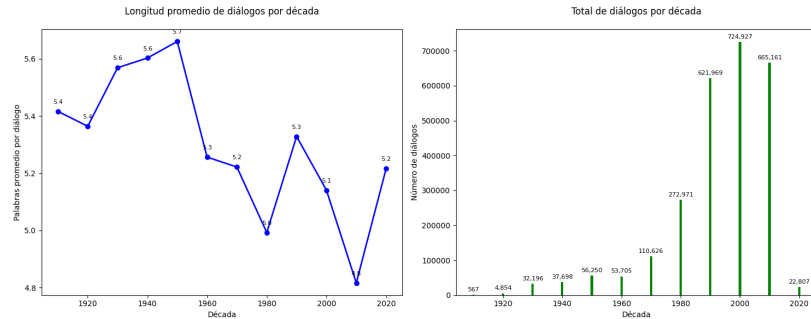


Figura 5: Análisis exploratorio de los diálogos por década

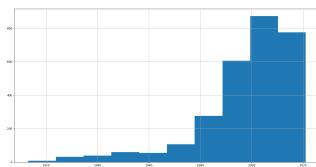


Figura 6: Distribución de las películas por año

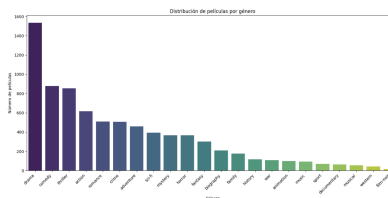


Figura 7: Distribución de las películas por género

5.4.1 Estructura de un fichero de guion utilizado El *Movie Scripts Corpus de Gufukuro* [23] reúne casi 3000 guiones cinematográficos en dos formatos principales, junto con metadatos descriptivos para cada película. A continuación se detallan los elementos que lo componen:

- Texto completo sin procesar (.txt): Un directorio con 2858 archivos de texto que contienen el guion íntegro tal como aparece en su fuente original. Cada fichero incluye tanto las indicaciones de escenario (slug lines) como los diálogos y las descripciones de acción, pero sin ningún marcado adicional.
- Anotaciones estructurales (.txt o .json): Un segundo conjunto de 2858 documentos con etiquetas por elementos narrativos (por ejemplo, SCENE HEADING, ACTION, CHARACTER, DIALOGUE, PARENTHETICAL). Estas anotaciones permiten segmentar automáticamente cada guion en unidades funcionales: escenas, descripciones, intervenciones de personajes, etc. Podemos ver un ejemplo para la película *The Matrix (1999)* en la tabla 1.
- Metadatos (.csv): Un único fichero CSV que acompaña al corpus. Cada fila corresponde a una película e incluye columnas como título, año, género, director, guionista, duración, país de origen y, en algunos casos, valoraciones de IMDb.

- Con esta estructura, el corpus facilita tanto el análisis “a distancia” de la narrativa (lectura cuantitativa de flujos y patrones) como la extracción de características específicas por escena o por elemento de guion.

movie_id	title	line_type	text
1234	The Matrix	scene heading	INT. NEBUCHADNEZZAR - MAIN DECK
1234	The Matrix	action	Neo watches the monitors closely.
1234	The Matrix	character	MORPHEUS
1234	The Matrix	dialogue	The Matrix is everywhere.

Tabla 1: Ejemplo de elementos en el archivo `structured_scripts.csv` relativo a la película *The Matrix (1999)*

5.5 Creación de los *scripts* necesarios

5.5.1 Implementación del extractor de información del diálogo Con el objetivo de analizar patrones narrativos y expresivos en guiones de cine, se diseñó un script en Python que permite procesar tanto los metadatos de las películas como sus diálogos. Este proceso incluye varias fases:

1. Carga y limpieza de metadatos
2. Mapeo entre películas y géneros
3. Extracción de diálogos
4. Filtrado léxico y construcción de un diccionario de frecuencia
5. Generación de n-gramas y frases frecuentes

Se parte de un conjunto de datos `movie_meta_data.csv` que contiene información sobre películas, incluyendo título, año y géneros. El script se encarga de cargar el archivo CSV con `pandas`, eliminar registros con títulos o géneros vacíos, filtrar películas publicadas desde el año 1900 y de normalizar los títulos (conversión a minúsculas, eliminación de caracteres no alfabéticos) para facilitar la coincidencia con los guiones.

5.5.2 Procesamiento de los guiones y detección de estructuras narrativas Con el objetivo de analizar la evolución narrativa a lo largo del desarrollo de los guiones, se ha procesado el archivo `structured_scripts.csv`, que contiene los guiones cinematográficos segmentados línea por línea, con anotaciones sobre su tipología textual (*scene heading*, *action*, *character*, *dialogue*, etc.).

El primer paso del procesamiento consistió en filtrar las líneas relevantes para el análisis narrativo, centrándose principalmente en los segmentos de tipo *action* y *dialogue*, al considerar que son los que mejor reflejan la progresión narrativa de la historia. Posteriormente, se

agruparon las líneas por película y se concatenaron en un único bloque textual por guion, preservando el orden original.

A partir de estos bloques textuales se aplicaron distintas técnicas de procesamiento del lenguaje natural (PLN), incluyendo:

- **Segmentación por escenas:** mediante las marcas de tipo *scene heading*, se dividieron los guiones en unidades narrativas equivalentes a escenas cinematográficas.
- **Análisis de similitud semántica y detección de temas:** se utilizaron representaciones vectoriales (*embeddings*) y modelos de clasificación para asignar a cada escena una o varias etiquetas narrativas, como por ejemplo *llamada a la aventura*, *clímax* o *resolución*.
- **Detección de momentos clave:** en base a las etiquetas obtenidas, se identificaron los momentos narrativos más representativos de cada guion. Estos momentos fueron posteriormente comparados con estructuras narrativas clásicas como la *Pirámide de Freytag* o el *Viaje del Héroe*, con el fin de evaluar el grado de alineación de los guiones con dichas estructuras.

Finalmente, se generaron visualizaciones temporales que representan gráficamente la aparición de los elementos narrativos a lo largo del desarrollo del guion. En lugar de utilizar marcas temporales absolutas, se empleó una escala relativa basada en el porcentaje de avance dentro del texto, permitiendo así comparar guiones de diferente longitud en una misma representación.

Estas visualizaciones fueron clave para el análisis comparativo entre películas y para observar patrones recurrentes en la estructura narrativa de los guiones estudiados.

5.6 Obtención de los patrones de estructuras narrativas

Para poder comparar los datos obtenidos de los guiones, será necesario obtener a partir de las estructuras narrativas que se exponen a continuación unas visualizaciones sobre las cuales mostrar los puntos obtenidos. La metodología a seguir, descrita en la figura ??.

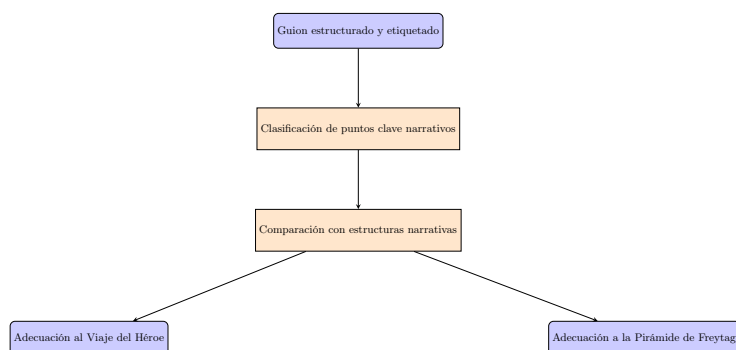


Figura 8: Metodología seguida para el análisis narrativo de guiones

5.6.1 El viaje del héroe El Viaje del Héroe o *Monomito*, es un modelo narrativo dividido en doce etapas que describen la evolución arquetípica de un protagonista desde su mundo ordinario hasta su transformación final. El término viaje del héroe como denominación de esta estructura fue acuñado y popularizado por Joseph Campbell en su obra *El héroe de las mil caras* (1949) [4], mientras que el término *Monomito* lo tomó de James Joyce y su obra *Finnegans Wake*, conocida por ser una de las novelas más difíciles de comprender de la literatura inglesa debido a su convulsa estructura. Posteriormente fue adaptado por Christopher Vogle [26] en su obra. Se estructura de la siguiente forma:

1. **Mundo ordinario:** El héroe se presenta en su entorno habitual, mostrando su vida cotidiana y limitaciones.
2. **Llamado a la aventura:** Surge un desafío o una invitación que rompa la rutina y propone un objetivo fuera de lo conocido.
3. **Rechazo de la llamada:** El héroe duda o rechaza inicialmente el desafío por miedo, inseguridad o lealtades previas.
4. **Encuentro con el mentor:** Un guía o figura sabia aporta consejo, herramientas o motivación para emprender el viaje.
5. **Cruce del primer umbral:** El héroe abandona su mundo ordinario y entra en el “mundo especial”, más peligroso y desconocido.
6. **Pruebas, aliados y enemigos:** Se suceden obstáculos y encuentros con compañeros y antagonistas que ponen a prueba al héroe.
7. **Acercamiento a la caverna más profunda:** El héroe se prepara para enfrentar su mayor desafío, física o emocionalmente, acercándose al corazón del conflicto.
8. **Prueba suprema:** Momento culminante en el que el héroe afronta el peligro más grande: puede ser la batalla final o una crisis interna.
9. **Recompensa:** Tras superar la prueba, obtiene el objeto deseado, una revelación o un poder que reconoce su valía.
10. **El camino de vuelta:** Inicia el retorno al mundo ordinario, a menudo con nuevos peligros o persecuciones.

11. **Resurrección:** Última confrontación donde el héroe demuestra su transformación al enfrentarse a un último reto que redefine su ser.
12. **Regreso con el elixir:** De vuelta en su mundo original, el héroe comparte el “elixir”, conocimiento, objeto o poder, beneficiando a su comunidad.

En la figura 9 podemos ver gráficamente su patrón visual.



Figura 9: Estructura narrativa del viaje del Héroe [27]

La Pirámide de Freytag, como se ve en la figura 1, es un modelo clásico de análisis de la estructura dramática desarrollado por el novelista y crítico literario alemán Gustav Freytag en su obra *Die Technik des Dramas* (1863) [28]. Inspirado por las tragedias griegas y las obras teatrales de Shakespeare, Freytag propuso una estructura en cinco actos que representa el desarrollo típico de la tensión dramática en una obra narrativa.

A diferencia del *Viaje del Héroe*, que enfatiza la evolución del protagonista a través de un ciclo vital arquetípico, la Pirámide de Freytag se centra en la progresión dramática de la historia como un todo, utilizando una forma triangular que simboliza el ascenso y descenso de la tensión narrativa.

Su estructura se divide en las siguientes cinco fases principales [6]:

1. **Exposición:** Introduce el contexto de la historia, incluyendo los personajes principales, el escenario y el conflicto inicial. Sirve como base para entender los eventos posteriores.
2. **Ascenso de la acción (rising action):** Se desarrollan los conflictos, aparecen obstáculos y se intensifica la tensión. Es la fase más extensa de la historia y marca la acumulación del drama.
3. **Clímax:** Punto culminante y de mayor tensión narrativa. Representa un giro decisivo que determina el destino de los personajes principales.

4. **Descenso de la acción (falling action):** Tras el clímax, se desencadenan las consecuencias de las decisiones tomadas. La tensión comienza a disminuir y se prepara el desenlace.
5. **Desenlace (denouement):** Se resuelven los conflictos y se restablece un nuevo equilibrio. La historia concluye, dejando al lector o espectador con una sensación de cierre.

Este modelo ha sido ampliamente utilizado en la crítica literaria, el análisis cinematográfico y la escritura de guiones, al proporcionar una visión clara de cómo se distribuye la tensión narrativa a lo largo del tiempo. Su simplicidad lo hace especialmente útil para comparar guiones entre sí y detectar patrones estructurales comunes.

Relación entre la Pirámide de Freytag y los Turning Points analizados
La Pirámide de Freytag ofrece un marco conceptual útil para comprender la progresión dramática y la distribución de la tensión narrativa en una obra. Los *turning points* o momentos clave detectados automáticamente en los guiones, como *opportunity*, *change of plans*, *point of no return*, *major setback* y *climax Final*, pueden ser interpretados como hitos narrativos que corresponden a fases concretas dentro de esta estructura clásica.

Por ejemplo, los momentos de *opportunity* suelen coincidir con la fase de **exposición** o el inicio del **ascenso de la acción**, donde se presentan las circunstancias que impulsan la historia. Los *change of plans* marcan puntos en los que el conflicto se intensifica o se redefine, típicamente durante el **ascenso de la acción**, aumentando la tensión narrativa.

El *point of no return* se corresponde con el acercamiento al **clímax**, cuando las decisiones de los personajes conducen a una situación irreversible que determinará el desenlace. Los *major setback* representan obstáculos significativos que podrían vincularse al **descenso de la acción**, momento en que la tensión comienza a resolverse y las consecuencias de los eventos anteriores se manifiestan.

Finalmente, el *climax Final* coincide con el **clímax** propiamente dicho o con el **desenlace**, donde se resuelven los conflictos y se restablece el equilibrio narrativo.

Este paralelismo entre los *turning points* identificados automáticamente y las fases de la Pirámide de Freytag valida el enfoque empleado, pues permite mapear la estructura narrativa detectada en los guiones con un modelo dramático clásico ampliamente aceptado. Además, facilita la interpretación y comparación de los resultados, proporcionando una visión clara y visual de cómo se distribuyen los momentos narrativos clave a lo largo del desarrollo de la historia.

6. Caso de estudio

El caso de estudio del presente trabajo se dividirá en dos partes, una primera donde empleando los métodos presentados para obtener turning points y un segundo apartado en el que se construirá una visualización acorde para representar los datos obtenidos, es decir, los *turning points per sé* además de intentar trasladar esta aproximación a la estructura narrativa *el viaje del héroe*.

6.1 Clasificación

Partimos del supuesto de que tomamos un guion de una película para tratar de analizar la distribución de puntos clave en la misma y cuanto se ajusta al patrón del *viaje del héroe* para obtener una buena visualización de la información que nos permita obtener conclusiones relevantes.

A modo de ejemplo, se selecciona la película *12 años de esclavitud*. Lo primero que se hará será mediante clasificación *zero-shot* mediante el modelo preentrenado de Facebook *bart-large-mnli* [29] sobre el conjunto de datos *MultiNLI*, que contiene una colección de 433000 pares de frases anotadas [30].

La estrategia de clasificación consiste en la división del guión por escenas, la posterior agrupación de los textos junto con las anotaciones de escenarios y de actuación y los resultados propuestos se pasan al clasificador *zero-shot*.

Los resultados de los momentos detectados clasificados según la confianza que otorga el modelo a la clasificación los podemos ver en la figura 10, aparecen representados inicialmente aquellos elementos que tengan una confianza superior a 0,5. Como podemos ver para este ejemplo concreto no se encuentran TP con una confianza superior al umbral de todos los tipos definidos.

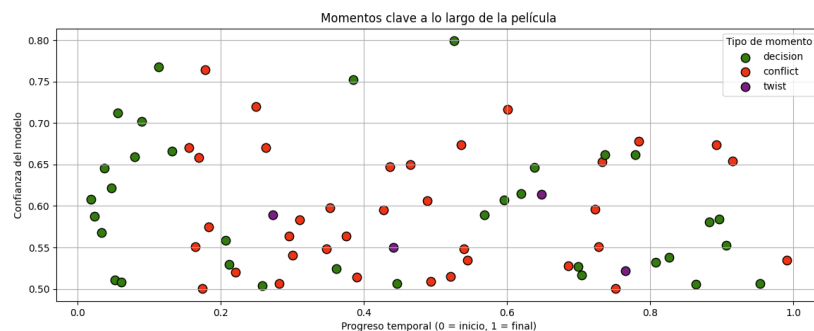


Figura 10: Momentos clave detectados en la película "12 años de esclavitud" (2013) detectados mediante la clasificación *zero-shot*

Al limitar la salida al mejor resultado clasificado de cada tipo, como se ve en la figura 11, podemos ver que aún faltando elementos, es capaz de clasificar correctamente en secuencia los turning points con mayor presencia en el guion.

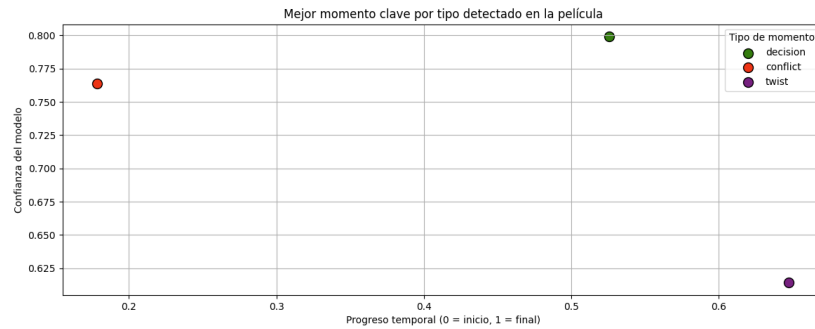


Figura 11: Filtrando a mejor resultado obtenido por clasificación *zero-shot*

Tras la ejecución del script se obtiene como salida un fichero `.json` que contiene todos los elementos clasificados como se indica a continuación:

1. **Scene Index:** Identificador numérico de la escena dentro del guion.
2. **Scene Heading:** Cabecera de descripción de la escena identificada
3. **Moment:** Etiqueta asignada por el modelo (por ejemplo, *decision*, *climax*, etc.). Para el caso del *viaje del héroe* se indicará como phase y la fase en sí.
4. **Score:** Puntuación de confianza o probabilidad asociada a la etiqueta, en rango $[0,1]$.
5. **Position (relative):** Posición del bloque dentro del guion, normalizada en $[0,1]$.

Tabla 2: Datos de un momento clave narrativo: Ejemplo de "12 años de esclavitud" (2013)

Campo	Valor
Scene Index	150
Scene Heading	"112 INT. EPPS'S PLANTATION/SLAVE SHACK - DAY 112"
Moment	decision
Score	0.5166
Position (relative)	0.7042

Tabla 3: Datos de una fase del *viaje del héroe*: Ejemplo de *12 años de esclavitud* (2013)

Campo	Valor
Scene Index	145
Scene Heading	INT. MASTER EPPS'S PLANTATION/SLAVE SHACKS - NIGHT"
Phase	point of no return
Score	0.5648
Position (relative)	0.6808

A continuación se emplean estos momentos clave como partida sobre el siguiente clasificador, que será el encargado de indicar a qué parte de la estructura narrativa se corresponde. Para extraer la secuencia más larga posible de momentos narrativos que cumpla simultáneamente dos condiciones —el orden predefinido de tipos de momentos y la progresión temporal dentro del guion— se implementó un algoritmo basado en programación dinámica. Dado un conjunto de momentos detectados en una película, cada uno con un tipo (*moment*) y una posición normalizada temporal (*position*), se busca la subsecuencia más larga tal que:

- El orden de los tipos de momentos respeta la jerarquía narrativa establecida, definida como:

opportunity → change of plans → point of no return →
major setback → climax final

- La posición de cada momento en la secuencia es estrictamente creciente.

Para el caso del *viaje del héroe* se procede de la siguiente manera:

- El orden de las fases respeta la secuencia narrativa establecida del *viaje del héroe*, definida como:

ordinary world → call to adventure → refusal of the
call → meeting with the mentor → crossing the first
threshold → tests, allies, enemies → approach to the
inmost cave → ordeal → reward → the road back →
resurrection → return with the elixir

- La posición de cada fase en la secuencia es estrictamente creciente.

6.1.1 Implementación El algoritmo considera cada momento ordenado según su posición temporal ascendente. Para cada momento i , se

almacena en una lista $dp[i]$ la subsecuencia más larga que termina en dicho momento y cumple las condiciones anteriores.

Para calcular $dp[i]$, se itera sobre todos los momentos j anteriores ($j < i$) y se verifica si:

- El tipo del momento j precede al tipo del momento i en el orden narrativo definido.
- La posición temporal del momento j es menor que la del momento i .
- Extender la subsecuencia almacenada en $dp[j]$ con el momento i genera una subsecuencia más larga que la almacenada actualmente en $dp[i]$.

Si estas condiciones se cumplen, se actualiza $dp[i]$ como $dp[j]$ concatenado con el momento i .

Finalmente, la subsecuencia más larga se obtiene seleccionando la lista con mayor longitud dentro de todas las $dp[i]$.

Este algoritmo no nos garantiza que siempre se vaya a encontrar una subsecuencia que incluya los 5 tipos de *moments*, pero sí que los que aparecen están en orden en cuanto a la progresión narrativa y a su orden de aparición en el guion.

Posteriormente, para cada película, se almacena la secuencia narrativa más larga calculada y se exporta un archivo JSON global que contiene, para cada película, la subsecuencia obtenida con su información completa (índice de escena, encabezado, tipo de momento, puntuación y posición).

De esta forma se garantiza una identificación estructurada y cronológicamente coherente de los momentos narrativos clave en los guiones analizados previamente por el clasificador *zero-shot*.

6.2 Visualización de la progresión narrativa mediante gráficos

Para representar gráficamente la progresión narrativa de los *turning points* detectados en cada guion, se desarrolló un script en Python que procesa un fichero JSON con la estructura de momentos clave y genera una imagen por cada guion, mostrando la posición relativa de cada momento narrativo dentro del guion.

El script realiza los siguientes pasos principales:

- **Carga de datos:** Se lee un fichero JSON que contiene, para cada guion, una lista de momentos narrativos con atributos tales como el índice de escena, encabezado, tipo de momento (*opportunity*, *change of plans*, etc.), puntuación de confianza y posición relativa normalizada dentro del guion.

- **Definición de rangos narrativos:** Se establecen rangos esperados de posición para cada tipo de momento narrativo según un modelo clásico de estructura narrativa basado en la Pirámide de Freytag que se aplica a los TP clasificados:
 - *opportunity*: posición entre 0.0 y 0.25
 - *change of plans*: posición entre 0.25 y 0.4
 - *point of no return*: posición entre 0.4 y 0.6
 - *major setback*: posición entre 0.6 y 0.8
 - *climax final*: posición entre 0.8 y 1.0
- **Definición de rangos narrativos:** Se establecen además los rangos esperados de posición para cada fase del *viaje del héroe*, dividiendo la progresión narrativa en 12 segmentos equivalentes:
 - *ordinary world*: posición entre 0.00 y 0.08
 - *call to adventure*: posición entre 0.08 y 0.17
 - *refusal of the call*: posición entre 0.17 y 0.25
 - *meeting with the mentor*: posición entre 0.25 y 0.33
 - *crossing the first threshold*: posición entre 0.33 y 0.42
 - *tests, allies, enemies*: posición entre 0.42 y 0.50
 - *approach to the inmost cave*: posición entre 0.50 y 0.58
 - *ordeal*: posición entre 0.58 y 0.67
 - *reward*: posición entre 0.67 y 0.75
 - *the road back*: posición entre 0.75 y 0.83
 - *resurrection*: posición entre 0.83 y 0.92
 - *return with the elixir*: posición entre 0.92 y 1.00
- **Generación del gráfico:** Para cada guion, se crea una figura con `matplotlib` que incluye:
 - Áreas coloreadas semi-transparentes que representan los rangos esperados para cada tipo de momento.
 - Marcadores circulares que indican la posición real de cada momento detectado en el guion, coloreados de acuerdo con su tipo.
 - Etiquetas con información relevante de cada punto, como el tipo de momento, la puntuación de confianza y el índice de escena correspondiente.
- **Configuración visual y guardado:** Se ajustan límites y etiquetas del eje horizontal para reflejar la posición relativa dentro del guion (entre 0 y 1), se ocultan los ejes verticales para mayor claridad, y se añade una leyenda para identificar cada tipo de momento. Finalmente, la gráfica se guarda como imagen PNG en un directorio específico, generando un archivo por cada guion analizado.

Este enfoque visual facilita la interpretación de la distribución y progresión de los momentos narrativos clave, permitiendo comparar la estructura detectada con los rangos narrativos teóricos y evaluar la coherencia narrativa de cada guion de forma intuitiva y visual. Podemos ver los resultados obtenidos en las siguientes figuras 12, 13, 14 y 15 :

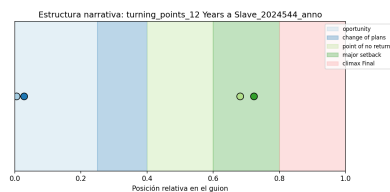


Figura 12: Gráfica de Turning Points de *12 años de esclavitud*

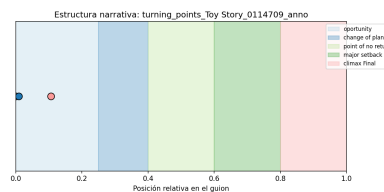


Figura 13: Gráfica de Turning Points de *Toy Story*

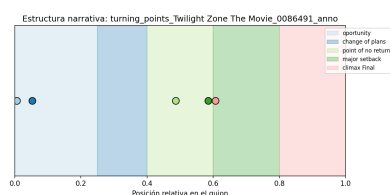


Figura 14: Gráfica de Turning Points de *Twilight Zone*

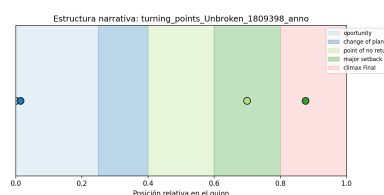


Figura 15: Gráfica de Turning Points de *Unbroken*

6.3 Aproximación al viaje del héroe

Para representar gráficamente la secuencia narrativa detectada en cada guion, se ha desarrollado una visualización que permite observar de forma simultánea la posición relativa de las fases del *viaje del héroe* y su correspondencia con la estructura teórica esperada. Esta visualización se genera automáticamente a partir del fichero `secuencias_maximas_hero_journey.json`, que contiene las fases narrativas detectadas con mayor coherencia en cada película analizada.

Cada figura consiste en una línea temporal normalizada (de 0 a 1) que representa el desarrollo del guion desde su inicio hasta su desenlace. Sobre ella se superponen rectángulos de color que indican los rangos esperados para cada una de las 12 fases del *viaje del héroe*, según una segmentación equitativa del eje temporal. Los puntos sobre la línea indican las fases narrativas detectadas por el modelo de clasificación, con su posición relativa en el guion. Cada punto se codifica por color (según la fase correspondiente) y se muestra con un contorno negro fino para resaltar su visibilidad sobre el fondo coloreado. La leyenda, ubicada debajo de la figura y organizada en dos columnas, permite identificar fácilmente a qué fase corresponde cada color.

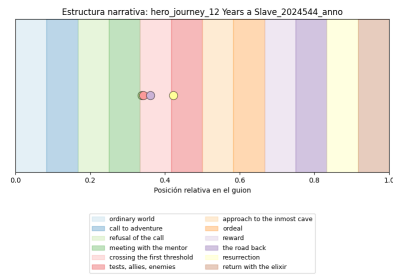


Figura 16: Estructura del *viaje del héroe* para 12 Years a Slave

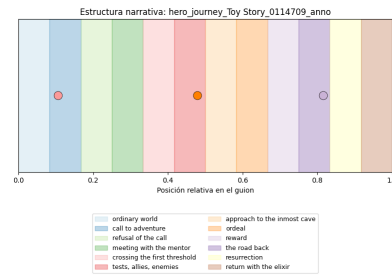


Figura 17: Estructura del *viaje del héroe* para Toy Story

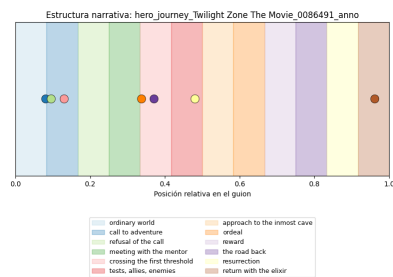


Figura 18: Estructura del *viaje del héroe* para Twilight Zone

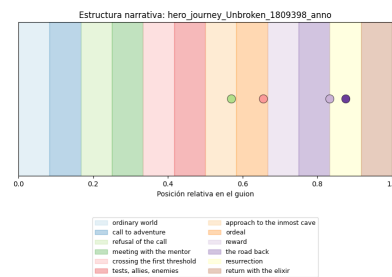


Figura 19: Estructura del *viaje del héroe* para Unbroken

7. Resultados y discusión

Si bien es cierto que es muy atractivo el hecho de poder emplear clasificación *zero shot* para detectar de una forma directa, sin necesidad de disponer de corpus anotados de entrenamiento, las escenas consideradas como *turning points*, los resultados obtenidos son en general poco precisos en cuanto a la localización de los puntos en el rango esperado, como se puede ver en las figuras 12, 13, 14 y 15. En el ámbito de la clasificación de *turning points* se han obtenido resultados bastante satisfactorios, sin embargo, al representar una estructura con más clases como lo son las fases del *viaje del héroe* el sistema se vuelve bastante más impreciso. Por tanto podemos considerar como un enfoque plausible el uso de clasificadores *zero-shot* para casos en los que no se disponga de corpus anotados. Podría ser especialmente útil para su implementación en lenguas diferentes al inglés donde la posibilidad de encontrar corpus anotados es más difícil.

Por la parte de la representación visual, se han conseguido obtener visualizaciones sencillas empleando gráficas que muestran los diferentes puntos claves que marcan la progresión narrativa de una historia.

El sistema es aplicable a diferentes estructuras narrativas y por tanto reutilizable.

7.1 Evaluación del sistema zero-shot para la *Pirámide de Freytag* / *TPs*

Dado que el sistema zero-shot empleado para la clasificación de momentos clave en los guiones cinematográficos no cuenta con etiquetas de referencia explícitas, se ha optado por utilizar métricas específicas para este tipo de modelos y análisis cualitativo basado en la coherencia narrativa. Podemos ver un resumen en la figura 20.

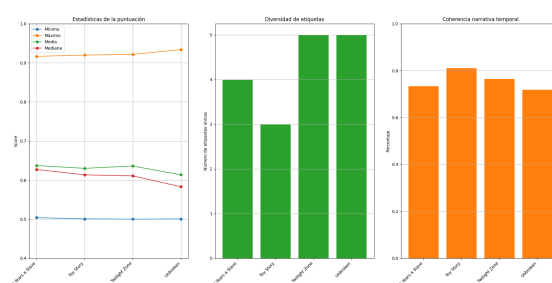


Figura 20: Evaluación del clasificador para *turning points*

Se aplicó un umbral de confianza de 0.5 sobre las predicciones obtenidas por el modelo, considerando únicamente aquellos momentos cuya puntuación superaba este valor para garantizar la fiabilidad de las asignaciones. Las métricas evaluadas y sus resultados promedio en diferentes guiones analizados son los siguientes:

- **Distribución de puntuaciones (Score statistics):** las puntuaciones asignadas mostraron un rango mínimo cercano al umbral ($\sim 0,50$) y máximos superiores al 0.91, con una media en torno a 0,63 y mediana cercana a 0,61, lo que indica una confianza media-alta en las predicciones.
- **Diversidad de etiquetas (Diversity):** el número de etiquetas distintas asignadas por guion varió entre 3 y 5, lo que refleja la capacidad del modelo para diferenciar múltiples tipos de momentos narrativos dentro de las escenas.
- **Coherencia narrativa temporal (Coherence):** la métrica de coherencia osciló entre aproximadamente 0,72 y 0,81, indicando que entre el 72 % y 81 % de los pares de escenas consecutivas respetan el orden esperado en la estructura narrativa, lo que sugiere una buena lógica temporal en la asignación de etiquetas.

Tabla 4: Estadísticas de calidad narrativa de los *turning points* detectados (`threshold = 0.5`)

Película	Min	Max	Media	Mediana	Diversidad	Coherencia
1	0.5039	0.9165	0.6373	0.6272	4	0.733
2	0.5007	0.9198	0.6300	0.6134	3	0.811
3	0.5001	0.9214	0.6360	0.6108	5	0.764
4	0.5005	0.9338	0.6136	0.5829	5	0.718

7.2 Evaluación de la clasificación zero shot para *el viaje del héroe*

Con el fin de evaluar la calidad de las fases del *viaje del héroe* detectadas en los guiones, se ha aplicado un algoritmo de programación dinámica para extraer, por cada película, la secuencia más larga de fases que respeta tanto el orden teórico de esta estructura narrativa, como la progresión temporal dentro del guion.

A continuación se muestran los resultados obtenidos para las cuatro películas analizadas:

- **12 años de esclavitud (2013)**: Se identifican cuatro fases consecutivas del *viaje del héroe*, comenzando con *refusal of the call* y terminando con *resurrection*, todas ubicadas entre las posiciones 0.33 y 0.42 del guion. Esta secuencia muestra una progresión coherente, aunque limitada en extensión y fases cubiertas.
- **Toy Story (1995)**: Se detecta una secuencia de tres fases: *crossing the first threshold*, *ordeal* y *reward*, distribuidas de forma amplia a lo largo del guion, desde el primer tercio hasta el clímax. Aunque la progresión narrativa respeta el orden teórico, se observa una omisión de las fases iniciales.
- **Twilight Zone: The Movie (1983)**: Esta película presenta la secuencia más extensa, con siete fases consecutivas que abarcan desde el *call to adventure* hasta el *return with the elixir*, siguiendo de forma notable la estructura canónica del *viaje del héroe*. Las fases están razonablemente distribuidas a lo largo del guion, con una cobertura narrativa amplia y una progresión coherente.
- **Invencible (Unbroken, 2014)**: La secuencia máxima consta de cuatro fases: *refusal of the call*, *crossing the first threshold*, *reward* y *the road back*. Aunque el orden de aparición es correcto, las fases iniciales y finales del *viaje del héroe* no se detectan, lo cual limita la cobertura de la estructura completa.

Tabla 5: Estadísticas de calidad narrativa de las fases detectadas (`threshold = 0.5`)

Película	Min	Max	Media	Mediana	Diversidad	Coherencia
1	0.5119	0.6853	0.5923	0.5860	3	0.667
2	0.5059	0.5497	0.5346	0.5482	2	0.500
3	0.5200	0.7425	0.5749	0.5528	5	0.556
4	0.5316	0.6221	0.5769	0.5769	2	1.000

Estas secuencias fueron posteriormente visualizadas en diagramas de línea temporal, permitiendo comprobar de forma gráfica el grado de alineación entre las fases detectadas y sus rangos esperados. En general, se ha observado que las fases centrales del *viaje del héroe* son las más detectadas por el modelo, mientras que fases como *ordinary world* o *meeting with the mentor* aparecen con menor frecuencia o no forman parte de las secuencias válidas. Esto puede estar relacionado con la ambigüedad semántica de estas fases o con limitaciones del clasificador para distinguir momentos más sutiles o introspectivos por no poder tener en cuenta el contexto del conjunto global de escenas.

En resumen, los resultados muestran que es posible identificar estructuras narrativas consistentes en los guiones analizados, aunque aún existen desafíos en cuanto a cobertura y precisión en las fases detectadas. La secuencia más completa encontrada pertenece a *Twilight Zone: The Movie*, lo que demuestra el potencial de esta técnica para analizar la progresión del *viaje del héroe* en relatos cinematográficos.

7.3 Interpretación de los resultados del *viaje del héroe*

Los resultados obtenidos a partir del análisis cuantitativo de las fases del Viaje del Héroe detectadas en los guiones revelan ciertas limitaciones y desafíos en la aplicación del enfoque *zero-shot* para esta tarea.

Las estadísticas de puntuación muestran que, aunque los valores mínimos se sitúan ligeramente por encima del umbral de confianza, que tuvo que ser reducido a 0.30, los valores máximos y las medias no alcanzan niveles altos (máximos entre 0.54 y 0.74, medias en torno a 0.53 a 0.59). Esto indica que la confianza del modelo en la asignación de fases del Viaje del Héroe es moderada-baja, lo que puede estar asociado a la complejidad de las escenas, la ambigüedad en la definición de estas fases, o a limitaciones inherentes del modelo para capturar con precisión los matices narrativos sin entrenamiento específico.

En cuanto a la diversidad de fases detectadas, que oscila entre 2 y 5 por guion de un total de 14, se observa una presencia limitada en algunos casos, lo que sugiere que el modelo puede no estar identificando completamente la riqueza y variedad esperadas en las estructuras na-

rrativas. Esto puede ser síntoma de una asignación conservadora o de confusión entre fases similares.

La métrica de coherencia narrativa temporal, que evalúa el respeto al orden esperado de las fases del Viaje del Héroe, presenta valores muy variables, desde 0.5 hasta 1.0. En varios casos, el índice se encuentra cercano a 0.5, lo que indica que en muchas secuencias el ordenamiento temporal no sigue fielmente la progresión teórica del modelo. Esto pone de manifiesto dificultades del sistema para captar la continuidad narrativa o refleja desviaciones reales en la estructura de algunos guiones analizados.

En conjunto, los resultados muestran que aunque el enfoque *zero-shot* permite una detección inicial de fases narrativas sin necesidad de datos anotados, la calidad y precisión de esta detección son limitadas. La confianza moderada, la diversidad reducida y la coherencia narrativa irregular indican que es necesario complementar esta metodología con técnicas adicionales, posiblemente supervisadas o que integren contexto más profundo, para mejorar la fidelidad y utilidad en el análisis de estructuras narrativas complejas como el Viaje del Héroe.

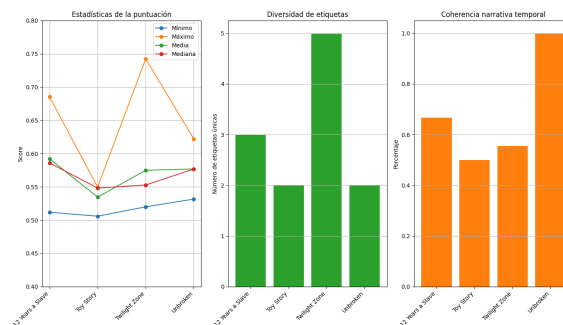


Figura 21: Evaluación de la clasificación para el *viaje del héroe*

7.3.1 Viaje del héroe Estos resultados, aunque no sustituyen una evaluación supervisada con datos anotados, ofrecen una visión cuantitativa y cualitativa valiosa sobre el desempeño y la fiabilidad del sistema *zero-shot* aplicado a la detección de momentos clave narrativos en guiones cinematográficos.

8. Limitaciones del estudio

A pesar de los resultados obtenidos y del valor que puede aportar el sistema desarrollado para el análisis estructural de guiones, este

trabajo presenta una serie de limitaciones que es necesario tener en cuenta tanto en la interpretación de los resultados como en su posible extensión futura.

En primer lugar, la detección de momentos clave narrativos se ha realizado mediante un enfoque *zero-shot*, lo cual permite evitar procesos de anotación manual extensivos, pero también introduce una fuerte dependencia del modelo lingüístico subyacente. Los modelos de lenguaje generalistas utilizados pueden no estar específicamente entrenados para reconocer estructuras narrativas formales como las propuestas por Field, McKee o Vogler, lo que puede afectar a la precisión de la detección.

En segundo lugar, el sistema asume una estructura narrativa estándar y lineal en los guiones analizados, lo cual no siempre se cumple en la práctica. Muchas películas emplean narrativas no lineales, estructuras fragmentadas o múltiples líneas temporales que desafían los esquemas convencionales. Esto limita la aplicabilidad directa del análisis a obras con estructuras más experimentales.

Otro aspecto a tener en cuenta es que las escenas están representadas únicamente por sus encabezados, diálogos y descripciones textuales, sin considerar elementos audiovisuales o semióticos clave que influyen en la progresión narrativa (como música, montaje o interpretación actoral). Esta simplificación puede reducir la fiabilidad de los resultados, especialmente en escenas con carga visual significativa y poco diálogo.

Además, cabe destacar la ambigüedad semántica inherente a muchos de los momentos narrativos definidos. En algunos casos, una misma escena podría interpretarse como un *change of plans* o un *point of no return*, en el caso de la clasificación para la estructura del *viaje del héroe*. Dependiendo del contexto narrativo general o de interpretaciones subjetivas. Esta ambigüedad semántica plantea un reto importante tanto para modelos automáticos como para la anotación manual, y limita la posibilidad de establecer una única "verdad" narrativa para cada guion. En consecuencia, incluso con buenos resultados cuantitativos, siempre existirán discrepancias interpretativas que deben ser tenidas en cuenta en el análisis.

Por último, el tamaño del corpus y su distribución por género o década puede condicionar los resultados. Además, el sistema no contempla las variaciones culturales o contextuales que pueden influir en la estructura narrativa según el país o el tipo de producción.

Estas limitaciones no invalidan los resultados obtenidos, pero sí invitan a interpretarlos con prudencia y a considerar mejoras metodológicas en futuros trabajos.

9. Conclusiones

En conclusión, el presente trabajo ha logrado cumplir con el objetivo principal planteado: desarrollar una herramienta capaz de generar una representación visual de la estructura narrativa de un guion cinematográfico, permitiendo representarla de forma explícita sobre un modelo narrativo. Esta visualización facilita la comprensión de la progresión dramática y el análisis estructural de las historias, aportando un recurso útil tanto para el estudio académico como para el ámbito creativo.

Asimismo, se han alcanzado los objetivos secundarios mediante un enfoque sistemático que ha combinado técnicas de procesamiento del lenguaje natural, clasificación automatizada y análisis narrativo. En particular, se ha implementado un proceso de detección de momentos clave a partir de guiones estructurados, utilizando modelos de clasificación *zero-shot* para asignar etiquetas narrativas sin necesidad de entrenamiento supervisado. Esta estrategia ha permitido una mayor flexibilidad y adaptabilidad del sistema, especialmente útil en contextos donde no se dispone de corpus etiquetados.

Los resultados obtenidos en la evaluación del sistema confirman la viabilidad y utilidad del enfoque propuesto. Se ha logrado una cobertura alta en la asignación de etiquetas a las escenas, mostrando que el modelo identifica momentos relevantes en buena parte del guion.

Además, la diversidad de etiquetas asignadas indica que el sistema es capaz de discriminar diferentes tipos de momentos narrativos, aportando riqueza y detalle al análisis estructural. Finalmente, la métrica de coherencia narrativa temporal, con valores entre 0.72 y 0.81, evidencia que las etiquetas respetan en gran medida el orden esperado en la progresión dramática, reforzando la validez del modelo para representar estructuras narrativas clásicas.

Estos hallazgos sugieren que la combinación de técnicas de procesamiento de lenguaje natural con modelos *zero-shot* es una vía prometedora para el análisis automatizado de guiones, especialmente en ausencia de datos anotados. No obstante, la evaluación realizada también pone de manifiesto la necesidad de futuras investigaciones que permitan integrar evaluaciones supervisadas y métricas cualitativas más avanzadas para mejorar la precisión y robustez del sistema.

En definitiva, este trabajo contribuye a ampliar las herramientas disponibles para el estudio y la creación narrativa, facilitando el acceso a análisis estructurados y visuales que pueden ser aprovechados tanto por investigadores como por profesionales del guion y la narrativa audiovisual.

10. Líneas de trabajo futuras

10.1 Líneas futuras

Finalmente, estos resultados abren múltiples vías de investigación en la automatización del análisis narrativo de guiones cinematográficos:

1. **Comparaciones por género y cultura.** Estudiar de forma sistemática cómo varían las estructuras narrativas entre géneros (drama, comedia, acción. . .) y contextos culturales, para identificar patrones y peculiaridades específicas.
2. **Métricas estructurales avanzadas.** Diseñar indicadores más objetivos —basados en alineamiento secuencial, teoría de grafos o distancias estructurales— que midan con precisión la semejanza entre guiones o la adaptación a modelos como la *Pirámide de Freytag*.
3. **Nuevos clasificadores y LLMs.** Explorar otras técnicas de aprendizaje automático, en particular grandes modelos de lenguaje capaces de procesar secuencias completas, para capturar relaciones contextuales y semánticas más profundas en la narrativa.
4. **Expansión multilingüe.** Crear y anotar corpus de guiones en castellano y otras lenguas, validando los métodos desarrollados en distintos entornos lingüísticos y ampliando su aplicabilidad global.
5. **Análisis multimodal.** Incorporar datos de vídeo y audio sincronizados con el texto del guion, enriqueciendo el análisis de la dinámica dramática y ofreciendo una visión integral del relato audiovisual.

En conjunto, estas líneas futuras potenciarán la lingüística computacional aplicada al cine, con repercusiones tanto en la investigación académica como en la práctica profesional de guionistas y productores.““

Referencias

1. J. E. Cutting, “Narrative theory and the dynamics of popular movies,” *Psychonomic Bulletin & Review*, vol. 23, pp. 1713–1743, Dec. 2016.
2. G. Papalampidi, S. Degaetano-Ortlieb, and J. Monti, “Turning point identification in tv series synopses,” in *Proceedings of ACL 2019*, (Florence, Italy), pp. 2135–2145, 2019.
3. M. N. Hoque, B. Ghai, K. Kraus, and N. Elmqvist, “Portrayal: Leveraging nlp and visualization for analyzing fictional characters,” in *Proceedings of the Designing Interactive Systems Conference (DIS)*, ACM, 2023.
4. J. Campbell, *El héroe de las mil caras*. Fondo de Cultura Económica, 1949. Traducción de Luisa Josefina Villar.
5. L. Yang, X. Xu, X. Lan, Z. Liu, S. Guo, Y. Shi, H. Qu, and N. Cao, “A design space for applying the freytag’s pyramid structure to data stories,” *IEEE Transactions on Visualization and Computer Graphics*, vol. 28, no. 1, pp. 922–932, 2022.

6. A. Harun, M. R. A. Razak, M. N. F. M. Nasir, and A. Ali, "Freytag's Pyramid: An Approach for Analyzing The Dramatic Elements and Narrative Structure in Filem Negara Malaysia's First Animated Cartoon," in *Proceedings of the 2013 IEEE Symposium on Humanities, Science and Engineering Research (SHUSER)*, (Penang, Malaysia), June 2013.
7. G. Papalampidi, J. Monti, and S. Degaetano-Ortlieb, "Latent narrative structure for enhanced tv synopsis summarization," in *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, (Online), pp. 1122–1133, 2020.
8. X. Zhao, S. Lee, and W. Liu, "Learning-by-narrating: Pretraining bart on dialog–synopsis pairs for zero-shot narrative classification," *Journal of Artificial Intelligence Research*, vol. 73, pp. 201–229, 2022.
9. A. J. Reagan, L. Mitchell, C. M. Danforth, P. S. Dodds, and D. Harris, "The emotional arcs of stories are dominated by six basic shapes," *EPJ Data Science*, vol. 5, no. 31, pp. 1–12, 2016.
10. D. Del Vecchio, M. Rossi, and A. Sorrentino, "Sentiment curves in screenplays: Predicting box-office success from emotional flow," *Entertainment Computing*, vol. 37, p. 100397, 2021.
11. P. Rao, J. Martinez, and T. Nguyen, "Scriptviz: Real-time visual referencing for screenwriters via large-scale cinematic databases," in *Proceedings of the 2024 ACM Symposium on Interactive Narratives*, (Vancouver, Canada), pp. 45–56, 2024.
12. F. Hoque, L. Zhang, and C. Pérez, "Portrayal: Interactive visualization of character arcs without manual annotation," in *Proceedings of the 2023 IEEE Conference on Visual Analytics Science and Technology (VAST)*, (Houston, TX, USA), pp. 78–89, 2023.
13. T. Han, M. Bain, A. Nagrani, G. Varol, W. Xie, and A. Zisserman, "Autoad: Movie description in context," *arXiv preprint arXiv:2303.16899*, 2023.
14. N. Oostdijk, *Corpus Linguistics and the Automatic Analysis of English*, vol. 6 of *Language and Computers*. Leiden: BRILL, 2024.
15. E. W. Pamungkas, V. Basile, and V. Patti, "A joint learning approach with knowledge injection for zero-shot cross-lingual hate speech detection," *Information Processing & Management*, vol. 58, p. 102544, July 2021.
16. M. Lewis, Y. Liu, N. Goyal, M. Ghazvininejad, A. Mohamed, O. Levy, V. Stoyanov, and L. Zettlemoyer, "Bart: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension," *arXiv preprint arXiv:1910.13461*, 2019.
17. Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, "Roberta: A robustly optimized bert pretraining approach," *arXiv preprint arXiv:1907.11692*, 2019.
18. A. Radford, K. Narasimhan, T. Salimans, and I. Sutskever, "Improving language understanding by generative pre-training," 2018. [Accedido: 2025-07-02](https://arxiv.org/abs/2025-07-02).
19. A. Piper, R. J. So, and D. Bamman, "Narrative Theory for Computational Narrative Understanding," in *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing (M.-F. Moens, X. Huang, L. Specia, and S. W.-t. Yih, eds.)*, (Online and Punta Cana, Dominican Republic), pp. 298–311, Association for Computational Linguistics, Nov. 2021.

20. L. Zhu, R. Zhao, L. Gui, and Y. He, “Are NLP Models Good at Tracing Thoughts: An Overview of Narrative Understanding,” Oct. 2023. arXiv:2310.18783 [cs].
21. P. Papalampidi, F. Keller, and M. Lapata, “Movie plot analysis via turning point identification,” in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)* (K. Inui, J. Jiang, V. Ng, and X. Wan, eds.), (Hong Kong, China), pp. 1707–1717, Association for Computational Linguistics, Nov. 2019.
22. B. Yao *et al.*, “Zero-shot story element classification with large language models,” *arXiv preprint arXiv:2302.10876*, 2023.
23. Gufukuro, “Movie scripts corpus.” <https://www.kaggle.com/datasets/gufukuro/movie-scripts-corpus>, 2023. Accessed: 2025-06-15.
24. J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “Bert: Pre-training of deep bidirectional transformers for language understanding,” *arXiv preprint arXiv:1810.04805*, 2019.
25. G. Langton and P. Goy, “Screenpy: A rule-based screenplay annotation tool.” <https://github.com/ScreenPyHQ/screenpy>, 2023. Accessed: 2025-07-02.
26. C. Vogler, *El viaje del escritor: Estructura mítica para escritores*. Barcelona: Robinbook, 2002. Traducción de Miguel Martínez-Lage.
27. D. Hy, “The i ching and the hero’s journey,” 2020. Último acceso: 9 de julio de 2025.
28. G. Freytag, *Die Technik des Dramas*. Leipzig: S. Hirzel, 1863. Introdujo la estructura narrativa conocida como la pirámide de Freytag.
29. Hugging Face, “BART model documentation.” https://huggingface.co/docs/transformers/model_doc/bart, 2025.
30. A. Williams, N. Nangia, and S. Bowman, “A broad-coverage challenge corpus for sentence understanding through inference,” in *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pp. 1112–1122, Association for Computational Linguistics, 2018.

Modelos de predicción para la mejora del control del alineamiento de un sistema láser de alta potencia

José David Pisonero Carabias, María Angélica González Arrieta,
Fabio Galán Prado, and Guillermo Hernández González

Departamento de Informática y Automática,
Facultad de Ciencias, Universidad de Salamanca
Plaza de los Caídos s/n (37008), Salamanca, España
Centro de Láseres Pulsados,
C. del Adaja 8, 37185 Villamayor, Salamanca, España
{jdpisonero,fgalan}@clpu.es

Resumen Este trabajo aborda la utilización de inteligencia artificial para asistir en el alineamiento del láser de alta potencia Vega del Centro de Láseres Pulsados (CLPU), una tarea realizada manualmente y propensa a errores. Se ha desarrollado un prototipo experimental utilizando cámaras y actuadores piezoeléctricos para capturar datos del centroide del haz en campo cercano y lejano. Numerosas medidas sobre este montaje han proporcionado un conjunto de datos posteriormente ampliado sintéticamente sobre el que se aplicaran modelos de aprendizaje supervisado tales como regresión polinómica, bosque aleatorio y perceptrón multicapa, optimizados mediante ajuste de hiperparámetros y validación cruzada siendo el perceptrón multicapa el que obtuvo los mejores resultados. Finalmente, se construyó un gemelo digital del sistema que permite simular su comportamiento con gran precisión. La validación en un entorno experimental demostró la viabilidad del enfoque, confirmando que el uso de inteligencia artificial puede mejorar el proceso de alineamiento de sistemas láser de alta potencia.

Keywords: Láser, CPA, Control alineamiento, Aprendizaje supervisado, Perceptrón multicapa, Gemelo digital.

1. Introducción

Desde que en 1960 T.H. Maiman construyera el primer láser [1], sus aplicaciones no tardaron en aparecer en el campo de la medicina [2], en procesos industriales de corte por láser [3], [4], en el mundo de las telecomunicaciones [5] y en aplicaciones de uso general para el gran público [6], [7].

Rápidamente, el láser estaba totalmente introducido en la sociedad. De igual forma a como se desarrollaron las aplicaciones, también

lo hicieron nuevas técnicas de generación láser para conseguir nuevas utilidades en el mundo científico.

En 1985, Dona Strickland desarrolló la técnica conocida como CPA por sus siglas en inglés *Chirped Pulse Amplification*, que marcaría el camino posterior para el desarrollo de los sistemas láser de alta potencia y que le haría merecedora del Premio Nobel de Física en 2018 [8]. Gracias a esta técnica, se pudieron conseguir fácilmente láseres pulsados que proporcionaban potencias pico del orden de gigavatios (GW) y teravatios (TW), proporcionando unas intensidades en foco de $10^{19}W/cm^2$, a partir de concentrar energías, E moderadamente altas, de decenas de Julios (J) en pulsos de duración temporal, Δt , extremadamente corta, del orden de femtosegundos (10^{-15} s, aprovechando la definición de la potencia P 1:

$$P = \frac{E}{\Delta t}. \quad (1)$$

Esta tecnología aprovecha la anchura espectral de los láseres cuyo medio activo es un cristal de titanio:zafiro, para estirar temporalmente el pulso hasta centenas de picosegundos en el elemento óptico compuesto por redes de difracción llamado estirador de pulsos, que aprovecha que cada longitud de onda del pulso que incide en ellas se difracta en un ángulo θ diferente dado por la ecuación de difracción 2

$$n\lambda = 2d\sin(\theta) \quad (2)$$

donde n es el orden de difracción, λ la longitud de onda y d la distancia entre las líneas de la red de difracción. Siguiendo la definición dada para la potencia en la ecuación 1, al aumentar la duración temporal del pulso, disminuye la potencia, por tanto, ahora es posible amplificar la energía de los pulsos propagándolos en aire, hasta valores de decenas de Julios. Tras obtener la energía deseada, el haz pasa de nuevo por otro sistema de redes de difracción que deshacen el estiramiento temporal hecho anteriormente, juntando de nuevo todas las longitudes de onda temporalmente, haciendo que la potencia se incremente hasta valores de teravatios (TW) o petavatios (PW).

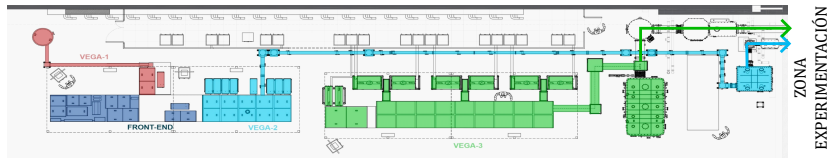


Figura 1: Esquema del láser Vega del CLPU [9]

El láser Vega del CLPU está basado en esta tecnología. En la figura 1 se muestra el plano de la instalación de dicho láser en la que se

marcan las distintas etapas para llevar a cabo la generación de pulsos ultracortos de alta potencia. Vega tiene tres salidas, cada una de ellas con una etapa de amplificación diferente, Vega 1 con energía máxima de $E = 600$ mJ, Vega 2 con energía máxima de $E = 6$ J y Vega 3 con energía máxima $E = 30$ J.

En estas etapas de amplificación el láser pasa múltiples veces por un cristal de titanio zafiro que es bombeado por láseres de neodimio:YAG, en un sistema óptico conocido como amplificador multipaso.

Una de las principales aplicaciones de los láseres pulsados basados en la tecnología CPA es la aceleración de partículas a partir de la interacción láser-materia, la cual se consigue al hacer incidir un pulso de alta intensidad en un objetivo en estado, normalmente, gaseoso. Históricamente, los experimentos relacionados con partículas se realizan en sincrotrones, aceleradores de forma circular y de grandes dimensiones, que utilizan campos magnéticos para acelerar protones o neutrones y hacerlos colisionar entre sí. Estas instalaciones pueden tener una circunferencia que va desde cientos de metros, como el sincrotrón ALBA en Barcelona con 270 m [10], hasta varios kilómetros, como el gran colisionador de hadrones (LHC por sus siglas en inglés) del CERN, con una circunferencia de 27 km ubicado en Ginebra, Suiza [11].

Construir este tipo de colisionadores de partículas es un gran reto técnico, pero también económico, por eso el desarrollo de los sistemas láseres de alta potencia utilizado para replicar esta clase de experimentos se presenta como un gran avance tanto para el mundo de la física de partículas como de la óptica física, pues son mucho más compactos y baratos de construir, permitiendo operar en salas de decenas de metros. Es en esta clase de aportaciones científicas donde entra en juego el láser Vega, que ya ha demostrado ser capaz de llevar a cabo experimentos de aceleración de partículas [12].

Sin embargo, para conseguir ser una alternativa real a los sincrotrones, el desarrollo tecnológico de los sistemas láser aún debe mejorar en varios aspectos, en concreto, para conseguir replicar la reproducibilidad y estabilidad de un sincrotrón. Es en esta clase de retos donde entran en juego los desarrollos que utilizan técnicas de inteligencia artificial, con el objetivo de maximizar la optimización de los diferentes parámetros del láser. Se han presentado ya múltiples estudios que utilizan redes neuronales para el control de la estructura temporal de los pulsos láser [13], para el control adaptativo del frente de onda [14] o para el control del alineamiento del sistema [15].

Es en este último escenario donde se va a centrar este trabajo cuyo punto de partida es optimizar, mediante técnicas de inteligencia artificial, el alineamiento del haz láser en la cadena de amplificación. En los amplificadores multipaso de las distintas salidas de Vega, el haz recorre decenas de metros, por lo que un pequeño cambio en el alineamiento a la entrada de estos amplificadores se incrementa rápidamente debido a la larga propagación que sufre. Este desvío en el alineamiento

puede afectar a la eficiencia de la amplificación, por lo que el control del alineamiento es fundamental para mantener la eficiencia tanto en esta etapa como en el apuntado posterior en el objetivo de impacto en la zona de experimentación.

La monitorización del alineamiento se realiza mediante sistemas ópticos que dividen el haz en dos réplicas, una para analizar el campo cercano o NF por sus siglas en inglés (*Near Field*) y otra para analizar el campo lejano o FF por sus siglas en inglés (*Far Field*). Ambos haces son mandados a cámaras que permiten ver el perfil espacial de los mismos y medir su centroide para determinar su posición y, por tanto, su alineamiento en la cadena del láser.

Este control de la direccionalidad del haz es uno de los grandes retos en esta clase de sistemas láser de alta potencia. Habitualmente el control del apuntado del láser se realiza de forma manual por personal experto, antes de empezar con experimentos que se llevan a cabo durante varias horas de operación, dificultando su optimización durante su realización. En este trabajo se propone una solución para aumentar la eficiencia del control del alineamiento del sistema láser, de manera que sirva de apoyo al personal técnico que realiza estas tareas, reduciendo el tiempo dedicado a ello y, por lo tanto, las posibilidades de error debido a la fatiga.

Para ello, el objetivo de este trabajo es desarrollar un montaje óptico paralelo al sistema Vega para poder recoger datos experimentales para entrenar y validar modelos basados en aprendizaje supervisado que sirvan para desarrollar un gemelo digital que se pueda utilizar como herramienta para la optimización y apoyo al personal experto en la tarea de alineamiento del láser y que después sea escalable al sistema Vega original.

2. Estado de la cuestión mediante un mapeo sistemático de la literatura.

El estado de la cuestión en el panorama científico en los temas tratados en este estudio se estudia mediante un mapeo sistemático de la literatura siguiendo el modelo PRISMA 2020 [16] en el que se define un protocolo de acción con una serie de criterios para determinar cuáles son los artículos que aportan valor para este trabajo. El objetivo de este mapeo sistemático es:

“Estudiar técnicas y métodos de control de alineamiento de un sistema láser aplicando técnicas de inteligencia artificial, de manera que se pueda comprender cuáles son los elementos básicos para construir estos sistemas y como aplicarlos en un sistema láser de alta potencia como el láser Vega del CLPU”.

En total, el Corpus del mapeo lo formaron 16 artículos distribuidos como se muestra en la figura 2 según las categorías dadas por las preguntas del mapeo (PM) que se responden a continuación.

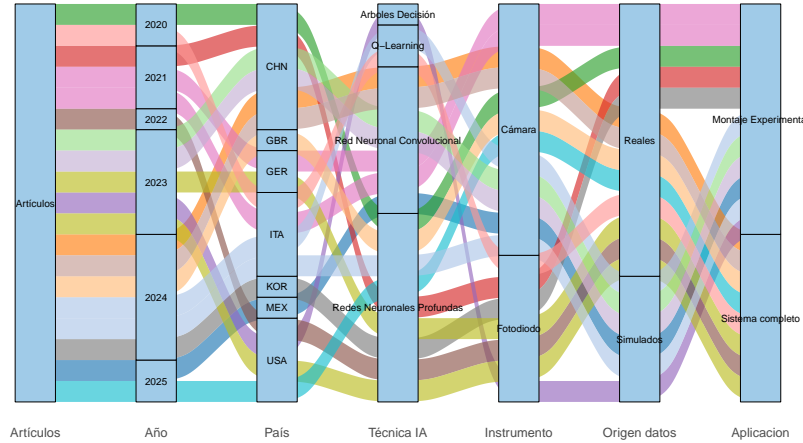


Figura 2: Resumen de la distribución de los artículos según los atributos estudiados y como se relacionan entre ellos.

2.0.1 PM1: ¿Cuántos estudios se han realizado desde 2020?

Si bien el año en el que se realizó el estudio se tuvo en cuenta como un criterio de inclusión, analizar cómo es la evolución de esta clase de estudios desde 2020, no solo da una respuesta cuantitativa de cuántos estudios se han realizado en los últimos años, sino que también puede dar una idea de la tendencia de investigar el uso de la inteligencia artificial en sistemas láser de alta potencia, que históricamente no ha sido un nicho de aplicaciones de este tipo.

Los resultados muestran que la tendencia es totalmente al alza, siendo el año 2024 el que presenta más artículos dentro de este trabajo. Además, en los primeros meses de 2025 se presentan ya dos de los artículos más interesantes encontrados en el mapeo. Por un lado, en [15] aplican redes neuronales con una arquitectura de perceptrón multicapa para corregir el alineamiento del sistema láser BELLA, que cuenta con las mismas características que la salida Vega 3 del CLPU, que proporciona pulsos de 1PW a 1Hz. En este estudio, describen que el sistema sufre perturbaciones mecánicas y ambientales que llevan a una deriva en el alineamiento del láser a la hora de realizar experimentos, por lo que desarrollan un modelo para poder corregirlos de forma preventiva, lo que es clave a la hora de realizar experimentos de aceleración de partículas.

En la otra contribución publicada en 2025 [17], se describe una combinación entre técnicas de inteligencia artificial como las redes neuronales convolucionales con sistemas PID clásicos, con la intención de que las virtudes de cada técnica compensen las carencias de la otra. En este caso, no se aplica a ningún láser de alta intensidad, pero el

sistema que se describe sería escalable a estas estructuras más grandes y delicadas.

Estos resultados demuestran que la aplicación de la inteligencia artificial en el alineamiento de láseres está completamente de actualidad y está justificado aplicar estas técnicas para esta tarea.

Sin embargo, también hay que fijarse que el número de estudios encontrados no es demasiado numeroso, lo que demuestra que aunque cada vez haya más, sigue siendo un nicho aún por explorar. Es por todo esto por lo que el estudio del control del alineamiento mediante técnicas de inteligencia artificial demuestra ser un tema de actualidad e interés para su desarrollo.

2.0.2 PM2: ¿En qué países se están realizando estos estudios?

La distribución geográfica de las instalaciones en las que se hacen estos estudios es importante, ya que no hay una gran cantidad de ellas a lo largo del planeta y por lo general en cada lugar están enfocadas a propósitos diferentes. China y Estados Unidos son los países que más contribuciones aportan a este estudio, pero Europa no queda lejos.

El estudio presentado en [18] muestra cómo aplicar redes neuronales para el alineamiento de interferómetros láser usados para detectar ondas gravitacionales en los observatorios LIGO (acrónimo del nombre *Laser Interferometry Gravitational-waves Observatory*), en Estados Unidos, y Virgo, en Alemania. En el caso del estudio de [19], se presenta un estudio que combina múltiples láseres de diodo, que son económicos, robustos y fáciles de construir, para obtener un sistema de alta potencia, aunque en este caso orientado al procesado láser.

Esta tecnología de combinación de múltiples láseres, se usa también por el estudio [24] realizado en China, que aplica redes neuronales convolucionales para poder monitorizar el alineamiento de múltiples láseres con una sola cámara para mejorar la eficiencia del efecto de combinación de láseres coherentes. Esta técnica también es la que se pretende mejorar mediante el uso de redes neuronales en el estudio [21] realizado en Estados Unidos, para mantener el alineamiento de múltiples láseres de fibra. Estas técnicas son muy importantes en experimentos de fusión nuclear por confinamiento láser, que principalmente son llevados a cabo en Estados Unidos y China, lo que explica por qué se realizan estos estudios en dichos países.

Por último, también cabe destacar cómo los aceleradores de partículas también impulsan en desarrollo en otras áreas, como en este caso es el desarrollo de sistemas láser. En el estudio [22] hecho en Italia y en [23] de Estados Unidos, se trata de mejorar el alineamiento de los sistemas, láser auxiliares de aceleradores de partículas y láseres de electrones libres. Como se ha comentado previamente, estas instalaciones tienen un tamaño muy grande y necesitan una gran inversión económica, que en este caso se ve como es devuelta en forma de apoyo a otras tecnologías.

En cuanto a los países de procedencia de los artículos encontrados, se puede concluir que China y Estados Unidos son los que aportan la mayoría de ellos, además aplicando las técnicas a distintas tecnologías como la detección de ondas gravitacionales [18], los aceleradores de partículas [23] y la fusión nuclear [24]. Por otro lado, los países europeos no quedan lejos, a base de fomentar los trabajos entre varios países y la conexión de redes conectadas para promover el conocimiento científico en el continente.

2.0.3 PM3: ¿Qué técnicas de inteligencia artificial se han empleado en el estudio? La gran mayoría de artículos emplea redes neuronales, incluyendo redes neuronales convoluciones, lo que es un indicativo que en esos artículos se pretende controlar el alineamiento del láser a partir de imágenes tomadas con cámaras.

En concreto, en [25] se emplean redes convolucionales para mejorar la estabilidad en sistemas de telecomunicaciones láser entre satélites, de manera que puedan ser sistemas más robustos frente a perturbaciones térmicas o atmosféricas que suelen ser habituales en estos sistemas. En [26] también se pretende mejorar sistemas de telecomunicaciones, en este caso submarinas, utilizando redes neuronales para predecir el comportamiento del láser frente a perturbaciones.

Asimismo, cabe mencionar el estudio [17], que no solo aplica redes convolucionales para el control del alineamiento de un láser, sino que lo combina con un PID. En este artículo presentan una técnica en la que primero se hace un primer alineamiento aplicando redes neuronales convolucionales y después se utiliza un PID, cuando ya está muy cerca del objetivo, para hacer un ajuste más fino. Además, en este caso se plantea un problema de clasificación y no de regresión, siendo el objetivo de la red convolucional determinar en que región de la cámara que mide el alineamiento se encuentra el láser, para actuar después con unos ajustes preestablecidos y ligados a esas regiones.

No solo se aplican redes neuronales, también se utilizan otras técnicas de predicción de valores por aprendizaje supervisado como en [23], que aplican modelos bayesianos para estudiar las deformaciones térmicas del sistema y máquinas de vectores de soporte para corregir las derivas que proceden de esas deformaciones. Además, usan las técnicas de bosque aleatorio o *Random Forest* y redes neuronales para los mismos cometidos con objeto de encontrar efectos que se expliquen con términos no lineales.

2.0.4 PM4: ¿Qué tipo de instrumentos se utiliza para medir el alineamiento del sistema láser? El alineamiento de un láser se controla con dos tipos de instrumentos, cámaras con sensores sensibles a las longitudes de onda del láser de estudio y fotodiodos de posición. El hecho de elegir cámaras está intrínsecamente relacionado con la elección previa de usar redes neuronales convolucionales. Sin embargo, no todos

los estudios aplican esta relación. En el estudio [15] utilizan cámaras para medir el centroide del láser y que este sea el parámetro a corregir por el sistema inteligente. Así se aplica también en [27] y [28]. Esto es debido a que en la cadena de los láseres de alta potencia, el control del alineamiento se realiza mediante cámaras, llevando el láser a una referencia establecida. Las cámaras tienen la ventaja de que no solo permiten controlar el alineamiento, sino controlar el perfil espacial del haz o estimar la energía de cada pulso láser. Además, la señal del centroide es mucho más sencilla de analizar que las señales eléctricas que proceden de un fotodiodo, que requieren montajes especiales dedicados para procesar estas señales.

2.0.5 PM5: ¿Se utilizaron datos reales o simulados para construir el sistema de alineamiento? A la hora de entrenar los modelos, se pueden recoger datos de un sistema láser real o se pueden simular y generar de forma sintética para después aplicar los resultados al sistema láser del laboratorio. Por la complejidad de los montajes ópticos de láseres de alta intensidad, es muy complicado implementar estos sistemas para la toma de datos fuera de los sistemas de control instalados, por ello se construyen pequeños montajes aislados para probar estos sistemas de control. Esto es lo que se hace en [17], [24] o [15] donde construyen un pequeño prototipo del sistema, pero pensando en escalarlo a un sistema más grande una vez los resultados fueron validados.

En este apartado destacan los estudios en [18] o [22] que utilizan datos históricos de las instalaciones previamente etiquetados para entrenar los modelos de predicción.

También destaca el caso del artículo [29], donde se aplica directamente en un oscilador, el cual tiene la ventaja de ser un sistema más pequeño y que se puede construir de cero en un laboratorio.

En conclusión, para un primer trabajo es recomendable construir un sistema óptico más pequeño y manejable que permita tomar datos reales de forma rápida para poder entrenar con ellos los modelos de predicción.

La última pregunta del mapeo es si se aplicaron los modelos construidos a un sistema láser de alta potencia o su funcionalidad quedó aislada a un sistema más pequeño construido únicamente para hacer el estudio que se presenta en los artículos. Como se explicó previamente, poder aplicar estas técnicas es muy complicado en láseres de alta intensidad, ya que son montajes muy artesanales, con sistemas de control cerrados y hechos a medida para los elementos originales de la configuración de la cadena del láser, donde incluir automatizaciones es muy complicado.

Hay casos como [25] o [23] que directamente son completamente simulados, aunque los resultados que se enseñan demuestran que estos estudios son un buen punto de partida para aplicarlo a sistemas reales.

Es por eso que destaca el caso de [15] que sí que aplica el modelo al láser BELLA (*Berkeley Lab Laser Accelerator*), aunque no lo hace a plena potencia, ya que lo aplican con pulsos atenuados a energías de pocos miliJulios (mJ).

2.1 Conclusiones del mapeo sistemático

En este mapeo sistemático de la literatura se ha hecho un estudio buscando encontrar estudios que combinan el uso de técnicas de inteligencia artificial aplicadas al control del alineamiento de un láser. Desde 2020, este tema ha cobrado interés dentro de la comunidad láser, tanto en los desarrollos orientados a la investigación como en otras materias como en el estudio de ondas gravitacionales [18], las telecomunicaciones [25], [26] y en la industria de procesado por láser [19].

Los estudios que tienen un objetivo similar al de este trabajo, controlar el alineamiento del láser a partir del control de unas monturas ópticas con actuadores motorizados [28], [?], emplean cámaras para medir el centroide y después aplican redes neuronales, aunque también se emplean otras técnicas como máquinas de vectores de soporte y árboles de decisión [23], por lo que en este trabajo se tratara de incluir este tipo de técnicas para comparar con las redes neuronales.

Además, los resultados obtenidos en este mapeo también demuestran la importancia de usar datos reales, a pesar de que después su implementación en un sistema láser de alta potencia sea complicada por la complejidad propia de esta clase de instalaciones. Es por eso por lo que la conclusión final orientada a este trabajo es que será necesario construir un sistema óptico para la toma de datos de entrenamiento de los diferentes modelos, con la intención de que después el método sea escalable y se pueda utilizar en el sistema Vega del CLPU. Los modelos que se construirán utilizaran redes neuronales, ya que son las que despiertan mayor consenso en la comunidad científica, pero también será importante comparar sus resultados con otras técnicas como los árboles de decisión.

3. Toma de datos de entrenamiento a través de un prototipo óptico.

La figura 3 muestra el montaje óptico construido para la toma de datos de entrenamiento. Los elementos principales son un, la fuente de luz, un láser de Helio - Neón ($\lambda = 633nm$) y cuatro actuadores piezoeléctricos modelo 8301NF de Newport distribuidos en dos monturas ópticas para el control de la posición en los ejes horizontal y vertical, manejados mediante el controlador 8742 de Newport. Por últimos se utilizan dos cámaras CCD modelo acA640-120gm de Basler para controlar la referencia de NF y modelo acA1440-73gm de Basler para el FF.

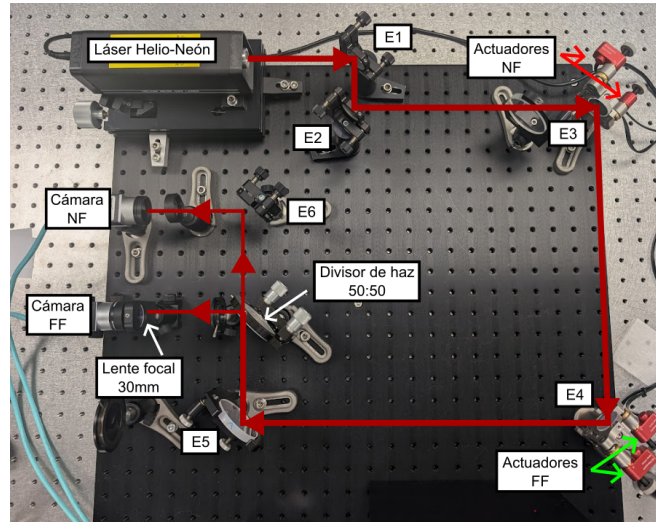


Figura 3: Montaje óptico experimental

El recorrido que sigue el láser es el siguiente. Sale de la fuente de emisión y es dirigido con los espejos E1 y E2 hacia el primer espejo con actuadores, E3, que será el que controle el alineamiento ajustando la referencia de la cámara de NF. Tras incidir en el espejo E3, el haz impacta en el espejo E4 cuya montura está controlada por los actuadores piezoeléctricos que llevarán el haz a la referencia de la cámara de FF. Tras estas dos monturas, que son las que controlan el alineamiento, el láser pasa por el espejo E5 que lo dirige al divisor de haz que divide el láser en 2 réplicas iguales. Una de ellas se refleja en el divisor de haz y pasa por una lente de focal $f = 30\text{mm}$ que focaliza el láser en la cámara de FF. Por otro lado, la parte que se transmite por el divisor de haz es reflejada en el espejo E6 y llega a la cámara de NF.

3.1 Recogida de datos para construir el dataset de entrenamiento de los modelos

La recogida de datos se realizan tomando los datos de los desplazamientos en horizontal y en vertical por separado.

Se parte del extremo superior izquierda de la cámara de NF y se recorre la fila hasta el extremo derecho. Acto seguido, se parte del extremo derecho y se recorre la fila en sentido inverso. Así se barre todo el espacio de la cámara de NF. Posteriormente, se realiza lo mismo con los movimientos verticales partiendo del mismo punto y yendo hacia el extremo inferior. Después, se desplaza el haz hasta un nuevo punto y se repite el proceso.

De esta manera se registraba el movimiento necesario de cada actuador para llegar a un punto concreto desde un punto de partida, que será diferente en cada caso.

Para el cálculo de los centroides con las cámaras de NF y FF se tomaban imágenes en cada punto y se aplicaban las ecuaciones 3 y 4 [30]:

$$C_x = \frac{\sum_{x,y} x \cdot I(x,y)}{\sum_{x,y} I(x,y)}, \quad (3)$$

$$C_y = \frac{\sum_{x,y} y \cdot I(x,y)}{\sum_{x,y} I(x,y)} \quad (4)$$

donde C_x y C_y son los centroides en los ejes horizontal y vertical y $I(x,y)$ es la intensidad del pixel en la posición (x,y) .

3.2 Aumento artificial del conjunto de datos

La información obtenida con los datos recogidos previamente está limitada a movimientos desde los extremos del sensor de la cámara. Se propone expandir de forma artificial los datos para obtener información de movimientos desde cualquier punto de la cámara, pues obtener esos datos del experimento real llevaría demasiado tiempo. Para ello se supone la lineal de los datos.

En el caso del eje horizontal, se tomaron varios puntos dentro de una misma línea, es decir, sin mover el actuador vertical, y en cada uno de ellos se recalculó la posición como si se partiera de ese punto utilizando los datos recogidos previamente. De esta manera, se consigue tener la información completa de todos los movimientos posibles en cada uno de estos puntos. El proceso se realiza de manera análoga en el caso del conjunto de datos para los movimientos verticales.

De esta manera, además de tener la información de los puntos dentro de la cámara de NF, se expande esta información a puntos que están fuera, es decir, posiciones en las que el láser está tan desalineado que está fuera del sensor de esta cámara, lo que expande aún más las posibilidades de los modelos que se van a generar.

4. Resultados y discusión

Los conjuntos de datos para los movimientos horizontal y vertical tras realizar la expansión detallada previamente se muestran en las figuras 4, 5 y 6, 7 respectivamente. Para el modelo horizontal se obtuvieron 3033 datos y para el vertical 3002 antes de la expansión y 70846 para el modelo horizontal y 68815 para el modelo vertical tras la expansión sintética.

Con estos nuevos conjuntos de datos se entrenan los modelos de aprendizaje supervisado. En este estudio se ha optado por utilizar

las técnicas de bosque aleatorio y de perceptrón multicapa. La razón de elegir estos modelos es que presentan buenos resultados cuando hay que captar comportamientos no lineales como son los de los actuadores piezoeléctricos, descartando otros modelos lineales, ya que el comportamiento del FF debido a la lente de focalización no es lineal.

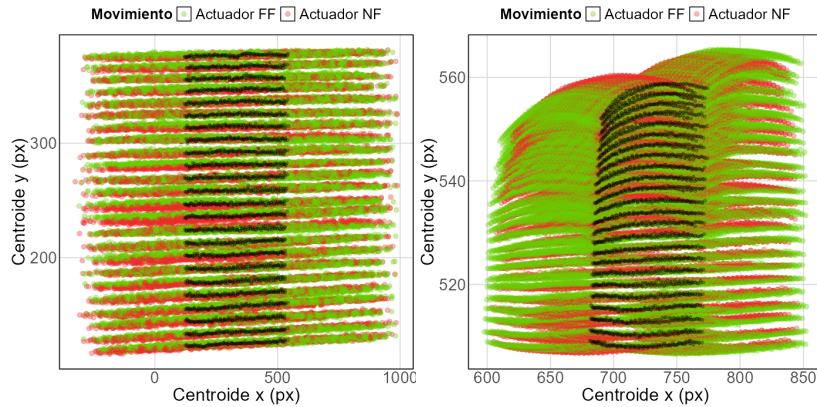


Figura 4: Datos del centroide de NF Figura 5: Datos del centroide de FF

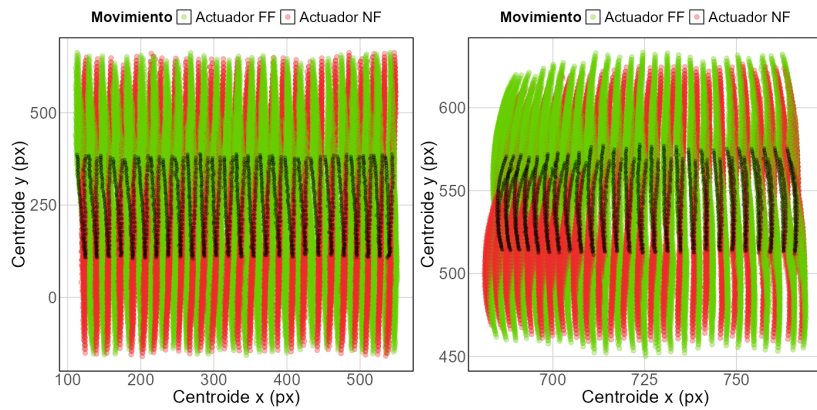


Figura 6: Datos del centroide de NF Figura 7: Datos del centroide de FF

Como el alineamiento se realiza separando movimientos horizontales y verticales, pero de manera que tanto el haz de NF, como el de FF lleguen a las referencias al mismo tiempo, los centroides de NF y FF se tendrán que calcular a la vez usando modelos con multisalida. Estos modelos puede aplicar técnicas que relacionan o no las dos salidas a

predecir. Cuando no se tiene en cuenta la relación entre ambas variables se utiliza `MultiOutputRegressor` [31] de la librería Scikit-learn [32] y cuando si tiene en cuenta esta relación, se aplica la técnica de encadenado de regresores mediante `RegressorChain` [33] de la misma librería. En este último caso, se asume que existe una dependencia entre la variable que se predice antes y la que se predice después. En este estudio, se calculará primero NF y después FF, pues es el procedimiento habitual al alinear manualmente un sistema óptico.

Las métricas obtenidas para estos modelos usando los conjuntos de datos expandidos se muestran en las tablas 1 para el caso de los movimientos en horizontal y la tabla 2 para los movimientos verticales. En estas tablas se puede apreciar que el valor de R^2 es prácticamente 1 en todos los casos, y que el modelo del bosque aleatorio obtiene mejores resultados que con el perceptrón multicapa.

En cuanto a la comparación entre la dependencia de las variables de salida, se puede ver como los datos son iguales para las métricas de NF, ya que en el caso dependiente se calcula primero y no le afecta la dependencia. Esta diferencia sí se aprecia en las métricas de FF, que son ligeramente mejores en los casos en los que se ha calculado esta salida usando el centroide de NF como variable dependiente. Esto se puede apreciar en las figuras 8 y 9 en el caso de movimientos horizontales y en las figuras 10 y 11 para los movimientos verticales, en las que se ha representado los valores MAE y RMSE descartando las otras dos métricas, R^2 por ser igual en todos los casos y MSE por no tener las mismas unidades que MAE y RMSE. En estas figuras, en concreto en 9 y 11 se puede ver claramente que el modelo con salidas dependientes obtiene mejores resultados en el modelo del perceptrón multicapa, siendo más sutiles en el caso del bosque aleatorio.

Tabla 1: Métricas obtenidas entrenando los modelos para predecir la posición horizontal de NF y FF.

Modelo		Métricas de NF				Métricas FF			
		R^2	MAE	MSE	RMSE	R^2	MAE	MSE	RMSE
Bosque aleatorio	Independiente	0.99	1.37	3.23	1.79	0.99	0.39	0.29	0.53
Bosque aleatorio	Dependiente	0.99	1.37	3.23	1.79	0.99	0.36	0.24	0.49
Perceptrón multicapa	Independiente	0.99	12.67	262.95	16.22	0.99	3.03	14.86	3.85
Perceptrón multicapa	Dependiente	0.99	12.67	262.95	16.21	0.99	2.94	14.00	3.74

Tabla 2: Métricas obtenidas entrenando los modelos para predecir la posición vertical de NF y FF.

Modelo		Métricas de NF				Métricas FF			
		R^2	MAE	MSE	RMSE	R^2	MAE	MSE	RMSE
Bosque aleatorio	Independiente	0.99	1.83	5.48	2.34	0.99	0.42	0.30	0.55
Bosque aleatorio	Dependiente	0.99	1.83	5.48	2.34	0.99	0.41	0.29	0.54
Perceptrón multicapa	Independiente	0.99	1.96	6.42	2.53	0.99	0.94	1.43	1.19
Perceptrón multicapa	Dependiente	0.99	1.96	6.42	2.53	0.99	0.87	1.22	1.11

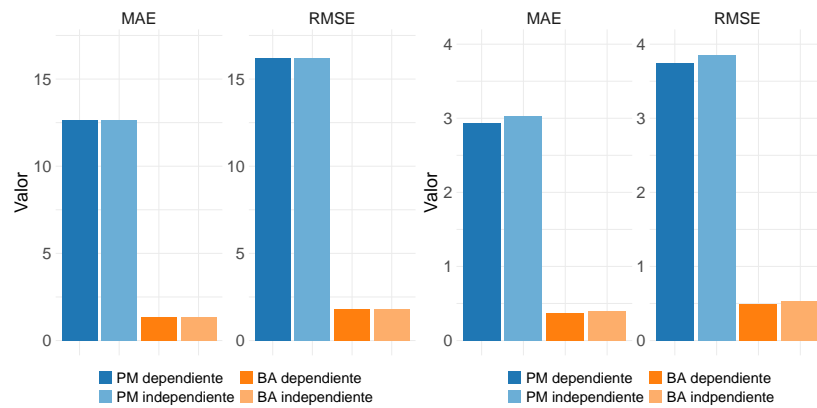


Figura 8: Métricas modelos para NF en eje horizontal. Figura 9: Métricas modelos para FF en eje horizontal.

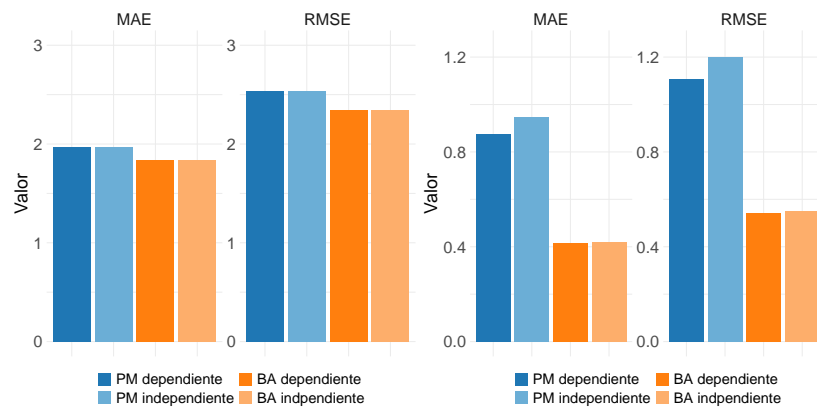


Figura 10: Métricas modelos para NF en eje vertical. Figura 11: Métricas modelos para FF en eje vertical.

4.1 Búsqueda de hiperparámetros

Tras concluir que los mejores modelos son aquellos que aplican la técnica de encadenado de regresores a los modelos de bosque aleatorio y de perceptrón multicapa, se optimizan mediante la búsqueda de hiperparámetros con validación cruzada. Los resultados de los mejores parámetros obtenidos son, para los modelos con el perceptrón multicapa:

- Neuronas en capas ocultas: (100,100) en ambos modelos.
- Función de activación: \tanh en ambos modelos.
- Regularización L2: 0.0001 para el modelo de movimientos horizontales y 0.001 para el modelo de movimientos verticales.

Y para los modelos de bosque aleatorio:

- Número de árboles: 500 en ambos modelos.
- Profundidad del árbol: 10 en ambos modelos.

Con estos parámetros entreno nuevos modelos y comparo las métricas obtenidas tras la evaluación de cada uno de ellos. Los resultados se muestran en la tabla 3 para el modelo de predicción de movimientos horizontales y en la tabla 4 para la predicción de movimientos verticales. Además, igual que se hizo previamente, se representan los errores MAE y RMSE en las figuras 12 y 13 para el modelo de movimientos horizontales y en las figuras 14 y 15 para el vertical.

Tabla 3: Métricas obtenidas tras hacer el ajuste de hiperparámetros en los modelos de predicción de movimientos horizontales.

Modelo	Métricas de NF				Métricas FF			
	R^2	MAE	MSE	RMSE	R^2	MAE	MSE	RMSE
Bosque aleatorio	0.99	8.61	180.33	13.43	0.99	1.60	6.48	2.54
Perceptrón	0.99	1.04	1.76	1.32	0.99	0.27	0.12	0.35

Tabla 4: Métricas obtenidas tras hacer el ajuste de hiperparámetros en los modelos de predicción de movimientos verticales.

Modelo	Métricas de NF				Métricas FF			
	R^2	MAE	MSE	RMSE	R^2	MAE	MSE	RMSE
Bosque aleatorio	0.99	6.57	89.95	9.48	0.99	1.35	3.71	1.92
Perceptrón multicapa	0.99	0.64	0.70	0.83	0.99	0.16	0.04	0.21

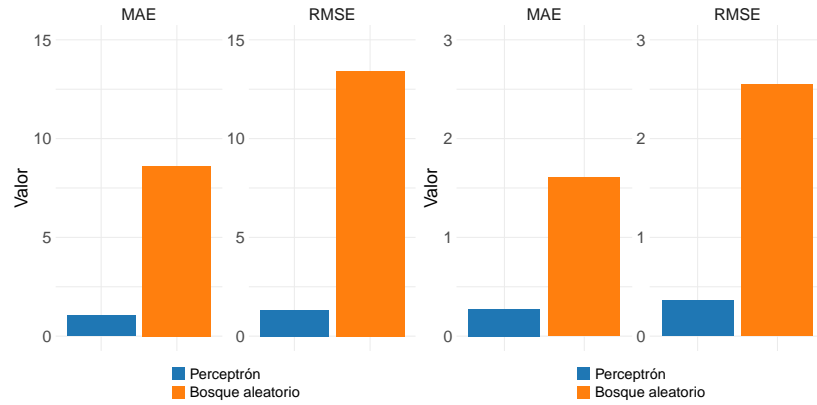


Figura 12: Métricas de NF en eje horizontal. Figura 13: Métricas de FF en eje horizontal.

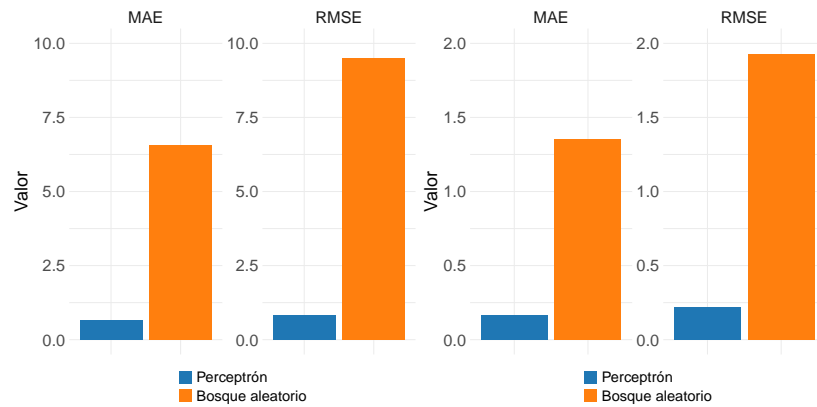


Figura 14: Métricas de NF en eje vertical. Figura 15: Métricas de FF en eje vertical.

Como se puede ver en las figuras 12-15, los resultados obtenidos con el modelo del perceptrón multicapa son muchísimo mejores que los obtenidos con la técnica del bosque aleatorio, aunque cabe mencionar que los valores obtenidos para este último son peores que los mostrados en las tablas 1 y 2, si bien los datos obtenidos en este último caso para el perceptrón multicapa son los mejores que se han obtenido en este trabajo.

Por tanto, se puede concluir que el modelo del perceptrón multicapa es el que mejor reproduce los movimientos del sistema óptico y que mejor servirá para poder simular el comportamiento de los centroides de NF y FF con los movimientos de los distintos actuadores.

4.2 Evaluación sobre el prototipo.

Una vez se determinó cuál era el modelo que mejores métricas obtenía con el conjunto de datos de entrenamiento, se valida directamente sobre el montaje óptico del laboratorio.

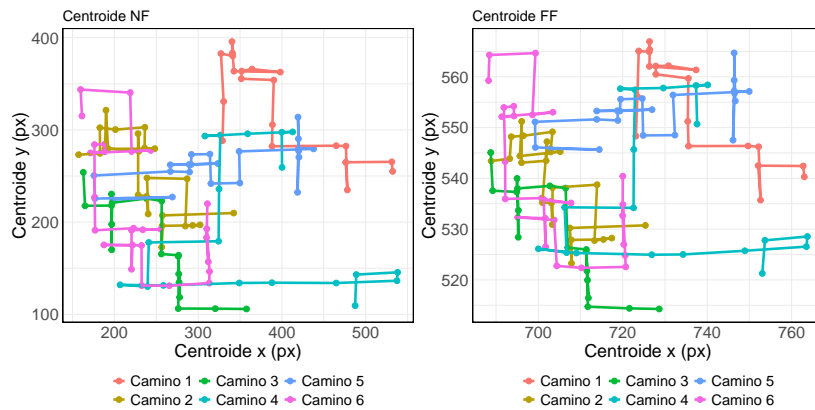


Figura 16: Caminos realizados por el centroide de NF. Figura 17: Caminos realizados por el centroide de FF.

Para hacer esta validación, se obtuvo un nuevo conjunto de datos del montaje real compuesto por seis subconjuntos de datos. En cada subconjunto se recorría un camino partiendo de un punto inicial diferente y se iban moviendo los actuadores en orden arbitrario, recogiendo el centroide antes y después de cada movimiento, así como el número de pasos movidos de cada actuador en cada iteración. De esta manera, se espera recorrer un rango lo más amplio posible dentro del sensor de la cámara y además no partir siempre desde el mismo punto inicial. Los centroides de estos movimientos se muestran en la figura 16 para el NF y 17 para el FF.

Una vez obtenidos los datos reales, se pasan al modelo previamente entrenado para predecir las posiciones finales de los centroides. Para evaluar los resultados obtenidos se recurre a las mismas métricas que para la evaluación anterior de los modelos que se muestran en la tabla 5. Además, en las figuras 18 y 19 muestran las comparativas en los datos de los centroides de NF reales frente a los predichos y en 20 y 21 los de FF.

Como se puede apreciar, los resultados son bastante buenos, aunque todavía mejorables, principalmente en el caso del centroide de FF. Se puede apreciar en la figura 20 como el modelo subestima el valor real del centroide de forma sistemática, además en la figura 21 se puede ver como los datos, están mucho más dispersos que en el caso del eje horizontal, de ahí que el valor de los errores MAE y RMSE sea mayor.

Tabla 5: Métricas obtenidas tras evaluar los valores de los centroides de NF y FF predichos.

Variable	R^2	MAE	MSE	RMSE
Centroide NFx	0.97	12.48	229.46	15.15
Centroide NFy	0.97	9.60	134.94	11.62
Centroide FFx	0.93	4.05	25.77	5.08
Centroide FFy	0.85	4.57	29.24	5.41

5. Conclusiones

En este trabajo se ha desarrollado un gemelo digital de un sistema de control de alineamiento de un láser basado en un montaje óptico que replica el sistema óptico de monitorización del alineamiento en la cadena de un láser de alta potencia.

El resultado obtenido tras realizar el mapeo sistemático de la literatura demuestra que el uso de modelos de inteligencia artificial en instalaciones láser cada vez va tomando más protagonismo, si bien todavía queda mucho trabajo por hacer en diversas tareas como el control de la energía, del frente de onda o en el tema que se trata este trabajo, el control del alineamiento del láser. Para el caso que trata este trabajo, la comunidad científica opta por construir sistemas auxiliares para no poner en peligro la integridad de las instalaciones láser, ya que pequeños daños y errores conllevarían grandes pérdidas económicas y dejarían el sistema obsoleto durante el tiempo de reparación.

Tras el estudio de los elementos más óptimos, se construyó el montaje óptico que replica el sistema de control de alineamiento con dos referencias, la de campo cercano o NF y la de campo lejano o FF, mediante el uso de cámaras para medir el centroide del haz y actuadores piezoeléctricos para determinar su posición.

Con este montaje se realizó la toma de datos del conjunto de datos para entrenar los modelos de predicción. Durante este proceso se decidió dividir la toma de datos en dos partes determinadas por los ejes horizontal y vertical, para entrenar dos modelos que predijeran los datos separados de la misma manera. Estos datos son expandidos sintéticamente posteriormente para aumentar la información que se pasa al modelo.

El resultado obtenido demostró la dependencia entre estas dos variables al aplicar el método de encadenado de regresores, por lo que este último fue el elegido para la construcción del gemelo digital del sistema de control de alineamiento.

Por último, para comprobar su funcionalidad, se realizó una prueba en el laboratorio, recreando los movimientos aleatorios dados por el montaje real, fuera del entorno simplificado de la toma de datos de entrenamiento, demostrando ser capaz de reproducir los mismos movimientos aunque con resultados mejorables.

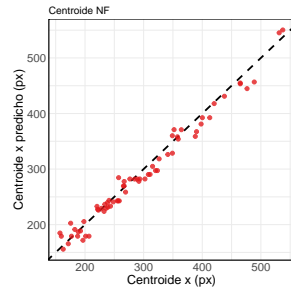


Figura 18: Comparativa centroide de NF horizontal.

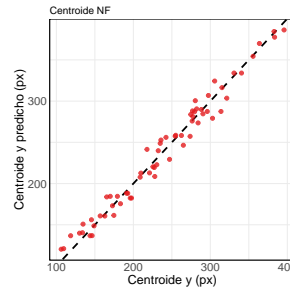


Figura 19: Comparativa centroide de NF vertical.

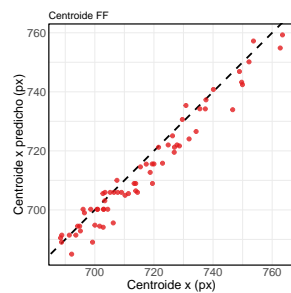


Figura 20: Comparativa centroide de FF horizontal.

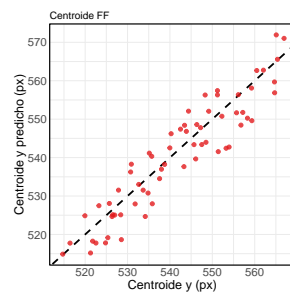


Figura 21: Comparativa centroide de FF vertical.

En conclusión, este trabajo demuestra la utilidad de los métodos de inteligencia artificial aplicados al control de alineamiento de un láser de alta potencia. Se ha conseguido generar un gemelo digital que recrea el sistema de monitorización de alineamiento mediante el uso de una red neuronal con arquitectura de perceptrón multicapa, obteniendo unos resultados buenos pero mejorables. Es esta mejora donde deberían de centrarse los trabajos futuros.

En esta memoria se ha descrito como se ha llevado a cabo la expansión del dataset de entrenamiento para cubrir la mayoría de posiciones dentro de las cámaras de monitoreo, por lo que una mejora sería aumentar el conjunto de datos utilizado con medidas reales en lugar de sintéticas.

Otro aspecto a mejorar sería la elección de los elementos utilizados en el sistema de control, como utilizar los actuadores en lazo cerrado para evitar esa no linealidad que existe al mover el actuador en un sentido u otro, si bien esto no solucionaría los problemas de precisión de sus movimientos.

El último punto para un trabajo futuro sería la implementación en la cadena del láser Vega de este gemelo digital para asistir en la tarea de alineamiento del sistema, ayudando a los operadores dedicados a ello y pudiendo operar en los casos en los que un operador no pueda hacer estas correcciones durante el proceso de experimentación.

Referencias

1. Maiman, T.H.: Stimulated Optical Radiation in Ruby Nature 187, 493–494 (1960).
2. Spetz, J.: Sources of Medical Technology: Universities and Industry. Institute of Medicine (US) Committee on Technological Innovation in Medicine; Rosenberg N, Gelijns AC, Dawkins H, editors. Washington (DC): National Academies Press (US); 1995.
3. Ready, J.F.: Effects Due to Absorption of Laser Radiation. Appl. Phys. 1 February 1965; 36 (2): 462–468
4. Laser Cutting History: Complete Timeline, Inventors and Evolution, <https://www.accurl.com/blog/laser-cutting-history/>
5. Goodwin, F.E.: A review of operational laser communication systems. Proceedings of the IEEE, vol. 58, no 10, pp: 1746-1752, 1970
6. Pioneering the Barcode. = <https://www.datalogic.com/eng/media-center/news/pioneering-the-barcode-nd2-8448.html>
7. M. G. Carasso, J.B.H. Peek, and J.P. Sinjou, The Compact disc digital audio' system, Philips Tech. Rev., (Special Issue), 40, 151-156 (1982).
8. Strickland, D. and Mourou, G.: Compression of amplified chirped optical pulses. Optics Communications, vol. 56, no. 3, pp. 219-221, 1985.
9. Laser Vega del CLPU. <https://www.clpu.es/laser-vega-pw/>
10. Biscari, C. and Aigner, E. and Attenkofer, K. and Casas, J. and Ferrer, S. and Matilla, O. and Nicolas, J. and Pascual, R. and Pérez, F. and Pont, M. and Sánchez, A.: ALBA Synchrotron and its future. The European Physical Journal Plus, vol. 138, no. 5, pp. 405, 2023.
11. De Roeck. Et. al.: From the LHC to future colliders. The European Physical Journal C, vol. 66, no. 3, pp. 525-583, 2010.
12. Volpe, L. and Fedosejevs et.al.: Generation of high energy laser-driven electron and proton sources with the 200 TW system VEGA 2 at the Centro de Laseres Pulsados. High Power Laser Science and Engineering, vol. 7, pp. E25, 2019
13. Capuano, Francesco and Peceli et.al.: TempoRL: laser pulse temporal shape optimization with Deep Reinforcement Learning. arXiv:2304.12187
14. Zheng, Yamin and Zhang: Deep learning enabled robust wavefront sensing for active beam smoothing with a continuous phase modulator High Power Laser Science and Engineering, Vol. 13, e19, 2025
15. . Amodio, Alessio and Wang: Pointing Stabilization of a 1Hz High Power Laser via Machine Learning High Power Laser Science and Engineering, vol vol. 13, e35, 2025
16. Page, Matthew J and McKenzie et.al.: The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ, vol. 372, n71, 2021
17. Oropeza-Gomez, Gregorio A. and Orozco-López et.al.: High-Precision Spatial Filtering with AI-driven control and multicolor laser testing for interferometric application. Integration, vol. 103, pp. 102406. July, 2025.

18. Mukund, N. and Lough, J.: Neural sensing and control in a kilometer-scale gravitational-wave observatory *Phys. Rev. Appl.* Vol. 20, no. 6, pp. 064041, 2023.
19. Khachikyan, Alexander and Pippione: Micro-optics assembly for fast axis collimation by means of convolutional neural network. *Opt. Express*, vol. 29, no. 17, pp. 26765—26774, 2021.
20. Li, Xunzheng and Peng: Deep learning assisted far-field multi-beam pointing measurement. *Opt. Express*, vol. 62, no. 8, pp. 086102, 2023.
21. Wang, Dan and Du: Machine-Learning Control of Coherent Combining of Fiber Lasers for Plasma Accelerators. *IEEE Advanced Accelerator Concepts*, pp. 1-5, 2022.
22. Bruchon, Niky and Fenu: Basic Reinforcement Learning Techniques to Control the Intensity of a Seeded Free- Electron Laser. *Electronics*, vol. 9, no. 5, pp. 781, 2020.
23. Mishra, Aashwin and Brennan: Machine Learning Based Alignment For LCLS-II-HE Optics- arXiv:2308.07521.
24. Li, Xunzheng and Peng: Deep learning assisted far-field multi-beam pointing measurement *Optical Engineering*, vol. 62, no. 8, pp. 086102, 2023
25. Liu, Aidong and Yang: A Novel Laser Spot Centroid Positioning Method in Satellite Communication Based on Deep Learning with Physical Information. *15th International Conference on Communication Software and Networks*, pp. 408-413, 2023.
26. Shin, Huicheol and Baek: Multidimensional Beam Optimization in Underwater Optical Wireless Communication Based on Deep Reinforcement Learning. *IEEE Internet of Things Journal*, vol. 11, no. 17, pp. 28623 – 28634, 2024
27. Robb, David A. and Risbridger, D.: Three Approaches to the Automation of Laser System Alignment and Their Resource Implications. arXiv:2409.11090.
28. Qin, Fangbo and Zhang: Laser Beam Pointing Control With Piezoelectric Actuator Model Learning. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 50, no. 3, pp.1024-1034, 2020
29. Guo, Pan and Fan: Intelligent Laser Emitting and Mode Locking of Solid-State Lasers Using Human-Like Algorithms. *Laser and Photonics Reviews*, vol. 18, no. 8, pp. 2301209, 2024.
30. González, Rafael C. and Woods, Richard E.: *Digital Image Processing* Pearson, ed. 4th, 2018.
31. MultiOutputRegressor <https://scikit-learn.org/stable/modules/generated/sklearn.multioutput.MultiOutputRegressor.html>
32. Librería Scikit-learn <https://scikit-learn.org/>
33. RegressorChain <https://scikit-learn.org/stable/modules/generated/sklearn.multioutput.RegressorChain.html>

Optimización energética de sistemas intralogísticos mediante técnicas de IA

Manal Jammal, Javier Parra Domínguez, and Juan Manuel Corchado Rodríguez

Departamento de Informática y Automática, Universidad de Salamanca,
España
mjammal@usal.es

Resumen La eficiencia energética se ha convertido en una prioridad estratégica en sistemas industriales automatizados. Esta investigación desarrolla una metodología híbrida que integra mapas autoorganizados (SOM) con redes LSTM bidireccionales y mecanismos de atención para establecer un marco predictivo de eficiencia energética. La contribución principal radica en la creación de un sistema que combina descubrimiento automático de patrones energéticos con predicción temporal avanzada, reduciendo la dependencia de conocimiento experto específico del dominio. El enfoque contribuye al avance de los objetivos de desarrollo sostenible 7, 9 y 13, facilitando la transición de gestión energética reactiva hacia estrategias predictivas en entornos industriales automatizados.

Keywords: Eficiencia energética, Mapas autoorganizados, Redes LSTM, Sistemas intralogísticos, Optimización industrial

1. Introducción

La demanda energética del sector industrial constituye un reto en la transición ecológica europea. En 2022, la industria fue responsable de aproximadamente el 25 % de la energía final consumida en la Unión Europea, según datos de Eurostat¹. Esta categoría hace referencia a la energía utilizada por los sectores, excluyendo las pérdidas por transformación y transporte. Para abordar esta situación, la Directiva (UE) 2023/1791 sobre eficiencia energética establece un objetivo vinculante de reducción del consumo de energía final del 11,7 % para el año 2030, respecto a las proyecciones de referencia, duplicando los compromisos anteriores y reforzando las obligaciones legales para los Estados miembros². En línea con esta política, el Plan Nacional Integrado de Energía

¹ Eurostat. (2024). Final energy consumption in industry – detailed statistics. European Commission.

² Unión Europea. (2023). Directiva (UE) 2023/1791 del Parlamento Europeo y del Consejo de 13 de septiembre de 2023 relativa a la eficiencia energética y por la que se modifica el Reglamento (UE) 2023/955. Diario Oficial de la Unión Europea.

y Clima 2021-2030 (PNIEC)³ identifica a la eficiencia energética industrial como uno de los pilares estratégicos para la descarbonización de la economía española, subrayando el potencial de mejora mediante tecnologías inteligentes de gestión energética y automatización.

En este contexto, las organizaciones industriales enfrentan presión creciente por parte de inversores, clientes y reguladores para demostrar compromiso con la reducción de emisiones de carbono, convirtiendo la eficiencia energética en un imperativo tanto económico como estratégico. Los sistemas de almacenamiento automatizados representan un ejemplo paradigmático de esta problemática. Estos sistemas, ampliamente utilizados en entornos industriales para gestión de inventarios y transporte automatizado de materiales, integran múltiples componentes electromecánicos que operan coordinadamente: cintas transportadoras, raíles verticales motorizados, sensores de posición y sistemas de control programable. La complejidad inherente resulta en patrones de consumo energético dinámicos y multivariados, donde la eficiencia global depende tanto del rendimiento individual de cada componente como de las interacciones temporales entre ellos. Paralelamente, el desarrollo de marcos regulatorios para autoconsumo energético renovable, como los sistemas de net-metering y net-billing implementados en diversos países [?], está configurando un ecosistema energético que favorece la integración de soluciones de optimización energética en entornos industriales. A pesar de que estos sistemas generan continuamente grandes volúmenes de datos operativos a través de sensores integrados, incluyendo mediciones de potencia, voltaje, posición y estados de activación, existe una brecha tecnológica fundamental entre la abundancia de datos disponibles y su utilización efectiva para optimización energética [?]. La problemática central radica en que los sistemas de control actuales se basan predominantemente en lógicas reactivas que responden a eventos ya ocurridos, careciendo de capacidades de anticipación que permitan ajustar proactivamente parámetros operativos para minimizar consumo energético sin comprometer funcionalidad. Esta limitación se agrava por la naturaleza estocástica de los patrones de consumo, donde transiciones entre estados operativos generan picos de demanda energética impredecibles [39].

La motivación para desarrollar soluciones inteligentes encuentra fundamento en tres pilares críticos: el económico, donde la optimización puede alcanzar reducciones de costes del 15–30 % según estudios recientes [40]; el regulatorio, que impone obligaciones legales crecientes con sanciones por incumplimiento; y el ambiental, donde Los sistemas inteligentes de gestión energética no solo mejoran la eficiencia, sino que también aportan valor estratégico. Su implementación refuerza la imagen corporativa, facilita el acceso a financiación sostenible y contribuye al cumplimiento de los Objetivos de Desarrollo Sostenible estableci-

³ Gobierno de España – Ministerio para la Transición Ecológica y el Reto Demográfico (MITECO). (2020). Plan Nacional Integrado de Energía y Clima (PNIEC) 2021–2030.

dos en la Agenda 2030, en particular los ODS 7, 9 y 13⁴. El contexto tecnológico actual presenta una convergencia única de factores que hacen viable la aplicación de técnicas avanzadas de inteligencia artificial para optimización energética industrial. La proliferación de sensores IoT de bajo coste y alta precisión ha democratizado el acceso a datos operativos en tiempo real, mientras el incremento exponencial de capacidad de procesamiento computacional hace posible la implementación de algoritmos de aprendizaje automático complejos en entornos industriales [?]. La disponibilidad de frameworks de aprendizaje automático como TensorFlow y PyTorch, junto con el creciente volumen de datos operativos generados por sensores IoT industriales, ha facilitado el desarrollo de sistemas basados en inteligencia artificial para análisis energético. Los mapas autoorganizados (SOM) permiten identificar patrones en datos multivariados y generar eventos discretos a partir de señales continuas [36]. En el contexto de sistemas energéticos complejos, esta capacidad de segmentación automática puede combinarse con redes neuronales LSTM que incorporan mecanismos de atención para predicción de secuencias temporales [44], y con algoritmos de análisis de grafos que facilitan el modelado de transiciones entre estados operativos.

La integración de estas técnicas en sistemas de optimización energética industrial constituye el objeto de investigación de este trabajo.

1.1 Objetivo del trabajo

El objetivo general de este trabajo es diseñar un sistema basado en inteligencia artificial que permita caracterizar y predecir el comportamiento energético en un entorno intralogístico automatizado, con el fin de identificar oportunidades de optimización que puedan traducirse en mejoras de eficiencia energética.

De forma específica, se persigue caracterizar los distintos estados energéticos del sistema a partir de datos reales de operación, identificar patrones de transición entre dichos estados, y desarrollar un modelo predictivo capaz de anticipar secuencias ineficientes mediante técnicas de aprendizaje profundo. Además, se pretende evaluar el impacto potencial de las estrategias propuestas en términos energéticos, económicos y ambientales.

2. Estado del arte

El presente apartado tiene la finalidad de revisar el estado actual del conocimiento sobre estrategias de optimización energética apli-

⁴ Naciones Unidas. (2015). Transformar nuestro mundo: la Agenda 2030 para el Desarrollo Sostenible. Resolución A/RES/70/1 de la Asamblea General de la ONU.

cadadas a sistemas de transporte intralogístico, como cintas transportadoras automatizadas empleadas en entornos industriales. El análisis se organiza en torno a dos ejes principales que reflejan la evolución reciente del campo:

1. En primer lugar, se examinan las soluciones convencionales de eficiencia energética, que abarcan mejoras mecánicas, eléctricas y estrategias de control lógico en sistemas de transporte continuo.
2. En segundo lugar, se analizan las técnicas basadas en inteligencia artificial, aplicadas a la caracterización del comportamiento energético, la predicción de la demanda y el control operativo de dichos sistemas.

2.1 Relevancia energética de los sistemas de transporte intralogístico

Los sistemas de transporte intralogístico, tales como cintas transportadoras, sistemas de manipulación automatizada y vehículos guiados, representan un componente crítico del consumo energético en entornos industriales. En plantas con automatización media, las cintas transportadoras pueden llegar a consumir hasta el 15 % del total energético, concentrando el 90 % de dicho consumo en la propia cinta motriz, mientras que los motores y convertidores de frecuencia representan una fracción menor [?].

Este alto consumo se agrava en condiciones de operación ineficiente, como los modos inactivos, donde la cinta sigue consumiendo entre el 51 % y el 92 % de su potencia nominal sin transportar material alguno, causando un desperdicio energético significativo [?].

Este tipo de ineficiencias ha motivado la adopción de mejoras en hardware como motores de alta eficiencia, variadores de velocidad (VFD) y tecnologías regenerativas. Un estudio reciente destaca que los VFD, al modular la velocidad en función de la carga, permiten ahorros energéticos de entre un 10 % y 75 %, con una media europea estimada del 2 % al 30 % en consumo eléctrico de motores [?, 4]. Estas conclusiones reflejan y amplían las estimaciones de investigaciones previas sobre la importancia de los sistemas motorizados en la transición energética industrial [?].

Paralelamente, se han implementado estrategias de control y monitorización avanzada. El uso de controladores inteligentes y sistemas de medición en tiempo real permite ajustar dinámicamente la operación según la demanda real. En cintas, estos sistemas han mostrado reducciones superiores al 60 % en pérdidas por operación en vacío [?]. En redes industriales, los sistemas de compensación reactiva y controladores adaptativos mejoran la calidad de la energía y reducen pérdidas [?, ?].

Para justificar inversiones, es fundamental cuantificar los ahorros y emplear indicadores de rendimiento energético. Técnicas de simulación de flujos de energía permiten estimar de forma precisa la desviación de consumo y emisiones antes y después de las intervenciones [?].

Además, estudios de adopción en PYMEs muestran cómo los análisis coste-beneficio y ROI son clave para impulsar la eficiencia energética industrial [?].

En conjunto, estos estudios refuerzan la importancia de abordar la eficiencia energética desde una perspectiva sistémica, incorporando tanto mejoras en hardware (motores, variadores, sistemas regenerativos) como estrategias avanzadas de control y monitorización. La cuantificación de los ahorros y el desarrollo de indicadores específicos permiten justificar económicamente las inversiones y alinearlas con objetivos de sostenibilidad y reducción de emisiones.

2.2 Mapeo sistemático de técnicas de inteligencia artificial para la optimización energética en sistemas intralogísticos

El presente mapeo sistemático se fundamenta en la guía PRISMA 2020 (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) [45] y sigue un procedimiento estructurado de selección en cinco fases secuenciales: búsqueda inicial, eliminación de duplicados, lectura de título y resumen, verificación del acceso al texto completo y evaluación de calidad metodológica. En la primera fase, se realizó una búsqueda sistemática en dos bases de datos científicas: Scopus y Web of Science. Para ambas plataformas se definió una ecuación de búsqueda general combinando términos relacionados con inteligencia artificial, eficiencia energética y sistemas industriales reales, con especial atención a entornos intralogísticos. La estrategia se limitó a publicaciones científicas revisadas por pares, escritas en inglés y publicadas entre 2020 y 2025. La ecuación de búsqueda principal fue la siguiente:

```
("artificial intelligence" OR "machine learning"
OR "deep learning" OR "self-organizing map" OR "LSTM")
AND ("energy efficiency" OR "energy optimization"
OR "energy consumption")
AND ("industrial system" OR "production"
OR intralogistics)
```

Esta fórmula se adaptó a los requisitos de cada base de datos. Se obtuvieron un total de 120 resultados: 53 en Scopus y 67 en Web of Science. A continuación, se eliminaron 35 estudios duplicados, lo que dio lugar a una colección depurada de 85 estudios únicos. En la tercera fase, se llevó a cabo una lectura crítica de los títulos y resúmenes para evaluar su pertinencia temática. Se aplicaron criterios de exclusión que descartaban aquellos trabajos que no abordaban la optimización energética, no utilizaban técnicas de inteligencia artificial o no estaban enmarcados en entornos industriales. También se eliminaron estudios excesivamente genéricos o conceptuales sin aplicación técnica. Como resultado de esta fase, se seleccionaron 38 estudios que cumplían los requisitos básicos de alineación temática con el objetivo de este trabajo. Posteriormente,

en la cuarta fase, se verificó la disponibilidad del texto completo de los 38 estudios preseleccionados. Esta revisión permitió confirmar el acceso íntegro a 36 estudios, mientras que 2 estudios fueron eliminados al no estar disponibles para su análisis en profundidad. Finalmente, en la quinta fase, se evaluó la calidad metodológica de los 36 estudios con acceso completo. Se analizaron aspectos como la claridad en la descripción técnica de los métodos empleados, la aplicabilidad práctica de los enfoques propuestos, y la presentación de resultados cuantificables o replicables. Se eliminaron 6 estudios que, pese a su aparente relevancia, no describían con suficiente detalle la técnica de IA utilizada. Así, el conjunto final quedó compuesto por 30 estudios, que constituyen la base del análisis desarrollado en el presente trabajo. A continuación, se presentan los criterios de inclusión y exclusión utilizados para la selección de estudios en esta revisión:

Criterios de inclusión
El estudio debía aplicar técnicas de inteligencia artificial para optimizar el consumo energético.
El contexto debía ser industrial.
El documento debía estar publicado entre 2020 y 2025.
Debía estar escrito en inglés.
Debía tratarse de un artículo científico o comunicación en congreso revisado por pares.
El texto completo debía estar disponible.
El estudio debía aportar métricas o resultados cuantificables relacionados con la eficiencia energética.

Tabla 1: Criterios de inclusión aplicados en la selección de estudios.

Criterios de exclusión
Estudios centrados en ámbitos no industriales.
Publicaciones que no aplican técnicas de inteligencia artificial o se limitan a enfoques clásicos.
Documentos sin acceso al texto completo.
Artículos que no presentan resultados relacionados con eficiencia energética.

Tabla 2: Criterios de exclusión aplicados en la selección de estudios.

Antes de incluir los estudios en el análisis final del mapeo, se aplicó una evaluación de calidad con el objetivo de asegurar que cada trabajo seleccionado aportara información útil y relevante para responder a las preguntas de investigación planteadas. Esta evaluación se llevó a cabo tras la lectura completa de los artículos que superaron la fase de selección inicial. Para ello, se definieron una serie de preguntas que permitieron comprobar si el estudio abordaba aspectos clave como: ¿Se

detallan los objetivos del estudio de forma clara y específica? ¿La metodología empleada es adecuada para alcanzar los objetivos planteados? ¿El estudio aplica una técnica de inteligencia artificial relacionada con la optimización energética? ¿Incluye métricas para evaluar el rendimiento energético o la eficiencia? ¿Hace referencia al entorno de aplicación industrial? ¿Se indican las herramientas o entornos utilizados para el modelado o implementación? ¿El estudio proporciona información útil para al menos una de las preguntas de investigación planteadas? Cada pregunta se formuló como un criterio binario (sí o no) y se aplicó a todos los estudios seleccionados para su revisión completa. Los estudios fueron incluidos en el análisis final si respondían afirmativamente a la tercera pregunta (uso efectivo de IA para optimización energética) y obtenían al menos tres respuestas afirmativas del total de siete criterios evaluados. Tras completar las fases de selección, exclusión y evaluación de calidad, se obtuvieron un total de 30 estudios que cumplen con los criterios establecidos. El desarrollo completo de este proceso se resume en la Figura 1.

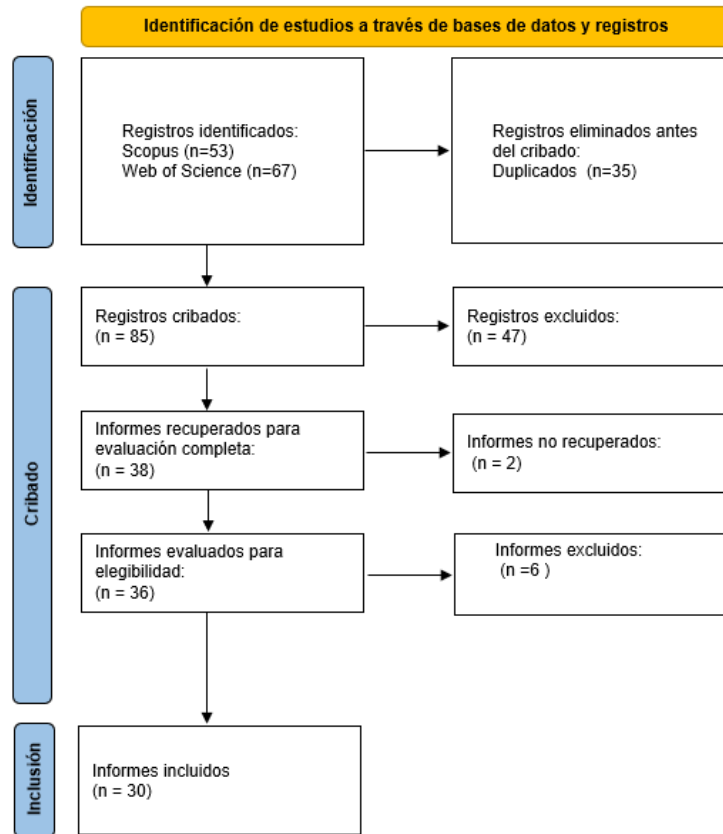


Figura 1: Diagrama de flujo basado en la metodología PRISMA 2020

El análisis del corpus seleccionado identifica un conjunto de trabajos que aplican técnicas de clustering para el análisis de patrones energéticos en sistemas industriales. Otros trabajos como [8] proponen un enfoque para la detección de anomalías en patrones de consumo energético en sistemas ciber-físicos industriales, aplicando técnicas de aprendizaje automático no supervisado. Su metodología se basa en la premisa de que los patrones de consumo energético anómalos pueden identificarse mediante análisis estadístico de datos históricos de operación. Los mapas auto-organizados (SOM) aparecen como una técnica recurrente en múltiples estudios del conjunto analizado. [36] Desarrollan una metodología que utiliza SOM para aprender autómatas temporales híbridos en ausencia de eventos discretos explícitos. Los autores abordan el problema de que las señales binarias o el conocimiento experto para generar eventos discretos no siempre están disponibles en plantas industriales, imposibilitando el aprendizaje de autómatas. Su propuesta combina SOM con transformaciones watershed como alternativa no

supervisada y no paramétrica para esta limitación. El trabajo de [44] extienden esta aproximación, estableciendo que los SOM pueden utilizarse para habilitar el aprendizaje de autómatas temporales híbridos cuando los eventos discretos no están disponibles. La metodología propuesta se orienta hacia sistemas de diagnóstico basado en modelos para identificación de anomalías y causas raíz en sistemas de producción ciber-físicos, donde la creación manual de modelos requiere esfuerzo considerable y no resulta adecuada para sistemas que cambian rápidamente.

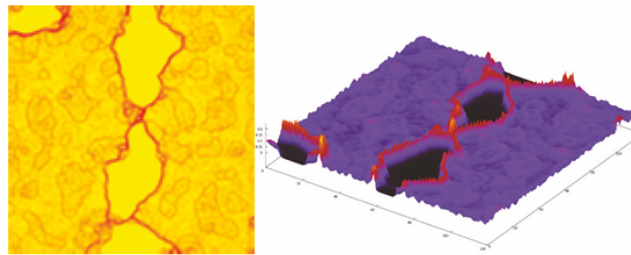


Figura 2: la matriz u de un SOM de 120×120 . Fuente: von Birgelen y Niggemann (2017).

De Santana et al. (2023) [9] aplican técnicas de minería de datos para mejoras de eficiencia energética en el suministro de servicios auxiliares de una planta petroquímica. Su estudio se centra en el análisis de datos operativos para identificar oportunidades de optimización energética mediante técnicas de clustering y análisis de patrones. Los autores reportan la implementación de su metodología en un entorno industrial real. [12] desarrollan una metodología que combina machine learning con algoritmos evolutivos para asignación energéticamente eficiente de buffers y tasas de servicio en sistemas manufactureros. Su trabajo incorpora técnicas de clustering como componente de un sistema más amplio de optimización energética, aplicado específicamente a sistemas de manejo de materiales. Los autores proponen algoritmos híbridos que integran múltiples técnicas de optimización, incluyendo métodos de aprendizaje automático para caracterización de estados operativos. La aplicación de técnicas de clustering en el contexto de sistemas de almacenamiento automatizados se observa en el trabajo de [5], quienes proponen algoritmos evolutivos multi-objetivo para programación integrada de producción y mantenimiento considerando aspectos energéticos. Su aproximación incluye técnicas de segmentación para identificación de configuraciones operativas, aunque el enfoque principal se orienta hacia optimización multiobjetivo más que hacia caracterización energética específica. Los trabajos analizados coinciden en identificar las técnicas de clustering no supervisado como metodologías viables para

análisis de patrones energéticos en sistemas industriales, particularmente cuando no se dispone de conocimiento previo estructurado sobre los estados operativos del sistema.

La aplicación de redes neuronales recurrentes, particularmente modelos LSTM, para predicción de consumo energético constituye una línea de investigación consolidada en el conjunto analizado. [?] implementan una red LSTM basada en mecanismo de atención para predicción de consumo energético HVAC a ultra-corto plazo. Los autores utilizan datos reales de consumo energético recolectados en el Kitakyushu Science Research Park en Japón para entrenamiento y validación del modelo. Su estudio incluye comparación con cinco modelos baseline: A-LSTM, LSTM, RNN, DNN y SVR, aplicando el algoritmo Tree-structured Parzen Estimators para optimización de hiperparámetros. El trabajo de [?] reporta que el modelo A-LSTM alcanza mejor rendimiento que los modelos baseline en múltiples perspectivas de análisis, incluyendo comparaciones en diferentes tamaños de conjunto de entrenamiento y diferentes modos de operación del sistema. [39] desarrollan un modelo que combina LSTM con optimización por enjambre de partículas para gestión energética utilizando datos de bombas de calor del Reino Unido. Su trabajo se enfoca en pronóstico de carga a corto plazo y compara LSTM con redes neuronales backpropagation, SARI-MA, Random Forest y suavizado exponencial Holt-Winters. Los autores reportan que LSTM alcanza el mejor MAPE promedio de 1.59 % para el conjunto de datos de prueba, superando a otros modelos comparados. El mismo estudio incluye validación estadística mediante test de rangos signados de Wilcoxon, estableciendo que la mejora en precisión de predicción de LSTM comparado con otros modelos es estadísticamente significativa, excepto cuando se compara con RF en datos de entrenamiento. Los autores investigan el impacto de la resolución de agregación de datos sobre la precisión del pronóstico, concluyendo que el modelo LSTM propuesto presenta mejor rendimiento en datos de resolución de 1 hora comparado con otros períodos temporales. [33] introducen una red LSTM informada por física (Physics-Informed LSTM) para identificación de flexibilidad en sistemas de refrigeración evaporativa. Su aproximación integra conocimiento físico del dominio en la arquitectura de la red neuronal, orientándose hacia aplicaciones en sistemas de control industrial. Los autores abordan el problema de modelado de sistemas complejos donde las relaciones físicas subyacentes pueden incorporarse explícitamente en el proceso de aprendizaje. La aplicación de técnicas de aprendizaje profundo para optimización energética en sistemas de robótica se observa en [18], quienes desarrollan metodologías de planificación de trayectorias energéticamente eficientes para robots industriales utilizando aprendizaje por refuerzo profundo paralelo. Su trabajo se centra en una clase específica de robots industriales, aplicando técnicas de optimización para reducir consumo energético durante operaciones de movimiento. Los autores proponen algoritmos que

consideran tanto eficiencia energética como rendimiento operativo en la planificación de trayectorias. [2] desarrollan un modelo de pronóstico mejorado para carga eléctrica a corto plazo en empresas papeleras orientado a optimización de líneas de producción. Su trabajo aborda específicamente el contexto de industrias intensivas en energía, donde la predicción precisa de demanda energética resulta crítica para optimización operativa. Los autores se enfocan en el desarrollo de modelos predictivos adaptados a las características específicas de procesos industriales de manufactura de papel. [21] implementan redes neuronales DeepESN para mantenimiento predictivo industrial mediante detección de anomalías basada en datos de energía de producción. Su aproximación se orienta hacia la identificación de patrones anómalos en consumo energético como indicadores de necesidades de mantenimiento. Los autores aplican su metodología a datos reales de producción industrial. La integración de mecanismos de atención en arquitecturas LSTM aparece como una tendencia recurrente en los trabajos analizados. Múltiples estudios incorporan estas técnicas para mejorar la capacidad de los modelos de enfocarse en aspectos temporales específicos relevantes para la predicción energética, aunque la implementación específica de estos mecanismos varía entre diferentes aproximaciones metodológicas.

La implementación de sistemas de control inteligente para optimización energética en entornos industriales constituye una línea de investigación prominente en el corpus analizado. [4] desarrollan metodologías de aprendizaje por refuerzo para control energéticamente eficiente de máquinas paralelas e idénticas. Su trabajo aborda el problema de programación de producción considerando simultáneamente aspectos de eficiencia energética y productividad operativa. Los autores implementan algoritmos de aprendizaje por refuerzo que aprenden políticas de control adaptativo para reducción de consumo energético sin comprometer objetivos de producción. [1] abordan análisis transitorio de desempeño productivo y consumo energético en sistemas de producción flexibles. Su estudio se centra en sistemas manufactureros donde la configuración geométrica afecta tanto la productividad como el consumo energético. Los autores desarrollan modelos que integran variables geométricas del sistema con métricas de eficiencia energética, adoptando un enfoque que considera múltiples objetivos operativos simultáneamente. [6] investigan el impacto de la inteligencia industrial sobre la intensidad energética utilizando datos de China. Su estudio analiza cómo la implementación de tecnologías de inteligencia artificial en el sector industrial afecta el consumo energético por unidad de producción. Los autores utilizan datos de empresas manufactureras chinas para examinar la relación entre adopción de tecnologías inteligentes y eficiencia energética sectorial, aunque no especifican las tecnologías específicas de IA analizadas. [10] examinan si la inteligencia artificial puede mejorar la eficiencia energética de empresas manufactureras, utilizando datos de empresas chinas. Su investigación se basa en datos

de encuestas integrales de empresas manufactureras en la provincia de Guangdong, midiendo la inteligencia artificial a través del uso de robots industriales. Los autores concluyen que la inteligencia artificial mejora la eficiencia energética de empresas manufactureras, principalmente mediante promoción del progreso tecnológico, mientras que el impacto sobre la eficiencia tecnológica de las empresas no resulta significativo. [7] desarrolla sistemas de soporte a decisiones impulsados por inteligencia artificial para gestión energética sostenible en ciudades inteligentes. Su trabajo se orienta hacia aplicaciones urbanas más que industriales específicas, pero incluye componentes relevantes para sistemas de gestión energética industrial. El autor propone marcos conceptuales para integración de múltiples fuentes de datos y algoritmos de IA para optimización energética en gran escala. [22] desarrollan algoritmos de enfrentamiento energético eficiente basados en aprendizaje para redes de sensores inalámbricos verdes en colaboración terminal-edge. Su trabajo, aunque orientado hacia redes de comunicación, incluye principios de optimización energética aplicables a sistemas industriales distribuidos. Los autores proponen algoritmos que balancean eficiencia energética con rendimiento de comunicación en entornos industriales conectados. [23] implementan un sistema de control de instalaciones basado en IA explicable para ahorro energético y reducción de carbono. Su trabajo se centra en el desarrollo de sistemas de control que no solo optimizan el consumo energético sino que también proporcionan explicaciones comprensibles sobre las decisiones de control tomadas. Los trabajos analizados en esta categoría reflejan una evolución hacia sistemas de control más sofisticados que integran múltiples objetivos operativos, incluyendo eficiencia energética, productividad y sostenibilidad ambiental. La tendencia observada sugiere una transición desde enfoques de optimización single-objetivo hacia metodologías multi-objetivo que consideran la complejidad inherente de los sistemas industriales modernos.

La aplicación de técnicas de inteligencia artificial para mantenimiento predictivo y monitorización de sistemas industriales representa una línea de investigación consolidada en el corpus analizado. [25] desarrollan un sistema de mantenimiento predictivo basado en machine learning para máquinas de hilado artificial. Su trabajo implementa algoritmos de aprendizaje automático para análisis de datos operativos con el objetivo de anticipar necesidades de mantenimiento y optimizar programas de intervención preventiva. [24] proponen un enfoque de machine learning para programación energéticamente eficiente de transporte inteligente en circunstancias dinámicas del mundo real. Su estudio aborda el problema de optimización de rutas y programación de vehículos considerando eficiencia energética como objetivo primario. Los autores desarrollan algoritmos que adaptan las decisiones de transporte a condiciones operativas cambiantes en tiempo real. [19] desarrollan un método de optimización para refrigeración industrial basado en datos que considera pronóstico de demanda. Su trabajo se centra en sistemas de

refrigeración industrial, implementando metodologías data-driven para optimización de consumo energético. Los autores integran modelos de predicción de demanda con sistemas de control para optimización operativa en procesos de refrigeración continua. [32] contribuyen con sensores suaves basados en aprendizaje profundo para mejora de flexibilidad en automatización industrial. Su trabajo aborda el desarrollo de sensores virtuales que utilizan técnicas de aprendizaje profundo para estimación de variables de proceso no medidas directamente. Los autores se enfocan en aplicaciones que requieren adaptabilidad y flexibilidad en entornos de automatización industrial. [34] examina el ahorro energético mediante innovación sinérgica con digitalización en el contexto de la industria de pulpa y papel. Su trabajo analiza cómo la implementación de tecnologías digitales puede contribuir al ahorro energético en procesos industriales específicos. El autor se centra en identificar oportunidades de sinergia entre innovación tecnológica y eficiencia energética. [41] analizan las oportunidades de la Industria 4.0 para sostenibilidad energética. Su trabajo examina cómo las tecnologías pueden contribuir a objetivos de sostenibilidad energética en entornos manufactureros. Los autores desarrollan un modelo estructural interpretativo para mapear las interrelaciones entre varias funciones de sostenibilidad energética de la industria 4.0.

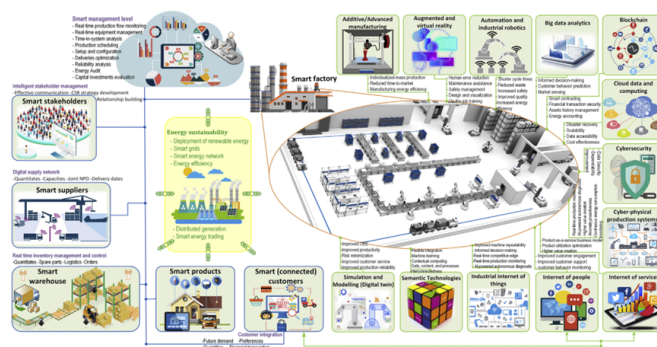


Figura 3: Presentación esquemática de la arquitectura de la Industria 4.0. Fuente: Pathak (2024).

La convergencia observada en estos trabajos hacia sistemas de monitorización y mantenimiento predictivo sugiere una evolución del campo hacia aproximaciones proactivas para gestión energética industrial, donde la anticipación de condiciones operativas permite optimización preventiva más que correctiva.

Para complementar el análisis sistemático desarrollado, se implementó un análisis bibliométrico del corpus seleccionado utilizando

técnicas de mapeo científico. El análisis de concurrencia de palabras clave se realizó mediante VOSviewer.

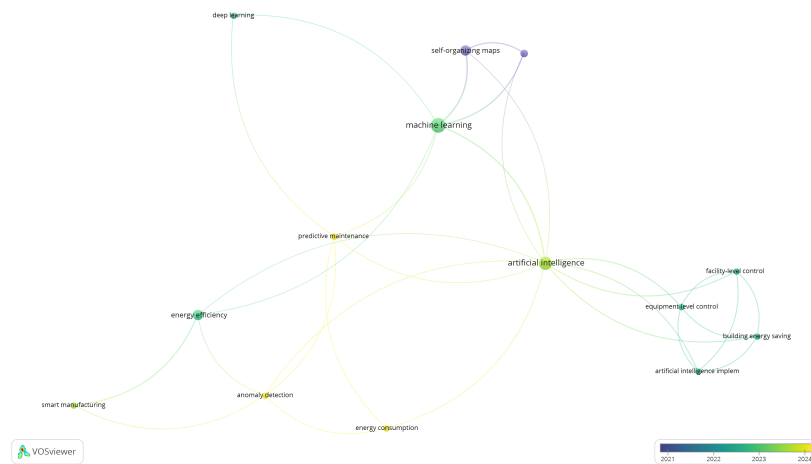


Figura 4: Mapa de co-ocurrencias de palabras clave del corpus analizado (2020-2025)

El área de color verde se articula en torno a términos metodológicos centrales como “machine learning”, “artificial intelligence”, “deep learning” y “neural networks”. Este cluster constituye el núcleo técnico del dominio, reflejando la consolidación de estas tecnologías como paradigmas fundamentales. La presencia de términos específicos como “LSTM” y “predictive maintenance” en estrecha proximidad propone la integración de técnicas de predicción temporal con mantenimiento preventivo. La concurrencia frecuente entre “deep learning” y términos específicos como “energy efficiency” sugiere la especialización de estas técnicas hacia aplicaciones energéticas, coherente con los hallazgos del mapeo sistemático que identifican una consolidación de arquitecturas LSTM para predicción temporal energética.

El área de color azul agrupa conceptos relacionados con técnicas de análisis no supervisado, destacando términos como “self-organizing maps”, “clustering”, “anomaly detection” y “energy efficiency”. La proximidad de estos términos indica su aplicación conjunta en metodologías de caracterización automática sin supervisión previa. La posición de “self-organizing maps” como nodo central dentro de este cluster refuerza la relevancia de esta técnica específica para caracterización energética, concordante con los precedentes establecidos por [36] y otros trabajos que implementan SOM para segmentación automática de estados opera-

tivos. La concurrencia entre “clustering” y “anomaly detection” sugiere la aplicación integrada de estas técnicas.

Los términos relacionados con sistemas de control sofisticados, como “equipment-level control”, “building energy saving” y conceptos emergentes como “explainable AI”. La emergencia de “explainable AI”, particularmente en trabajos recientes como [23], refleja la tendencia identificada hacia sistemas que proporcionan comprensión sobre decisiones operativas. La proximidad entre términos de control y explicabilidad sugiere la consolidación de aproximaciones que integran optimización de rendimiento con interpretabilidad operativa, validando arquitecturas metodológicas que combinen técnicas automáticas con capacidades explicativas.

La presencia de términos como industrial “automation” y “smart manufacturing” en proximidad con energy “consumption” indica la integración de objetivos de eficiencia energética en sistemas de automatización industrial más amplios. La aparición de “facility-level control” como término emergente sugiere una evolución hacia enfoques de control más granulares y específicos por nivel de sistema, coherente con la tendencia observada en el mapeo hacia aplicaciones industriales reales validadas en sistemas operativos.

3. Metodología

3.1 Descripción del sistema HRSS

Este trabajo utiliza el conjunto de datos High Rack Storage System (HRSS), un sistema desarrollado por SmartFactoryOWL en colaboración con el instituto inIT, ambos situados en Lemgo (Alemania). El sistema funciona como demostrador de tecnologías de la Industria 4.0 y fue diseñado con el propósito de analizar la eficiencia operativa y detectar anomalías en procesos industriales automatizados. El proceso está compuesto por cuatro cintas transportadoras (BLO, BHL, BHR, BRU) y dos raíles verticales (HR, HL). Las cintas centrales (BHL y BHR) están montadas sobre los raíles y pueden moverse verticalmente, mientras que BLO y BRU permanecen fijas. Cada cinta está equipada con tres sensores de inducción que permiten detectar el paso y la posición de los objetos transportados. El ciclo estándar de funcionamiento comienza con la colocación de la carga en el extremo derecho de la cinta BRU. Desde ahí, se desplaza horizontalmente hasta alcanzar el extremo derecho de la cinta BHR. A continuación, las cintas BHR y BHL ascienden en paralelo mientras la carga permanece sobre BHR. Una vez alcanzada la altura deseada, la carga es transferida horizontalmente desde BHR al extremo derecho de la cinta BHL, continuando el recorrido hasta llegar al extremo opuesto de BLO. El archivo HRSS_normal_standard.csv contiene 25.907 registros correspondientes al funcionamiento normal, compuesto por 20 variables que recogen mediciones sincronizadas de

distintos componentes del sistema, incluyendo variables temporales, de posición, potencia y voltaje.

3.2 Modelado de estados con SOM

El objetivo del SOM dentro del sistema de optimización energética consiste en crear una representación topológica de los estados operativos normales del sistema HRSS, clasificando cada estado según su índice de desempeño energético. Esta representación sirve como base para establecer un mapa de referencia del comportamiento energético normal, identificar regiones de alta y baja eficiencia energética, y proporcionar una segmentación automática de estados operativos. El modelo se entrenó con las variables potencia, voltaje y posición correspondientes a los seis componentes principales. El preprocesamiento consistió en la imputación de valores nulos mediante interpolación lineal, la detección de valores atípicos mediante el criterio de rango intercuartílico, y una normalización Z-score para asegurar escalas homogéneas entre las variables. La arquitectura del SOM se configuró como una rejilla rectangular bidimensional cuyas dimensiones se calcularon mediante una heurística adaptativa basada en el tamaño del conjunto de datos. El número total de neuronas se estableció siguiendo la regla empírica de $5\sqrt{n}$, donde n representa el número de muestras de entrenamiento. Para los 25.907 registros disponibles, esto resultó en aproximadamente 806 neuronas, distribuidas en una rejilla de 32×26 neuronas. El proceso

completo de entrenamiento puede resumirse en el siguiente algoritmo:

Algorithm 1: Algoritmo de Entrenamiento SOM

Input: $X = \{x_1, x_2, \dots, x_n\} \in \mathbb{R}^{n \times d}$ (datos HRSS normalizados),
grid_size = (rows, cols), $\sigma_0, \alpha_0, \text{max_iter}$
Output: $W = \{w_{i,j}\} \in \mathbb{R}^{\text{rows} \times \text{cols} \times d}$ (matriz de pesos entrenada)

- 1 Inicializar matriz de pesos W aleatoriamente desde distribución uniforme $[-1, 1]$;
- 2 Calcular constantes de decaimiento: $\tau_\alpha \leftarrow \frac{\text{max_iter}}{4}$,
 $\tau_\sigma \leftarrow \frac{\text{max_iter}}{8}$;
- 3 **for** $t = 1$ **to** max_iter **do**
- 4 $\alpha(t) \leftarrow \alpha_0 \cdot \exp\left(-\frac{t}{\tau_\alpha}\right)$;
- 5 $\sigma(t) \leftarrow \sigma_0 \cdot \exp\left(-\frac{t}{\tau_\sigma}\right)$;
- 6 Seleccionar muestra aleatoria $x(t)$ de X ;
// Encontrar Best Matching Unit **forall** neuronas (i, j) **do**
- 7 $\text{dist}[i, j] \leftarrow \|x(t) - w_{i,j}(t)\|_2$;
- 8 $(c_1, c_2) \leftarrow \text{arg m\u00edn}(\text{dist})$;
- 9 // Actualizar vecindario **forall** neuronas (i, j) **do**
- 10 $d_{\text{grid}} \leftarrow \|(i, j) - (c_1, c_2)\|_2$;
- 11 $h \leftarrow \exp\left(-\frac{d_{\text{grid}}^2}{2\sigma(t)^2}\right)$;
- 11 $w_{i,j}(t+1) \leftarrow w_{i,j}(t) + \alpha(t) \cdot h \cdot [x(t) - w_{i,j}(t)]$;
- 12 **return** W ;

La evaluaci\u00f3n del desempe\u00f1o energ\u00e9tico por estado se fundamenta en el c\u00e1lculo de un \u00edndice de desempe\u00f1o energ\u00e9tico que relaciona el consumo observado con el baseline energ\u00e9tico establecido. El \u00edndice de desempe\u00f1o energ\u00e9tico para cada componente c se define como:

$$\text{IPE}_c = \frac{\text{baseline}_c}{\text{potencia_media}_c} \quad (1)$$

El \u00edndice global del sistema se calcula como la media ponderada de los \u00edndices individuales:

$$\text{IPE}_{\text{sistema}} = \frac{\sum_c (\text{baseline}_c \cdot \text{IPE}_c)}{\sum_c \text{baseline}_c} \quad (2)$$

3.3 Modelo LSTM para predicci\u00f3n de ineficiencias energ\u00e9ticas

La fase de predicci\u00f3n implementa un predictor basado en redes neuronales LSTM bidireccionales con mecanismo de atenci\u00f3n multi-cabeza para la predicci\u00f3n preventiva de ineficiencias energ\u00e9ticas en

secuencias temporales. El proceso comienza con la extracción de secuencias de estados energéticos previamente identificados, donde cada muestra temporal del dataset HRSS se asigna a uno de los clusters energéticos mediante el cálculo de BMU correspondiente.

La arquitectura neuronal implementada emplea un diseño jerárquico de tres componentes para capturar dependencias temporales en secuencias de estados energéticos. Comienza con una capa LSTM bidireccional de 128 unidades que procesa secuencias en ambas direcciones temporales, seguida de una capa LSTM unidireccional de 64 unidades que actúa como mecanismo de compresión de características temporales. El componente principal de innovación es un mecanismo de atención multi-cabeza con 4 cabezas que identifica automáticamente los timesteps más relevantes para la predicción, cuya salida se procesa mediante capas densas de 64 y 32 unidades antes de la neurona de salida con activación sigmoide. Para el entrenamiento, se utilizan 25898 secuencias generadas por ventanas deslizantes de 10 timesteps del dataset original, divididas estratificadamente en 60 % entrenamiento, 20 % validación y 20 % testing, manteniendo la distribución equilibrada de clases (55,4 % eficientes, 44,6 % ineficientes). El proceso de entrenamiento emplea optimizador Adam con learning rate inicial de 0,001, función de pérdida binary cross entropy, y se extiende hasta 150 épocas con early stopping y reducción adaptativa del learning rate para garantizar convergencia estable y prevenir overfitting.

4. Resultados y Evaluación

La aplicación del algoritmo Watershed sobre la matriz-U resultó en la identificación automática de 52 regiones teóricas, de las cuales 48 presentaron asignaciones de muestras suficientes para el análisis estadístico posterior, constituyendo los clusters efectivos del modelo. Esta segmentación automática establece una taxonomía operativa del sistema HRSS que captura la variabilidad energética de los seis componentes principales. La estructura topológica aprendida se visualiza mediante tres representaciones complementarias (Figura 5). La Matriz-U bidimensional revela fronteras de alta disimilitud (amarillo) que separan configuraciones operativas diferenciadas de los componentes BLO, BHL, BHR, BRU, HR y HL, mientras que las regiones de baja disimilitud (azul) agrupan estados con patrones de consumo similares. Adicionalmente, la representación tridimensional identifica valles topológicos que corresponden a configuraciones estables del sistema, proporcionando una perspectiva espacial de la distribución energética. Por su parte, el mapa de eficiencia energética clasifica cada cluster según su índice de desempeño respecto al baseline energético establecido, donde las regiones verdes representan configuraciones que operan cerca del baseline óptimo y las rojas identifican estados con gaps energéticos elevados. El SOM ha aprendido que el sistema HRSS presenta configuraciones ope-

rativas heterogéneas con diferentes niveles de eficiencia energética. En particular, la predominancia de clusters en rangos de eficiencia media-baja confirma los gaps energéticos identificados previamente, especialmente en componentes BHL, HR y HL. Por tanto, la estructura topológica obtenida establece la base para la identificación automática de transiciones hacia estados energéticamente eficientes, permitiendo que el modelo discrimine entre configuraciones operativas que mantienen la funcionalidad del sistema con diferentes demandas energéticas.

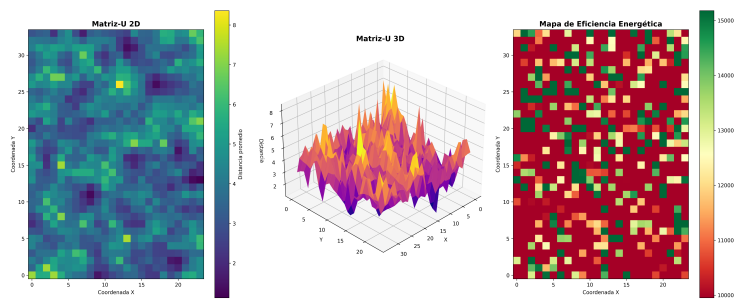


Figura 5: Matriz-U y mapa de eficiencia energética del SOM entrenado.

Una vez establecida la taxonomía energética de estados median- te el SOM, el análisis se enfoca en caracterizar las dinámicas tempora- les de transición entre estas configuraciones operativas identificadas. La construcción del Grafo de Transiciones entre estados (Figura 6) se reali- zó mediante el análisis secuencial de las asignaciones BMU de las 25907 muestras temporales a los clusters identificados por el SOM, generando cadenas de estados discretos que permitieron cuantificar las dinámicas de cambio operativo del sistema HRSS. El grafo resultante mostró una red compuesta por 48 nodos y 929 aristas, con una densidad de cone- xión de 0,41. La estructura del grafo presenta una alta conectividad en la región central, indicando estados operativos hub desde los cuales el sistema puede transitar hacia múltiples configuraciones alternativas. Se observó una mayor conectividad entre estados con eficiencia energética similar, mientras que las transiciones entre categorías opuestas fueron menos frecuentes. La matriz de probabilidades de transición (Figura 6) muestra una estructura diagonal dominante con probabilidades de auto-transición superiores a 0,8 en la mayoría de los estados, indicando una fuerte inercia operativa donde el sistema tiende a permanecer en su configuración actual. Las probabilidades de transición entre estados distintos se concentran mayoritariamente por debajo de 0,2, confirman- do que los cambios de configuración constituyen eventos relativamente infrecuentes en la operación normal del sistema HRSS.

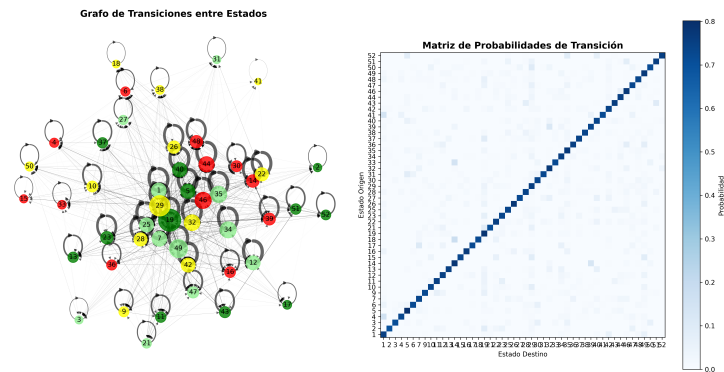


Figura 6: Grafo de Transiciones entre estados y matriz de probabilidades de transición

Los estados centrales dentro del grafo de transiciones representan configuraciones desde las cuales el sistema accede a múltiples trayectorias operativas, convirtiéndolos en puntos estratégicos para intervenciones orientadas a la mejora de eficiencia. La identificación de estos nodos críticos delimita regiones del sistema con mayor capacidad de influencia sobre el comportamiento energético global.

El análisis de ciclos operativos mencionado en la metodología previamente completa la caracterización del comportamiento temporal del sistema HRSS, presentando patrones de duración y variabilidad en las secuencias operativas identificadas mediante el grafo de transiciones. El coeficiente de variación del 71,7% caracteriza una alta inconsisten-

Tabla 3: Características de ciclos operativos del sistema HRSS

Métrica	Valor
Ciclos operativos completos detectados	183
Duración promedio (muestras)	141,1
Desviación estándar (muestras)	101,2
Coefficiente de variación	71,7 %
Ciclo más frecuente (muestras)	240
Duración mínima observada (muestras)	25
Duración máxima observada (muestras)	>325

cia en los tiempos de ejecución de la secuencia estándar de transporte, donde los ciclos prolongados indican permanencias excesivas en estados intermedios con consumo energético activo pero sin progreso efectivo de carga. Los ciclos excepcionalmente cortos (25 muestras) corresponden a transferencias directas que minimizan el tiempo en configuraciones de espera, mientras que los ciclos largos (>325 muestras) reflejan de-

tenciones prolongadas en componentes específicos como BHL, HR y HL, que coinciden con los componentes identificados previamente como generadores de los mayores gaps energéticos. Esta variabilidad temporal proporciona una base cuantitativa para priorizar intervenciones en las secuencias que exceden el umbral de 240 muestras, donde el tiempo adicional se traduce directamente en consumo energético improductivo.

4.1 Dinámica operativa y transiciones

Una vez establecida la taxonomía energética de estados median- te el SOM, el análisis se enfoca en caracterizar las dinámicas temporales de transición entre estas configuraciones operativas identificadas.

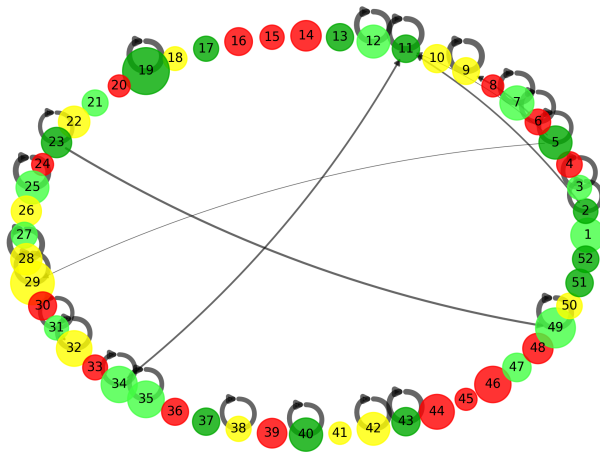


Figura 7: Grafo de transiciones con probabilidades superiores a 0,1

El grafo de transiciones con probabilidades superiores a 0,1 (Figura 7) muestra que los nodos, codificados por colores según su clasificación energética (verde oscuro para OPTIMAL, verde claro para EFFICIENT, amarillo para NORMAL, naranja para SUBOPTIMAL y rojo para INEFICIENT), no forman clústeres cohesivos por tipo de eficiencia, sino que se distribuyen de manera heterogénea en la topología. Esta disposición indica que el sistema HRSS carece de rutas preferenciales directas hacia estados de alta eficiencia, requiriendo transitar frecuentemente por estados de menor eficiencia para alcanzar configuraciones óptimas.

4.2 Evaluación del predictor LSTM de eficiencia energética

El análisis de convergencia del entrenamiento muestra estabilización del loss desde valores iniciales de 0,14 hasta 0,0064 en validación

tras 65 épocas. Esta convergencia indica que el modelo ha aprendido a distinguir entre secuencias temporales de los 6 componentes del HRSS (BLO, BHL, BHR, BRU, HR, HL) que resultan en configuraciones energéticamente eficientes versus ineficientes, basándose en los patrones de transición identificados en la fase SOM. El accuracy alcanza 99,81% en validación, significando que el modelo clasifica correctamente 9981 de cada 10000 secuencias temporales de 10 timesteps del sistema HRSS. Esta precisión en la clasificación permite identificar cuándo las transiciones entre los 48 clusters efectivos del mapa SOM conducen hacia estados de mayor o menor eficiencia energética. La métrica AUC-ROC de 0,999 indica que el modelo discrimina adecuadamente entre secuencias eficientes e ineficientes en diferentes umbrales de decisión. Para el sistema HRSS, esto significa capacidad de detectar transiciones problemáticas antes de que se materialicen en consumos energéticos elevados en los componentes del sistema de almacenamiento.

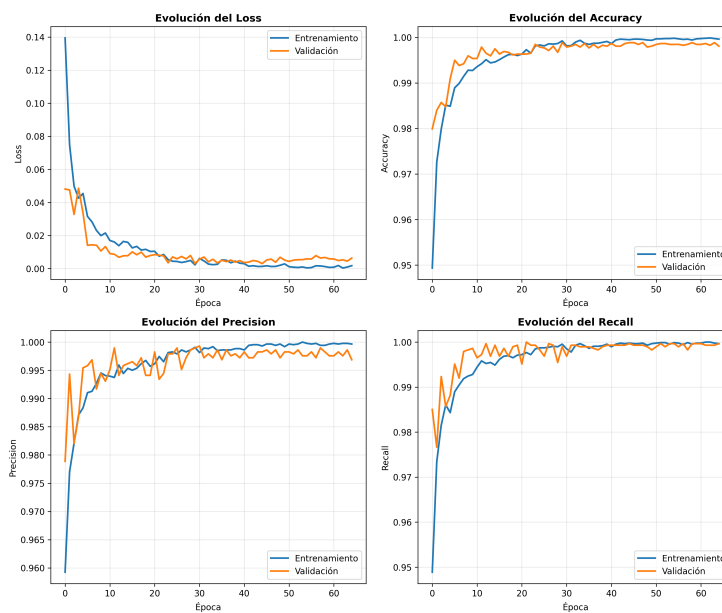


Figura 8: Evolución de las métricas de rendimiento del modelo LSTM

La precisión de 99,69% y recall de 99,97% en validación indican que el modelo identifica 9969 de cada 10000 secuencias ineficientes reales (recall) y que 9969 de cada 10000 predicciones de ineficiencia son correctas (precisión). En términos del HRSS, esto significa detección confiable de secuencias temporales que conducen a estados de alto consumo energético en las cintas transportadoras y raíles de elevación, minimizando falsas alarmas y secuencias ineficientes no detectadas.

Métrica	Entrenamiento	Validación	Test
Loss	0,0064	0,0064	0,0064
Accuracy	99,81 %	99,81 %	99,88 %
Precision	99,69 %	99,69 %	99,83 %
Recall	99,97 %	99,97 %	99,91 %
AUC-ROC	0,999	0,999	0,998

Tabla 4: Métricas finales del predictor LSTM Bi-direccional con atención

La matriz de confusión del conjunto de test muestra 2306 secuencias ineficientes correctamente identificadas y 4 falsos positivos, junto con 2868 secuencias eficientes correctamente clasificadas y 2 falsos negativos. Esto representa que de 5180 secuencias temporales de 10 timesteps del sistema HRSS evaluadas, el modelo clasifica incorrectamente solo 6 secuencias (0,12%).

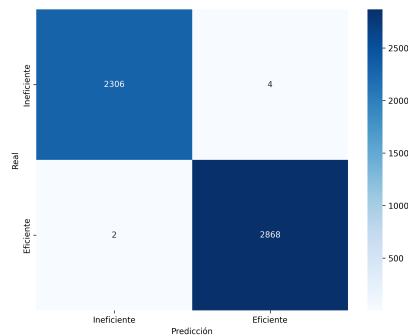


Figura 9: Matriz de confusión del conjunto de test

Los resultados del sistema de alertas muestran estimaciones de desperdicio entre 5,0,kW y 18,0,kW por secuencia ineficiente detectada. Estas estimaciones se calculan aplicando factores de ineficiencia preestablecidos basados en el consumo diferencial entre estados eficientes e ineficientes del sistema HRSS. Las alertas nivel CRÍTICO indican secuencias donde el consumo energético proyectado supera significativamente el baseline establecido para las configuraciones óptimas de cintas transportadoras y raíles de elevación.

El sistema de alertas automáticas genera notificaciones para las secuencias ineficientes detectadas, calculando estimaciones de desperdicio energético mediante la aplicación de factores de ineficiencia preestablecidos por categoría de estado.

5. Conclusiones

Este trabajo presnta la viabilidad de aplicar técnicas de inteligencia artificial para la identificación automática de ineficiencias energéticas en sistemas intralogísticos automatizados. La metodología desarrollada establece un marco replicable que combina clustering no supervisado con predicción temporal para análisis energético industrial.

5.1 Contribuciones principales

La integración de mapas auto-organizados con redes LSTM bidireccionales proporciona una aproximación dual que complementa la caracterización de estados energéticos con capacidades predictivas. Esta combinación permite tanto la comprensión del comportamiento del sistema como la anticipación de configuraciones ineficientes, superando las limitaciones de enfoques puramente reactivos empleados en sistemas de control convencionales.

La automatización del proceso de identificación de ineficiencias limita la dependencia de conocimiento experto específico del sistema, facilitando la transferencia metodológica a otros entornos industriales. El enfoque no supervisado del SOM permite descubrir patrones energéticos, mientras que la arquitectura LSTM presenta capacidad suficiente para discriminar secuencias operativas problemáticas.

La cuantificación de gaps energéticos por componente proporciona una base objetiva para la priorización de intervenciones de optimización.

5.2 Implicaciones teóricas

Los hallazgos confirman que la hipótesis de que se puede optimizar todo el proceso de manera uniforme no se sostiene empíricamente. La variabilidad energética observada sugiere que los factores de implementación, configuración y degradación tienen un impacto más significativo que las características nominales de diseño.

La efectividad de las técnicas de aprendizaje automático para la intralogística valida la aplicabilidad de estas metodologías desarrolladas para otros campos y para la optimización energética industrial.

5.3 Limitaciones y restricciones

La validación experimental se limita a un único sistema operando bajo condiciones controladas, lo que restringe la validez externa de los resultados. La escalabilidad de la metodología a sistemas con configuraciones diferentes o patrones operativos más complejos requiere verificación empírica adicional.

La evaluación se centra exclusivamente en la detección de ineficiencias sin abordar los mecanismos de implementación de mejoras

identificadas. Esta limitación separa el diagnóstico de la intervención, reduciendo la utilidad práctica inmediata de los resultados obtenidos.

5.4 Trabajo futuro

La validación en múltiples sistemas de almacenamiento automatizado constituye la extensión más crítica para establecer la robustez metodológica. La implementación de estudios longitudinales en entornos operativos permitiría evaluar tanto la estabilidad temporal de los patrones identificados, como la efectividad real de las intervenciones propuestas.

La integración con sistemas de autoconsumo industrial presenta oportunidades para la optimización energética que considere tanto la eficiencia de consumo como la gestión de generación distribuida. Esta dirección requiere el desarrollo de modelos que incorporen variabilidad renovable y estrategias de gestión de demanda.

El desarrollo de interfaces de control en tiempo real representa un paso necesario para transformar capacidades de diagnóstico en sistemas de optimización operativa. Esta línea de investigación debe abordar aspectos de fiabilidad, seguridad operativa y degradación elegante bajo condiciones de fallo.

Referencias

1. Wang, Z., Xu, Z., Gao, Z., Zhang, K., Liu, L. (2024). Prediction and Online Control for Process Parameters of Vanadium Nitrogen Alloys Production Based on Digital Twin. *SUSTAINABILITY*, 16(17). DOI: 10.3390/su16177545.
2. Lai, C., Wang, Y., Fan, K., Cai, Q., Ye, Q., Pang, H., Wu, X. (2022). An improved forecasting model of short-term electric load of papermaking enterprises for production line optimization. *ENERGY*, 245. DOI: 10.1016/j.energy.2022.123225.
3. Wang, X., Dai, Y., Jia, Z., Wang, G. (2024). Transient analysis of production performance and energy consumption in geometric flexible production systems. *EXPERT SYSTEMS WITH APPLICATIONS*, 245. DOI: 10.1016/j.eswa.2023.123065.
4. Loffredo, A., May, M. C., Schaefer, L., Matta, A., Lanza, G. (2023). Reinforcement learning for energy-efficient control of parallel and identical machines. *CIRP JOURNAL OF MANUFACTURING SCIENCE AND TECHNOLOGY*, 44, 91-103. DOI: 10.1016/j.cirpj.2023.05.007.
5. Zhou, B., Li, X., Liu, W. (2021). Hybrid multi-objective opposite-learning evolutionary algorithm for integrated production and maintenance scheduling with energy consideration. *NEURAL COMPUTING & APPLICATIONS*, 33(5), 1587-1605. DOI: 10.1007/s00521-020-05075-3.
6. Zhang, X., Liu, P., Zhu, H. (2022). The Impact of Industrial Intelligence on Energy Intensity: Evidence from China. *SUSTAINABILITY*, 14(12). DOI: 10.3390/su14127219.
7. Ma, N. (2024). Artificial Intelligence-Driven Decision Support Systems for Sustainable Energy Management in Smart Cities. *INTERNATIONAL JOURNAL OF ADVANCED COMPUTER SCIENCE AND APPLICATIONS*, 15(9), 523-529.
8. Mendia, I., Gil-Lopez, S., Grau, I., Del Ser, J. (2024). A novel approach for the detection of anomalous energy consumption patterns in industrial cyber-physical systems. *EXPERT SYSTEMS*, 41(2). DOI: 10.1111/exsy.12959.
9. de Santana, D. M., Lourenco, S. R., Cassiano, D. A. (2023). Data mining approach for energy efficiency improvements in a utilities supply on a petrochemical plant. *EVOLVING SYSTEMS*, 14(6), 1071-1081. DOI: 10.1007/s12530-023-09515-y.
10. Mohammadpour Velni, J. (2022). GOALI/Collaborative Research: Control-Oriented Modeling and Predictive Control of High Efficiency Low-emission Natural Gas Engines. (Sin revista, solo título y año).
11. Kumar, A. R., Nagabhooshanam, N., Kumar, M. N., Kumar, D. S., Babu, G. R., Chakravarthy, A. K., Praveen, N., Ganeshan, P., Kannan, S. (2023). Energy optimization in additive manufacturing based on cluster-aware enhanced spider monkey optimization. *INTERNATIONAL JOURNAL OF ADVANCED MANUFACTURING TECHNOLOGY*. DOI: 10.1007/s00170-023-11846-8.
12. Gao, S. X., Liu, H., Ota, J. (2024). Energy-efficient buffer and service rate allocation in manufacturing systems using hybrid machine learning and evolutionary algorithms. *ADVANCES IN MANUFACTURING*, 12(2), 227-251. DOI: 10.1007/s40436-023-00461-1.

13. da Ponte, R. M. (2023). *Optimization of the New Alarm System Implemented on PowerStudio SCADA*. (Libro/Tesis).
14. Leyne, T. (2021). *Uptime Energy Reduction Methodology in Industrial Settings*. (Libro/Tesis).
15. Oliveira, R. J. E. (2021). *Energy Monitoring of Intralogistics Systems*. (Libro/Tesis).
16. Karakaya, E. (2020). *Container Handling and Layout Optimization in Empty Container Depots*. (Libro/Tesis).
17. Aydoğan, C. (2021). *Energy Efficient Scheduling in Flow-Shop and Parallel Machine Robotic Cells*. (Libro/Tesis).
18. Wang, X., Cao, J., Cao, Y., Zou, F. (2025). Energy-efficient trajectory planning for a class of industrial robots using parallel deep reinforcement learning. *Nonlinear Dynamics*, 113(8), 8491 – 8511. DOI: 10.1007/s11071-024-10510-4.
19. Cirera, J., Carino, J. A., Zurita, D., Ortega, J. A. (2020). A data-driven-based industrial refrigeration optimization method considering demand forecasting. *Processes*, 8(5). DOI: 10.3390/PR8050617.
20. Ning, M. A. (2024). Artificial Intelligence-Driven Decision Support Systems for Sustainable Energy Management in Smart Cities. *International Journal of Advanced Computer Science and Applications*, 15(9), 523 – 529. DOI: 10.14569/IJACSA.2024.0150953.
21. Bonci, A., Fredianelli, L., Kermenov, R., Longarini, L., Longhi, S., Pompei, G., Prist, M., Verdini, C. (2024). DeepESN Neural Networks for Industrial Predictive Maintenance through Anomaly Detection from Production Energy Data. *Applied Sciences (Switzerland)*, 14(19). DOI: 10.3390/app14198686.
22. Guo, H., Wu, R., Ma, Y., Li, Y., Li, S., Qi, B., Sun, Y. (2024). Learning-Based Terminal-Edge Collaborative Energy-Efficient Routing Algorithm for Green RWSN. *IEEE Transactions on Industrial Informatics*, 20(6), 9037 – 9047. DOI: 10.1109/TII.2024.3379641.
23. Tieng, H., Lai, C. Y., Fan, S. X., Wu, T. Q. (2025). Explainable AI-based facility control system for energy saving and carbon reduction. *Journal of Mechanical Science and Technology*, 39(4), 2301 – 2310. DOI: 10.1007/s12206-025-0348-5.
24. Mou, J., Gao, K., Duan, P., Li, J., Garg, A., Sharma, R. (2023). A Machine Learning Approach for Energy-Efficient Intelligent Transportation Scheduling Problem in a Real-World Dynamic Circumstances. *IEEE Transactions on Intelligent Transportation Systems*, 24(12), 15527 – 15539. DOI: 10.1109/TITS.2022.3183215.
25. Akyaz, T., Engin, D. (2024). Machine Learning-Based Predictive Maintenance System for Artificial Yarn Machines. *IEEE Access*, 12, 125446 – 125461. DOI: 10.1109/ACCESS.2024.3454548.
26. Lee, D., Huang, H. Y., Lee, W. S., Liu, Y. (2020). Artificial intelligence implementation framework development for building energy saving. *International Journal of Energy Research*, 44(14), 11908 – 11929. DOI: 10.1002/er.5839.
27. Buratti, Y., Sowmya, A., Dumbrell, R., Dwivedi, P., Trupke, T., Hameiri, Z. (2022). Automated efficiency loss analysis by luminescence image reconstruction using generative adversarial networks. *Joule*, 6(6), 1320 – 1332. DOI: 10.1016/j.joule.2022.05.001.

28. Emam, W., Waqas, H. M., Mahmood, T., Rehman, U., Pamucar, D. (2025). AI-driven energy management system based on hesitant bipolar complex fuzzy Hamacher power aggregation operators and their applications in MADM. *Scientific Reports*, 15(1). DOI: 10.1038/s41598-025-94340-3.
29. Abdelsameia, T., Abd El-Aal, M. F. (2025). Optimizing Energy Consumption Efficiency in Global Industrial Systems Using the Random Forest Algorithm. *International Journal of Energy Economics and Policy*, 15(3), 239 – 244. DOI: 10.32479/ijeep.18796.
30. Li, J., Dai, J., Issakhov, A., Almojil, S. F., Souri, A. (2021). Towards decision support systems for energy management in the smart industry and Internet of Things. *Computers and Industrial Engineering*, 161. DOI: 10.1016/j.cie.2021.107671.
31. Huang, T. W., Lin, D. L., Lin, C. X., Lin, Y. (2022). Taskflow: A Light-weight Parallel and Heterogeneous Task Graph Computing System. *IEEE Transactions on Parallel and Distributed Systems*, 33(6), 1303 – 1320. DOI: 10.1109/TPDS.2021.3104255.
32. Ayadi, R., Abd El-Aziz, R. M., Taloba, A. I., Aljuaid, H., Hamed, N. O., Khder, M. A. (2022). Deep Learning-Based Soft Sensors for Improving the Flexibility for Automation of Industry. *Wireless Communications and Mobile Computing*, 2022. DOI: 10.1155/2022/5450473.
33. Lahariya, M., Karami, F., Develder, C., Crevecoeur, G. (2023). Physics-Informed LSTM Network for Flexibility Identification in Evaporative Cooling System. *IEEE Transactions on Industrial Informatics*, 19(2), 1484 – 1494. DOI: 10.1109/TII.2022.3173897.
34. Pathak, C. (2024). SAVING ENERGY THROUGH SYNERGISTIC INNOVATION WITH DIGITALIZATION. *IPPTA: Quarterly Journal of Indian Pulp and Paper Technical Association*, 36(1), 145 – 148.
35. Xu, Z., Han, G., Liu, L., Martinez-Garcia, M., Wang, Z. (2021). Multi-energy scheduling of an industrial integrated energy system by reinforcement learning-based differential evolution. *IEEE Transactions on Green Communications and Networking*, 5(3), 1077 – 1090. DOI: 10.1109/TGCN.2021.3061789.
36. Von Birgelen, A., Niggemann, O. (2017, June). Using self-organizing maps to learn hybrid timed automata in absence of discrete events. En *IEEE International Conference on Emerging Technologies and Factory Automation, ETFA* (pp. 1-8). IEEE. DOI: 10.1109/ETFA.2017.8247695.
37. Wang, W., Yang, H., Zhang, Y., Xu, J. (2018). IoT-enabled real-time energy efficiency optimisation method for energy-intensive manufacturing enterprises. *International Journal of Computer Integrated Manufacturing*, 31(4-5), 362-379. DOI: 10.1080/0951192X.2017.1337929.
38. Rehman, U., Faria, P., Gomes, L., Vale, Z. (2025). Future of Energy Management Models in Smart Homes: A Systematic Literature Review of Research Trends, Gaps, and Future Directions. *Process Integration and Optimization for Sustainability*, 1-30. DOI: 10.1007/S41660-025-00506-X.
39. Cao, Z., Han, X., Lyons, W., O'Rourke, F. (2021). Energy management optimisation using a combined Long Short-Term Memory recurrent neural network – Particle Swarm Optimisation model. *Journal of Cleaner Production*, 326. DOI: 10.1016/J.JCLEPRO.2021.129246.
40. Liu, J., Qian, Y., Yang, Y., Yang, Z. (2022). Can Artificial Intelligence Improve the Energy Efficiency of Manufacturing Companies? Evidence

- from China. *International Journal of Environmental Research and Public Health*, 19(4), 2091. DOI: 10.3390/IJERPH19042091.
41. Ghobakhloo, M., Fathi, M. (2021). Industry 4.0 and opportunities for energy sustainability. *Journal of Cleaner Production*, 295. DOI: 10.1016/J.JCLEPRO.2021.126427.
 42. Alex, B., Johnson, M. (2025, February). A Framework for IoT-Enabled Smart Manufacturing for Energy and Resource Optimization. (Artículo en arXiv).
 43. Xu, Y., Gao, W., Qian, F., Li, Y. (2021). Potential Analysis of the Attention-Based LSTM Model in Ultra-Short-Term Forecasting of Building HVAC Energy Consumption. *Frontiers in Energy Research*, 9, 730640. DOI: 10.3389/FENRG.2021.730640/BIBTEX.
 44. von Birgelen, A., Niggemann, O. (2018). Enable learning of Hybrid Timed Automata in Absence of Discrete Events through Self-Organizing Maps. En: *Service-Oriented Cyber-Physical Production Systems* (pp. 37-54). Springer Vieweg, Berlin, Heidelberg. DOI: 10.1007/978-3-662-57805-6_3.
 45. Yepes-Nuez, J. (2021). The Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) statement, published in 2009, was designed to help systematic reviewers transparently report why the review was done, what the authors did, and what they found. (Incomplete/Resumen de PRISMA 2020).

Visión e Inteligencia Artificial: Comparativa de Modelos en Detección de Violencia

Marcelo Hernández Lemos, Vidal Moreno Rodilla, and Belén Curto
Diego

Universidad de Salamanca
Departamento de Informática y Automática
Facultad de Ciencias.
Plaza de los Caídos s/n
37008 Salamanca, España
`{marceloh,vmoreno,bcurto}@usal.es`

Resumen La detección automática de violencia en video requiere un equilibrio entre rendimiento, eficiencia y robustez. Este estudio compara cuatro arquitecturas de aprendizaje profundo (I3D, SlowFast, TSM y ViViT) que representan distintas filosofías en el reconocimiento de acciones. Tras un ajuste fino en los datasets RWF-2000 y RLVS, y una evaluación de generalización en Hockey Fights, los resultados revelan un claro compromiso.

ViViT alcanza el mejor rendimiento en dominios conocidos (F1-Score de hasta 0.992), pero a costa de un alto coste computacional y una robustez limitada. Por otro lado, TSM surge como un modelo que tiene un rendimiento competitivo y una eficiencia radicalmente superior. SlowFast destaca como la arquitectura más robusta, generalizando eficazmente a dominios dispares. Se concluye que la elección del modelo no es universal, y dependiendo de los requisitos específicos de cada aplicación, se deberá elegir una configuración que priorice rendimiento, costo computacional o robustez según el caso de uso.

Palabras Clave: Detección de Violencia, Visión Artificial, Aprendizaje Profundo, Redes Neuronales Convolucionales, Transformers, ViViT, TSM, SlowFast, I3D, Reconocimiento de Acciones.

1. Introducción

La detección de violencia en video es un reto clave en seguridad y vigilancia. El crecimiento exponencial del contenido audiovisual presenta nuevos desafíos. Por ejemplo, en 2022 se estimaba que existían casi 10 mil millones de videos en YouTube, con un promedio de 7

millones de subidas diarias [1]. Esta enorme cantidad hace inviable la supervisión manual de estos videos. Lo mismo ocurre en videovigilancia; el aumento de cámaras de seguridad hace muy difícil que los operadores humanos puedan detectar violencia durante largas jornadas de trabajo. En [2] se demuestra una alta ineficiencia de los operadores al trabajar con múltiples pantallas: con 4 pantallas, el 20 % de las anomalías no se detectaron, y con 9 pantallas, este porcentaje se disparó al 60 %. Esto demuestra que es imprescindible desarrollar sistemas automáticos e inteligentes.

El problema presenta un doble desafío: el reto técnico de interpretar complejas dinámicas espacio-temporales, y el otro desafío es el conceptual, de definir un término tan subjetivo como ‘violencia’. Para establecer un marco de estudio riguroso, este trabajo acota la tarea a las peleas físicas, abordándola como un problema de clasificación binaria entre ‘violencia’ y ‘no violencia’.

Este estudio aborda la problemática mediante un estudio comparativo de cuatro arquitecturas de vanguardia, seleccionadas por representar diferentes filosofías en el reconocimiento de acciones en video. Se parte desde una base sólida dentro de las Redes Neuronales Convolucionales 3D (CNNs 3D) con el modelo *Inflated 3D ConvNet* (I3D) [3], se explora el enfoque de vías lentas y rápidas, inspiradas en la biología, propuesto en *SlowFast Networks* [4], se analiza un algoritmo eficiente como *Temporal Shift Module* (TSM) [5], y se contrastan dichas arquitecturas basadas en CNNs 3D con la arquitectura emergente de los Transformers, representada por el modelo *Video Vision Transformer* (ViViT) [6].

Para llevar a cabo este TFM, se diseñará un riguroso protocolo experimental con la finalidad de analizar no solo el rendimiento de los modelos de aprendizaje profundo resultantes, sino también identificar el impacto que tiene el propio conjunto de datos de entrenamiento en los resultados y en la capacidad de generalización de los modelos.

Este documento se estructura de la siguiente manera: el apartado 13 revisa el estado del arte en la detección de violencia, describiendo los conjuntos de datos y las arquitecturas fundamentales. El apartado 13 detalla la metodología, el marco experimental y métricas utilizadas. El apartado 13 presenta y analiza los resultados cuantitativos obtenidos en ambas fases experimentales. Se realiza la discusión global de los hallazgos en el apartado 13, y por último, el apartado 13 expone las conclusiones finales del trabajo y líneas de trabajo a futuro.

2. Estado del Arte

2.1 Conceptualización de la Detección de Violencia

Un desafío de la detección de violencia se encuentra en la complejidad y subjetividad que contiene el concepto de violencia. La tecnología actual depende de grandes volúmenes de datos y se encuentra



Figura 1: Ejemplos de fotogramas de los datasets RLVS, RWF-2000 y Hockey Fights

La fila superior muestra ejemplos de una etiqueta violenta, mientras que la inferior muestra escenarios no violentos. A la izquierda, se observa un ejemplo de RLVS, en el medio de RWF-2000 y a la derecha de Hockey Fights.

propensa a sesgos indeseados, como puede ser la definición misma del concepto de violencia, como se muestra en revisiones sistemáticas relacionadas [7], [8].

Este desafío genera una brecha entre la demanda de soluciones y las que actualmente hay. Según Avellaneda, et al. [9] y algunos de los artículos que fueron analizados en su estudio [10], [11], los algoritmos de aprendizaje profundo entrenados con conjuntos de datos de violencia no tienen el poder de generalización suficiente para poder detectar eventos violentos en el mundo real; existe una escasez de material sobre violencia en diferentes escenarios. Esto conlleva a buscar nuevas soluciones que permitan una mejor generalización de los modelos, así como también a generar conjuntos de datos más amplios con distintos escenarios y variados tipos de violencia.

2.1.1 Conjuntos de Datos de Violencia. Los conjuntos de datos a describir se basan en distintas modalidades de *aprendizaje débilmente supervisado*, donde a distintos niveles de granularidad se determinan las clases que aparecen en el video, como el etiquetado a nivel de clip/video, o etiquetados con anotaciones temporales para saber cuándo ocurre la acción.

Se detallan algunos de los conjuntos con contenido violento y no violento más utilizados por la literatura hasta la fecha.

1. Hockey Fights Dataset [12]: Contiene 1.000 videoclips de corta duración (500 peleas, 500 no-peleas) de la Liga Nacional de Hockey (NHL). Es un entorno controlado y específico, lo que lo hace menos desafiante para modelos modernos y limita su aplicabilidad a la violencia general [10]. Este conjunto es de los más utilizados respecto a agresiones físicas según la revisión de literatura realizada por

Negre et al [7]. En la figura 1, a la derecha se observan fotogramas de ejemplo para ambas clases.

2. Datasets de Violencia en Películas: Estos conjuntos se basan en películas de Hollywood. El primer conjunto destacable fue presentado en el artículo de Hockey Fights [12], llamado ‘*Movie Fights*’. Este contiene 200 clips de acciones violentas y no violentas. El ‘*Violent Scenes Dataset (VSD)*’ [13] y su versión ‘*VSD2014*’ [14], ampliaron la idea con un mayor volumen y diversidad de acciones (nuevas clases además de ‘violencia’ o ‘pelea’). La principal limitación de estos conjuntos se basa en la brecha de dominio con la violencia en el mundo real (por ejemplo respecto a videos de cámaras de seguridad), además de su gran volumen de datos, que exige potentes recursos computacionales.
3. RWF-2000 (Real-World Fighting) [15]: Incluye 2.000 videoclips de corta duración equilibrados (1.000 violentos/1.000 no-violentos) de cámaras de vigilancia del mundo real (obtenidos de YouTube). Presenta desafíos como baja calidad de imagen, oclusiones, multitudes y movimientos de cámara, siendo muy utilizado para evaluar el rendimiento en escenarios realistas. En la parte central de la figura 1, se puede observar un fotograma del video de una pelea (arriba), y debajo un fotograma de un video no violento.
4. RLVS (Real-Life Violence Situations Dataset): También incluye 2.000 videoclips de corta duración de YouTube, pero con una variedad más amplia de escenarios y puntos de vista. La distinción de clases puede ser muy marcada, lo que a veces genera muy buenos rendimientos de clasificación, pero malas generalizaciones en otros dominios. A la izquierda de la figura 1, se puede observar un ejemplo de cada etiqueta.

Además de los conjuntos vistos, existen otros conjuntos como UCF-Crime Dataset [16], que extiende el número de clases y contiene videos de larga duración, XD-Violence [17], que aborda el desafío de violencia en formato multimodal (audio y video), o el Atomic Visual Actions (AVA) [18], que realiza anotaciones temporales con mayor densidad, buscando una mayor comprensión de las acciones humanas, incluyendo aquellas que pueden ser violentas. A pesar de la existencia de múltiples conjuntos de datos, existen limitaciones persistentes en todos ellos [10]. Los datasets a menudo carecen de realismo en comparación con los escenarios de vigilancia del mundo real (mala iluminación, baja resolución), presentan sesgos demográficos y de contexto, y no abarcan una diversidad suficiente de situaciones, especialmente las más sutiles. A estos problemas de datos se suma la dificultad de la anotación, que es inherentemente subjetiva, y el elevado coste computacional para entrenar algunos modelos, especialmente los Transformers, que demandan grandes cantidades de datos [19]. Estas limitaciones impactan la generalización y la fiabilidad de los modelos de detección de violencia.

2.2 Arquitecturas para Clasificación de Acciones en Video

Las soluciones para el reconocimiento de acciones en video han evolucionado desde métodos basados en características construidas manualmente, hacia modelos de aprendizaje profundo. A continuación, se describen distintas arquitecturas que actualmente obtienen resultados de vanguardia y se utilizan hasta la fecha.

2.2.1 CNNs 3D. Estas redes utilizan filtros tridimensionales ($D \times k \times k$) para capturar las características espaciales y temporales del video de forma conjunta, como se propuso en los primeros estudios sobre este tipo de arquitectura [20], [21].

Para evitar pasar por procesos tan largos de entrenamiento que conlleven este tipo de arquitecturas, surgió la arquitectura Inflated 3D ConvNet (I3D) [3]. Se propuso ‘inflar’ los filtros de una CNN 2D, como pueden ser los modelos de ResNet [22] o Inception [23], modelos que obtienen resultados de vanguardia en clasificación de imágenes (como ImageNet [24]), para convertirlos en filtros 3D. Este método de transferencia de conocimiento fue fundamental para entrenar redes 3D profundas y robustas de manera eficiente, evitando el costoso entrenamiento desde cero.

La transferencia de conocimiento, combinada con el entrenamiento en el masivo dataset de acciones Kinetics [25], estableció a I3D como una de las arquitecturas de referencia para el reconocimiento de acciones en video. En particular, estudios como el realizado por David Freire-Obregón et al. (2021) [26], utilizan I3D para clasificar videos de violencia, obteniendo resultados de vanguardia, dejando claro el impacto de las CNNs 3D y, en particular, de las I3D en lo que respecta a la clasificación de videos.

A pesar de su buen rendimiento, I3D y otros tipos de CNNs 3D profundas tienen limitaciones por el alto costo computacional y consumo de memoria tanto en entrenamiento como en despliegue.

2.2.2 Arquitecturas Eficientes y Enfoques Alternativos. Se describirán dos arquitecturas que permiten mejorar la eficiencia y obtener iguales o mejores resultados que las CNNs 3D tradicionales.

Redes SlowFast. Inspiradas en la percepción visual humana, la arquitectura SlowFast [4] procesa el video a través de dos caminos paralelos. Contiene un camino lento (slow pathway), que captura la semántica espacial y contextual para obtener información espacial y semántica detallada, operando a una baja velocidad de fotogramas, mientras que el camino rápido (fast pathway) trabaja a una alta resolución temporal, capturando los movimientos rápidos con una menor resolución espacial y capacidad de canales. Las conexiones laterales entre ambas vías fusionan su información, obteniendo un reconocimiento más robusto.

Este método requiere menos operaciones en punto flotante (FLOPs) por clip que modelos basados en CNNs 3D, como C3D o I3D, demostrando ser más eficiente, sumado a un igual o mejor rendimiento en el reconocimiento de acciones.

Temporal Shift Module (TSM). Esta técnica [5] permite a una CNN 2D estándar modelar relaciones temporales con una carga computacional prácticamente nula (cero FLOPs y cero parámetros adicionales). Esto se logra mediante el desplazamiento de canales. Antes de aplicar las convoluciones 2D de cada capa, una fracción de los canales del tensor de activaciones se desplaza en la dimensión temporal. De esta forma, cuando el filtro 2D opera sobre un fotograma, implícitamente tiene acceso a la información de los fotogramas adyacentes.

Los autores también mencionan que su rendimiento de clasificación en benchmarks de reconocimiento de acciones iguala y supera en algunos casos a otras arquitecturas como I3D; se muestran exactitudes superiores en un dataset de reconocimiento de acciones, sin consumir tantos recursos computacionales en términos del número de parámetros, FLOPs e incluso mostrando una latencia menor en los entrenamientos realizados.

2.2.3 Arquitecturas Basadas en Transformers para Video. El paradigma más reciente en el reconocimiento de acciones proviene de la arquitectura Transformer que revolucionó el Procesamiento de Lenguaje Natural (NLP) [27]. Esta arquitectura ha impulsado una evolución importante en NLP y actualmente también se ha incursionado en la visión artificial. La adaptación al reconocimiento de acciones en video proviene de modelos como el Video Vision Transformer (ViViT) [6].

ViViT extrae del video ‘tubelets’ espacio-temporales, que son volúmenes de píxeles (por ejemplo, de 16x16 píxeles a lo largo de 2 fotogramas). Estos tubelets se proyectan linealmente para crear una secuencia de tokens. Esta secuencia se procesa mediante un codificador Transformer estándar, utilizando el mecanismo de autoatención, donde el modelo aprende las relaciones entre todos los tubelets del video, sin importar la separación de espacio o tiempo.

Siguiendo la teoría, la capacidad de ViViT para modelar relaciones globales y a largo plazo es excelente para la detección de violencia, según confirma el estudio realizado por Singh, et al. (2022) [28]. Sin embargo, el mecanismo de autoatención impone unos altos requisitos computacionales y la necesidad de conjuntos de datos exageradamente grandes para alcanzar su máximo potencial. Son necesarias múltiples técnicas de preprocesamiento y de aumento de datos para sobreponerse a este requerimiento.

2.2.4 Ventajas e Inconvenientes de Transformers para Video Frente a CNNs. Las CNNs y los Transformers representan paradig-

mas diferentes para el procesamiento de datos visuales, especialmente en videos. La principal distinción radica en sus sesgos inductivos, que son las asunciones que un modelo hace sobre la naturaleza de los datos para facilitar el aprendizaje.

Las CNNs basan su éxito en dos sesgos inductivos que son adecuados para imágenes y videos. El primero es la Localidad, donde las capas convolucionales procesan la imagen aplicando filtros en conjuntos de píxeles adyacentes. Esto proviene por diseño de las CNNs, donde se asume que los píxeles cercanos están muy correlacionados y que las características visuales como bordes o texturas son locales. El otro es la Invariancia a la traslación, donde un mismo filtro convolucional se aplica a lo largo de todo el fotograma o imagen. Esto asume que si una característica es importante en una parte de la imagen, también lo será en otra.

Por otro lado, los Transformers carecen de estos sesgos inductivos. Estos dividen la imagen o el video en una secuencia de parches o ‘tubelets’ y procesan las relaciones entre ellos. El mecanismo de autoatención les permite modelar dependencias a larga distancia entre cualquier par de parches o tubelets en la imagen o en el video. Esta flexibilidad, sin embargo, conlleva dos desafíos críticos: la ineficiencia de datos (‘Data Hungry’) exige un preentrenamiento masivo, y un alto coste computacional. Este dilema entre la eficiencia de las CNNs y el poder expresivo de los Transformers es de gran importancia para la detección de acciones complejas como la violencia, y justifica la necesidad de una evaluación comparativa para determinar qué modelo es más conveniente dependiendo de la situación.

2.3 Resultados de Distintas Arquitecturas en Detección de Violencia

Distintas revisiones de literatura sobre detección de violencia en videos revelan tendencias claras tanto en el rendimiento de clasificación como en la eficiencia de las arquitecturas.

2.3.1 Rendimiento de Clasificación. Según lo visto en la bibliografía, se realizó la tabla 1, que resume el rendimiento en validación de varias arquitecturas representativas del estado del arte.

Existen grandes avances en lo que respecta a la clasificación de videos de violencia, con un modelo basado en Transformers (CrimeNet) con los mejores resultados en cuanto al rendimiento en todos los conjuntos de datos presentados.

2.3.2 Costo Computacional y Eficiencia. Se realizó la tabla 2, que compara distintas arquitecturas con información obtenida de la bibliografía. Los modelos que fueron analizados tienen configuraciones particulares: para los modelos basados en CNNs, la configuración se

Dataset	Arquitectura	Exactitud	Referencia
Hockey Fights	ViViT	97.14 %	Singh et al. (2022) [28]
Hockey Fights	SlowFast	97.30 %	Diao et al. (2022) [29]
Hockey Fights	I3D (RGB)	98.50 %	Cheng et al. (2020) [15]
Hockey Fights	CrimeNet	100 %	Rendón-Segador et al. (2024) [30]
RLVS	I3D (RGB)	87.60 %	Diao et al. (2022) [29]
RLVS	SlowFast	91.20 %	Diao et al. (2022) [29]
RLVS	CNN 3D preent. (X3D)	98.00 %	Santos et al. (2021) [31]
RLVS	CrimeNet	100 %	Rendón-Segador et al. (2024) [30]
RWF-2000	CNN 3D preent. (X3D)	84.75 %	Santos et al. (2021) [31]
RWF-2000	I3D (RGB)	85.75 %	Cheng et al. (2020) [15]
RWF-2000	CNN 2D Bi-LTMA + TSM	90.25 %	Wang et al. (2024) [32]
RWF-2000	CrimeNet	99.98 %	Rendón-Segador et al. (2024) [30]

Tabla 1: Rendimiento de arquitecturas representativas en datasets de violencia (publicado en la bibliografía)

basa en ResNet-50 (R-50), con un número de muestreo de fotogramas particular por clip. Para el ViViT, se utiliza su modelo ‘Base’, con tubelets de 16x16 píxeles y 2 fotogramas (16x2), en un clip muestreado a 32 frames. Esta comparación muestra el número de parámetros a ajustar y la velocidad en operaciones de punto flotante por segundo (GFLOPs), según los datos reportados por la bibliografía.

Arquitectura	Configuración	Parámetros (M)	GFLOPs ¹	Referencia
I3D	R-50, 32 frames	28	76.5	Carreira y Zisserman (2017) [3]
TSM	R-50, 8 frames	24.3	33	Lin et al. (2019) [5]
SlowFast	R-50, 8x8 frames	32.9	65.7	Feichtenhofer et al. (2019) [4]
ViViT	Base - 16x2, 32fr.	88.9	455.2	Arnab et al. (2021) [6]

Tabla 2: Coste Computacional de Arquitecturas Representativas (Valores de Referencia)

Entre las arquitecturas destacadas, se revela que TSM es la arquitectura más ligera por diseño. En el otro extremo, se observa que ViViT tiene el mayor número de parámetros y GFLOPs por amplia diferencia. Estos datos sirven para mostrar un contexto donde se han desarrollado modelos que alcanzan muy alta precisión, pero a menudo implica un alto costo computacional. El equilibrio óptimo entre rendimiento y eficiencia sigue siendo, entonces, un área de investigación activa y relevante.

3. Metodología

En este apartado se describe la metodología seguida para cumplir con el objetivo de este trabajo, que es realizar una evaluación comparativa de cuatro arquitecturas de aprendizaje profundo para la tarea de clasificación binaria de videos como ‘violentos’ o ‘no violentos’, que será llevada a cabo en dos fases, cada una bajo el ajuste fino de un conjunto de datos diferente. Asimismo, también se realizarán pruebas en dos configuraciones con recursos computacionales distintos.

Según la revisión del estado del arte (apartado 13), los datasets elegidos, que contienen agresiones físicas como forma de violencia, son

Real-Life Violence Situations (RLVS) y Real-World Fighting (RWF-2000) para realizar el ajuste fino de cada modelo, y el Hockey Fights Dataset, con su dominio controlado y específico, se utilizará solamente para probar la robustez de los modelos.

Se seleccionaron cuatro arquitecturas que representan diferentes filosofías del reconocimiento de acciones, basadas en el análisis del estado del arte (apartado 13): I3D como una sólida base de CNNs 3D, SlowFast como un enfoque complejo de dos vías inspirado en la biología, TSM como el algoritmo con gran eficiencia, y ViViT como el representante de la emergente arquitectura Transformer.

Respecto a la configuración, las arquitecturas basadas en CNN (I3D en su versión RGB, SlowFast y TSM) tienen una ResNet-50 de backbone. Por el lado de ViViT, mantiene la arquitectura básica presentada en su artículo (no tiene factorizaciones). Todos los modelos fueron preentrenados en Kinetics-400 y se les cambió la cabeza de clasificación (capas finales de la red que transforman las representaciones profundas en probabilidades de clase) a 2 clases (violencia y no violencia).

El diseño experimental se estructura en dos etapas principales: una primera etapa donde se utilizarán elementos de computación de prestaciones estándar, y una segunda etapa con dispositivos con mayor capacidad de computación, permitiendo realizar una comparativa tanto en entornos reducidos como en entornos más potentes.

3.1 Primera Etapa: Marco Común de Experimentación

Para llevar a cabo la comparativa, se establece un marco común, adaptando algunos hiperparámetros y configuraciones, buscando también un buen rendimiento de cada arquitectura. Se utilizan recursos computacionales estándar, en los que se incluye la GPU NVIDIA GeForce RTX 3060 Ti con 8 GB de VRAM, un Intel(R) Core(TM) i5-12600 y 32 GB de RAM.

3.1.1 Preentrenamiento y Ajuste Fino. Para establecer el mismo contexto de trabajo, todos los modelos partirán de pesos preentrenados en el dataset Kinetics-400, y posteriormente se les realizará un ajuste fino con distintos parámetros y configuraciones que se describen a continuación. Para todos los modelos, se utilizó el optimizador AdamW [33] por su mejora en el rendimiento, y la función de pérdida Cross Entropy Loss, que es la comúnmente utilizada para estos casos. Respecto a los hiperparámetros, la tabla 3 muestra los valores de configuración.

Modelo	LR	Weight Decay	Épocas	Batch Size	Frames Muestreados	Frame Step
ViViT	2×10^{-5}	1×10^{-2}	5	2	32	4
I3D	1×10^{-4}	1×10^{-5}	10	2	32	4
SlowFast	1×10^{-4}	1×10^{-5}	10	2	32 (fast), 8 (slow)	4
TSM	1×10^{-5}	1×10^{-5}	10	2	8	8

Tabla 3: Comparativa de hiperparámetros de ajuste fino

La selección de estos hiperparámetros se basó en una combinación de recomendaciones de la literatura, las configuraciones de los modelos preentrenados, el tamaño de los conjuntos de datos para hacer el ajuste fino (respecto a Kinetics-400) y las limitaciones del hardware disponible.

En la configuración de los hiperparámetros se siguen los lineamientos para el ajuste fino de cada arquitectura, donde a su vez se hereda el uso de un número de frames de entrada y el salto entre frames. El tamaño de lote se fijó en 2 debido a restricciones de memoria en ViViT.

Se buscó Optimizar el uso del hardware mediante la utilización de la Precisión Mixta Automática (AMP), pues reduce el consumo de memoria y también disminuye el tiempo en operaciones aritméticas y de memoria en redes neuronales profundas al reducir la precisión numérica [34]. ViViT además cuenta con el ‘Gradient Checkpointing’ para obtener un mayor ahorro en la memoria de la GPU. Esta técnica guarda algunas activaciones clave (checkpoints) y descarta otras durante la retropropagación, permitiendo ahorrar memoria (a costa de un tiempo mayor de computación) [35].

Respecto al Preprocesamiento, se realizó una normalización y se toman los frames correspondientes según la configuración de hiperparámetros. Si el video no permite obtener el número de frames requerido, se aplican estrategias de padding, que se basan en repetir el último frame, ya sea copiándolo o haciendo una relectura.

Se emplearon técnicas básicas de Aumento de datos en la fase de entrenamiento con el objetivo de incrementar la diversidad del conjunto de datos. Las técnicas a destacar son dos: el Muestreo Temporal Aleatorio del Clip, donde los frames de cada video se extraen desde una posición de inicio aleatoria y la segunda es el Redimensionamiento y Recorte Aleatorio, donde a cada frame extraído del video se le implementó un redimensionamiento a un tamaño de 256x256 píxeles, para posteriormente realizar un recorte aleatorio controlado y así pasar a 224x224 píxeles con distinto posicionamiento entre épocas.

Cabe remarcar que todas estas técnicas solo se aplican en la fase de entrenamiento para evitar un rápido sobreajuste y buscar mejor generalización de los modelos. En validación, se evita aumentar los datos para que se pueda evaluar correctamente el entrenamiento realizado en cada época.

3.2 Segunda etapa: Marco Común de Experimentación Revisado

En la Segunda Etapa, se utilizó una configuración con mayores recursos computacionales, en los que se incluye la GPU NVIDIA GeForce RTX 4090 con 24 GB de VRAM, un Intel(R) Core(TM) i9-13900KF y 64 GB de RAM.

A continuación se destacarán las principales diferencias respecto de la primera etapa, dada la mayor capacidad computacional. En la tabla 4 aparecen los nuevos hiperparámetros. Como se observa, ahora el Tamaño de lote (Batch Size) es de 16 para todos los modelos; aunque no sea ideal, se mantiene el mismo para observar tanto métricas de rendimiento como de eficiencia. Respecto a las épocas, ahora cada modelo cuenta con 30 épocas, que, acompañado de un LR y Weight Decay adaptados a esta situación, se realiza un ajuste fino más prolongado para observar el funcionamiento más a fondo de cada modelo.

Modelo	Learning Rate (LR)	Weight Decay	Épocas	Batch Size
ViViT	2×10^{-5}	1×10^{-2}	30	16
I3D	1×10^{-5}	2×10^{-5}	30	16
SlowFast	1×10^{-5}	2×10^{-5}	30	16
TSM	1×10^{-5}	1×10^{-3}	30	16

Tabla 4: Hiperparámetros de la segunda etapa

Se incluye además un Planificador (Scheduler) para ajustar la tasa de aprendizaje durante el entrenamiento. Se utiliza la técnica del ‘decaimiento del coseno’ (Cosine Decay), una técnica moderna y muy adoptada para el entrenamiento de redes neuronales profundas [36]. En cuanto al aumento de datos, se implementó el volteo horizontal. Esto está ligado con el número mayor de épocas y la búsqueda de que los modelos puedan generalizar mejor la información recibida de los clips.

3.3 Métricas de Evaluación

Se utilizarán tanto métricas de rendimiento en la clasificación, como métricas de eficiencia y del costo computacional. De esta manera, se podrá comparar con más profundidad el funcionamiento de cada modelo en estudio.

Las métricas de rendimiento de clasificación servirán para evaluar y comparar el rendimiento en la clasificación realizada por cada modelo. Se basan en los resultados de clasificación: Verdadero Positivo (TP), Falso Positivo (FP), Falso Negativo (FN) y Verdadero Negativo (TN). Con estos valores se realiza la matriz de confusión, y también se calcularán 4 métricas:

- $Recall$ (Sensibilidad) = $\frac{TP}{TP+FN}$
 Permite visualizar la sensibilidad para detectar positivos correctamente.
- $Precisión$ = $\frac{TP}{TP+FP}$
 Muestra la precisión obtenida respecto a los elementos positivos.
- $Accuracy$ (Exactitud) = $\frac{TP+TN}{TP+TN+FP+FN}$
 La exactitud calcula las predicciones correctas, tanto positivas como negativas, sobre todos los elementos.
- $F1-Score$ = $\frac{2 \times (Precision \times Recall)}{Precision + Recall}$

Esta métrica pondera la Precisión y el Recall (sensibilidad) de un sistema calculando su media armónica. Cuanto mayor F1-Score, mejor equilibrio entre el número de falsos positivos y de falsos negativos.

Respecto a las métricas que permiten comparar el costo computacional y eficiencia, se establecieron 4 métricas, cuyos valores, entre otros factores, dependen del hardware que se utilice durante la experimentación.

- Velocidad de inferencia (FPS): Se calculará cuántos fotogramas son los que puede procesar en un segundo cada modelo.
- Tiempo de entrenamiento total por modelo: Se contabilizará el tiempo de entrenamiento total en cada modelo, resumido en minutos.
- Operaciones de Punto Flotante por Segundo (FLOPs): Se hará una estimación de FLOPs utilizando la librería *fvcore*.
- Número de parámetros del modelo: Estos parámetros son los que los modelos ajustan durante el entrenamiento, lo que dictamina la complejidad del mismo. También se obtendrá con *fvcore*.

3.4 Protocolo de Experimentación

El diseño experimental de este trabajo se estructura en las dos etapas descritas en los apartados 12 y 13, una con los recursos computacionales estándar, y la segunda con un hardware de mayor capacidad computacional. Se ejecuta el mismo procedimiento experimental en cada etapa para evaluar el rendimiento y la robustez. Este procedimiento tiene dos fases:

- Fase 1: Ajuste fino y evaluación con RWF-2000 como dataset primario.
- Fase 2: Ajuste fino y evaluación con RLVS como dataset primario.

Ambas fases siguen el mismo procedimiento, que incluye los siguientes pasos:

1. Entrenamiento y Validación Intra-Dataset: Cada modelo será entrenado y validado mediante un ajuste fino en el conjunto de datos según la fase (RWF-2000 o RLVS), partiendo de sus pesos preentrenados en Kinetics-400.
2. Selección del Mejor Modelo: Se guarda el checkpoint del modelo que alcance el mejor F1-Score en el conjunto de validación del dataset elegido durante el entrenamiento. El F1-Score asegura que el modelo seleccionado sea aquel que ofrece el mejor compromiso entre Precisión y Recall.
3. Estimación de Costo Computacional: Se hará un cálculo de las métricas de costo computacional y eficiencia presentadas en el apartado 13.

4. Evaluación en Conjunto Secundario: Los modelos ajustados harán la inferencia en conjuntos de datos secundarios para medir el rendimiento y así observar la capacidad de generalización.
5. Análisis de resultados: Se realizará un análisis comparativo de todos los modelos según los resultados, basado en el rendimiento, el costo computacional, la robustez, la evolución entre ambas etapas y también se buscará determinar el balance entre rendimiento y eficiencia.

4. Resultados

Este capítulo detalla los resultados cuantitativos derivados del protocolo experimental, primero para la primera etapa y luego para la segunda etapa, en cada una de sus fases (ajuste fino en RWF-2000 y RLVS, respectivamente).

4.1 Resultados de la primera etapa

Esta etapa se realizó utilizando una GPU NVIDIA RTX 3060 Ti con 8 GB de VRAM, lo que impuso restricciones en hiperparámetros como el tamaño del lote (2) y distintas configuraciones de los modelos.

La tabla 5 muestra las métricas de rendimiento del ajuste fino en RWF-2000, con la época con el mejor F1-Score obtenido.

Modelo	Mejor Época	F1-Score	Precision	Exactitud	Recall
ViViT	4/5	0.876	0.927	0.882	0.83
I3D	2/10	0.841	0.880	0.847	0.805
SlowFast	5/10	0.846	0.794	0.835	0.905
TSM	6/10	0.845	0.908	0.855	0.790

Tabla 5: Métricas de Rendimiento en validación en RWF-2000

Observando la tabla se puede inferir que el mayor F1-Score lo obtiene ViViT (0.876) en su cuarta época, mostrando la precisión (0.927) y exactitud (0.882) más altas entre todos los modelos.

La tabla 6 refleja una comparativa de eficiencia en términos de tiempo total de entrenamiento, número de parámetros, FLOPs y velocidad de inferencia (en FPS) en el hardware especificado. Antes de analizar estas métricas, hay que remarcar que están directamente ligadas al número de frames y al tamaño de la entrada de video.

Modelo	Parámetros (Millones)	GFLOPs	Inferencia (FPS)	Tiempo de Entr. total (min)
ViViT	88.6	270.8	5	111.8 ^b
I3D	27.2	114.2	16	177.7
SlowFast	33.6	50.6	38	167.7
TSM ^r	23.5	32.9	73	51.0

Tabla 6: Comparativa de Eficiencia y Coste Computacional en Fase 1

El modelo TSM se destaca como el más eficiente. En el extremo opuesto, se encuentra ViViT, que alcanzó el F1-Score más alto a costa de ser el modelo más pesado y lento (88.6M de parámetros y 5 FPS). Esta comparativa subraya el claro compromiso existente entre rendimiento y eficiencia. TSM ofrece un equilibrio excepcional, mientras que ViViT exige un costo computacional muy elevado para alcanzar su máximo potencial de clasificación.

Para observar la capacidad de generalización de los modelos ante cambios de dominio, los checkpoints entrenados en RWF-2000 se evaluaron directamente en los datasets RLVS y Hockey Fights. Los resultados se pueden observar en la tabla 7.

Modelo	Dataset	F1-Score	Exactitud	Precisión	Recall
ViViT	RLVS	0.887	0.884	0.861	0.914
I3D	RLVS	0.825	0.811	0.769	0.890
SlowFast	RLVS	0.741	0.674	0.614	0.936
TSM	RLVS	0.867	0.859	0.820	0.920
ViViT	Hockey Fights	0.396	0.604	0.833	0.260
I3D	Hockey Fights	0.643	0.716	0.865	0.512
SlowFast	Hockey Fights	0.755	0.788	0.893	0.654
TSM	Hockey Fights	0.723	0.759	0.851	0.628

Tabla 7: Inferencia Cruzada en RLVS y Hockey Fights (Modelos entrenados en RWF-2000)

ViViT muestra la mejor capacidad de generalización al dominio de RLVS, liderando en F1-Score (0.887), exactitud y precisión, además de obtener un muy buen Recall. Le sigue de cerca TSM, con un F1-Score de 0.867 y mostrando un rendimiento muy robusto. Se destaca también el muy alto recall de SlowFast (0.936).

En Hockey Fights el escenario es distinto; SlowFast se posiciona como el más robusto, obteniendo el mejor rendimiento en todas las métricas principales. Se observa que ViViT resulta muy deficiente: tiene un F1-Score de 0.396. Esto evoca la idea de que es incapaz de reconocer la gran mayoría de peleas en este dispar contexto visual. I3D y TSM, por su parte, muestran una generalización más sólida que ViViT en este complejo escenario.

Se hará el mismo análisis que el realizado para RWF-2000, pero ahora aplicado a los modelos ajustados sobre RLVS como dataset primario.

En la tabla 8 se pueden observar las métricas de rendimiento en la mejor época de cada modelo bajo el ajuste fino en RLVS.

Los resultados en este caso son sobresalientes. ViViT tiene las mejores métricas, con un F1-Score de 0.992, gracias a una precisión de 0.995 y un recall de 0.993. El resto de los modelos también demuestran un gran rendimiento: ninguna métrica se encuentra por debajo de 0.96, lo cual indica un rendimiento casi perfecto en todos los modelos.

⁶ ViViT tiene la mitad de épocas (5) que el resto de los modelos (10).

⁷ TSM tiene 8 frames de entrada en lugar de 32.

Modelo	Mejor Época	F1-Score	Precision	Exactitud	Recall
ViViT	4/5	0.992	0.995	0.993	0.990
I3D	10/10	0.970	0.961	0.970	0.980
SlowFast	5/10	0.967	0.975	0.968	0.960
TSM	9/10	0.983	0.980	0.983	0.985

Tabla 8: Métricas de Rendimiento en validación en RLVS

Respecto a las métricas de eficiencia, se observó que el ajuste fino en RLVS fue generalmente más rápido que en RWF-2000. El tiempo total para las 10 épocas de TSM fue de aproximadamente 44 minutos, mientras que I3D y SlowFast tomaron alrededor de 140 y 135 minutos, respectivamente. ViViT completó sus 5 épocas en unos 94 minutos.

Para observar la robustez, la tabla 9 refleja el rendimiento de los modelos entrenados en RLVS al realizar una inferencia cruzada en los conjuntos de datos RWF-2000 y Hockey Fights.

Modelo	Dataset	F1-Score	Exactitud	Precisión	Recall
ViViT	RWF-2000	0.775	0.760	0.729	0.827
I3D	RWF-2000	0.692	0.709	0.735	0.654
SlowFast	RWF-2000	0.770	0.790	0.848	0.705
TSM	RWF-2000	0.700	0.624	0.582	0.877
ViViT	Hockey Fights	0.696	0.571	0.539	0.982
I3D	Hockey Fights	0.681	0.572	0.543	0.912
SlowFast	Hockey Fights	0.796	0.811	0.864	0.738
TSM	Hockey Fights	0.684	0.539	0.520	0.996

Tabla 9: Resultados de Inferencia Cruzada en RWF-2000 y Hockey Fights (Modelos entrenados en RLVS)

Los resultados de la inferencia cruzada en RWF-2000, muestran a ViViT con el mejor F1-Score (0.775), seguido muy de cerca por SlowFast, que obtuvo un F1-Score de 0.770 con la exactitud y precisión más altas (0.790 y 0.848, respectivamente). También se destaca el recall de 0.877 obtenido por TSM, indicando que omite menos peleas, a costa de una precisión más baja.

En la inferencia cruzada con Hockey Fights, se aprecia un destacable resultado de SlowFast, con el F1-Score (0.796), exactitud (0.811) y precisión (0.864) más altos entre los 4 modelos. Por parte de ViViT, se observa una respuesta más robusta en la generalización de RLVS a Hockey Fights y, en conjunto con TSM, logran un Recall casi perfecto, indicando que ambos modelos no se pierden casi ninguna pelea, pese a generar falsas alarmas (por la baja precisión se clasifican muchas escenas no violentas como si lo fuesen).

4.2 Resultados de la segunda etapa

Se muestran los resultados de la experimentación realizada bajo un hardware con mayor capacidad computacional (GPU NVIDIA RTX 4090 de 24 GB de VRAM), permitiendo un mayor tamaño de lote (16) y 30 épocas de entrenamiento. En esta etapa, dado lo extenso del entrenamiento, se podrá observar una evolución más prolongada y así

interpretar cómo afecta la variación de hiperparámetros como el tamaño de lote en la generalización.

La tabla 10 muestra las métricas de rendimiento bajo el ajuste fino en RWF-2000.

Modelo	Mejor Época	F1-Score	Precision	Exactitud	Recall
ViViT	25/30	0.897	0.904	0.898	0.890
I3D	16/30	0.801	0.853	0.813	0.755
SlowFast	26/30	0.875	0.860	0.873	0.890
TSM	6/30	0.856	0.837	0.853	0.875

Tabla 10: Métricas de Rendimiento en validación en RWF-2000

Las métricas muestran a ViViT alcanzar el mejor valor tanto de F1-Score (0.897) como del resto de las métricas, posicionándolo como el modelo que mejor clasificó los videos del conjunto RWF-2000. De cerca lo sigue SlowFast (F1-Score de 0.875) que además obtiene el más alto recall junto a ViViT (0.890). TSM e I3D les siguen con buenos resultados (F1-Scores de 0.856 y 0.801 respectivamente).

Ahora que se tiene una capacidad computacional más grande, resultó interesante visualizar nuevamente las métricas de eficiencia y costo computacional.

Modelo	Parámetros (Millones)	GFLOPs	Inferencia (FPS)	Tiempo de Entr. total (min)
ViViT	88.6	270.4	22	190.8
I3D	27.2	149.1	46	189.7
SlowFast	33.6	50.6	102	192.7
TSM	23.5	32.9	201	48.3

Tabla 11: Comparativa de Eficiencia y Coste Computacional en Fase 1

En la tabla 11 se pueden observar los resultados obtenidos con el hardware de la segunda etapa. Si comparamos lo visto con la tabla 6 (eficiencia mostrada en la primera etapa), el cambio más interesante es visible en lo que respecta a los FPS: los modelos pueden inferir aproximadamente entre 3 y 4 veces más fotogramas por segundo. Otra métrica que destaca es el tiempo de entrenamiento: en esta etapa, debido al tamaño de lote y las capacidades computacionales, el tiempo de entrenamiento por época es alrededor de 3 veces inferior al de la primera etapa.

Para observar la robustez, se hizo una inferencia cruzada en los conjuntos RLVS y Hockey Fights. Se pueden observar resultados en la tabla 12. Observando la inferencia cruzada en RLVS, se encuentran buenos rendimientos. Este dataset tiene un dominio similar al de RWF-2000, y a su vez, sus clases se encuentran bien definidas.

ViViT mantiene su liderazgo, con el F1-Score (0.890), la exactitud (0.883) y la precisión (0.839) más altos, igual que en la primera etapa. Lo mismo se observa de SlowFast, con el recall más alto (0.967), pero que ahora obtiene mejores valores en el resto de métricas, mostrando una mejor generalización en esta etapa.

Modelo	Dataset	F1-Score	Exactitud	Precisión	Recall
ViViT	RLVS	0.890	0.883	0.839	0.948
I3D	RLVS	0.826	0.805	0.746	0.925
SlowFast	RLVS	0.860	0.843	0.775	0.967
TSM	RLVS	0.830	0.812	0.758	0.916
ViViT	Hockey Fights	0.507	0.628	0.752	0.382
I3D	Hockey Fights	0.685	0.718	0.775	0.614
SlowFast	Hockey Fights	0.791	0.760	0.700	0.910
TSM	Hockey Fights	0.730	0.679	0.630	0.870

Tabla 12: Inferencia Cruzada en RLVS y Hockey Fights (Modelos entrenados en RWF-2000)

En el caso de Hockey Fights, ViViT muestra una mejora en la generalización, que obtiene un aumento de alrededor del 11% en F1-Score respecto a la primera etapa. Aun así, sigue siendo el modelo que peor generaliza en este dominio. Nuevamente se destaca SlowFast, que obtiene los mejores valores de F1-Score (0.791), exactitud (0.760) y recall (0.910).

Se mostrarán los resultados del ajuste fino con RLVS de todos los modelos en la segunda etapa.

Respecto a los resultados de rendimiento, la tabla 13 muestra las métricas de rendimiento en RLVS en esta segunda etapa.

Modelo	Mejor Época	F1-Score	Precision	Exactitud	Recall
ViViT	10/30	0.990	0.995	0.990	0.985
I3D	23/30	0.971	1.000	0.973	0.945
SlowFast	9/30	0.985	0.990	0.985	0.980
TSM	9/30	0.971	0.952	0.970	0.990

Tabla 13: Métricas de Rendimiento en validación en RLVS

Nuevamente se visualiza un rendimiento casi perfecto: ViViT obtiene el mejor F1-Score y exactitud (0.990 en ambas métricas). Se destaca también la mejora de I3D, con una precisión del 100%. TSM y SlowFast también obtienen valores casi perfectos, al igual que ocurrió en la etapa anterior.

Respecto a la eficiencia de esta fase, el tiempo de entrenamiento es similar al de RWF-2000 y el costo computacional de los modelos es el ya visto anteriormente.

En la tabla 14 se muestran los resultados de robustez mediante la inferencia cruzada en los datasets RWF-2000 y Hockey Fights.

Los resultados de la inferencia cruzada en RWF-2000 de esta etapa, al igual que la anterior, muestran a ViViT con el mejor F1-Score (0.784), seguido por SlowFast (0.746) y también se vuelve a observar a TSM con alto recall (ahora de 0.922) en esta evaluación. La más alta precisión (0.914) es obtenida por I3D.

En Hockey Fights, los resultados confirman el desafío del cambio de dominio, y esto es observable en el decaimiento de las métricas de rendimiento. Se vuelve a observar a SlowFast con el más alto F1-Score. I3D muestra un peor F1-Score (0.473) que en la primera etapa, que aunque obtiene la mejor precisión entre los 4 modelos (0.833), tiene

Modelo	Dataset	F1-Score	Exactitud	Precisión	Recall
ViViT	RWF-2000	0.784	0.760	0.713	0.871
I3D	RWF-2000	0.694	0.754	0.914	0.560
SlowFast	RWF-2000	0.746	0.758	0.784	0.712
TSM	RWF-2000	0.715	0.633	0.584	0.922
ViViT	Hockey Fights	0.695	0.577	0.543	0.966
I3D	Hockey Fights	0.473	0.632	0.833	0.330
SlowFast	Hockey Fights	0.758	0.686	0.617	0.984
TSM	Hockey Fights	0.684	0.540	0.521	0.996

Tabla 14: Segunda etapa - Resultados de Inferencia Cruzada en RWF-2000 y Hockey Fights (Modelos entrenados en RLVS)

un pésimo recall (0.330), clasificando a la mayoría de eventos como violentos y generando dicho F1-Score. Esto muestra un fuerte sobreajuste sobre el conjunto de datos original, impidiendo la generalización de este modelo en otros dominios.

5. Discusión

Los resultados de este estudio muestran que no existe una única arquitectura superior para la detección de violencia. En su lugar, se observa un compromiso entre el rendimiento de clasificación, la eficiencia computacional y la robustez ante cambios de dominio. La elección del modelo óptimo, por tanto, es una decisión estratégica que depende de los requisitos específicos de la aplicación.

La figura 2 sirve como resumen de los resultados presentados y analizados previamente. El panel izquierdo muestra el rendimiento de los modelos en el dominio de entrenamiento (RWF-2000), donde ViViT se consolida como el del mejor rendimiento. Fiel a la potencia teórica de los Transformers, alcanza el F1-Score más alto, demostrando su capacidad para aprender los patrones complejos de este desafiante dataset.

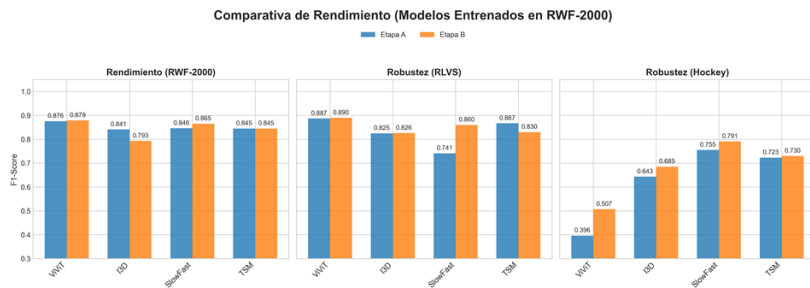


Figura 2: Rendimiento de los modelos (F1-Score) en cada Etapa en RWF-2000

Sin embargo, a la derecha de dicha figura se observa la fragilidad de ViViT a la hora de generalizar. Al ser evaluado en un dominio

visualmente distinto como el de Hockey Fights, el rendimiento de ViViT colapsa. En este escenario, SlowFast es el modelo que resulta ser el más robusto, superando en rendimientos al cambiar de dominio a las demás arquitecturas. Esto sugiere que su diseño de dos vías es superior para capturar las dinámicas de movimiento universales, independientemente del contexto visual.

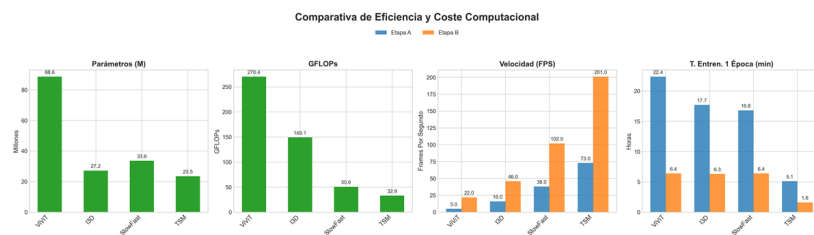


Figura 3: Comparativa de la eficiencia de los modelos

Además, en la sección de resultados también se constató que la naturaleza del dataset de entrenamiento (el desafiante RWF-2000 frente al más limpio RLVS) es un factor tan determinante como la propia arquitectura.

Respecto a la eficiencia, como se observa en los resultados y en la figura 3, TSM es indiscutiblemente el modelo más eficiente y que igualmente mantiene buenas métricas de rendimiento. Ofreció un rendimiento muy competitivo, pero con una fracción del coste computacional de sus rivales, siendo el modelo más ligero, rápido y estable. Su sobresaliente equilibrio lo convierte en la opción más lógica para sistemas del mundo real con restricciones de recursos. El estudio también demuestra que el entorno experimental es un factor muy importante, ya que la segunda etapa permitió no solo mejores métricas de rendimiento en modelos como ViViT, sino que también mejoró la eficiencia.

6. Conclusiones

Este estudio comparativo concluye que no existe una única arquitectura superior para la tarea de detección de violencia. En su lugar, los resultados muestran un compromiso entre el rendimiento de clasificación, la eficiencia computacional y la robustez ante cambios de dominio. La elección del modelo, por tanto, no es una cuestión de superioridad absoluta, sino una decisión estratégica que debe alinearse con los requisitos de la aplicación final.

La investigación posiciona a ViViT como el de mejor rendimiento en dominios conocidos; a TSM como el modelo pragmático y más eficiente; y a SlowFast como la arquitectura más robusta ante escenarios impredecibles. Se demostró, además, que factores como la complejidad

del dataset de entrenamiento y la configuración del hardware son críticos e interactúan directamente con la arquitectura para determinar el resultado final.

A pesar de estos hallazgos, el estudio presenta ciertas limitaciones que abren nuevas líneas de investigación. Se utilizaron las versiones básicas de los modelos, por lo que un trabajo futuro clave sería evaluar variantes optimizadas (como ViViT con factorización) para reevaluar el equilibrio de eficiencia. Asimismo, una búsqueda de hiperparámetros más exhaustiva y el uso de técnicas de aumento de datos más avanzadas podrían mejorar la robustez general, especialmente en los modelos basados en Transformers. Finalmente, extender esta comparativa a datasets con formas más sutiles de violencia y explorar arquitecturas híbridas que combinen, por ejemplo, la eficiencia de TSM con la robustez de SlowFast, son direcciones prometedoras para avanzar hacia soluciones más completas y versátiles.

Referencias

1. R. McGrady, K. Zheng, R. Curran, *et al.*, “Dialing for Videos: A Random Sample of YouTube,” *Journal of Quantitative Description: Digital Media*, vol. 3, 2023.
2. N. Sulman, T. Sanocki, D. Goldgof, *et al.*, “How effective is human video surveillance performance?,” in *2008 19th International Conference on Pattern Recognition*, pp. 1–3, 2008.
3. J. Carreira and A. Zisserman, “Quo Vadis, Action Recognition? A New Model and the Kinetics Dataset,” feb 2018.
4. C. Feichtenhofer, H. Fan, J. Malik, *et al.*, “SlowFast Networks for Video Recognition,” oct 2019.
5. J. Lin, C. Gan, and S. Han, “Tsm: Temporal shift module for efficient video understanding,” 2019.
6. A. Arnab, M. Dehghani, G. Heigold, *et al.*, “Vivit: A video vision transformer,” 2021.
7. P. Negre, R. S. Alonso, A. González-Briones, *et al.*, “Literature review of deep-learning-based detection of violence in video,” *Sensors*, vol. 24, no. 12, 2024.
8. B. Omarov, S. Narynov, Z. Zhumanov, *et al.*, “State-of-the-art violence detection techniques in video surveillance security systems: a systematic review,” *PeerJ Computer Science*, vol. 8, 2022.
9. J. A. Avellaneda Gonzalez, T. Matsukawa, and E. Suzuki, “Leveraging vision language models for understanding and detecting violence in videos,” in *Proceedings of the 20th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications - Volume 2: VISAPP*, pp. 99–113, INSTICC, SciTePress, 2025.
10. N. Mumtaz, N. Ejaz, S. Habib, *et al.*, “An overview of violence detection techniques: Current challenges and future directions,” 2022.
11. J.-H. Park, M. Mahmoud, and H.-S. Kang, “Conv3d-based video violence detection network using optical flow and rgb data,” *Sensors*, vol. 24, no. 2, 2024.
12. E. Bermejo Nievas, O. Deniz Suarez, G. Bueno García, *et al.*, “Violence Detection in Video Using Computer Vision Techniques,” in *Computer Analysis of Images and Patterns* (P. Real, D. Diaz-Pernil, H. Molina-Abril, A. Berciano, and W. Kropatsch, eds.), (Berlin, Heidelberg), pp. 332–339, Springer Berlin Heidelberg, 2011.
13. C.-H. Demarty, C. Penet, M. Soleymani, *et al.*, “VSD, a public dataset for the detection of violent scenes in movies: design, annotation, analysis and evaluation,” *Multimedia Tools and Applications*, vol. 74, pp. 7379–7404, sep 2015.
14. M. Schedi, M. Sjöberg, I. Mironică, *et al.*, “Vsd2014: A dataset for violent scenes detection in hollywood movies and web videos,” in *2015 13th International Workshop on Content-Based Multimedia Indexing (CBMI)*, pp. 1–6, 2015.
15. M. Cheng, K. Cai, and M. Li, “Rwf-2000: An open large scale video database for violence detection,” in *2020 25th International Conference on Pattern Recognition (ICPR)*, pp. 4183–4190, 2021.
16. W. Sultani, C. Chen, and M. Shah, “Real-world anomaly detection in surveillance videos,” 2019.

17. P. Wu, J. Liu, Y. Shi, *et al.*, “Not only look, but also listen: Learning multimodal violence detection under weak supervision,” 2020.
18. C. Gu, C. Sun, D. A. Ross, *et al.*, “Ava: A video dataset of spatio-temporally localized atomic visual actions,” 2018.
19. M. H. Habeb, M. Salama, and L. A. Elrefaei, “Enhancing video anomaly detection using a transformer spatiotemporal attention unsupervised framework for large datasets,” *Algorithms*, vol. 17, no. 7, 2024.
20. S. Ji, W. Xu, M. Yang, *et al.*, “3d convolutional neural networks for human action recognition,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 1, pp. 221–231, 2010.
21. M. Baccouche, F. Mamalet, C. Wolf, *et al.*, “Sequential deep learning for human action recognition,” in *Human Behavior Understanding*, (Berlin, Heidelberg), pp. 29–39, Springer Berlin Heidelberg, 2011.
22. K. He, X. Zhang, S. Ren, *et al.*, “Deep residual learning for image recognition,” 2015.
23. C. Szegedy, W. Liu, Y. Jia, *et al.*, “Going deeper with convolutions,” 2014.
24. J. Deng, W. Dong, R. Socher, *et al.*, “Imagenet: A large-scale hierarchical image database,” in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 248–255, 2009.
25. W. Kay, J. Carreira, K. Simonyan, *et al.*, “The kinetics human action video dataset,” 2017.
26. D. Freire-Obregón, P. Barra, M. Castrillón-Santana, *et al.*, “Inflated 3D ConvNet context analysis for violence detection,” *Machine Vision and Applications*, vol. 33, p. 15, dec 2021.
27. A. Vaswani, N. Shazeer, N. Parmar, *et al.*, “Attention is all you need,” 2017.
28. S. Singh, S. Dewangan, G. S. Krishna, *et al.*, “Video vision transformers for violence detection,” 2022.
29. X. Diao and Y. Xu, “A slowfast-based violence recognition method,” in *2022 6th Asian Conference on Artificial Intelligence Technology (ACAIT)*, pp. 1–6, 2022.
30. F. J. Rendón-Segador, J. A. Álvarez García, and L. M. Soria-Morillo, “Transformer and adaptive threshold sliding window for improving violence detection in videos,” *Sensors*, vol. 24, no. 16, 2024.
31. F. Santos, D. Duraes, F. Marcondes, *et al.*, *Efficient Violence Detection Using Transfer Learning*, pp. 65–75. Berlin, Heidelberg: Springer Berlin Heidelberg, 09 2021.
32. J. Wang, D. Zhao, H. Li, and D. Wang, “Lightweight violence detection model based on 2d cnn with bi-directional motion attention,” *Applied Sciences*, vol. 14, no. 11, 2024.
33. I. Loshchilov and F. Hutter, “Decoupled weight decay regularization,” 2019.
34. P. Micikevicius, S. Narang, J. Alben, *et al.*, “Mixed precision training,” 2018.
35. T. Chen, B. Xu, C. Zhang, *et al.*, “Training deep nets with sublinear memory cost,” 2016.
36. I. Loshchilov and F. Hutter, “Sgdr: Stochastic gradient descent with warm restarts,” 2017.

Incorporación de técnicas de IA para asistir al preprocesamiento de datos para usuarios no expertos

Marta de Castro Leira, Alicia García Holgado, and Andrea Vázquez Ingelmo

Universidad de Salamanca
Departamento de Informática y Automática
Facultad de Ciencias.
Plaza de los Caídos s/n
37008 Salamanca, España
{mardecalei,aliciagh,andreavazquez}@usal.es

Resumen El preprocesamiento de datos es una etapa crítica en el análisis y aprendizaje automático, pero suele resultar compleja y poco accesible para usuarios sin experiencia técnica. Este trabajo propone un sistema interactivo que asiste el preprocesamiento mediante técnicas de Inteligencia Artificial (IA), orientado a usuarios no expertos. Para fundamentar el desarrollo, se realizó una revisión sistemática de la literatura (SLR), que permitió identificar las técnicas más relevantes y los principales vacíos de accesibilidad y explicabilidad. Con estos hallazgos se diseñó un prototipo que combina un backend en Flask y un frontend en Angular, integrando procesos automáticos de limpieza, normalización, selección de características y evaluación con AutoML. La validación práctica con usuarios no técnicos permitió identificar mejoras y refinar la herramienta para que resulte más accesible y comprensible. Los resultados sugieren que esta solución no solo automatiza el preprocesamiento, sino que también democratiza su uso, convirtiéndolo en una herramienta educativa y práctica. Como trabajo futuro, se plantean mejoras generales que fortalezcan aún más la accesibilidad y la personalización del proceso.

Keywords: Preprocesamiento automatizado · Inteligencia Artificial · Usuarios no expertos · Herramientas accesibles · AutoML · Interfaz gráfica

1. Introducción

El análisis de datos y el aprendizaje automático se han convertido en herramientas fundamentales para la toma de decisiones en diversos ámbitos. Sin embargo, el éxito de cualquier proyecto de análisis depende en gran medida de la calidad y preparación de los datos utilizados. Esta fase inicial, conocida como preprocesamiento de datos, resulta

crítica para transformar los datos en un formato adecuado, eliminando errores, inconsistencias y valores atípicos que puedan comprometer los resultados.

El preprocesamiento abarca tareas como la limpieza de datos, la normalización y la selección de características. Estas tareas, aunque esenciales, suelen requerir conocimientos técnicos avanzados. Para profesionales de otras áreas —como la salud o la educación—, la falta de formación específica en ciencia de datos convierte el preprocesamiento en un obstáculo importante. Además, los métodos manuales que suelen emplearse son propensos a errores, consumen mucho tiempo y dependen del criterio subjetivo de cada persona, lo que puede afectar la calidad final de los modelos predictivos y los análisis.

En respuesta a estos desafíos, la automatización del preprocesamiento mediante técnicas de inteligencia artificial ha surgido como una solución prometedora. Estas técnicas permiten reducir la intervención humana y minimizar los errores, facilitando la preparación de datos con mayor eficiencia y precisión. Sin embargo, muchas de las herramientas basadas en IA siguen orientadas a perfiles expertos, limitando su accesibilidad y comprensión para usuarios no técnicos.

Ante esta problemática, el presente trabajo propone el desarrollo de una herramienta interactiva que combina un flujo de preprocesamiento automático con explicaciones visuales y una interfaz sencilla, diseñada específicamente para usuarios sin experiencia técnica avanzada. Esta solución no solo busca facilitar el acceso a técnicas avanzadas de preprocesamiento, sino también mejorar la calidad de los conjuntos de datos y, en consecuencia, la fiabilidad de los análisis y modelos resultantes.

Para fundamentar este desarrollo, se ha realizado una revisión sistemática de la literatura que identifica las técnicas de IA más relevantes y accesibles para el preprocesamiento de datos, así como las principales barreras de accesibilidad y explicabilidad existentes. A partir de estos hallazgos, se diseña e implementa un sistema que integra dichas técnicas en un flujo automatizado, ofreciendo explicaciones comprensibles y fomentando la participación del usuario. Finalmente, se valida el sistema mediante pruebas con usuarios no expertos, analizando su experiencia, comprensión y utilidad percibida.

1.1 Objetivos

El objetivo general de este trabajo es simplificar el preprocesamiento de datos mediante técnicas de IA, orientado a usuarios no expertos. Para ello, se plantean los siguientes objetivos específicos:

- Analizar las principales técnicas de IA para el preprocesamiento de datos, identificando las más relevantes y accesibles para usuarios sin conocimientos técnicos.

- Diseñar e implementar una herramienta que combine un backend robusto con una interfaz gráfica intuitiva, permitiendo procesar datos de forma sencilla.
- Evaluar la usabilidad y efectividad de la herramienta mediante pruebas prácticas con usuarios no expertos.

Por último, el resto de este artículo se organiza de la siguiente forma: la Sección 2 describe la revisión sistemática de la literatura que fundamenta el desarrollo de la herramienta; la Sección 3 presenta la metodología seguida y el diseño de la solución propuesta; la Sección 4 discute los resultados obtenidos y la validación con usuarios; y, finalmente, la Sección 5 recoge las conclusiones principales y plantea líneas futuras de mejora.

2. Revisión sistemática de la literatura

Para fundamentar el desarrollo de la herramienta propuesta, se llevó a cabo una revisión sistemática de la literatura centrada en técnicas de preprocesamiento de datos asistidas por IA, con especial atención a herramientas accesibles para usuarios sin conocimientos técnicos. Esta revisión siguió las directrices metodológicas de Kitchenham y Charters [1] y PRISMA 2020 [2], con el propósito de sintetizar la evidencia científica reciente (2020–2025) y detectar vacíos y oportunidades de mejora.

Se definieron dos bloques de preguntas para guiar y estructurar el análisis: las preguntas de mapeo (MQ), que permiten caracterizar el panorama de publicaciones, y las preguntas de investigación (RQ), orientadas a un análisis cualitativo y crítico de las herramientas y técnicas encontradas. Para complementar este marco, se empleó el PICOC, que define la población (usuarios no expertos), la intervención (técnicas de IA para limpieza, normalización y selección), los resultados esperados (herramientas accesibles y efectivas) y el contexto de aplicación (entornos sin formación técnica especializada).

La búsqueda se realizó en las bases de datos Scopus y Web of Science (WoS), ampliamente reconocidas por su cobertura en el ámbito de la IA y las ciencias computacionales, usando términos clave relacionados con el preprocesamiento, la IA y la accesibilidad para usuarios no técnicos. Se aplicaron criterios de inclusión (estudios relevantes y publicados en inglés o español, recientes y completas) y de exclusión (duplicados, sin acceso o fuera de tema). Además cada estudio se evaluó con siete criterios de calidad, reteniendo sólo los que superaron un umbral mínimo (4.5/7 puntos).

El proceso completo quedó reflejado en el diagrama PRISMA (Figura 1): de los 666 registros iniciales se eliminaron 133 duplicados y se cribaron los 533 restantes aplicando los criterios de inclusión, exclusión y calidad. Finalmente, se seleccionaron 78 estudios que constituyen la base para el análisis y la discusión de este artículo.

El proceso completo se refleja en el diagrama PRISMA (Figura 1): de los 666 registros iniciales se eliminaron 133 duplicados y se cribaron los 533 restantes aplicando los criterios de inclusión, exclusión y calidad. Finalmente, se seleccionaron 78 estudios que constituyen la base para el análisis y la discusión del artículo. Los datos detallados y el seguimiento de cada etapa de filtrado pueden consultarse en un archivo Excel alojado en Zenodo: <https://zenodo.org/records/15631262>.

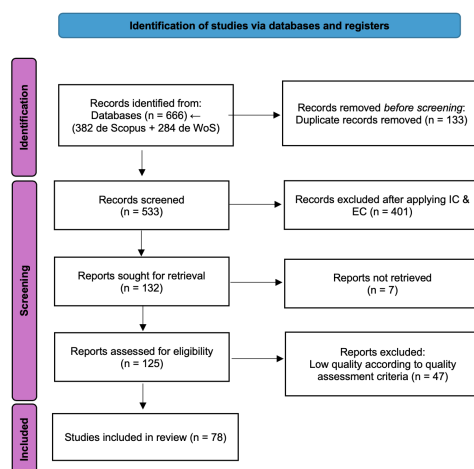


Figura 1: Diagrama PRISMA: flujo de búsqueda, selección y revisión de los estudios.

2.1 Resultados y análisis

El análisis de los 78 estudios finales confirma un crecimiento sostenido en el interés por herramientas de preprocesamiento automático con IA accesibles a usuarios no técnicos, desde 2020 en adelante. Se identificaron autores recurrentes como Wang D., Zhang J. o McCombe N. [3–5], y foros de publicación diversos, destacando *CEUR Workshop Proceedings* [6, 7], *Diagnostics* [8, 9] y *Scientific Reports* [10, 11]. Los dominios de aplicación más representados fueron salud y biomedicina (40 estudios) [4, 6, 12, 13], seguidos de ciencia de datos, agricultura, ingeniería y otros.

La mayoría de herramientas priorizan la usabilidad y la explicabilidad (con interfaces gráficas y módulos visuales [14–17]), y emplean técnicas como KNN, Random Forest y autoencoders [17–20]. También se identificó un uso creciente de AutoML [4] y de modelos preentrenados (ResNet, VGG [20, 21]). En limpieza y normalización, predominan estrategias como Min-Max Scaling y métodos de imputación como KNN

e Imputación Iterativa [7, 22, 23]. Para la selección de características, se utilizan generalmente métodos basados en Random Forest [19, 24].

En cuanto a la evaluación de la accesibilidad y usabilidad de estas herramientas, la mayoría de estudios (48) se centraron en la facilidad de uso y la interfaz intuitiva, mientras que sólo 10 utilizaron métricas cuantitativas y 3 realizaron pruebas de usabilidad en entornos reales [15, 25–27], reflejando la necesidad de más validación empírica y controlada.

Entre las principales ventajas destacan la automatización completa del flujo de trabajo [4, 15, 22, 28], la interfaz gráfica intuitiva y accesible para usuarios no técnicos [5, 14, 16] y la mejora del rendimiento frente a los métodos tradicionales [21, 29]. Sin embargo, persisten limitaciones como el alto coste computacional [4, 22], la falta de transparencia en algunos modelos [16], lo que genera desconfianza, y la necesidad de grandes volúmenes de datos [20, 21]. Por ello, la combinación de la IA con prácticas tradicionales resulta clave para garantizar un análisis de datos más robusto y confiable. También se identificaron carencias en la forma de hacer más comprensible el proceso (explicabilidad) y en adaptar las herramientas a distintos perfiles de usuario, así como en ofrecer interfaces accesibles [23, 30, 31].

En conjunto, esta revisión sistemática destaca el crecimiento y la diversidad de enfoques en el uso de la IA para el preprocesamiento de datos, así como la necesidad de diseñar herramientas más accesibles, comprensibles y ajustadas a las necesidades reales de los usuarios no expertos. Se concluye que la combinación de soluciones basadas en IA con estrategias centradas en la explicabilidad y la participación del usuario representa un camino esencial para democratizar el análisis de datos en entornos donde el conocimiento técnico es limitado.

3. Metodología

Esta sección presenta el enfoque metodológico seguido para diseñar y construir el sistema de preprocesamiento automatizado de datos, concebido como una solución accesible y fiable para usuarios sin formación técnica avanzada. El desarrollo se abordó de manera progresiva: primero se analizaron y seleccionaron las técnicas más adecuadas para cada etapa del flujo de preprocesamiento; luego se construyó un prototipo inicial en Python para validar su eficacia; posteriormente se desarrolló una interfaz web basada en Angular y Flask que permite interactuar fácilmente con el sistema; y, por último, se realizó una evaluación práctica con usuarios reales, incorporando las mejoras resultantes para obtener una versión final robusta y adaptada a las necesidades detectadas.

3.1 Selección y análisis de algoritmos

A partir de la revisión sistemática (Sección 14), se identificaron las técnicas más eficaces y accesibles para cada fase clave del flujo. Todas las técnicas seleccionadas están documentadas y validadas en la literatura, lo que garantiza su idoneidad y fiabilidad para su uso en el sistema propuesto.

Para la **limpieza de datos**, se seleccionaron tres técnicas complementarias que representan enfoques distintos: KNN (estimación local a partir de vecinos más cercanos), Imputación Iterativa (modelos basados en boosting para predecir valores faltantes) y Random Forest (predicción no lineal mediante árboles de decisión). Esta variedad garantiza robustez y adaptabilidad ante diferentes patrones de datos incompletos.

En la **normalización**, se utilizó una combinación de escalado MinMax (rango $[0,1]$) y un autoencoder no supervisado. El escalado asegura compatibilidad con redes neuronales y modelos de aprendizaje profundo, mientras que el autoencoder reconstruye los datos de forma más estable y reduce el ruido presente.

La **selección de características** se basa en modelos de árboles (Random Forest y Extra Trees), que asignan una puntuación de importancia a cada variable. Se aplicó un umbral fijo (0.02) para retener solo las variables con contribución significativa.

Finalmente, para la **evaluación automática de modelos** se adoptó H2O AutoML. Este sistema automatiza la selección de algoritmos y la validación, comparando métricas como AUC para clasificación binaria, R^2 para regresión y error medio por clase para clasificación multiclase. Se convierte el mejor resultado a una escala 0–10 para facilitar su interpretación, brindando transparencia y comprensión incluso a perfiles sin formación técnica.

3.2 Prototipo inicial

Antes de la integración en la interfaz web, se desarrolló un prototipo en Python ejecutable desde terminal. Este prototipo permitió verificar la coherencia y efectividad de la secuencia completa de preprocesamiento, integrando las técnicas seleccionadas y validando la solidez técnica de cada fase.

La Figura 2 ilustra de forma esquemática el flujo de procesamiento implementado en este módulo inicial.

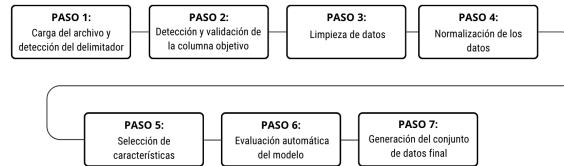


Figura 2: Diagrama del flujo de procesamiento del módulo inicial en Python.

El prototipo facilitó la depuración de cada paso y la identificación de posibles inconsistencias, sirviendo de base para la posterior construcción de la interfaz web. Además, permitió asegurar que el sistema final combinara precisión técnica y facilidad de uso para usuarios no expertos.

3.3 Diseño e implementación del sistema web

Tras validar las técnicas seleccionadas mediante un prototipo inicial en Python, se diseñó un sistema accesible desde la web (Figura 3), orientado a usuarios sin experiencia técnica. La arquitectura cliente-servidor combina un frontend en Angular, con una interfaz clara e interactiva, y un backend en Flask, que orquesta automáticamente el flujo completo: limpieza, normalización, selección de características y evaluación con AutoML.



Figura 3: Pantalla inicial del sistema.

El **frontend** permite a los usuarios cargar sus propios archivos CSV, seleccionar la columna objetivo y visualizar el progreso de cada fase en tiempo real. Incluye explicaciones accesibles, gráficos y botones interactivos, facilitando la comprensión de las técnicas utilizadas y los resultados obtenidos, incluso para usuarios sin formación técnica. La

Figura 4 muestra la selección de la columna objetivo, la Figura 5 la barra de progreso y la secuencia de pasos en tiempo real, y la Figura 6 el resumen final con la puntuación y opciones de descarga.

El **backend** expone una API RESTful que permite al frontend iniciar el preprocesamiento (POST /procesar), consultar el progreso en tiempo real mediante Server-Sent Events (GET /progreso/{id}) y recuperar los resultados o descargar el archivo final (GET /resultado/{id}, GET /descargar/{nombre}). Los eventos SSE envían actualizaciones automáticas y continuas al navegador, evitando la necesidad de recargar la página y permitiendo al usuario ver el progreso de forma dinámica.

Selecciona la columna objetivo

La columna objetivo es la variable que deseas predecir o clasificar utilizando inteligencia artificial. Es crucial seleccionarla correctamente, ya que define el propósito del modelo que se va a entrenar. Hemos sugerido una columna automáticamente (si se detectó alguna), pero puedes cambiarla si prefieres otra.

Sugerida automáticamente: Otherinfo1

Columna objetivo

Otherinfo1

¿Por qué es importante elegir bien la columna objetivo?

- Determina lo que el modelo aprenderá a predecir.
- Un error en esta selección puede generar resultados incorrectos.
- Debe ser coherente con el problema que deseas resolver (por ejemplo, ingresos, clasificaciones, diagnósticos, etc.).

Confirmar y continuar

Figura 4: Selección de la columna objetivo, con explicación visual de su importancia.

Sube tu CSV para comenzar

Examinar... query.output.exome_summary(1).csv

Procesar con IA

Procesando archivo con IA, por favor espera...

- Limpieza de datos completada
- Normalización con autencoder aplicada
- Selección de características realizada
- Iniciando evaluación del CSV preprocesado con modelo Random Forest...

17% completado de la evaluación del CSV preprocesado...

Figura 5: Visualización en tiempo real del progreso del flujo de preprocesamiento.

¡Preprocesamiento Completado!

Tu archivo ha sido preprocesado y evaluado automáticamente.

Rendimiento del Preprocesamiento

10/10

Esta puntuación refleja la calidad del preprocesamiento realizado. Cuanto más alta sea la puntuación, mejor se han preparado tus datos para ser utilizados en inteligencia artificial. El sistema ha entrenado automáticamente un modelo de aprendizaje automático para calcular esta puntuación, analizando la precisión y el rendimiento del modelo entrenado con tus datos preprocesados.

Descargar CSV Procesado Preprocesar Otro Archivo

Ver Más Detalles

Figura 6: Resumen final con puntuación y opciones para descargar o reiniciar.

Además, se habilitó la política de CORS (Cross-Origin Resource Sharing), lo que permite al cliente web (Angular) acceder sin res-

tricciones a los recursos del backend, incluso si se ejecutan en distintos puertos o dominios. Esto asegura un flujo de trabajo transparente y sin bloqueos en entornos de desarrollo o producción.

Gracias a esta arquitectura, el sistema ofrece una experiencia accesible, fluida y técnicamente robusta para usuarios no expertos. Las técnicas de IA y las etapas del preprocesamiento se integran en un flujo visual y comprensible, eliminando barreras y simplificando el uso de herramientas avanzadas desde un simple navegador.

3.4 Despliegue del sistema

El sistema se validó primero en un entorno local, con el frontend en Angular en el puerto 4200 y el backend en Flask en el puerto 5000, probando la interacción y la gestión de sesiones. Posteriormente, se desplegó en producción en <https://smartcsv.grial.eu>, utilizando Nginx como proxy inverso y servidor de archivos estáticos, y Unicorn para ejecutar Flask en múltiples procesos. Esta arquitectura garantiza escalabilidad y disponibilidad, ofreciendo una experiencia de usuario fluida incluso con múltiples usuarios concurrentes. La separación modular facilita el mantenimiento y la ampliación futura del sistema.

3.5 Evaluación

Para validar la utilidad, comprensión y robustez del sistema propuesto, se realizaron dos tipos de evaluaciones complementarias: pruebas técnicas con distintos conjuntos de datos abiertos y una evaluación con usuarios reales. Esta sección detalla ambos enfoques, destacando los hallazgos y la adaptación del sistema a diferentes escenarios y necesidades.

3.5.1 Pruebas técnicas con distintos conjuntos de datos Se realizaron pruebas técnicas con varios datasets de diferentes dominios, disponibles en repositorios abiertos como datos.gob.es y UCI Machine Learning Repository. Estas pruebas permitieron validar la robustez del sistema y su capacidad de adaptarse a flujos de datos diversos, incluyendo problemas de clasificación binaria, clasificación multiclase y regresión.

En resumen, estas pruebas demostraron que el sistema es capaz de adaptarse a diferentes tamaños, dominios y características de los datos, manteniendo la consistencia y ofreciendo resultados confiables.

3.5.2 Evaluación con usuarios Para validar la usabilidad, claridad y utilidad del sistema, se realizó una evaluación cualitativa con dos grupos focales (cinco participantes en total): uno con dos residentes clínicos sin conocimientos técnicos en ciencia de datos y otro con tres

expertos en análisis de datos en el ámbito sanitario. Aunque el sistema está dirigido principalmente a usuarios sin experiencia técnica, los expertos ayudaron a detectar posibles carencias en la transparencia y comprensión de los resultados.

La evaluación consistió en que los usuarios utilizaran el sistema completo y respondieran las siguientes preguntas clave y la Tabla 1 resume las principales respuestas:

1. ¿Te genera confianza que el sistema tome decisiones automáticamente por ti durante el preprocesamiento de datos?
2. ¿Preferirías tener la opción de tomar tú mismo las decisiones en cada paso?
3. ¿Comprendes claramente qué ha hecho el sistema con tus datos?
4. ¿Sientes que falta información importante en el informe de resultados?
5. ¿Echas en falta alguna funcionalidad o característica en el sistema?

Usuario	Confianza en decisiones automáticas	Preferencia por decisiones manuales	Comprensión del preprocesamiento	Falta información en informe	Faltan funcionalidades o mejoras
Residente 1	Valora la automatización por su falta de conocimientos técnicos.	Le gustaría poder forzar la inclusión de algunas columnas.	Entiende la selección de características, pero no algoritmos como KNN o autoencoder.	Echa en falta más información básica sobre los algoritmos.	Sugiere mostrar gráficamente columnas eliminadas y explicar mejor cómo funcionan los algoritmos.
Residente 2	Agradece la automatización, pero desea más resúmenes intermedios.	Quiere más control sobre lo que ocurre en cada etapa.	Identifica las variables seleccionadas, pero no comprende el umbral ni el score final.	Considera confusa la explicación del score.	Propone más explicaciones claras y la posibilidad de revisar el progreso en detalle.
Experto 1	Necesita ver qué se hizo para confiar.	Prefiere revisar y modificar decisiones.	No entiende cómo se aplicó la normalización.	Falta explicación de la normalización y los rangos usados.	Sugiere avisar sobre nulos en la columna objetivo.
Experto 2	Acepta la automatización si se explican bien las fases.	No le interesa ver todos los pasos, pero sí el resultado final.	No entiende cómo se calcula el score ni la representación de las variables.	Falta explicación del score y del error de reconstrucción.	Quiere ver detalladamente qué se hizo en cada fase.
Experto 3	Confía si puede revisar el flujo al final.	Sí, mejor si puede avanzar paso a paso.	Quiere más detalle de las variables seleccionadas.	Confusa la explicación del score.	Quiere entender mejor la gestión de valores nulos y el cálculo del score final.

Tabla 1: Resumen de respuestas a la evaluación con usuarios.

En conjunto, los usuarios valoraron la automatización y la facilidad de uso, pero señalaron áreas clave de mejora relacionadas con la claridad del flujo interno, la explicación de los algoritmos empleados, el significado del score final y la gestión de la variable objetivo cuando presenta valores nulos. Estos resultados ofrecen una base sólida para

continuar refinando el sistema y ajustarlo a las necesidades reales de los usuarios finales.

3.6 Versión final con mejoras aplicadas

Tras la evaluación con usuarios, se rediseñó el sistema para priorizar la claridad y la comprensión de los resultados, especialmente para usuarios sin conocimientos técnicos. Las principales mejoras incluyeron:

- **Gestión de columnas objetivo con nulos:** se implementó un modal de advertencia (Figura 7) que informa al usuario cuando la columna objetivo contiene valores faltantes, ofreciendo tres opciones claras: cambiar de columna, continuar con la actual o subir otro archivo. Esta mejora evita errores en el entrenamiento y fomenta la toma de decisiones informadas.

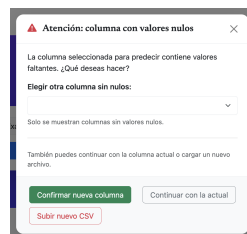


Figura 7: Modal de advertencia cuando la columna objetivo contiene valores nulos.

- **Explicación progresiva del flujo:** se introdujeron botones de “ver más” en cada etapa del flujo, ofreciendo explicaciones visuales y accesibles sobre qué ocurre en cada fase, qué técnicas se usan y por qué. La Figura 8 muestra esta funcionalidad. Por ejemplo, la Figura 9 muestra cómo se explica la limpieza de datos.

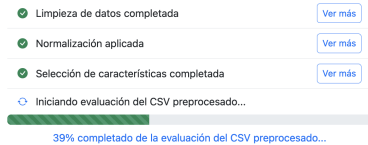


Figura 8: Vista del flujo de pasos con botones de “ver más” para cada etapa.

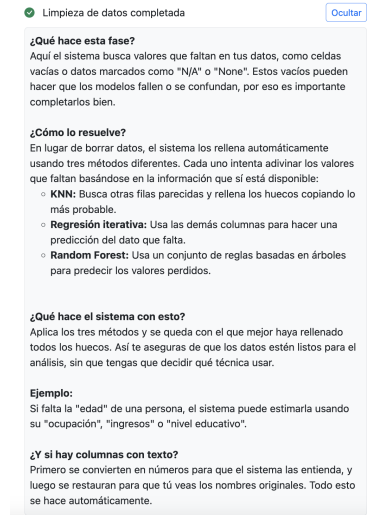


Figura 9: Explicación accesible durante la fase de limpieza de datos.

- Reorganización del informe final:** el modal de resultados se mejoró la presentación de resultados con iconos interactivos e información detallada sobre las variables seleccionadas y los modelos generados. Se muestran comparativas de limpieza, gráficos completos con todas la características seleccionadas, no solo las mejores (Figura 10) y explicaciones claras del score y las métricas empleadas.
- Clarificación del score final:** se eliminó de la vista principal y ahora solo se muestra en el informe detallado, explicado como el rendimiento del mejor modelo de AutoML según la métrica correspondiente (AUC, R^2 o error medio por clase), dependiendo del tipo de problema identificado.

La Figura 11 muestra la nueva vista sin el score en la pantalla principal.



Figura 10: Vista detallada de la selección de características.

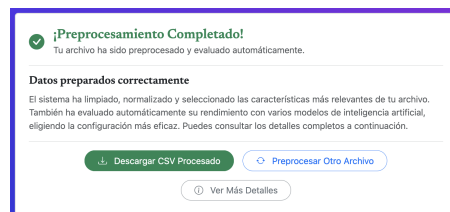


Figura 11: La pantalla principal ya no muestra el score final para evitar confusión.

En conjunto, estas mejoras han permitido consolidar un sistema que mantiene la potencia técnica del preprocesamiento automático, pero con un enfoque más comprensible y accesible para perfiles no técnicos. Gracias a las explicaciones interactivas, la transparencia en las decisiones y la gestión clara de los datos críticos como la columna objetivo, el sistema final ofrece ahora una experiencia de uso más segura, informativa y confiable.

4. Discusión

La versión final del sistema combina un flujo de preprocesamiento automático basado en IA con una interfaz accesible y centrada en usuarios no expertos, validada tanto técnica como prácticamente. Desde la revisión sistemática previa (ver Sección 14), se identificaron necesidades clave que guían este desarrollo: la falta de claridad y transparencia en los procesos de preprocesamiento, que dificulta a los usuarios no expertos entender cómo afectan las decisiones a los datos y resultados [5, 16, 32]; la ausencia de herramientas accesibles que permitan visualizar y explicar estos pasos de forma comprensible [33, 34]; y la carencia de flujos guiados e iterativos que ofrezcan control y opciones al usuario durante el análisis, fomentando su participación y comprensión [7, 23, 35].

Para solventar estos vacíos, el sistema incorpora explicaciones visuales e iconos informativos que describen cada técnica y muestran de forma intuitiva los cambios que se producen en cada fase del flujo [5, 16, 35]. Además, la interfaz gráfica incluye barra de progreso y avisos que guían al usuario paso a paso, con una pausa activa en caso de encontrar valores nulos en la columna objetivo, siguiendo recomendaciones de diseño iterativo y adaptado [7, 14, 22, 23, 36]. Por último, se integra una evaluación automática adaptada al tipo de problema (con métricas como AUC, R^2 , MAE o error medio por clase), así como un sistema de gestión de errores con mensajes claros para el usuario [4, 15, 22, 28, 29].

A nivel técnico, las pruebas demostraron que KNN e Imputación Iterativa lograron completar siempre los valores nulos, mientras que la imputación con Random Forest no lo consiguió de forma consistente. Esto sugiere que podría excluirse esta técnica en futuras versiones para simplificar y reforzar la robustez del flujo. La normalización, basada en MinMax y autoencoder, mostró un comportamiento estable en todos los conjuntos de datos, preservando las variables categóricas. Por otro lado, la selección de características mediante Random Forest y Extra Trees funcionó bien en la mayoría de los casos con el umbral fijo de 0,02, aunque en conjuntos con muchas variables irrelevantes a veces quedaba con pocas características. Una mejora futura podría ser la adopción de umbrales adaptativos para garantizar al menos varias características relevantes. Finalmente, la evaluación automática (AutoML) fue eficaz, alcanzando puntuaciones transformadas entre 7 y 9 sobre 10 en las pruebas con distintos conjuntos de datos. Este rendimiento elevado no solo demuestra la efectividad del modelo final, sino que también refleja la calidad y consistencia del preprocesamiento aplicado previamente.

La validación práctica con los cinco usuarios complementó las pruebas técnicas y permitió confirmar la utilidad, comprensibilidad y accesibilidad del sistema final. Todos completaron el flujo completo y comentaron su experiencia, destacando la claridad de la interfaz y la

reducción del esfuerzo técnico. Los residentes valoraron especialmente poder usarlo sin ayuda externa, mientras que los expertos reclamaron más transparencia y control sobre decisiones automáticas.

Las observaciones recogidas sirvieron para identificar mejoras clave: se añadió una verificación de la columna objetivo si contiene nulos, se incorporaron iconos informativos desplegados para explicar técnicas como KNN o Autoencoder, y se permitió revisar explicaciones de fases anteriores mientras se ejecuta AutoML. También se aclaró el significado de la puntuación final, evitando que se interpretara como una calificación global del archivo. Estas adaptaciones no solo mejoraron la comprensión y confianza en el sistema, sino que reflejan la importancia de incluir a usuarios reales en la validación de sistemas con enfoque humanocéntrico. Como resultado, el sistema final combina automatización robusta con explicaciones accesibles, haciendo el preprocesamiento comprensible y confiable para perfiles no expertos.

En conjunto, los resultados confirman que el sistema logra equilibrar la potencia técnica del preprocesamiento automático con la claridad y accesibilidad requeridas por perfiles no técnicos. Las decisiones de rediseño y las mejoras aplicadas demuestran la importancia de validar con usuarios reales para garantizar herramientas comprensibles, seguras y confiables, reforzando así el compromiso con la democratización de la IA para la preparación de datos.

5. Conclusiones y líneas futuras

Este trabajo ha diseñado, desarrollado y validado un sistema interactivo de preprocesamiento de datos asistido por IA, accesible para usuarios sin conocimientos técnicos avanzados. Se han alcanzado los objetivos definidos (Sección 1), mediante un enfoque sólido y bien fundamentado.

Primero, la revisión sistemática de la literatura permitió identificar técnicas de IA relevantes, buenas prácticas y vacíos metodológicos, cumpliendo el primer objetivo específico (Sección 2). A partir de estos hallazgos, se diseñó e implementó una herramienta funcional que integra imputación, normalización, selección de características y AutoML en una interfaz intuitiva, respondiendo al segundo objetivo (Sección 3).

Finalmente, la validación práctica con usuarios no expertos (técnicos y clínicos) mostró una alta tasa de comprensión, usabilidad y utilidad. Se incorporaron mejoras clave como explicaciones visuales y gestión de errores, que fortalecen la comprensión y confianza de los usuarios (Sección 3).

En conjunto, el sistema desarrollado es una solución viable, accesible y replicable para el preprocesamiento automático de datos, contribuyendo a la democratización de la IA para el análisis avanzado.

De cara a próximas versiones, se identifican varias oportunidades de mejora que pueden reforzar la experiencia de uso y la confianza

de los usuarios. Entre ellas, permitir la modificación manual de decisiones automáticas (por ejemplo, la revisión de las columnas seleccionadas tras la generación del informe) y ofrecer un tratamiento más flexible para columnas con alto porcentaje de nulos, mediante umbrales configurables. Asimismo, se propone dar al usuario la posibilidad de ajustar dinámicamente el umbral de selección de características, garantizando que no se excluyan variables relevantes y manteniendo la calidad del modelo.

Otras posibles líneas de mejora incluyen la introducción de explicaciones adaptadas al perfil del usuario (modo experto o no experto) y la simplificación del módulo de limpieza de datos, eliminando la imputación con Random Forest que mostró menor consistencia. Además, se identificó la necesidad de ampliar la compatibilidad con otros formatos de archivo (por ejemplo, `.data` y `.xlsx`) y desplegar el sistema en contextos reales, para recopilar retroalimentación directa de usuarios finales en entornos prácticos. Estas mejoras fortalecerían la accesibilidad y la comprensión, equilibrando la automatización con la participación del usuario, y consolidando la herramienta como un recurso fiable y adaptable a distintas necesidades.

En resumen, el sistema desarrollado cumple su propósito de acercar técnicas avanzadas de IA al preprocesamiento de datos, haciéndolas comprensibles y accesibles para usuarios no expertos. Combina automatización con explicaciones claras y visuales, demostrando que la IA puede actuar como facilitador y aliado, no como reemplazo, del conocimiento humano. Así, se amplían las posibilidades de análisis de datos en ámbitos donde antes era difícil o inaccesible.

Referencias

1. B. Kitchenham and S. Charters, "Guidelines for performing systematic literature reviews in software engineering," tech. rep., EBSE Technical Report, 2007.
2. M. J. Page, J. E. McKenzie, P. M. Bossuyt, D. Moher, I. A. Boutron, T. D. Turner, L. Shamseer, D. G. Altman, P. T. Higgins, J. M. G. Westwood, *et al.*, "The prisma 2020 statement: an updated guideline for reporting systematic reviews," *BMJ*, vol. 372, p. n71, 2021.
3. Y. Peng, X. Zhang, Y. Li, Q. Su, S. Wang, F. Liu, C. Yu, and M. Liang, "Mvpani: A toolkit with friendly graphical user interface for multivariate pattern analysis of neuroimaging data," *Frontiers in Neuroscience*, vol. 14, p. 576015, 2020.
4. R. Ooms and M. Spruit, "Self-service data science in healthcare with automated machine learning," *Applied Sciences (Switzerland)*, 2020.
5. S. Malkusch, L. Hahnefeld, R. Gurke, and J. Lötsch, "Visually guided preprocessing of bioanalytical laboratory data using an interactive r notebook (pguimp)," *CPT: Pharmacometrics and Systems Pharmacology*, 2021.
6. Y. Song, J. Zhang, Y.-D. Zhang, Y. Hou, X. Yan, Y. Wang, M. Zhou, Y.-F. Yao, and G. Yang, "Feature explorer (fae): A tool for developing and

- comparing radiomics models,” *PLoS ONE*, vol. 15, no. 8, p. e0237587, 2020.
7. J. Gerolami, J. J. M. Wong, R. Zhang, T. Chen, T. Imtiaz, M. Smith, T. Jamaspishvili, M. Koti, J. I. Glasgow, P. Mousavi, N. Renwick, and K. Tyryshkin, “A computational approach to identification of candidate biomarkers in high-dimensional molecular data,” *Diagnostics*, 2022.
 8. J. Y. An, H. Seo, Y.-G. Kim, K. E. Lee, S. Kim, and H.-J. Kong, “Codeless deep learning of covid-19 chest x-ray image dataset with knime analytics platform,” *Healthcare Informatics Research*, 2021.
 9. K.-H. Lai, D. Zha, G. Wang, J. Xu, Y. Zhao, D. Kumar, Y. Chen, P. Zumkhawaka, M. Wan, D. Martinez, and X. Hu, “Tods: An automated time series outlier detection system,” in *35th AAAI Conference on Artificial Intelligence (AAAI 2021)*, 2021.
 10. M. Karatzoglidi, P. Kerasiotis, and V. Kantere, “Automated energy consumption forecasting with enforce,” in *Proceedings of the VLDB Endowment*, 2021.
 11. H. Mendes, S. E. Quincozes, and V. E. Quincozes, “A web user interface tool for metaheuristics-based feature selection assessment for idss,” in *2022 6th Cyber Security in Networking Conference (CSNet 2022)*, 2022.
 12. N. McCombe, X. Ding, G. Prasad, P. Gillespie, D. P. Finn, S. Todd, P. L. McClean, and K. Wong-Lin, “Alzheimer’s disease assessments optimized for diagnostic accuracy and administration time,” *IEEE Journal of Translational Engineering in Health and Medicine*, 2022.
 13. F.-T.-Z. Khanam, A. Al-Naji, A. G. Perera, D. Wang, and J. Chahl, “Non-invasive and non-contact automatic jaundice detection of infants based on random forest,” *Computer Methods in Biomechanics and Biomedical Engineering: Imaging and Visualization*, 2023.
 14. M. Harth, C. Schorr, and R. Krieger, “A hierarchical multi-level product classification workbench for retail,” in *CEUR Workshop Proceedings*, 2020.
 15. F. Assunção, N. Lourenço, B. Ribeiro, and P. Machado, “Evolution of scikit-learn pipelines with dynamic structured grammatical evolution,” in *Lecture Notes in Computer Science*, Springer, 2020.
 16. M. Ma, C. Chen, D. Chen, H. Zhang, X. Du, Q. Sun, L. Fan, H. Kong, X. Chen, C. Cao, and X. Wan, “A machine learning-based prediction model for acute kidney injury in patients with community-acquired pneumonia: Multicenter validation study,” *Journal of Medical Internet Research*, 2024.
 17. N. Nanthini, D. Aishwarya, A. Simon, N. Baby Vishnupriya, and K. Jeyalakshmi, “A novel approach for prediction of the lung disease using deep learning,” in *Proceedings - 2024 8th International Conference on Inventive Systems and Control, ICISC 2024*, 2024.
 18. P. Washington, Q. Tariq, E. Leblanc, B. Chrisman, K. Dunlap, A. Kline, H. Kalantarian, Y. Penev, K. Paskov, C. Voss, N. Stockham, M. Varma, A. Husic, J. Kent, N. Haber, T. Winograd, and D. P. Wall, “Crowdsourced privacy-preserved feature tagging of short home videos for machine learning asd detection,” *Scientific Reports*, 2021.
 19. X. Wu, F. Zhai, A. Chang, J. Wei, Y. Guo, and J. Zhang, “Application of machine learning algorithms to predict osteoporosis in postmenopausal women with type 2 diabetes mellitus,” *Journal of Endocrinological Investigation*, 2023.

20. E. Yuzer, V. Dogan, V. Kilic, and M. Sen, "Smartphone embedded deep learning approach for highly accurate and automated colorimetric lactate analysis in sweat," *Sensors and Actuators B: Chemical*, 2022.
21. I. Shafi, M. Sajad, A. Fatima, D. G. Aray, V. Lipari, I. D. L. T. Diez, and I. Ashraf, "Teeth lesion detection using deep learning and the internet of things post-covid-19," *Sensors*, 2023.
22. M.-B. Bindilă and M. Negru, "End-to-end automated machine learning system for supervised learning problems," in *Proceedings of the 2021 IEEE 17th International Conference on Intelligent Computer Communication and Processing (ICCP 2021)*, 2021.
23. T. Petricek, G. J. J. V. D. Burg, A. Nazabal, T. Ceritli, E. Jimenez-Ruiz, and C. K. I. Williams, "Ai assistants: A framework for semi-automated data wrangling," *IEEE Transactions on Knowledge and Data Engineering*, 2023.
24. R. Subha, B. R. Nayana, R. Radhakrishnan, and P. Sumalatha, "Computational intelligence for early detection of infertility in women," *Engineering Applications of Artificial Intelligence*, 2024.
25. N. McCombe, X. Ding, G. Prasad, D. P. Finn, S. Todd, P. L. McClean, and K. Wong-Lin, "Multiple cost optimisation for alzheimer's disease diagnosis," in *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society, EMBS*, 2022.
26. E. Afgan, A. Nekrutenko, B. A. Grünig, D. Blankenberg, J. Goecks, M. C. Schatz, A. E. Ostrovsky, A. Mahmoud, A. J. Lonie, A. Syme, ..., and P. J. Briggs, "The galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2022 update," *Nucleic Acids Research*, 2022.
27. M. N. Ibrahim, R. Blazquez-Garcia, A. Lightstone, F. Meng, M. Bhat, A. El Kaffas, and E. Ukwatta, "Automated fatty liver disease detection in point-of-care ultrasound b-mode images," *Journal of Medical Imaging*, 2023.
28. S. Liang, Z. Fei, J. Wu, and X. Lin, "Tree-based pipeline optimization-based automated-machine learning model for performance prediction of materials and structures: Case studies and ui design," *Structural Control & Health Monitoring*, 2024.
29. R. Shukla and T. R. Singh, "Alzgenpred - catboost-based gene classifier for predicting alzheimer's disease using high-throughput sequencing data," *Scientific Reports*, 2024.
30. L. M. Matos, J. Azevedo, A. Matta, A. Pilastrri, P. Cortez, and R. Mendes, "Categorical attribute transformation environment (cane): A python module for categorical to numeric data preprocessing," *Software Impacts*, 2022.
31. Y. Zhong, Y. Dan, Y. Cai, J. Lin, X. Huang, O. Mahmoud, E. S. Hald, A. Kumar, Q. Fang, and S. S. Mahmoud, "Efficient malaria parasite detection from diverse images of thick blood smears for cross-regional model accuracy," *IEEE Open Journal of Engineering in Medicine and Biology*, 2023.
32. J. K. Dewangan, H. Patel, A. Sinha, and D. Das, "Graphical user interface aided intelligent diagnosis of stator faults in induction motors," in *INDICON 2024 - 5th IEEE India Council International Subsections Conference: Science, Technology and Society*, 2024.

33. E. A. Rodríguez Martínez, O. Polezhaeva, F. Marcellin, E. Colin, L. Boyaval, F.-R. Sarhan, and S. Dakpé, “Deepsmile: Anomaly detection software for facial movement assessment,” *Diagnostics*, 2023.
34. E. Serin, N. Vaidya, H. Walter, and J. D. Kruschwitz, “Nbs-predict: An easy-to-use toolbox for connectome-based machine learning,” in *Neuro-methods*, Humana Press, 2025.
35. G. Zhou, V. Aggarwal, M. Yin, and D. Yu, “Video-based ai decision support system for lifting risk assessment,” in *2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, 2021.
36. N. Uddin, M. K. Uddin Ahamed, M. A. Uddin, M. M. Islam, M. A. Talukder, and S. Aryal, “An ensemble machine learning based bank loan approval predictions system with a smart application,” *International Journal of Cognitive Computing in Engineering*, 2023.

Sistemas de IDS/IPS en redes utilizando deep learning

Óscar de las Heras García¹[0009-0004-8006-7009], Angélica González
Arrieta¹[0000-0002-4726-7103], and Álvaro
Bustos-Tabernero¹[0000-0003-4215-8933]

Universidad de Salamanca
Departamento de Informática y Automática
Facultad de Ciencias.
Plaza de los Caídos s/n
37008 Salamanca, España
{mosky,angelica,alvarob97}@usal.es



Resumen La fase de reconocimiento es crítica en cualquier ciberataque ya que permite a un atacante detectar posibles vulnerabilidades, por lo que anticiparse a estas actividades pronto es vital para prevenir intrusiones. En este trabajo, se ha propuesto e implementado un sistema de detección y prevención de intrusiones basado en aprendizaje profundo, orientado a identificar escaneos activos, los cuales son un indicio temprano de intrusión. El sistema cubre el ciclo completo: desde la construcción del laboratorio en GNS3, la generación y etiquetado de tráfico, preprocesamiento, entrenamiento de un modelo MLP y despliegue en tiempo real.

A diferencia de otros IDS/IPS tradicionales, este analiza el comportamiento del tráfico, lo que permite detectar ataques desconocidos y patrones anómalos con mayor precisión. El sistema alcanzó una precisión superior al 98 % en un entorno controlado. Además, para que fuera un sistema completo, se incluyen mecanismos de respuesta automática aplicando reglas iptables y se integra Wazuh como plataforma SIEM para monitorizar las alertas con la matriz MITRE ATT&CK. Por último, este diseño permite extender el sistema hacia fases más avanzadas de la secuencia de ataque, facilitando su evolución hacia una solución más integral de detección basada en comportamiento.

Keywords: Reconocimiento activo · Aprendizaje profundo · Análisis de comportamiento · Sistema de detección y prevención de intrusiones

1. Introducción

La detección temprana de amenazas en redes es una prioridad en la ciberseguridad, especialmente por el crecimiento del tráfico en Internet y la aparición de nuevas técnicas de ataque. En este contexto, el proceso de identificar actividades sospechosas antes de que lleguen a comprometer los sistemas es algo crítico. Por ello, los sistemas de detección y prevención de intrusiones (IDS/IPS) son muy decisivos, pues permiten identificar y mitigar comportamientos maliciosos. Sin embargo, muchos de estos siguen basándose en firmas o reglas estáticas, lo que aumenta su fragilidad frente a ataques desconocidos o patrones anómalos [1].

Además, con el aumento del volumen y complejidad del tráfico, técnicas de evasión como el cifrado o la ofuscación, dificultan cada vez más la capacidad de los IDS para adaptarse y reconocer tráfico malicioso que aparenta ser legítimo [2]. Esto ha impulsado el desarrollo de enfoques alternativos, como los modelos basados en aprendizaje profundo.

El objetivo principal de este trabajo es diseñar e implementar un IDS/IPS basado en aprendizaje profundo, capaz de reconocer patrones de comportamiento de la técnica T1595 de la matriz MITRE ATT&CK, evitando el uso de atributos como direcciones IP o puertos. Para ello, se ha creado un conjunto de datos realista en un entorno de laboratorio controlado con GNS3. A partir de este dataset, se entrenó un modelo MLP que se integra en un sistema funcional con Zeek y Suricata como sensores de red, Kafka como middleware y Wazuh como plataforma SIEM para la monitorización de alertas y correlación de eventos. Este enfoque no solo mejora la precisión, sino que permite respuestas automáticas y extensibilidad hacia fases más avanzadas de la secuencia de ataque.

2. Revisión de la literatura

La detección de intrusiones en redes constituye un área en continua exploración dentro de la ciberseguridad, con enfoques que han progresado desde sistemas basados en firmas hasta modelos que emplean aprendizaje profundo. En primer lugar, se presentan los fundamentos de los sistemas IDS, definiendo sus tipos y el flujo típico de soluciones basadas en aprendizaje profundo. A continuación, se analiza la fase de reconocimiento en la secuencia de ataque, destacando la técnica T1595. Finalmente, se revisan las arquitecturas de aprendizaje profundo aplicadas a la detección de amenazas en redes.

2.1 Ciberseguridad y detección de intrusiones

Los sistemas de detección y prevención de intrusiones (IDS/IPS) son clave en la protección de redes, ya que permiten identificar

comportamientos maliciosos antes de que comprometan la infraestructura [5]. Desde sus inicios con Anderson en los años 70, estos sistemas han evolucionado hacia soluciones más sofisticadas, incorporando técnicas de aprendizaje automático y profundo [3].

Entre sus variantes destacan los NIDS (Network Intrusion Detection System), que analizan tráfico de red, y los HIDS (Host-based Intrusion Detection System), centrados en eventos internos del sistema. Recientemente, los DL-IDS (Deep Learning-based Intrusion Detection System), han ganado relevancia al aplicar modelos de aprendizaje profundo sobre flujos de red, lo que les permite detectar amenazas incluso desconocidas [4].

El flujo de trabajo típico de un DL-IDS tiene varias etapas: (1) recolección de datos en bruto, como el tráfico de red; (2) almacenamiento eficiente; (3) parsing o estructuración de los datos; (4) resumen de la información para preservar los eventos críticos; (5) detección de patrones anómalos mediante modelos entrenados; y (6) investigación de las alertas generadas para comprender el alcance del ataque y tomar decisiones de respuesta o mitigación [4].

2.2 Reconocimiento activo en MITRE ATT&CK

La mayoría de los ciberataques siguen una secuencia estructurada formada por varias fases, desde el reconocimiento inicial hasta la exfiltración de datos. Detectar las etapas iniciales, como el escaneo activo, es fundamental para detectar a un atacante y detener su progresión. Modelos como la Cyber Kill Chain [6] o la matriz MITRE ATT&CK [7] han estandarizado estas fases, facilitando su análisis y respuesta.

MITRE ATT&CK organiza el conocimiento sobre amenazas en forma de Tácticas, Técnicas y Procedimientos (TTPs). Las tácticas reflejan el objetivo del atacante, las técnicas describen cómo se logra y los procedimientos detallan su implementación concreta. Asociar las alertas a técnicas específicas, como *Active Scanning* (T1595) dentro de la táctica de *Reconnaissance* (TA0043), ayuda a interpretar el comportamiento del atacante y anticiparse a su siguiente movimiento [8,9].

El escaneo activo es una de las fases iniciales más frecuentes y detectables dentro de la secuencia de ataque. Técnicas como T1595 *Active Scanning*, T1046 *Network Service Discovery* o T1018 *Remote System Discovery* suelen producir mucho ruido, lo que las convierte en un objetivo ideal para los IDS/IPS. Detectarlas a tiempo permite bloquear o ralentizar a un atacante, reduciendo de manera drástica la probabilidad de que un ciberataque evolucione a fases más destructivas como la explotación [10]. Sin embargo, aún son pocas las soluciones que mapean con precisión los eventos de red a técnicas del MITRE ATT&CK, lo que refuerza la importancia de modelos orientados a TTPs como el que se propone en este trabajo [11].

2.3 Aprendizaje profundo aplicado a seguridad de red

Los modelos de aprendizaje profundo (DL) han demostrado una capacidad superior a los algoritmos clásicos de aprendizaje automático (ML) en tareas de detección de intrusiones. Mientras que enfoques como *Random Tree* o *Naive Bayes* presentan buen rendimiento con bajo coste computacional, las redes neuronales profundas destacan por su habilidad para modelar relaciones complejas y adaptarse a entornos cambiantes, alcanzando precisiones de entre 97 % y 98 % en ciertos estudios [12,13].

Además, su capacidad para detectar anomalías permite identificar ataques desconocidos con menores tasas de falsos positivos, una ventaja clave frente al creciente volumen y sofisticación del tráfico de red [14, 15]. Entre las arquitecturas más comunes en DL-IDS se encuentran las redes multicapa (MLP), las convolucionales (CNN) y las recurrentes (LSTM), cada una con ventajas particulares según el tipo de datos y contexto [16,17].

2.4 Conclusión de la revisión

En los últimos años, se han propuesto diversas soluciones para la detección de intrusiones en redes, avanzando notablemente gracias a la integración de modelos de aprendizaje profundo y al uso de estándares como MITRE ATT&CK. Sin embargo, muchos de estos trabajos se centran en mecanismos de detección basados en firmas estáticas, lo que limita su capacidad para adaptarse a nuevas amenazas. Esta investigación propone un enfoque orientado al comportamiento, centrado en la técnica T1595 *Active Scanning* de la matriz MITRE ATT&CK, sin depender de atributos como direcciones IP o puertos específicos, permitiendo así una detección más flexible y generalizable.

3. Métodos

En esta sección se describe la metodología utilizada para establecer la base del sistema IDS/IPS propuesto. Se explica en detalle el diseño y topología del entorno de laboratorio, la generación de tráfico de red para construir un conjunto de datos realista y el entrenamiento del Perceptrón Multicapa (MLP) utilizado.

3.1 Diseño y topología del entorno de laboratorio

Para generar de manera controlada todo el tráfico de red se diseñó un entorno virtualizado con GNS3, que ha permitido simular una red realista compuesta por diversos nodos, como se puede ver en la Fig. 1. En consecuencia, se ha conseguido un control total sobre el flujo del tráfico y la posibilidad de capturar paquetes en diferentes segmentos de red. La topología está formada por los siguientes nodos:

- Kali Linux (10.1.1.10) utilizado como nodo atacante.
- Metasploitable2 (10.1.3.30) como primera máquina víctima.
- Windows 10 (10.1.2.20) como segunda máquina víctima.
- IDS/IPS (10.1.4.40 / 20.1.1.10) Ubuntu con el sistema IDS/IPS propuesto.
- Routers Cisco R1 y R2 permiten segmentar el tráfico.
- Víctima (20.1.2.20) máquina objetivo final.

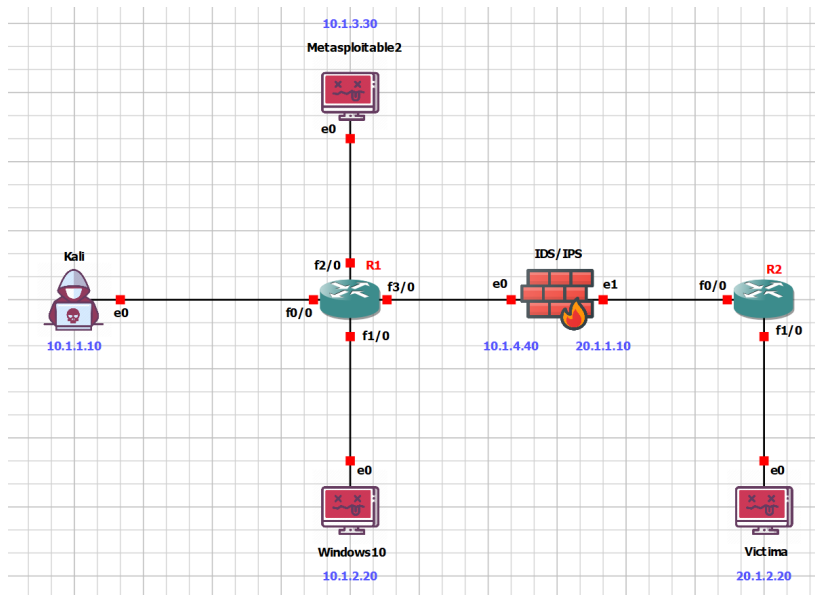


Figura 1: Topología del entorno de laboratorio implementado en GNS3.

La red se divide en dos segmentos: una red externa, formada por los nodos Kali Linux, para lanzar los ataques, junto con Metasploitable2 y Windows10 como máquinas objetivo, interconectadas mediante un router R1; y una red interna a proteger, conectada a través de un segundo router R2 y con el sistema IDS/IPS ubicado como pasarela entre ambas redes. Este sistema, funciona sobre una máquina Ubuntu y permite capturar los flujos del tráfico completos entre los segmentos 10.1.x.z y 20.1.x.z, lo que ha permitido evaluar la capacidad del sistema propuesto a la hora de detectar amenazas.

3.2 Generación del conjunto de datos

El conjunto de datos se generó de forma controlada en el entorno de laboratorio. Para ello, se simuló tráfico malicioso mediante la

técnica T1595 del MITRE ATT&CK con Nmap y tráfico legítimo utilizando servicios como FTP, SSH, HTTP, SMB, RDP, ICMP y DNS. Los paquetes se capturaron en formato .pcap desde GNS3 a través de Wireshark. Posteriormente, se realizó la extracción y el etiquetado de los flujos utilizando Zeek, pues permite enriquecerlos con las flags TCP y el estado de conexión, procesando de esta manera las capturas y etiquetando los flujos según su categoría de ataque o normal.

Tabla 1: Número de flujos por tipo de tráfico.

Tipo de tráfico (Attack_type)	Número de flujos
nmap_ack	232524
nmap_connect	158678
nmap_fin	215995
nmap_null	229679
nmap_syn	220834
nmap_xmas	227094
normal_dns	650
normal_ftp	1532
normal_http	4079
normal_icmp	392
normal_rdp	2820
normal_scp	17
normal_smb	185
normal_ssh	3376

Como se puede observar en el Cua. 1, el número de flujos generados para tráfico normal es significativamente menor que el de los ataques. Esto es debido a que la generación de tráfico legítimo requiere realizar interacciones auténticas que simulen comportamientos de usuario. Por el contrario, los escaneos pueden automatizarse fácilmente y generar una gran cantidad de flujos. Esta diferencia justifica el desequilibrio en el número total de muestras por clase y se abordará más adelante mediante técnicas de balanceo y estratificación.

3.3 Preprocesado

Una vez etiquetado el tráfico capturado, los conjuntos de datos de entrenamiento y prueba, definidos independientemente, fueron preprocesados para asegurar la calidad. Las partes más importantes de este proceso fueron:

1. Eliminar cualquier flujo con campos esenciales vacíos (SrcAddr, DstAddr, Sport, Dport, Proto). De esta manera, se garantiza la integridad y coherencia del conjunto de datos, dados que estos campos

son imprescindibles para representar de forma correcta un flujo de red.

2. Muestrear de manera estratificada el conjunto de ataques con el objetivo de mantener la diversidad interna del tráfico malicioso y mejorar la capacidad de generalización del modelo. Además, esto permite reducir el sobreajuste a patrones más frecuentes.
3. Muestrear de manera proporcional el conjunto de tráfico normal para preservar la distribución natural del tráfico y mantener el conjunto lo más realista posible.
4. Derivar métricas útiles con la finalidad de enriquecer el conjunto de datos (ingeniería de características), pues permiten capturar comportamientos que no son directamente observables con los campos en bruto (por ejemplo el campo `history`), facilitando el aprendizaje de patrones más complejos.

3.3.1 Conjunto de entrenamiento Para el entrenamiento del modelo, se utilizó un conjunto de datos realista y variado para que el este fuese capaz de generalizar. A continuación, en las Fig. 2 y Fig. 3 se observa la distribución por clase y por subtipo de tráfico para el conjunto de entrenamiento.

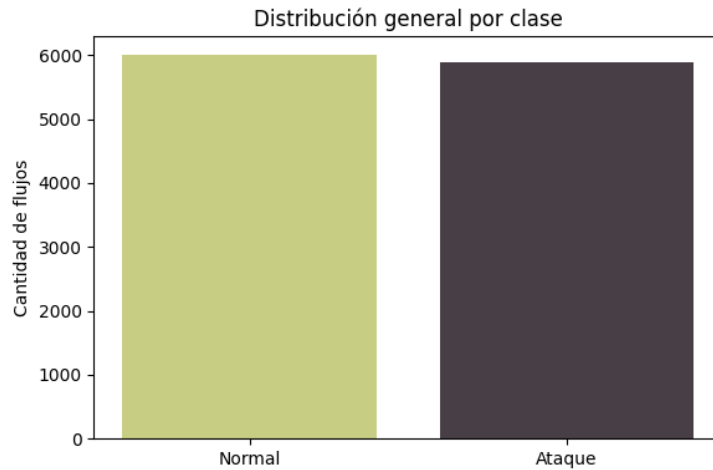


Figura 2: Distribución por clase en el conjunto de entrenamiento.

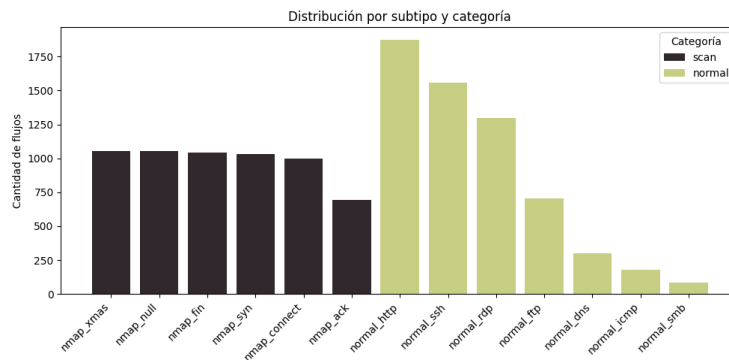


Figura 3: Distribución por subtipo de tráfico y categoría en el conjunto de entrenamiento.

3.3.2 Conjunto de prueba Para la evaluación del modelo, se empleó un conjunto de prueba con nuevos flujos siguiendo el mismo proceso de etiquetado y preprocesamiento, pero con menor volumen. A continua-

ción, en las Fig. 4 y Fig. 5 se observa la distribución por clase y por subtipo de tráfico para el conjunto de prueba.

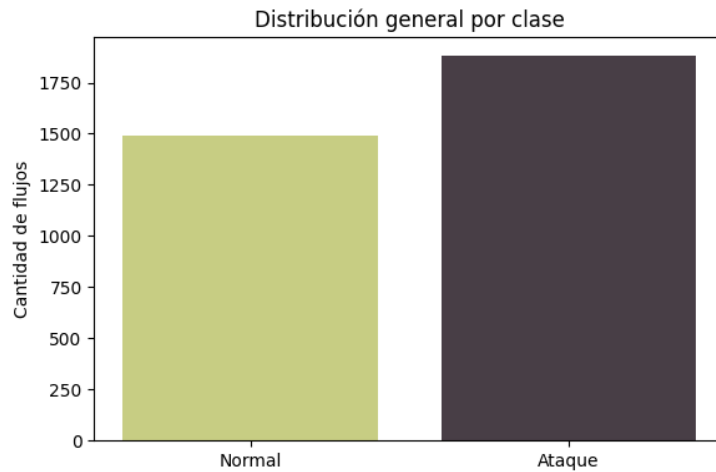


Figura 4: Distribución por clase en el conjunto de prueba.

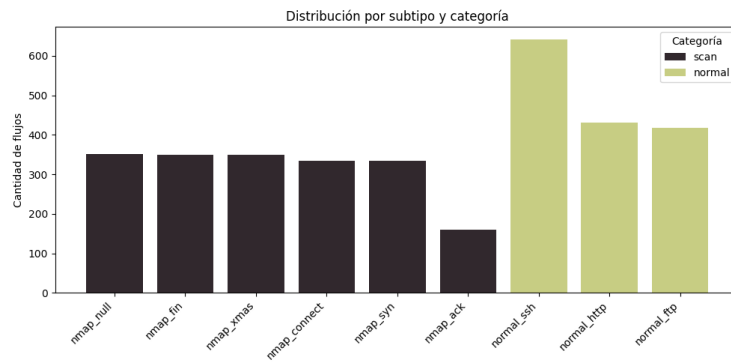


Figura 5: Distribución por subtipo de tráfico y categoría en el conjunto de prueba.

3.4 Características del dataset

Las características extraídas del tráfico representan tanto las propiedades cuantitativas del flujo como información categórica codificada. A continuación, en el Cua. 2 se recogen las métricas generales y flags, mientras que en el Cua. 3 se incluyen las variables codificadas

mediante one-hot encoding, tanto para el protocolo como para el estado de la conexión según Zeek.

3.5 Entrenamiento y validación

Una vez preprocesado el conjunto de datos, se eligió un Perceptrón Multicapa (MLP) para clasificar el tráfico como normal o scan. Este modelo se seleccionó por su eficiencia y ligereza, lo que lo hace ideal para la implementación futura en el sistema IDS/IPS.

Para el entrenamiento, se eliminaron las columnas irrelevantes del conjunto de datos y las etiquetas de salida se codificaron en formato one-hot. Las características se dividieron en tres tipos (numéricas, flags TCP binarios y categóricas one-hot) y se preprocesaron usando un ColumnTransformer para evitar escalar columnas binarias o categóricas, pues podría dañar la interpretación del modelo, por consecuencia, que no generalice bien en el despliegue. A continuación, se observa un fragmento de código del ColumnTransformer:

```

1 # ColumnTransformer separado por tipo
2 preprocessor = ColumnTransformer(
3     transformers=[
4         ("num", Pipeline([
5             ("imputer", SimpleImputer(missing_values
6                 =-1, strategy="mean")),
7             ("scaler", StandardScaler())
8         ]), numeric_real),
9         ("flags", "passthrough", flags),
10        ("cat", "passthrough", categorical)
11    ]
12 )

```

La validación del modelo se realizó mediante validación cruzada estratificada de 5 folds para evaluar su estabilidad. Se configuró un MLP con dos capas ocultas, incorporando técnicas como regularización L2, *Dropout* y *BatchNormalization*. Se usó el optimizador Adam y *callbacks* como *EarlyStopping* y *ReduceLROnPlateau* para optimizar el entrenamiento.

Tras la validación, se analizó la importancia relativa de las características observando los pesos absolutos de la primera capa del modelo final. Gracias a esto, se puede identificar qué variables tiene un impacto mayor en la toma de decisiones del modelo.

A continuación, en la Fig. 6 se puede visualizar mediante un gráfico de barras ordenado por importancia que las características más importantes son aquellas relacionadas con el tamaño medio de los paquetes (sMeanPktSz o dMeanPktSz) y las relaciones entre los bytes/paquetes enviados y recibidos (PktRatio o ByteRatio), revelando que la estructura y distribución del tráfico son claves en la detección.

Tabla 2: Características generales del dataset: métricas de flujo y flags TCP.

Característica	Significado
SrcAddr	Dirección IP de origen del flujo
DstAddr	Dirección IP de destino del flujo
Sport	Puerto de origen
Dport	Puerto de destino
Dur	Duración total del flujo en segundos
TotPkts	Total de paquetes en el flujo
TotBytes	Total de bytes en el flujo
SrcBytes	Bytes enviados por el host de origen
DstBytes	Bytes enviados por el host de destino
SrcPkts	Paquetes enviados por el host de origen
DstPkts	Paquetes enviados por el host de destino
Rate	Tasa de transferencia total (bytes por segundo)
SrcRate	Tasa de envío desde el origen (bytes/segundo)
DstRate	Tasa de recepción en destino (bytes/segundo)
Load	Carga total de transferencia (Rate / Dur)
SrcLoad	Carga de envío del origen ($\text{SrcRate} / \text{Dur}$)
sMeanPktSz	Tamaño medio de los paquetes enviados
dMeanPktSz	Tamaño medio de los paquetes recibidos
ByteRatio	Relación entre bytes enviados y recibidos
PktRatio	Relación entre paquetes enviados y recibidos
PktsPerSec	Paquetes por segundo durante el flujo
BytesPerSec	Bytes por segundo durante el flujo
has_SYN	Presencia del flag TCP SYN
has_RST	Presencia del flag TCP RST
has_FIN	Presencia del flag TCP FIN
has_ACK	Presencia del flag TCP ACK
has_PSH	Presencia del flag TCP PSH
has_URG	Presencia del flag TCP URG
has_ECE	Presencia del flag TCP ECE
has_CWR	Presencia del flag TCP CWR
attack_type	Subtipo específico de ataque o tráfico normal
attack_category	Categoría general del flujo (normal, scan)

Tabla 3: Características categóricas codificadas mediante one-hot encoding.

Característica	Significado
Proto_icmp	El flujo utiliza el protocolo ICMP
Proto_tcp	El flujo utiliza el protocolo TCP
Proto_udp	El flujo utiliza el protocolo UDP
State_SF	Conexión establecida y finalizada correctamente
State_S0	Intento de conexión observado, sin respuesta del destino
State_S1	Conexión establecida, pero no finalizada (sin FIN/RST)
State_S3	Conexión establecida, el destino intentó cerrarla pero sin respuesta
State_REJ	Intento de conexión rechazado por el destino
State_RSTO	Conexión establecida, el origen abortó mediante RST
State_RSTR	El destino respondió con un RST
State_RSTRH	El destino envió SYN-ACK seguido de RST, sin haber visto SYN
State_OTH	No se observó SYN, solo tráfico intermedio
State_SH	El origen envió SYN seguido de FIN, sin recibir SYN-ACK
State_SHR	El destino envió SYN-ACK seguido de FIN, sin haber recibido SYN

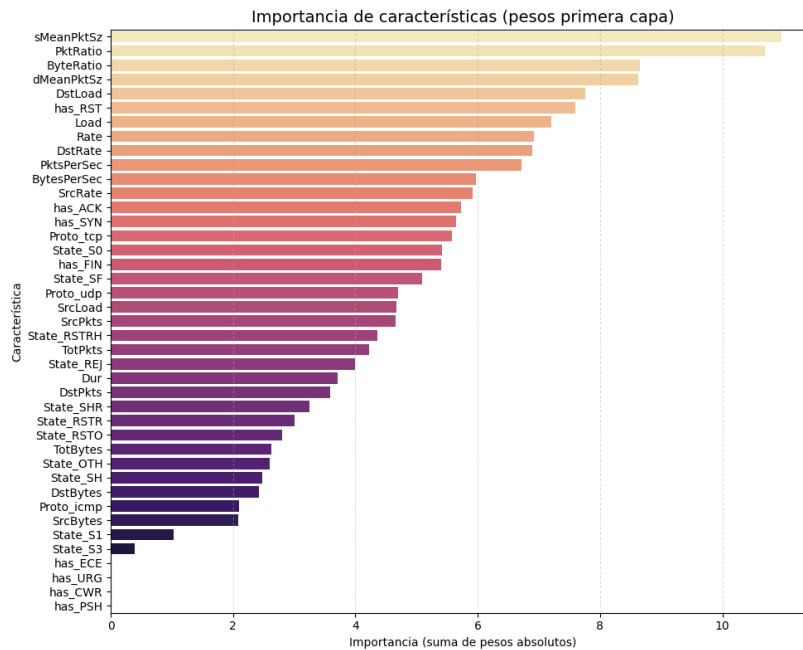


Figura 6: Importancia relativa primera capa.

4. Sistema propuesto

El sistema IDS/IPS propuesto está diseñado para la detección y prevención de intrusiones basada en el comportamiento del tráfico de red. Su arquitectura, detallada en la Fig. 7, garantiza una clara separación de responsabilidades, lo que facilita su despliegue, mantenimiento y escalabilidad.

Como se observa en la Fig. 7, el sistema funciona sobre una base de Apache Kafka como middleware, permitiendo el procesamiento de grandes volúmenes de datos en tiempo real. Zeek y Suricata actúan como sensores de red, generando eventos de tráfico que son recolectados y enviados a Kafka mediante Fluent Bit.

Un script desarrollado (`myVerbraucher.py`) consume estos eventos de Kafka, combinando los datos de ambos sensores y extrayendo las características necesarias. Posteriormente, aplica el modelo MLP previamente entrenado para clasificar el tráfico como normal o scan.

Basándose en la predicción del modelo, el sistema puede iniciar mecanismos de respuesta automáticos aplicando reglas iptables a la dirección IP de origen para bloquear tráfico malicioso. Adicionalmente, se generan alertas específicas que son volcadas a un archivo de logs y posteriormente integradas en Wazuh, una plataforma SIEM, para la visualización y monitorización de eventos, contextualizando las alertas bajo la matriz MITRE ATT&CK.

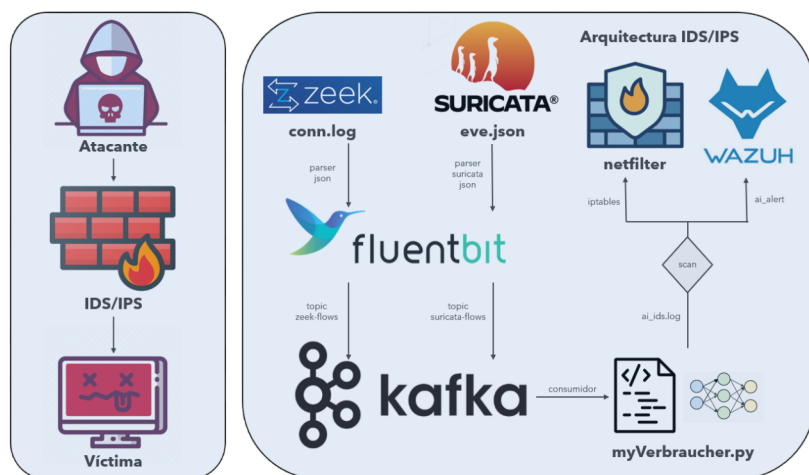


Figura 7: Arquitectura general del sistema IDS/IPS propuesto.

5. Resultados

Los resultados del sistema IDS/IPS se presentan tras evaluar dos modelos MLP distintos sobre un conjunto de datos de prueba. La evaluación se centró en métricas estándar como *Accuracy*, *Precision*, *Recall* y *F1-score*, calculando sus variantes *macro* y *weighted* para una visión completa, especialmente dada la naturaleza desbalanceada de las clases. En el contexto de un IDS/IPS, la capacidad de detectar correctamente todos los ataques es crítica, por lo que se puso especial atención en el recall. Además, para facilitar el análisis y tener una evaluación más clara del rendimiento por tipo de ataque, se ha agrupado el tráfico legítimo bajo una sola clase (normal).

5.1 Métricas de evaluación

Para la evaluación de los modelos, se emplearon las siguientes métricas clave en problemas de clasificación:

Accuracy representa el porcentaje total de flujos clasificados de manera correcta.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Donde TP son Verdaderos Positivos, TN Verdaderos Negativos, FP Falsos Positivos y FN Falsos Negativos.

F1 macro calcula la media entre precisión y recall para cada clase por separado y luego promedia los resultados (útil para clases desbalanceadas como mi caso).

$$\text{F1}_{\text{macro}} = \frac{1}{N} \sum_{i=1}^N \frac{2 \cdot \text{Precision}_i \cdot \text{Recall}_i}{\text{Precision}_i + \text{Recall}_i}$$

F1 weighted calcula lo mismo que F1 macro pero ponderando según el número de muestras de cada clase (tiene en cuenta la distribución real del dataset).

$$\text{F1}_{\text{weighted}} = \sum_{i=1}^N \frac{n_i}{n} \cdot \text{F1}_i$$

donde n_i es el número de muestras de la clase i , y n el total de muestras.

Recall macro es la media del recall (verdaderos positivos) de todas las clases.

$$\text{Recall}_{\text{macro}} = \frac{1}{N} \sum_{i=1}^N \frac{TP_i}{TP_i + FN_i}$$

Precisión macro es la media de la precisión (porcentaje de verdaderos positivos entre las predicciones positivas) para todas las clases.

$$\text{Precisión}_{\text{macro}} = \frac{1}{N} \sum_{i=1}^N \frac{TP_i}{TP_i + FP_i}$$

Clasificación por attack_category El modelo de clasificación binaria demostró un rendimiento muy bueno, con una precisión general muy alta, alcanzando una *Accuracy* de 98 %, así como altos valores para *F1 macro* (0.9804) y *Recall macro* (0.9792), como se detalla en el Cua. 4. En la figura Fig. 8 se observa la matriz de confusión absoluta. Esta matriz muestra que de un total de 1491 instancias normales, 1440 fueron clasificadas de manera correcta, mientras que solo 51 se predijeron erróneamente como *scan* (falsos negativos). Asimismo, de 1880 instancias de *scan*, 1866 fueron identificadas de manera correcta, con solo 14 instancias clasificadas erróneamente como *normal* (falsos positivos).

Tabla 4: Métricas de evaluación para el modelo con etiqueta `attack_category`.

Métrica	Valor
Accuracy	0.98
F1 macro	0.9804
F1 weighted	0.9807
Recall macro	0.9792
Precisión macro	0.9812

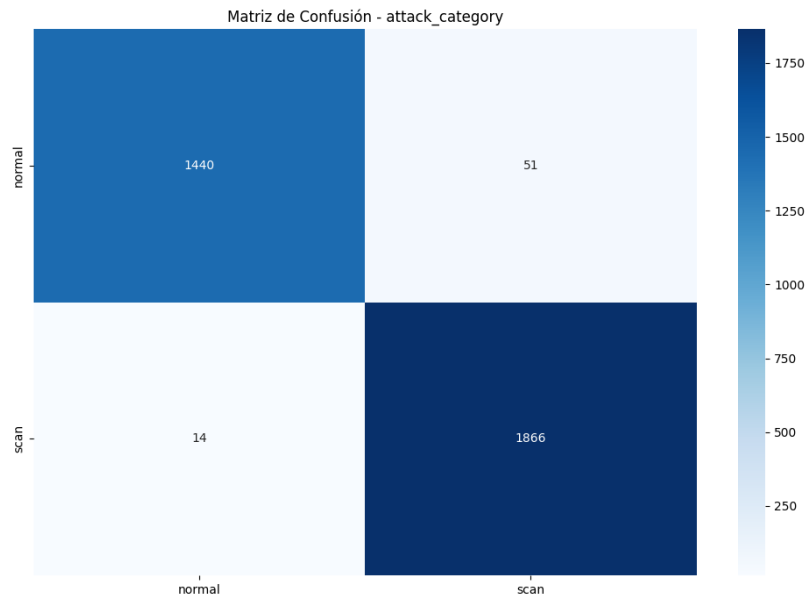


Figura 8: Matriz de confusión - *attack_category*.

Clasificación por *attack_type* Aunque el problema de clasificación por tipo de ataque es más granular, el modelo obtuvo resultados aceptables, como se observa en el Cua. 5. Además, en la figura Fig. 9 se observa la matriz de confusión absoluta. Esta revela que, si bien la mayoría de los tipos de ataque fueron clasificados correctamente, existe una dificultad particular a la hora de diferenciar ciertos sondeos. Específicamente, se ha observado una falla significativa en la clasificación de los escaneos *nmap_xmas*, donde 344 instancias fueron erróneamente predichas como *nmap_fin*. Por otro lado, el tráfico normal fue clasificado con alta precisión, con un número mínimo de falsos negativos.

Tabla 5: Métricas de evaluación para el modelo con etiqueta *attack_type*.

Métrica	Valor
Accuracy	0.88
F1 macro	0.7863
F1 weighted	0.8440
Recall macro	0.8321
Precisión macro	0.7632

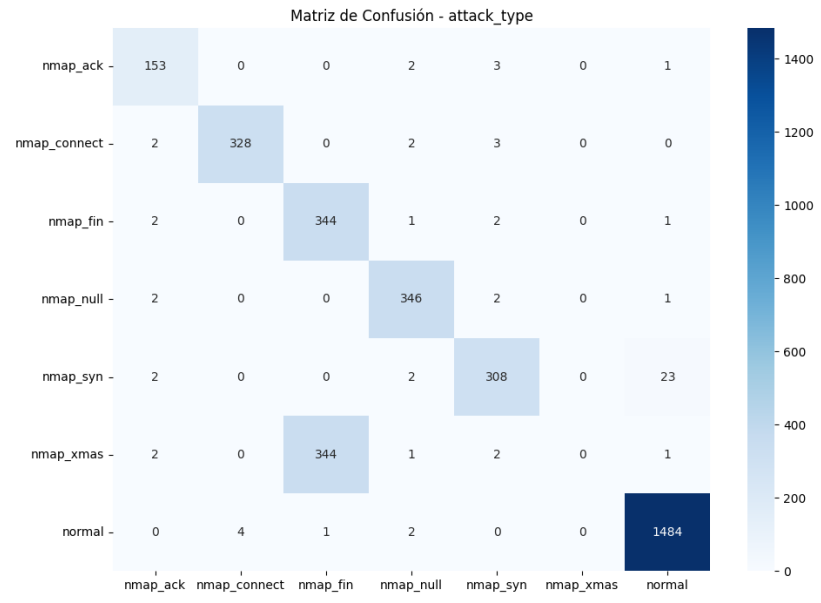


Figura 9: Matriz de confusión - *attack_type*.

5.2 Resultados en el entorno controlado

Para validar el modelo en un contexto más cercano a la realidad, se desplegó el sistema completo en el entorno de laboratorio descrito en la §15. A continuación, se presentan ejemplos selectos que demuestran la efectividad del IDS/IPS, mostrando su capacidad para detectar correctamente distintos tipos de ataques y para identificar el tráfico legítimo de manera precisa.

5.2.1 Detección de escaneo NULL Un sondeo TCP Null aprovecha una indefinición en la RFC de TCP¹ que diferencia los puertos abiertos y cerrados. En este caso, no se fija ningún bit (la cabecera de flags TCP es 0).

En la Fig. 10 se muestra el comando utilizado en Kali para lanzar un escaneo NULL a la máquina víctima sobre los puertos 21, 22, 23.

¹ <https://www.rfc-editor.org/rfc/rfc793.txt>

```
(root@kali)-[~/home/mosky]
└─# nmap -sN -p 21,22,23 20.1.2.20
Starting Nmap 7.95 ( https://nmap.org ) at 2025-06-28 19:12 CEST
Nmap scan report for 20.1.2.20
Host is up (0.073s latency).

PORT      STATE      SERVICE
21/tcp    open|filtered ftp
22/tcp    open|filtered ssh
23/tcp    open|filtered telnet

Nmap done: 1 IP address (1 host up) scanned in 14.97 seconds
```

Figura 10: Escaneo NULL desde Kali.

Por otro lado, la Fig. 11 muestra cómo los flujos correspondientes han sido clasificados de manera correcta por el IDS.

```
[+] [2025-06-28T19:17:39.625397] | 10.1.1.10:61057 -> 20.1.2.20:21 | Clase predicha: scan
1/1 ----- 0s 21ms/step
[+] [2025-06-28T19:17:39.683596] | 10.1.1.10:61057 -> 20.1.2.20:23 | Clase predicha: scan
1/1 ----- 0s 30ms/step
[+] [2025-06-28T19:17:39.753590] | 10.1.1.10:61057 -> 20.1.2.20:22 | Clase predicha: scan
1/1 ----- 0s 19ms/step
[+] [2025-06-28T19:17:40.621014] | 10.1.1.10:61059 -> 20.1.2.20:22 | Clase predicha: scan
1/1 ----- 0s 20ms/step
[+] [2025-06-28T19:17:40.681488] | 10.1.1.10:61059 -> 20.1.2.20:21 | Clase predicha: scan
1/1 ----- 0s 20ms/step
[+] [2025-06-28T19:17:40.740425] | 10.1.1.10:61059 -> 20.1.2.20:23 | Clase predicha: scan
```

Figura 11: Detección en tiempo real del escaneo NULL.

5.2.2 Detección de escaneo FIN Un sondeo FIN, al igual que el anterior, aprovecha la misma indefinición² en la RFC de TCP. Este, fija únicamente el bit TCP FIN en la cabecera de flags TCP.

En la Fig. 12 se muestra el comando utilizado en Kali para lanzar un escaneo FIN a la máquina víctima sobre los puertos 21, 22, 23 y 111.

² Los paquetes enviados sin fijar los bits SYN, RST, ACK deben ser ignorados si el puerto está abierto, pero generan una respuesta RST si el puerto está cerrado

```
(root@kali)-[~/home/mosky]
└─# nmap -sF -p 21,22,23,111 20.1.2.20

Starting Nmap 7.95 ( https://nmap.org ) at 2025-06-28 18:55 CEST
Nmap scan report for 20.1.2.20
Host is up (0.060s latency).

PORT      STATE      SERVICE
21/tcp    open|filtered ftp
22/tcp    open|filtered ssh
23/tcp    open|filtered telnet
111/tcp   open|filtered rpcbind

Nmap done: 1 IP address (1 host up) scanned in 14.84 seconds
```

Figura 12: Escaneo FIN desde Kali.

Por otro lado, en la Fig. 13 se puede observar los flujos correspondientes correctamente clasificados.

```
[+] [2025-06-28T18:56:29.625169] | 10.1.1.10:33569 -> 20.1.2.20:23 | Clase predicha: scan
1/1 [2025-06-28T18:56:29.625169] | 10.1.1.10:33569 -> 20.1.2.20:23 | Clase predicha: scan
0s 21ms/step
[+] [2025-06-28T18:56:32.627731] | 10.1.1.10:33571 -> 20.1.2.20:22 | Clase predicha: scan
1/1 [2025-06-28T18:56:32.627731] | 10.1.1.10:33571 -> 20.1.2.20:22 | Clase predicha: scan
0s 21ms/step
[+] [2025-06-28T18:56:32.692094] | 10.1.1.10:33569 -> 20.1.2.20:111 | Clase predicha: scan
1/1 [2025-06-28T18:56:32.692094] | 10.1.1.10:33569 -> 20.1.2.20:111 | Clase predicha: scan
0s 19ms/step
[+] [2025-06-28T18:56:36.623204] | 10.1.1.10:33571 -> 20.1.2.20:21 | Clase predicha: scan
1/1 [2025-06-28T18:56:36.623204] | 10.1.1.10:33571 -> 20.1.2.20:21 | Clase predicha: scan
0s 21ms/step
[+] [2025-06-28T18:56:36.684366] | 10.1.1.10:33571 -> 20.1.2.20:23 | Clase predicha: scan
1/1 [2025-06-28T18:56:36.684366] | 10.1.1.10:33571 -> 20.1.2.20:23 | Clase predicha: scan
0s 20ms/step
[+] [2025-06-28T18:56:37.631525] | 10.1.1.10:33569 -> 20.1.2.20:21 | Clase predicha: scan
1/1 [2025-06-28T18:56:37.631525] | 10.1.1.10:33569 -> 20.1.2.20:21 | Clase predicha: scan
0s 21ms/step
[+] [2025-06-28T18:56:37.689937] | 10.1.1.10:33569 -> 20.1.2.20:22 | Clase predicha: scan
1/1 [2025-06-28T18:56:37.689937] | 10.1.1.10:33569 -> 20.1.2.20:22 | Clase predicha: scan
0s 19ms/step
[+] [2025-06-28T18:56:38.623827] | 10.1.1.10:33571 -> 20.1.2.20:111 | Clase predicha: scan
```

Figura 13: Detección en tiempo real del escaneo FIN.

6. Discusión

Los resultados presentados demuestran la efectividad general del sistema IDS/IPS propuesto en la detección de intrusiones de red. La alta *Accuracy* y los valores de *F1-macro* y *Recall-macro* obtenidos en la clasificación por *attack_category* (superiores al 0.98) son un indicador robusto de la capacidad del modelo para distinguir eficazmente entre el tráfico normal y las actividades maliciosas. Este rendimiento es particularmente relevante en el contexto de un IDS/IPS, donde la minimización de falsos negativos (priorizando el *Recall*) es crucial para no pasar por alto ataques. La agrupación del tráfico legítimo bajo una única clase simplificó el análisis y confirmó la fuerte discriminación del modelo para la clasificación binaria.

La validación del sistema en un entorno de laboratorio controlado, mediante ejemplos concretos de detección de escaneos NULL y FIN, refuerza la viabilidad práctica del enfoque y su capacidad para operar en escenarios cercanos a los reales. Esta demostración práctica complementa los resultados cuantitativos al ilustrar cómo el modelo traduce su rendimiento teórico en detecciones efectivas en tiempo real, un paso fundamental para su despliegue en infraestructuras reales.

No obstante, el análisis de los resultados por *attack_type* (con una *Accuracy* del 0.88 y un *F1-macro* de 0.7863) revela una mayor complejidad inherente a la clasificación granular de ataques. La principal limitación identificada reside en la dificultad del modelo para diferenciar entre ciertos sondeos con características de tráfico muy similares, específicamente la confusión entre los escaneos *nmap_xmas* y *nmap_fin*. Este hallazgo sugiere que, si bien el modelo es competente para identificar categorías amplias de amenazas, la especificidad en la detección de variantes de ataques muy parecidas sigue siendo un desafío que podría requerir características de entrada más discriminatorias o arquitecturas de modelo aún más especializadas.

En comparación con sistemas tradicionales basados en firmas, el enfoque propuesto, al ser orientado al comportamiento y centrado en la técnica T1595 de la matriz MITRE ATT&CK, muestra una ventaja intrínseca en la adaptabilidad frente a amenazas desconocidas o variantes. Los resultados sugieren que este método tiene el potencial de ofrecer una defensa más proactiva y flexible que los sistemas dependientes de bases de datos de firmas predefinidas, los cuales suelen ser menos efectivos contra ataques *Zero-day* o nuevas variantes. La integración completa del sistema (sensores de red, Kafka, Wazuh) subraya la capacidad de desplegar soluciones de detección basadas en aprendizaje profundo en un *pipeline* operativo, lo cual es una contribución significativa en la aplicación práctica de estas técnicas.

7. Conclusiones

En este trabajo se ha desarrollado un sistema IDS/IPS basado en aprendizaje profundo capaz de detectar diversos ataques de red en tiempo real, cubriendo el ciclo completo desde el diseño del laboratorio y la generación de datos, hasta la integración con herramientas de monitorización como Wazuh. Las principales contribuciones incluyen el diseño de un entorno de laboratorio realista en GNS3, la generación sistemática de tráfico (malicioso y normal), la construcción y preprocesamiento de conjuntos de datos propios, y el entrenamiento y validación de modelos MLP que lograron resultados notables ($F1 > 0.98$ en *attack_category* y $F1 > 0.78$ en *attack_type*). Además, se logró la integración del modelo con Kafka para inferencia en tiempo real, utilizando Zeek y Suricata como sensores, completando el sistema con visualización y alertas en

Wazuh, demostrando su capacidad para identificar ataques con alta precisión y bajos falsos positivos en el entorno de laboratorio.

Como trabajo futuro, se propone ampliar la cobertura a nuevas técnicas de la matriz MITRE ATT&CK (como ataques de fuerza bruta o denegación de servicio), mejorar la precisión de detección sustituyendo sensores como Suricata por soluciones más adaptadas (libpcap en C), y enriquecer el SIEM con Wazuh mediante notificaciones automáticas y Active Responses. También sería beneficioso implementar un TAP virtual para un monitoreo pasivo sin latencia y explorar el entrenamiento incremental en tiempo real, incluyendo el uso de aprendizaje federado y técnicas de *eXplainable AI (xAI)*. Finalmente, se sugiere extender la revisión de la literatura con arquitecturas avanzadas como *transformers* o modelos fundacionales para mejorar la detección de *Zero-Days*.

7.0.1 Acknowledgments. Agradecer a la Universidad de Salamanca por los recursos proporcionados para la realización de este trabajo. En especial a mis tutores, la Dra. Angélica González Arrieta y el Dr. Álvaro Bustos Taberero por su orientación y apoyo durante la realización de este Trabajo de Fin de Máster.

7.0.2 Disclosure of Interests. Los autores declaran que no tienen ningún conflicto de intereses relevante para el contenido de este artículo.

Referencias

1. Khenwar, M., Nawal, M.: Challenges and Limitations of IDS: A Comprehensive Assessment and Future Perspectives. SKIT Research Journal, Vol. 14, No. 1, 35–39 (2024) <https://doi.org/10.47904/IJSKIT.14.1.2024.35-39>
2. Qiu, W., Ma, Y., Chen, X., Yu, H., Chen, L.: Hybrid intrusion detection system based on Dempster-Shafer evidence theory. Computers & Security, Vol. 117, 102709 (2022) <https://doi.org/10.1016/j.cose.2022.102709>
3. Anderson, J.P.: Computer Security Technology Planning Study. Technical Report ESD-TR-73-51, Vol. II, Deputy for Command and Management Systems, HQ Electronic Systems Division (AFSC), L.G. Hanscom Field (1972)
4. Xu, Z., Wu, Y., Wang, S., Gao, J., Qiu, T., Wang, Z., Wan, H., Zhao, X.: Deep Learning-based Intrusion Detection Systems: A Survey. Journal of the ACM 1(1), 1 (2025) <https://doi.org/10.48550/arXiv.2504.07839>
5. Liao, H.J., Lin, C.H.R., Lin, Y.C., Tung, K.Y.: Intrusion Detection System: A Comprehensive Review. Journal of Network and Computer Applications 36(1), 16–24 (2013). <https://doi.org/10.1016/j.jnca.2012.09.004>
6. Hutchins, E.M., Cloppert, M.J., Amin, R.M.: Intelligence-Driven Computer Network Defense Informed by Analysis of Adversary Campaigns and Intrusion Kill Chains. Lockheed Martin Corporation White Paper (2011)
7. MITRE Corporation: MITRE ATT&CK[®] for Enterprise (2023)
8. Pittman, J.M.: Machine Learning and Port Scans: A Systematic Review. arXiv preprint arXiv:2301.13581 (2023) <https://doi.org/10.48550/arXiv.2301.13581>

9. Jiang, Y., Meng, Q., Shang, F., Oo, N., Minh, L.T.H., Lim, H.W., Sikdar, B.: MITRE ATT&CK Applications in Cybersecurity and The Way Forward. arXiv preprint arXiv:2502.10825 (2025) <https://doi.org/10.48550/arXiv.2502.10825>
10. Muñoz Calle, J., Estepa Alonso, R., Estepa Alonso, A., Díaz Verdejo, J.E., Castillo Fernández, E., Madinabeitia, G.: A Flexible Multilevel System for MITRE ATT&CK Model-driven Alerts and Events Correlation in Cyberattacks Detection. *Journal of Universal Computer Science*, Vol. 30, No. 9, 1184–1204 (2024) <https://doi.org/10.3897/jucs.131686>
11. Roy, S., Panaousis, E., Noakes, C., Laszka, A., Panda, S., Loukas, G.: SoK: The MITRE ATT&CK Framework in Research and Practice. arXiv preprint arXiv:2304.07411 (2023) <https://doi.org/10.48550/arXiv.2304.07411>
12. Ali, M.L., Thakur, K., Schmeelk, S., DeBello, J., Dragos, D.: Deep Learning vs. Machine Learning for Intrusion Detection in Computer Networks: A Comparative Study. *Applied Sciences* 15(4), 1903 (2025) <https://doi.org/10.3390/app15041903>
13. Zhao, Q., Wang, F., Wang, W., Zhang, T., Wu, H., Ning, W.: Research on Intrusion Detection Model Based on Improved MLP Algorithm. *Scientific Reports* 15, 5159 (2025) <https://doi.org/10.1038/s41598-025-89798-0>
14. Lichy, A., Bader, O., Dubin, R., Dvir, A., Hajaj, C.: When a RF Beats a CNN and GRU, Together—A Comparison of Deep Learning and Classical Machine Learning Approaches for Encrypted Malware Traffic Classification. *Computers & Security* 124, 103000 (2023) <https://doi.org/10.1016/j.cose.2022.103000>
15. Kimanzi, R., Kimanga, P., Cherori, D., Gikunda, P.K.: Deep Learning Algorithms Used in Intrusion Detection Systems – A Review. arXiv preprint arXiv:2402.17020 (2024) <https://doi.org/10.48550/arXiv.2402.17020>
16. Arsalan, M., Mubeen, M., Bilal, M., Abbasi, S.F.: 1D-CNN-IDS: 1D CNN-based Intrusion Detection System for IIoT. In: *Proc. of the 29th Int. Conf. on Automation and Computing (ICAC)*, pp. 1–4. IEEE, Sunderland, United Kingdom (2024) <https://doi.org/10.1109/ICAC61394.2024.10718772>
17. Laghrissi, F., Douzi, S., Douzi, K., Hssina, B.: Intrusion Detection Systems Using Long Short-Term Memory (LSTM). *Journal of Big Data* 8, 65 (2021) <https://doi.org/10.1186/s40537-021-00448-4>

Evaluación teórico-práctica de modelos de detección de objetos en imágenes

Rubén Herrero Pérez, Vivian Félix López Batista, and Gabriel Villarrubia González

Universidad de Salamanca
Departamento de Informática y Automática
Facultad de Ciencias.
Plaza de los Caídos s/n
37008 Salamanca, España
{ruherrero,vivian,gvig}@usal.es

Resumen Este estudio presenta un análisis comparativo de modelos de detección de objetos en imágenes. Se centra en una evaluación experimental utilizando diversos conjuntos de datos a los que aplicamos distintas técnicas de preprocesado y filtrado. El objetivo principal es identificar cómo las variaciones en los datos de entrada influyen en el rendimiento de cada modelo y determinar cuál se adapta mejor, evaluando la robustez de los modelos frente a estas variaciones. El estudio concluye con un análisis integral de los resultados, identificando el mejor desempeño según las características de los datos y el contexto de aplicación. Además, se destacan implicaciones prácticas y se proporciona una guía para la selección de modelos basada en la eficiencia, adaptabilidad y rendimiento de los modelos, contribuyendo al avance de la visión artificial y la implementación de soluciones escalables y robustas en el análisis de imágenes.

Keywords: Visión artificial, modelos de detección de objetos, preprocesado de los datos, rendimiento, robustez, YOLO, Faster R-CNN, SSD, RT-DETR.

1. Introducción

El rápido crecimiento de la visión por computadora ha convertido la detección de objetos en una tecnología clave con un impacto significativo en múltiples sectores. Su capacidad para identificar y localizar elementos en imágenes y videos ha impulsado aplicaciones en seguridad, transporte, comercio y medicina, donde la rapidez y precisión son fundamentales. Sin embargo, la diversidad de modelos existentes plantea un desafío relacionado con la elección del detector según el contexto de uso. Cada modelo está desarrollado con arquitecturas avanzadas que buscan equilibrar precisión, velocidad y adaptabilidad a entornos cambiantes. Mientras algunos están diseñados para aplicaciones en tiempo

real, como vehículos autónomos, otros priorizan la detección de detalles complejos en diagnósticos médicos. Además, el creciente volumen de datos visuales generados por sensores y dispositivos requiere modelos eficientes que puedan operar bajo restricciones computacionales y condiciones adversas, como baja iluminación o ruido visual. La comparación exhaustiva de modelos es esencial para mejorar su aplicación y evolución en distintos escenarios tecnológicos. El objetivo principal de este estudio consiste en la realización de una evaluación práctica entre los detectores, utilizando dos conjuntos de datos a los cuales se les aplican varios tipos de procesados o filtros. Esto permite evaluar la robustez de los modelos ante variaciones en la calidad de los datos de entrada. Con los datos obtenidos, se realiza una comparación para identificar los modelos con mejor desempeño según el procesado de las imágenes empleado. Este estudio pretende ofrecer información valiosa sobre el uso de ciertos modelos en determinados conjuntos de datos con una disminución en su calidad debido a la naturaleza de un problema. Otros estudio relacionados están orientados en mejorar la calidad de las imágenes usando algoritmos de Súper Resolución [1] o en examinar los desafíos y estrategias para identificar elementos pequeños usando varios detectores [2]. Esta investigación pretende aportar un enfoque relacionado con el rendimiento de los modelos sobre conjuntos de datos con degradaciones en la calidad de las imágenes, comprobando la robustez de los modelos e identificando aquellos que son menos sensibles. Los resultados obtenidos en esta investigación pueden ser útiles en la inferencia de los modelos sobre conjuntos de datos, que debido a la naturaleza de los mismos, no presentan una calidad óptima. Otro de los aspectos diferenciadores de este estudio es la comparación de *Convolutional Neural Networks* (CNN) [3–5] y *Vision Transformers* (ViT) [3, 4, 6, 7], ya que muestran características diferentes que pueden ser interesantes para la evaluación y comparación del rendimiento de los detectores. También se aportan algunos modelos muy novedosos, como *You Only Look Once* YOLO12 [8, 9] y YOLO-World [10, 11], siendo este último un detector de vocabulario abierto.

2. Estado del arte

La detección de objetos es un campo de la visión artificial que ha experimentado un auge gracias a los avances en el aprendizaje profundo y el aumento de la disponibilidad de datos y potencia computacional. La esencia de estos modelos reside en su capacidad para procesar información visual compleja y extraer características distintivas que permitan diferenciar los objetos de interés del fondo y de otros objetos presentes en la escena. Para lograrlo, estos modelos emplean técnicas de aprendizaje automático, que les permiten aprender patrones y representaciones visuales a partir de grandes conjuntos de datos de imágenes anotadas.

Para realizar tareas de visión artificial es muy común utilizar redes neuronales. Son sistemas computacionales inspirados en el cerebro, compuestos por capas de nodos interconectados que procesan información de forma distribuida [3]. Permiten tareas complejas como el reconocimiento de imágenes o la predicción de datos. Aprenden ajustando sus conexiones internas (pesos) mediante algoritmos matemáticos, optimizando su precisión con cada nuevo dato. Son clave en la inteligencia artificial y el aprendizaje profundo por su capacidad de adaptación y habilidad para encontrar patrones en grandes volúmenes de información.

Las CNN son un tipo de redes neuronales ideales para datos con estructura de cuadrícula, como imágenes o vídeos [3]. A diferencia de las redes tradicionales, las CNN utilizan conexiones locales y compartidas en sus capas, lo que les permiten procesar pequeñas regiones de datos de entrada, permitiendo extraer automáticamente características jerárquicas de los datos [12]. Esto significa que detectan elementos simples (bordes, texturas) en capas iniciales y patrones más complejos (objetos, escenas) en capas más profundas. Según el mapeo sistemático realizado con la herramienta *Parsifal*, se ha encontrado información crucial para el desarrollo de este estudio, revelando un panorama en constante evolución en el campo de la detección de objetos. Se han identificado varias categorías principales de modelos, así como modelos concretos que destacan por su rendimiento y popularidad.

Los modelos YOLO son una familia de algoritmos de visión artificial diseñados para la detección de objetos en imágenes, clasificación, segmentación y reconocimiento de poses en vídeos en tiempo real [13–15]. Su enfoque revolucionario divide la imagen en cuadrículas y predice directamente las coordenadas de los objetos sin necesidad de múltiples pasos, como lo hacen otros modelos. Gracias a su rapidez y rendimiento, YOLO se ha convertido en un estándar en la investigación y la industria, ofreciendo una detección robusta en entornos dinámicos.

YOLOv11 representa un avance significativo en la tecnología de detección de objetos, mejorando las bases establecidas por las versiones anteriores [16]. La versión 12 es la última versión del modelo actualmente, innovando y diferenciándose con respecto a otras versiones al centrarse en mecanismos de atención, superando las limitaciones de velocidad de arquitecturas previas basadas en CNN mientras mantiene una alta precisión [8] [9].

YOLO-World es un modelo de detección de objetos de vocabulario abierto [10]. Dicha característica permite que el modelo pueda identificar categorías descritas a través de entradas de texto, sin importar que el modelo haya sido entrenado con dichas categorías. Este sistema destaca en la detección de una gran cantidad de objetos usando la técnica *zero-shots* con alta eficiencia [17] [18]. Dicha técnica se utiliza en un modelo que tiene que generar una salida de una entrada con la que nunca se ha entrenado, por lo que el modelo está preparado

para relacionar diferentes conceptos y generar respuestas basadas en su conocimiento previo.

Single Shot MultiBox Detector (SSD) es un detector de objetos que predice *bounding boxes* y clases directamente desde mapas de características [19, 20]. Utiliza la técnica *one-shot*, la cual se refiere a su enfoque de una sola pasada para la detección de objetos [21]. SSD predice las ubicaciones y clases de los objetos usando un grupo de cajas predefinidas como referencia directa en la imagen, dividiendo la imagen en celdas o cuadrículas para obtener el mapa de características y así identificar el objeto correctamente [22].

Region-based Convolutional Neural Network (Faster R-CNN) es un detector de objetos de dos etapas el cual primero propone regiones de interés y luego las clasifica y mejora la precisión de las posiciones detectadas [20, 23]. Su principal innovación reside en la combinación eficiente de dos módulos clave, lo que le permite lograr una alta precisión en la detección de objetos en tiempo real. La primera etapa *Region Proposal Network* (RPN) [24] genera diversas zonas de interés para posteriormente refinar y terminar de realizar la detección en la segunda etapa detector Fast R-CNN [25]. Esta arquitectura en dos pasos repercute significativamente en la velocidad del modelo, a diferencia de otros modelos de detección de una etapa como YOLO o SSD. Una de las características más importante en su arquitectura es la capacidad de compartir capas convolucionales entre los dos módulos. Esto permite utilizar el mismo mapa de características y mejorar el rendimiento del modelo, al evitar realizar un proceso de transformación de los datos entre los dos módulos.

Real Time Detection Transformer RT-DETR representa un modelo de detección de objetos que utiliza la arquitectura de los ViT diseñado para optimizar la eficiencia y velocidad de detección [18, 26]. Una de las principales motivaciones para el desarrollo de RT-DETR es solventar el alto costo computacional de los detectores basados en *Transformers* tradicionales. RT-DETR busca extender los beneficios de los detectores basados en *Transformers* al ámbito de la detección en tiempo real, logrando un equilibrio entre alta velocidad y precisión [27–29].

3. Metodología

En esta sección se propone la metodología utilizada en la realización de este estudio, con el objetivo de proporcionar una evaluación y comparación del rendimiento de diversos modelos de detección de objetos en imágenes. El objetivo es medir su desempeño ante escenarios adversos (baja iluminación, baja resolución, ruido digital, etc.), presentando los resultados en gráficas para analizar cómo mantienen su rendimiento frente a variaciones en la calidad de los datos de entrada. Para lograr este propósito, la investigación se estructura en las siguientes etapas clave:

1. **Selección de los modelos:** Elección de un conjunto representativo de modelos de detección de objetos para la evaluación. Se han considerado factores como el uso de diferentes arquitecturas, niveles de complejidad y popularidad. El factor determinante en la selección de los modelos está determinada por los resultados del mapeo sistemático, en los que se han observado los principales modelos: YOLO, Faster R-CNN, SSD, RT-DETR. También se añadió el modelo basado en vocabulario abierto YOLO-World, por tener una arquitectura y funcionamiento atractivos.
2. **Selección de métricas:** La definición de métricas de evaluación claras y objetivas será crucial para cuantificar el rendimiento de los modelos en términos de precisión, velocidad y eficiencia. Las métricas utilizadas son aquellas que suelen ser usadas al evaluar un modelo de detección:
 - **Tiempo total:** Tiempo total de inferencia del un modelo a lo largo de todas las imágenes del conjunto de datos. Se mide en segundos.
 - **Confianza media:** Valor promedio de las puntuaciones de confianza asignadas por el modelo a las detecciones que realiza. Le asigna una puntuación de confianza que indica qué tan seguro está de que esa detección es correcta y pertenece a la clase predicha. La “confianza media” sería el promedio de todas esas puntuaciones. Se mide en porcentaje.
 - **Tiempo medio de inferencia:** Tiempo medio que tarda un modelo en realizar la inferencia sobre una imagen. Se mide en milisegundos.
 - **Exactitud (*Accuracy*):** Mide la proporción de predicciones correctas realizadas por un modelo [11, 12, 14, 30]. Se mide en porcentaje.

$$Accuracy = \frac{VerdaderosPositivos(TP) + VerdaderosNegativos(TN)}{TP + TN + FalsosPositivos(FP) + FalsosNegativos(FN)} \quad (1)$$

- **Precisión (*Precision*):** Proporción de los casos positivos correctos y los que el modelo ha identificado como correctos [11, 12, 14, 30]. Se mide en porcentaje.

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

- **Exhaustividad (*Recall*):** Proporción de los casos positivos correctos y los casos positivos reales [11, 12, 14, 30]. Se mide en porcentaje.

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

- **F1-Score:** Es la media armónica de la precisión y el recall, proporcionando una medida equilibrada del rendimiento [11,

12, 14, 30]. Se mide en porcentaje.

$$F1\text{-score} = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (4)$$

- **Media IoU:** Media de la métrica *Intersection Over Union* [11, 12, 14, 30]. Compara la superposición entre la caja predicha por el modelo y la caja real (*ground truth*). Cuanto mayor sea el valor de IoU, mejor es la precisión del modelo en la predicción de la ubicación del objeto.

$$IoU = \frac{\text{Área de la intersección}}{\text{Área de la unión}} \quad (5)$$

- **mean Average Precision (mAP50):** Es una métrica común en la detección de objetos, que evalúa la precisión del modelo utilizando un umbral de IoU del 50% [31]. Se calcula determinando la precisión y el *recall* de las detecciones ordenadas por confianza, comparando las *bounding boxes* predichas con las reales.

$$mAP_{50} = \frac{1}{n} \sum_{i=1}^n AP_i \quad (6)$$

Posteriormente se obtiene el *Average Precision* (AP) para cada clase, calculando el área bajo la curva de *precision-recall*. Finalmente, el mAP50 se obtiene realizando un promedio de los AP de todas las clases, tal como se observa en la Ecuación 6.

3. **Diseño de la evaluación práctica:** Para evaluar el rendimiento de los modelos de detección de objetos, se ha diseñado un caso de estudio utilizando modelos preentrenados con el conjunto de datos *Common Objects in Context* (COCO) 2017. Usar modelos preentrenados con un *dataset* tan reconocido asegura que los resultados de esta investigación tengan una mayor utilidad y generalidad, sirviendo como punto de referencia para futuros estudios al no estar limitados a un contexto específico. Para la realización de la evaluación y comparación práctica, se han elegido dos conjuntos de datos, cuyas categorías se encuentran en el *dataset* COCO: Un conjunto de datos etiquetados con la clase “persona” para evaluar el rendimiento de los modelos detectando objetos de menor tamaño y elementos que no se muestran de forma completa en las imágenes. El objetivo de este *dataset* es evaluar el rendimiento de los modelos detectando objetos de menor tamaño y elementos que no se muestran de forma completa en las imágenes. El otro conjunto está formado por imágenes de aves y presenta etiquetas de la clase “pájaro”, con el objetivo de estudiar el impacto en el rendimiento de los modelos sobre objetos con una mayor diversidad de colores y formas. En cuanto a la aplicación de técnicas para reducir la calidad de las imágenes,

se han utilizado las siguientes: Poco brillo, escala de grises, baja resolución, ruido y desenfoque gaussiano.

4. **Implementación y Experimentación:** El diseño experimental de este estudio se centra en la evaluación y comparación exhaustiva de la robustez de diversos modelos de detección de objetos frente a variaciones en la calidad de las imágenes de entrada. Esta investigación puede ser relevante para entender la aplicabilidad de los modelos en escenarios del mundo real donde la calidad de las imágenes puede no ser óptima. Los resultados obtenidos permiten identificar los modelos que presentan una degradación mayor en el rendimiento ante ciertas alteraciones en la calidad de los datos. Se ha creado una aplicación de escritorio para facilitar el proceso de carga de conjuntos de imágenes y evaluación de los diferentes modelos. La implementación de los experimentos se ha utilizado Python como lenguaje de programación, el cual dispone de múltiples bibliotecas en tareas relacionadas con redes neuronales y tratamiento de grandes conjuntos de datos. Tkinter para la interfaz gráfica y Vega-Altair para la creación de las visualizaciones y gráficas. Las evaluaciones prácticas de los modelos se han realizado a través de un *hardware* con las siguientes especificaciones: Procesador Intel® Core™ Ultra 7 155H (Desde 1.4GHz, 16 núcleos, caché 24MB), 16 GB de RAM LPDDR5 (Doble Canal, max 6400 MHz), tarjeta gráfica Intel(R) Arc(TM) Graphics 8GB VRAM y una unidad de disco SSD 512GB M.2 (NVMe Gen.4).
5. **Evaluación y comparación de métricas:** Una vez finalizado el proceso de obtención de métricas, se han generado gráficas para comprender los resultados y comparar el rendimiento de los modelos ante degradaciones en las imágenes de entrada.

4. Resultados

En esta sección se examinarán los resultados obtenidos tras evaluar los modelos utilizando los dos conjuntos de datos y los filtros ya mencionados. Se analizarán las métricas y gráficas con los resultados de las inferencias de los detectores, determinando qué modelos son mejores o peores según cada filtro aplicado. Esta sección busca responder la incógnita principal de este estudio, relacionada con la robustez de los diferentes modelos de detección sobre variaciones en la calidad de las imágenes de entrada.

4.1 Experimento 1: Conjunto de imágenes de personas

En este experimento se ha realizado la evaluación de los modelos sobre un conjunto de imágenes de partidos de fútbol. El conjunto de datos se ha obtenido de *Roboflow Universe*¹ y consta de 1200 imágenes,

¹ <https://universe.roboflow.com/yinguo/soccer-data>

divididas en dos clases: Personas que poseen el balón y personas que no lo tienen. Se han aplicado los cinco filtros (poco brillo, escala de grises, baja resolución, ruido y desenfoco gaussiano) a las imágenes para realizar su comparación y determinar qué variaciones afectan de forma más severa al rendimiento de los detectores. El objetivo principal de realizar la comparación sobre este conjunto en concreto es evaluar la detección de elementos de menor tamaño dentro de las imágenes, siendo las personas el objeto de estudio. Al tratarse de elementos que pueden aparecer parcialmente en la imagen, se busca determinar la capacidad de los modelos para identificar objetos que no se muestran en su totalidad.

Se ha realizado una primera evaluación con el conjunto de datos sin aplicar ningún filtro. Es crucial establecer un rendimiento base de los modelos usando los datos originales y sin modificar. Esta evaluación inicial sirve como punto de referencia esencial. Al comparar los resultados de los modelos bajo situaciones adversas con este rendimiento base, se puede medir el impacto exacto de las alteraciones, determinando si las variaciones se deben a la falta de robustez o a la naturaleza propia del modelo.

La Figura 1 muestra los resultados de la evaluación sobre el conjunto de imágenes de personas sin filtros: El modelo más rápido en procesar las imágenes fue SSD, tanto en tiempo total como por imagen individual, mientras que Faster R-CNN resultó ser significativamente el más lento, siendo el tiempo de ejecución aproximadamente el doble que la media (línea roja). Las velocidades de estos modelos se deben posiblemente a la naturaleza de sus arquitectura: SSD es un modelo con una arquitectura de una etapa, lo que permite realizar la inferencia de una imagen a una velocidad mayor. En el caso contrario, Faster R-CNN necesita generar una predicción a nivel de píxel, perjudicando el tiempo de inferencia. Su arquitectura en dos etapas también repercute negativamente en la velocidad de inferencia de las imágenes. El otro gran problema de este modelo es la precisión: genera una gran cantidad de cajas delimitadoras propuestas en zonas de interés que considera el modelo, con el objetivo de garantizar la detección de todos los elementos de la imagen (minimizar los falsos negativos). Sin embargo, esto causa que el número de falsos positivos aumente, y como consecuencia, métricas como la exactitud, la precisión y el *F1-Score* disminuyan de forma considerable.

YOLO-World, YOLO11 y YOLO12 logran un buen equilibrio entre velocidad y calidad porque usan una arquitectura en una sola etapa similar a SSD, pero optimizada con técnicas avanzadas de procesamiento de características. Esto les permite alcanzar alta *precision* y *accuracy* sin perder demasiada velocidad. RT-DETR sobresale en *recall*, ya su arquitectura está diseñada para detectar la mayor cantidad de objetos posible, enfocándose en minimizar las omisiones (falsos negativos). Sin embargo, esta estrategia suele afectar a la *precision* al generar un

mayor número de falsos positivos, lo que explica su menor rendimiento en esta métrica.

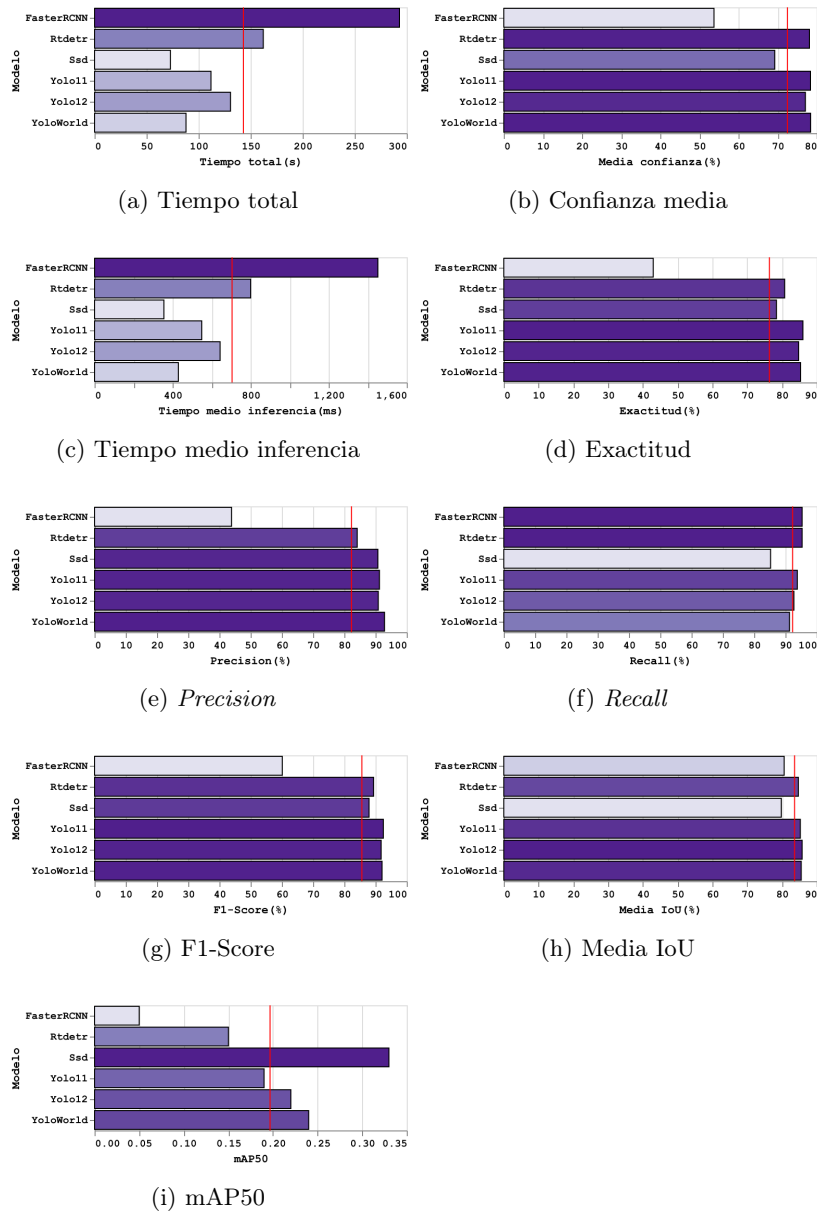


Figura 1: Gráficas resultantes de las métricas de los modelos evaluados con el conjunto de datos de personas sin aplicar filtros

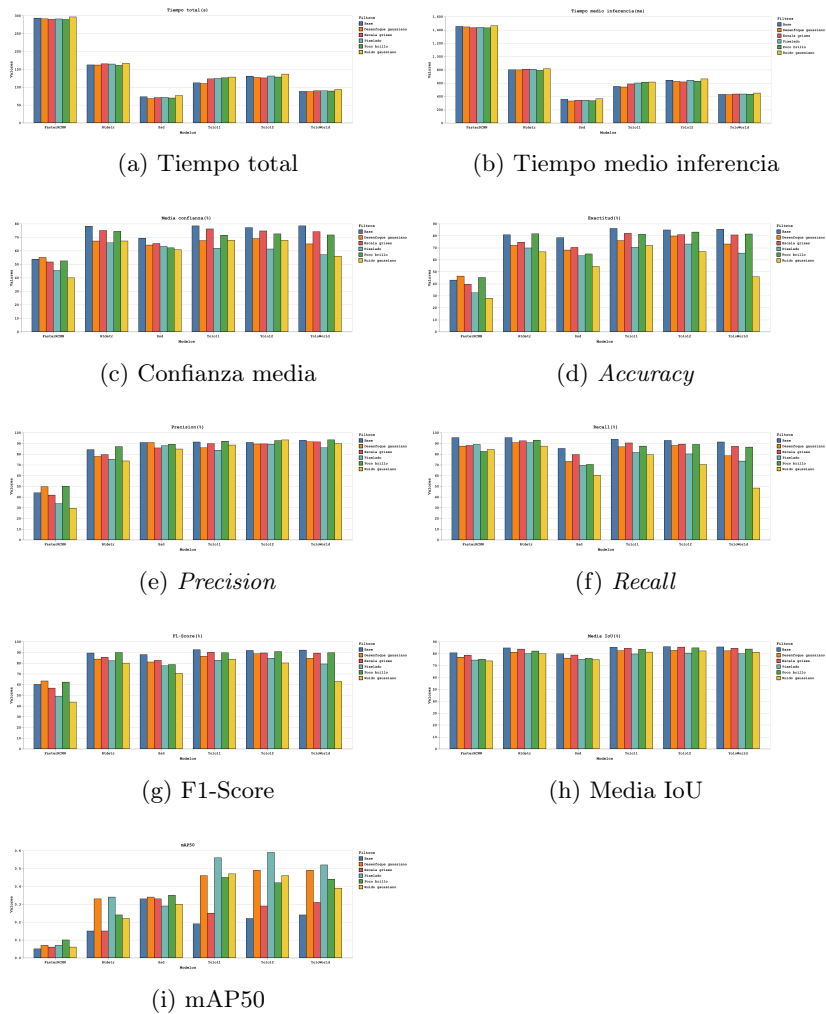


Figura 2: Gráficas comparativas de las métricas de los modelos evaluados con el conjunto de datos de personas bajo diferentes filtros

Tras observar el rendimiento de los modelos según las variaciones en el conjunto de datos de entradas; a continuación se realiza el análisis de los resultados obtenidos con respecto al caso base. El objetivo es comprobar el nivel de degradación de los valores de las métricas teniendo como referencia los resultados de la evaluación sobre las imágenes sin modificar.

En cuanto al desempeño general de los modelos observado en la Figura 2, considerando el tiempo total de ejecución y calidad en las detecciones; los modelos YOLO11, YOLO12, YOLO-World y SSD muestran consistentemente los mejores resultados en métricas de rendimiento (exactitud, precisión, *recall*, *F1-Score*, media IoU, mAP50), con valores significativamente más altos que Faster R-CNN y RT-DETR en la mayoría de los escenarios. Además, tienden a ser más rápidos en cuanto a la inferencia de las detecciones. Faster R-CNN y RT-DETR son notablemente más lentos (sobre todo Faster R-CNN) y generalmente exhiben menores valores en las métricas de precisión, lo que afecta a la calidad de la detección. RT-DETR es un ViT que generalmente necesita más potencia computacional para funcionar con el rendimiento esperado, siendo posiblemente el motivo de unos tiempos de inferencia más lentos.

A continuación se analiza el impacto que han tenido los filtros en el proceso de detección de los modelos, ya sea por un aumento en el tiempo de inferencia o en la reducción de los valores de las métricas relacionadas con la calidad de la detección:

- **Ruido gaussiano (mayor impacto negativo):** El filtro de ruido gaussiano tiene un impacto significativo en el rendimiento de los modelos de detección, afectando de manera más severa a la mayoría de detectores. Se observa una caída generalizada en el desempeño de todos los modelos, especialmente en métricas como la confianza media, exactitud y *recall*. Este deterioro se debe a la interferencia generada por el ruido aleatorio, que altera los patrones visuales en la imagen y complica la identificación precisa de los elementos por parte del modelo. Este obstáculo en la detección de los objetos provoca que los modelos reconozcan menos personas en este caso, aumentando el número de falsos negativos, y en consecuencia, disminuyendo el *recall*. Aunque este filtro dificulta la detección de los objetos, aquellos que sí son identificados correctamente por los modelos presentan una media IoU bastante similar a otras variaciones en las imágenes. Esto significa que a pesar de la inserción de elementos que dificultan la predicción, los modelos son capaces de establecer correctamente las cajas delimitadoras de los objetos detectados. También destaca el tiempo medio de inferencia, que es muy levemente superior al resto de perturbaciones de este experimento. Esto resalta una mayor dificultad a la hora de identificar los elementos en las imágenes. Aunque los modelos YOLO también se

ven afectados, muestran una mayor resistencia ante este tipo de perturbación, lo que sugiere que su arquitectura es más robusta frente a la presencia de ruido en las imágenes. Sin embargo, ninguno de los modelos logra mantener un rendimiento sobresaliente bajo esta condición, ya que la introducción de ruido gaussiano reduce la calidad de las detecciones y afecta negativamente a la capacidad de los sistemas para realizar predicciones confiables. Este fenómeno pone en evidencia cómo las perturbaciones visuales pueden comprometer seriamente la eficacia de los modelos de detección, resaltando la importancia de desarrollar algoritmos con mayor capacidad de adaptación a condiciones adversas.

- **Poco brillo y escala de grises (menor impacto):** Al aplicar este filtro en las imágenes, no se muestra un descenso significativo en el rendimiento de los modelos. La mayoría de métricas relacionadas con la calidad de las predicciones presentan valores altos. El uso de poca luminosidad en las imágenes no ha afectado a la velocidad de inferencia. De manera similar, la detección de imágenes en escala de grises no ha perjudicado apenas el rendimiento de los modelos. Los resultados del experimento permiten concluir que los modelos evaluados presentan una capacidad generalizada y robusta para la predicción de imágenes con estas dos variaciones. El filtro de poco del brillo aplicado a las imágenes de los conjuntos de datos redujo su luminosidad en un 90 %, provocando una disminución drástica en la visibilidad de los objetos dentro de la escena. A simple vista, esta alteración representa un desafío considerable para la percepción humana, ya que los detalles quedan casi ocultos en la oscuridad, haciendo extremadamente difícil identificar formas, bordes y contornos con claridad. Dado este escenario, era de esperar que los modelos de detección sufrieran una caída significativa en sus métricas de rendimiento. La falta de luz suele comprometer la capacidad de los algoritmos para extraer información relevante, afectando aspectos clave como el *recall*. Sin embargo, los resultados obtenidos revelaron que los modelos de visión artificial no solo conservaron un desempeño aceptable, sino que en varios casos superaron con creces la percepción humana en condiciones de baja iluminación. El hecho de que los modelos mantengan un rendimiento sólido en estas condiciones adversas pone en un primer plano la ventaja tecnológica de la inteligencia artificial en tareas de visión. Mientras que la percepción humana depende fuertemente de la cantidad de luz disponible, los algoritmos pueden procesar información más allá de las limitaciones físicas del ojo humano, lo que los convierte en herramientas esenciales para entornos donde la visibilidad es reducida.
- **Desenfoco gaussiano y baja resolución (impacto variable):** Estos filtros ofrecen resultados similares entre ellos. Dependiendo del modelo afecta de manera más o menos severa. En los modelos de la familia YOLO, la confianza en las detecciones se reduce en

mayor medida que en otros filtros. La exactitud y el *recall* también se resienten ligeramente. Una métrica a destacar es el mAP50 en los modelos RT-DETR, YOLO11, YOLO12 y YOLO-World. Esta métrica experimenta valores más elevados utilizando los filtros del desenfoque gaussiano y baja resolución. Esto significa que aunque el valor de la confianza es menor en estos detectores y con esos filtros, los valores de las confianzas son menos dispersos y por tanto superan los umbrales de confianza establecidos para el cálculo de mAP50. Es decir, en vez de tener valores probabilísticos dispersos entre 0% y 100%; se encuentran alrededor del 60%.

4.2 Experimento 2: Conjunto de imágenes de pájaros

Para este segundo experimento se ha seleccionado un conjunto de datos en los que se muestra un pájaro por imagen. Se obtuvo del repositorio *Kaggle*² y está formado por 2010 imágenes. La elección de un conjunto de datos de pájaros para evaluar el rendimiento de diversos modelos de detección de objetos se debe principalmente a la alta diversidad en la apariencia de estas criaturas. La variedad en el aspecto de estos animales puede presentar desafíos para los modelos de detección, ofreciendo una evaluación robusta de su capacidad para generalizar y detectar objetos en entornos complejos. Una de las principales razones para seleccionar pájaros es la inmensa gama de colores que manifiestan, presentando una paleta cromática muy diversa. Los modelos también deben ser capaces de distinguir los pájaros que presentan colores que se mimetizan con el entorno natural. Los hábitats de las aves suelen ser variados, con texturas, sombras y recargado de elementos naturales que pueden confundir a los detectores de objetos. El análisis y evaluación de un conjunto de datos tan colorido se ha elegido específicamente para verificar la importancia que tiene la característica del color en los modelos de detección cuando se trata de obtener características de una imagen. Además de los colores, la variedad de formas entre las aves es también destacable. Pueden presentarse variaciones en la morfología entre especies, cambios en la postura, la acción realizada (posado, volando, buscando alimento) o el ángulo de visión. Evaluar modelos con este tipo de datos contribuye a verificar que son capaces de reconocer objetos sin importar su orientación o forma física en un momento determinado, lo cual es crucial para aplicaciones del mundo real. Aunque en este experimento solo se detecta la clase “pájaro” (*dataset* COCO 2017), se consideró importante utilizar un conjunto de imágenes con diferentes especies de aves para comprobar si los modelos pueden detectar pájaros de diferentes formas y colores. Se quería estudiar esa distinción ya que en el experimento anterior sobre el conjunto de imágenes de personas, todos los elementos tenían una apariencia y forma muy similar.

² <https://www.kaggle.com/datasets/gpiosenka/birdies>

Al igual que en el experimento anterior, es esencial establecer un desempeño inicial de los modelos de detección de objetos sobre los conjuntos de datos originales y sin modificaciones. Las gráficas de la Figura 3 muestran de manera visual las diferentes métricas obtenidas en esta evaluación base sin alteraciones en las imágenes. En este caso concreto, la mayoría de detectores (exceptuando Faster R-CNN) ofrecen resultados muy similares ya que han identificado correctamente la inmensa mayoría de elementos, independientemente de la especie de pájaro y otras características como el color o la forma.

Al tratarse de objetos grandes y sólo uno por imagen, se estimó que la capacidad de los modelos de detectar los elementos de las imágenes era alta. A través de esta evaluación, se comprueba que los detectores entrenados con el *dataset* de COCO 2017 son capaces de reconocer objetos de la clase “pájaro” aunque estos difieran en una gran variedad de colores, posiciones y especies.

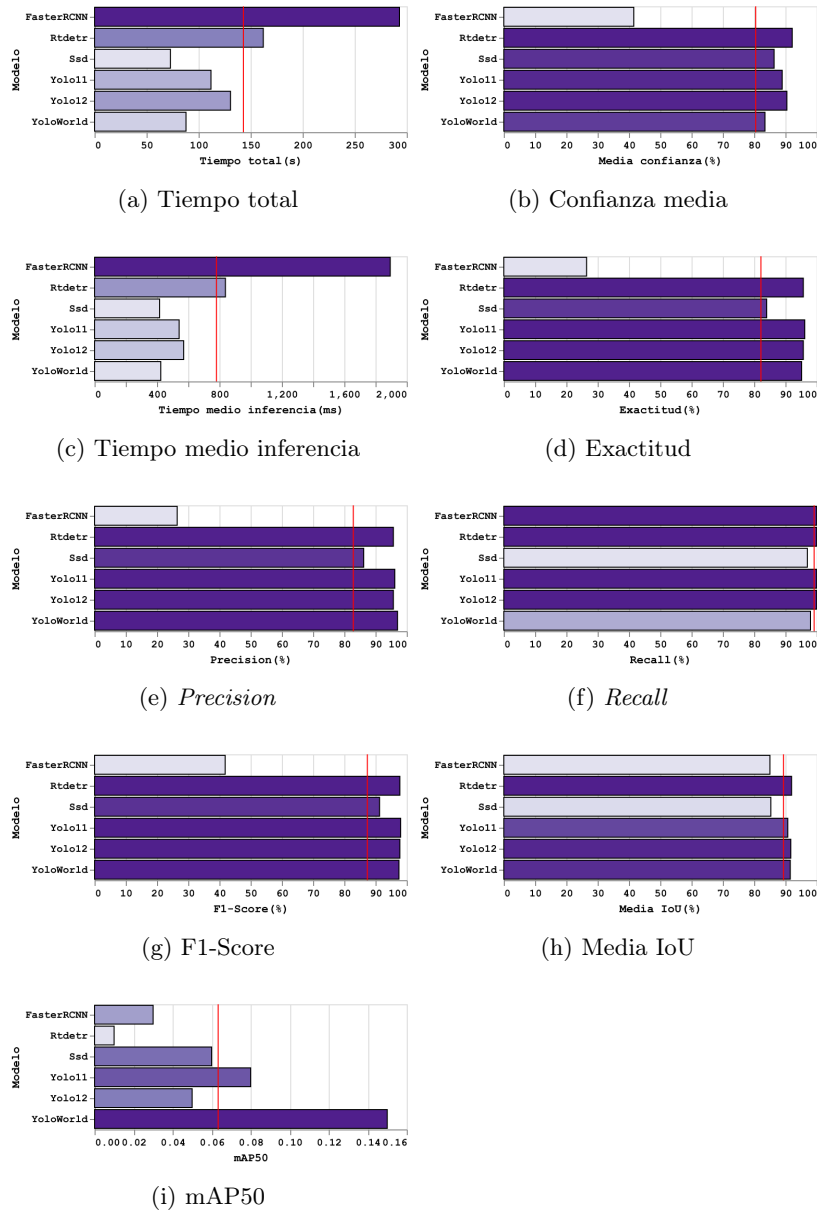


Figura 3: Gráficas resultantes de las métricas de los modelos evaluados con el conjunto de datos de pájaros sin aplicar filtros

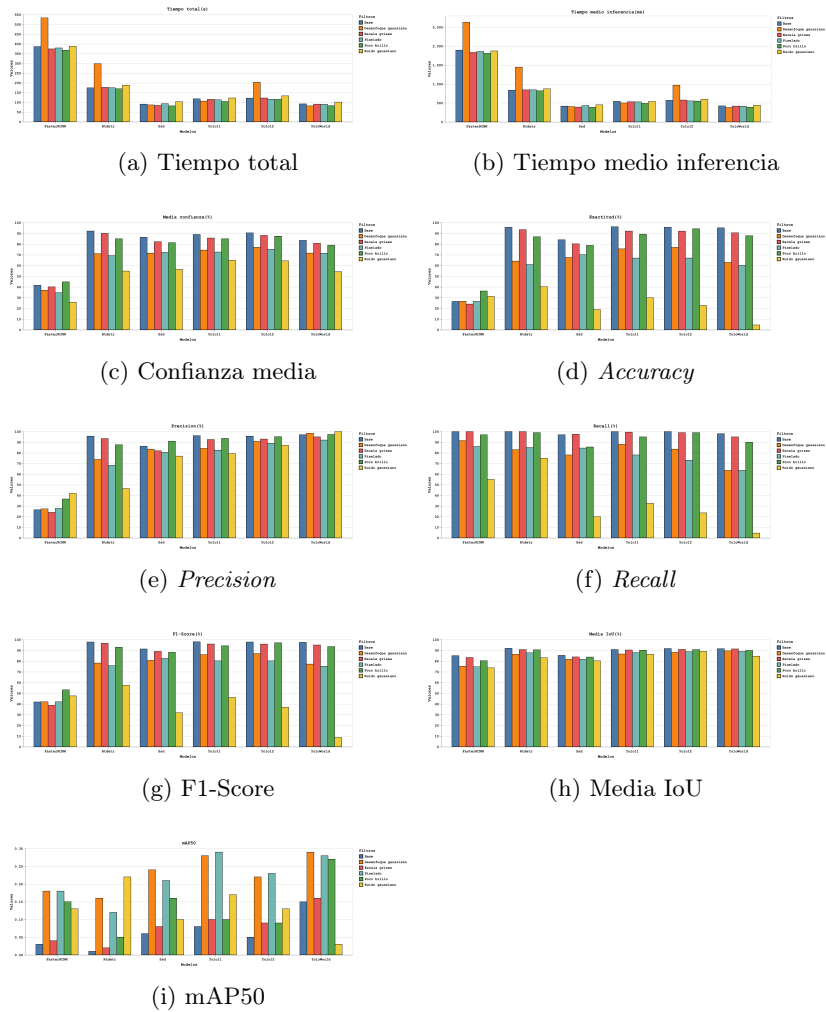


Figura 4: Gráficas comparativas de las métricas de los modelos evaluados con el conjunto de datos de pájaros bajo diferentes filtros

Al igual que en el Experimento 1, se procede a realizar un análisis de los resultados obtenidos con respecto al caso base, con el fin de comprobar cómo ha afectado al rendimiento de los modelos (YOLO11, YOLO12, YOLO-World, SSD, Faster R-CNN y RT-DETR) el uso de los diferentes filtros aplicados (Poco brillo, escala de grises, baja resolución ruido gaussiano y desenfoque gaussiano), teniendo como referencia los resultados de la evaluación sobre las imágenes sin modificar. La Figura 4 representa la comparación de los diferentes modelos y filtros utilizados en la evaluación, englobados por métricas. En ellas se pueden apreciar de forma directa y visual, qué modelos destacan o cuáles se ven merendados en su rendimiento por el uso de ciertas variaciones en la calidad de las imágenes del conjunto de pájaros.

Los modelos YOLO11 y YOLO12 muestran resultados muy equilibrados y robustos. Sus métricas de exactitud, precisión, exhaustividad y *F1-Score* son consistentemente altas, a menudo superando el 80-90%. La media IoU también es muy alta, indicando una buena localización de las detecciones. La velocidad del modelo es intermedia, comparándola con otros detectores de esta evaluación.

YOLO-World destaca por su alta velocidad consistente, independientemente del filtro aplicado, lo que lo convierte en una opción interesante para aplicaciones que requieren una rápida respuesta del modelo. Su precisión es sobresaliente, a menudo la más alta del grupo, llegando a 100% en alguna evaluación, lo que significa que sus detecciones son muy fiables. La media IoU también es generalmente alta. Su principal debilidad reside en la inconsistencia de su *recall*, que es notablemente más bajo al aplicar los filtros de ruido gaussiano, desenfoque gaussiano y baja resolución.

SSD es el modelo más veloz, aunque en este experimento la diferencia no es tan notable como en la evaluación con el conjunto de imágenes de personas. Su precisión es generalmente buena, superando el 80% en la mayoría de los casos. Sin embargo, otras métricas como el *recall* y el *accuracy* son más inconsistentes dependiendo del filtro aplicado a las imágenes.

Faster R-CNN es, sin lugar a dudas el modelo más lento de todos, con tiempos totales de procesamiento y de inferencia notablemente más altos que cualquier otro, llegando a más de 500 segundos y 2600 ms respectivamente en algunos casos. A excepción del *recall*, que presenta algunos de los mayores valores en esta métrica; sus valores de exactitud, precisión, *F1-Score* y Confianza media son las más bajas. Esto significa que, aunque detecta muchos objetos, produce una cantidad muy alta de falsos positivos.

RT-DETR demuestra un *recall* alto, en algunos casos alcanzando el 100%. Es uno de los modelos más consistentes en esta métrica, aunque existen variaciones según el filtro aplicado en las imágenes. Dicha variación en la métrica *recall* es generalizada en todos los modelos. Sus métricas de exactitud, precisión y F1-Score son generalmente muy

altas, lo que indica un rendimiento sólido. Su principal desventaja es la velocidad, siendo el segundo modelo más lento después de Faster R-CNN. La precisión también es alta, pero se resiente en los filtros ruido, desenfoque gaussiano y baja resolución.

En los modelos Faster R-CNN, RT-DETR y YOLO12 destaca el tiempo total y de inferencia al inferir imágenes con el filtro de desenfoque gaussiano, el cual es notablemente alto. Este resultado indica que los modelos han experimentado dificultades en la identificación y clasificación de los objetos presentes en las imágenes del conjunto de datos empleado.

A continuación se analizan los diferentes filtros utilizados según su capacidad de deteriorar el rendimiento de los modelos:

- **Ruido gaussiano (peor impacto):** Tiene un impacto negativo elevado en todas las métricas de detección, incluyendo la confianza media, la exactitud, la precisión, el exhaustividad y F1-Score. En todas las gráficas, las barras correspondientes a esta perturbación muestran los valores más bajos. Esto evidencia la extrema dificultad que enfrentan los modelos para detectar objetos, clasificarlos correctamente o generar cajas delimitadoras con precisión cuando se introduce ruido gaussiano. La generación de variaciones aleatorias en la intensidad de los píxeles parece ocultar características importantes de los objetos, dificultando la tarea de extracción de patrones relevantes.
- **Desenfoque gaussiano y baja resolución (impacto variable):** Estos filtros parecen afectar dependiendo del modelo utilizado. Ha repercutido de manera negativa y general en la métrica *accuracy* de los modelos. También han disminuido los valores de *recall* y F1-Score de la mayoría todos los modelos. En términos absolutos siguen siendo unos buenos valores (rondan el 80% de *recall*), aunque en comparación con la evaluación con otro filtros, se ha producido un ligero descenso.
- **Escala de grises y poco brillo (menor impacto):** Estos dos filtros son los que menos han repercutido en los resultados de la evaluación. Vuelve a destacar el buen funcionamiento de los modelos en condiciones de práctica oscuridad, superando la capacidad humana en este aspecto. Los resultados de todos los modelos son bastante similares a los observados en el caso base (imágenes sin filtros). Los resultados obtenidos también muestran, que aunque el conjunto de datos original sea muy colorido, los modelos han sido capaz de localizar y clasificar correctamente los elementos del conjunto utilizando un filtro de escala de grises. Este hecho demuestra la robustez de los modelos ante variaciones en los colores de las imágenes de entrada.

5. Conclusiones

En este estudio se ha realizado la evaluación y comparación de diversos modelos de detección de objetos en imágenes utilizando una aplicación de escritorio diseñada a medida, aplicando diferentes filtros para analizar su comportamiento ante variaciones en la calidad de las imágenes. Los resultados se presentaron gráficamente para una clara interpretación, mejorando la comprensión de las métricas evaluadas. Los modelos han sido evaluados satisfactoriamente sobre dos conjuntos de datos, aplicando distintos filtros a las imágenes de entrada con el fin de analizar el comportamiento de los modelos ante variaciones en la calidad de las imágenes.

Este estudio evalúa varios modelos de detección de objetos, destacando SSD por su rapidez y precisión, ideal para tareas que requieren respuestas rápidas y baja tasa de falsos positivos. Faster R-CNN es desaconsejado debido a su bajo rendimiento general y lenta inferencia. Aunque RT-DETR es lento, su alto *recall* lo hace útil para aplicaciones críticas como la detección de tumores, donde minimizar los falsos negativos es crucial.

Los modelos de la familia YOLO demuestran un excelente equilibrio entre velocidad y rendimiento, lo que los hace versátiles para diversas aplicaciones de visión artificial. Se incluyeron modelos recientes como YOLO12 (febrero de 2025) y YOLO-World, este último especializado en detección de vocabulario abierto. Aunque YOLO12 no mostró una mejora sustancial respecto a su versión anterior, YOLO-World sí ofreció un gran equilibrio entre rendimiento y tiempo de inferencia, siendo muy robusto ante variaciones en la calidad de las imágenes.

La evaluación también identificó que el ruido gaussiano es el filtro más perjudicial para la robustez de los modelos, dificultando la extracción de características relevantes. Otros filtros como el desenfoque gaussiano y la reducción de resolución tuvieron un impacto intermedio, mientras que las alteraciones en el brillo y los canales de color apenas afectaron el rendimiento de los detectores, demostrando una alta robustez a este tipo de variaciones.

En resumen, el estudio proporciona información clave para seleccionar el detector de objetos más adecuado según los requisitos de la tarea y las condiciones de la imagen, contribuyendo a la resolución de problemas prácticos y futuras investigaciones en visión artificial.

A continuación se proponen varias mejoras o líneas de investigación para ser desarrolladas en un futuro, mejorando así el alcance de este estudio:

Las futuras líneas de investigación propuestas para este estudio incluyen la realización de una evaluación similar utilizando modelos preentrenados en otros conjuntos de datos públicos conocidos, como PASCAL o ImageNet, para analizar y comparar los modelos en diversos contextos de interés. Asimismo, se plantea cargar los modelos en

dispositivos de borde o *Edge Computing* para comparar su rendimiento en sistemas con menor potencia computacional, aprovechando la reducción de ancho de banda y latencia que ofrecen estos entornos al procesar datos localmente y enviar sólo la información relevante. Otra dirección de trabajo futura es la utilización de un equipo con una *Neural Processing Unit* (NPU), dado que estas unidades están específicamente diseñadas para la inferencia de modelos de inteligencia artificial, lo que podría revelar un incremento significativo en el rendimiento de los modelos al comparar con los resultados actuales. Finalmente, se propone efectuar un estudio similar centrado en múltiples *Vision Transformers*. A diferencia del presente trabajo donde sólo se incluyó un ViT, una evaluación comparativa exhaustiva de varios de estos modelos representaría un enfoque novedoso y menos explorado en la literatura.

Referencias

1. A. B. Varol Malkocoglu and R. Samli, "A novel model for higher performance object detection with deep channel attention super resolution," *Engineering Science and Technology, an International Journal*, vol. 64, p. 102003, 2025.
2. H. Wang and P. Gao, "Survey of small object detection methods based on deep learning," in *2024 9th International Conference on Intelligent Informatics and Biomedical Sciences (ICIIBMS)*, vol. 9, pp. 221–224, 2024.
3. I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.
4. M. A. Elhassan, C. Zhou, A. Khan, A. Benabid, A. B. Adam, A. Mehmood, and N. Wambugu, "Real-time semantic segmentation for autonomous driving: A review of cnns, transformers, and beyond," *Journal of King Saud University - Computer and Information Sciences*, vol. 36, no. 10, p. 102226, 2024.
5. S. S. A. Zaidi, M. S. Ansari, A. Aslam, N. Kanwal, M. Asghar, and B. Lee, "A survey of modern deep learning based object detection models," *Digital Signal Processing*, vol. 126, p. 103514, 2022.
6. O. Elharrouss, Y. Himeur, Y. Mahmood, S. Alrabae, A. Ouamane, F. Bensaali, Y. Bechqito, and A. Chouchane, "Vits as backbones: Leveraging vision transformers for feature extraction," *Information Fusion*, vol. 118, p. 102951, 2025.
7. I. Pacal, B. Ozdemir, J. Zeynalov, H. Gasimov, and N. Pacal, "A novel cnn-vit-based deep learning model for early skin cancer diagnosis," *Biomedical Signal Processing and Control*, vol. 104, p. 107627, 2025.
8. Y. Tian, Q. Ye, and D. Doermann, "Yolov12: Attention-centric real-time object detectors," 2025.
9. M. A. R. Alif and M. Hussain, "Yolov12: A breakdown of the key architectural features," 2025.
10. T. Cheng, L. Song, Y. Ge, W. Liu, X. Wang, and Y. Shan, "Yolo-world: Real-time open-vocabulary object detection," enero 2024.
11. Ultralytics, "Yolo-world (real-time open-vocabulary object detection)," octubre 2024.
12. S. Rabanser, O. Shchur, and S. Günnemann, "Introduction to tensor decompositions and their applications in machine learning," noviembre 2017.
13. Ultralytics, "Ultralytics YOLO11 Tasks — docs.ultralytics.com."
14. R. Sapkota, Z. Meng, M. Churuvija, X. Du, Z. Ma, and M. Karkee, "Comprehensive performance evaluation of yolo11, yolov10, yolov9 and yolov8 on detecting and counting fruitlet in complex orchard environments," julio 2024.
15. J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection." arXiv.org, junio 2015.
16. R. Khanam and M. Hussain, "Yolov11: An overview of the key architectural enhancements," 2024.
17. X. Gu, T. Lin, W. Kuo, and Y. Cui, "Open-vocabulary object detection via vision and language knowledge distillation." arXiv.org, abril 2021.
18. S. Rahman, S. Khan, and F. Porikli, "A unified approach for conventional zero-shot, generalized zero-shot, and few-shot learning," *IEEE Transactions on Image Processing*, vol. 27, no. 11, pp. 5652–5667, 2018.

19. R. Wang, Y. Duan, M. Hu, X. Liu, Y. Li, Q. Gao, T. Tong, and T. Tan, "Lightr-yolov5: A compact rotating detector for sars-cov-2 antigen-detection rapid diagnostic test results," *Displays*, vol. 78, 2023.
20. A. Saleh, D. Jones, D. Jerry, and M. R. Azghadi, "A lightweight transformer-based model for fish landmark detection," *arXiv preprint arXiv:2209.05777v1*, 2022.
21. I. Sucholutsky and M. Schonlau, "'less than one'-shot learning: Learning n classes from m lt; n samples," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 35, pp. 9739–9746, May 2021.
22. W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Fu, and A. C. Berg, "Ssd: Single shot multibox detector," 2016.
23. S. Angara, N. R. Mannuru, A. Mannuru, and S. Thirunagaru, "A novel method to enhance pneumonia detection via a model-level ensembling of cnn and vision transformer," *arXiv preprint arXiv:2401.02358v1*, 2024.
24. E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," 2016.
25. S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks." arXiv.org, junio 2015.
26. Y. Zhao, W. Lv, S. Xu, J. Wei, G. Wang, Q. Dang, Y. Liu, and J. Chen, "Detrs beat yolos on real-time object detection," 2024.
27. M. Kang, C.-M. Ting, F. F. Ting, and R. Phan, "Cst-yolo: A novel method for blood cell detection based on improved yolov7 and cnn-swin transformer," *arXiv preprint arXiv:2306.14590v2*, 2023.
28. J. Cani, I. Mademlis, A. A. R. Chrysochoou, and G. T. Papadopoulos, "Illicit object detection in x-ray images using vision transformers," *arXiv preprint arXiv:2403.19043v2*, 2024.
29. W. He, Y. Zhang, T. Xu, T. An, Y. Liang, and B. Zhang, "Object detection for medical image analysis: Insights from the rt-detr model," *arXiv preprint arXiv:2501.16469v1*, 2025.
30. C. Janiesch, P. Zschech, and K. Heinrich, "Machine learning and deep learning," *Electronic Markets*, vol. 31, no. 3, pp. 685–695, 2021.
31. J. Hui, "map (mean average precision) for object detection — jonathan-hui.medium.com," 2018.

Autores

Arturo Álvarez Sánchez, es investigador en formación en la Universidad de Salamanca. Es licenciado en Ingeniería Informática y tiene un máster en Sistemas Inteligentes por la Universidad de Salamanca. Durante sus estudios de grado y máster, colaboró como investigador con el Laboratorio de Sistemas Expertos y Aplicaciones (ESALab) de la misma universidad, donde participó en varios proyectos de investigación relacionados principalmente con el blockchain y el desarrollo de software.



Belén Curto Diego, Licenciada en Ciencias Físicas (Electrónica y Automática) y Doctora en Ciencias por la Universidad de Salamanca. Tras un periodo de tres años en el sector privado en la empresa ENUSA Industrias Avanzadas, se incorpora a la Universidad de Salamanca, por su interés en el ámbito docente e investigador. Perteneció al área de Ingeniería de Sistemas y Automática, donde imparte docencia en titulaciones de Grado y Máster en materias relacionadas con los Sistemas Inteligentes, Robots Autónomos, Automática y Control y Computadores. Ha ocupado cargos de dirección académica de Postgrado, Máster y Doctorado. Respecto a su labor investigadora, es codirectora del GIR: Robótica y Sociedad", donde ha dirigido más de 7 Tesis Doctorales y TFMs sobre Robótica y Sistemas Inteligentes. Ha sido directora y ha participado en multitud de proyectos de investigación de ámbito nacional y autonómico financiado por entidades públicas. Además, ha sido directora y componente del equipo investigador de más de 40 proyectos de transferencia tecnológica e innovación industrial con empresas privadas del sector cárnico, nuclear, médico, seguridad, medioambiental, etc. Cuenta con tres (3) sexenios de investigación reconocidos y uno (1) de Transferencia. Es autora de más de 100 publicaciones en revistas, comunicaciones a congresos, presentaciones en jornadas, etc. así como revisora de diversas revistas y editora de 2 números especiales relacionados con la Robótica Colaborativa. También, es autora de tres (3) patentes en explotación.



Pablo Díez Gastón, es un desarrollador de Software especializado en Backend graduado en ingeniería informática en sistemas de información en la Escuela Politécnica Superior de Zamora por la Universidad de Salamanca y titulado en el Máster en Sistemas Inteligentes.



Francisco José García Peñalvo, es Catedrático de Universidad del Departamento de Informática y Automática en la Universidad de Salamanca (USAL), con 4 sexenios de investigación, 1 sexenio de transferencia y 5 quinquenios docentes reconocidos. Recibió el premio Gloria Begué a la excelencia docente en 2019 y el premio María de Maeztu a la excelencia investigadora en

2023. Además, fue Profesor Distinguido de la Escuela de Humanidades y Educación del Tecnológico de Monterrey, México entre 2016 y 2018 y es Investigador de Impacto Internacional de la Universidad Nacional San Agustín, Arequipa, Perú. Desde 2006 es el director del Grupo de Investigación Reconocido por la USAL GRIAL (GRupo de investigación en InterAcción y eLearning), grupo que es Unidad de Investigación Consolidada de la Junta de Castilla y León (UIC 81). Ha dirigido 30 tesis doctorales. Ha sido Vicedecano de Innovación y Nuevas Tecnologías de la Facultad de Ciencias de la USAL entre 2004 y 2007 y Vicerrector de Innovación Tecnológica de esta Universidad entre 2007 y 2009. Actualmente es el Subdirector del Instituto Universitario de Ciencias de la Educación (IUCE), el Delegado del Rector para la Docencia Digital y el Coordinador del Programa de Doctorado en Formación en la Sociedad del Conocimiento de la USAL. Es Editor Jefe de la revista Education in the Knowledge Society. Editor Asociado en Computers in Human Behavior, Computers in Human Behavior Reports, IEEE Access, entre otras, habiendo sido editor asociado de las revistas IEEE RITA e IEEE Transactions on Learning Technologies. Ha publicado más de 100 artículos en revistas indexadas en el JCR SCIE/SSCI (61 Q1).

Marcelo Hernández, es Ingeniero en Electrónica (Universidad ORT Uruguay) y Máster en Sistemas Inteligentes (Universidad de Salamanca). Su perfil combina la experiencia práctica en la industria, gestionando sistemas de telemetría y hardware en ST Consultores, con una sólida formación de posgrado. Su interés profesional se centra en la intersección entre el hardware, las comunicaciones y la inteligencia artificial, con un enfoque en el desarrollo de soluciones de IA robustas y eficientes.





Rubén Herrero Pérez, Ingeniero Informático por la Universidad de Salamanca y Máster Universitario en Sistemas Inteligentes. Durante su formación del grado y máster, ha colaborado como Personal Investigador en el grupo de investigación ESALAB, donde ha participado en varios proyectos de investigación relacionados principalmente con blockchain y desarrollo software.

Joel Jimenez Valido, es un profesional Full Stack con una sólida base académica internacional y dominio de las tecnologías más actuales; su formación incluye una licenciatura en Ingeniería Informática obtenida en Cuba, complementada con un Máster en Sistemas Inteligentes por la Universidad de Salamanca (USAL); actualmente, demuestra una gran capacidad de gestión y adaptabilidad al desempeñar dos posiciones laborales concurrentes: un puesto a jornada completa en Argentina y otro a media jornada en España; esta doble función subraya su compromiso, su versatilidad profesional y su habilidad para manejar responsabilidades significativas en entornos geográficos diversos.



Vivian López Batista, es profesora titular de la Universidad de Salamanca en el área de Ciencias de la Computación e Inteligencia Artificial. Doctorada en Informática por la Universidad de Valladolid en 1996. Miembro del Grupo de Minería de Datos. Ha realizado investigación en diferentes campos como procesamiento del lenguaje natural, redes neuronales y minería de datos. Tiene 80 artículos publicados en revistas de reconocido prestigio, talleres y actas de conferencias, 20 libros y capítulos de libros y 20 informes técnicos, la mayoría de ellos en estos temas. Miembro del comité organizador y científico de varios simposios internacionales. Fue directora del Máster en Sistema Inteligente y del Programa de Doctorado en Informática y Automática de la Universidad de Salamanca desde junio de 2010 hasta octubre de 2012.





Álvaro Lozano Murciego, es actualmente Profesor Contratado Doctor en el área de Lenguajes y Sistemas Informáticos. En cuanto su formación académica, es Ingeniero Técnico en Informática de Sistemas, Graduado en Ingeniería Informática, ha realizado el Máster de Sistemas Inteligentes y es Doctor en Ingeniería Informática. Ha sido beneficiario de contratos predoctorales y postdoctorales en su etapa formativa y ha participado en diferentes proyectos Europeos Nacionales y regionales a lo largo de los últimos años. Es miembro del grupo de investigación Expert Systems and Application Lab (ESALAB) y colaborador en el Grupo de Minería de Datos (MIDA). Sus líneas de investigación incluyen el machine learning aplicado a diferentes áreas, sistemas de recomendación, sistemas de optimización, desarrollo de sensores inalámbricos en IoT y edge computing.

Giannina Milagros Miguel Jurado, Ingeniera geógrafa con experiencia en SIG y teledetección, orientada al análisis y procesamiento de datos espaciales e imágenes satelitales para proyectos de investigación y consultoría. Con participación en la producción de cartografía y series históricas de cobertura/uso de suelo con Google Earth Engine, así como en la creación de variables para modelos de conectividad espacial y apoyo a diagnósticos de gestión del riesgo. Complementa su perfil con competencias en programación (Python, R, JavaScript), ciencia de datos y aprendizaje automático y desarrollo WebGIS, con un enfoque actual en el desarrollo de soluciones inteligentes y sistemas de alerta temprana.



Diego Marcelo Miranda Rodriguez es Ingeniero de Sistemas Computacionales. Ha obtenido la Maestría en Sistemas Inteligentes por la Universidad de Salamanca. Sus áreas de interés e investigación se centran en la minería de datos, machine learning y el desarrollo de soluciones inteligentes aplicadas a la gestión y análisis de información. Se desempeña como Ingeniero de Sistemas, orientando su labor profesional hacia la mejora continua de procesos tecnológicos y la aplicación práctica de sus conocimientos en sistemas inteligentes.

Maria N. Moreno García, es Catedrática de Universidad en el Departamento de Informática y Automática de la Universidad de Salamanca. Actualmente es directora del Grupo de Investigación en Minería de Datos (mida.usal.es) y de la Cátedra Extraordinaria Viewnext- Universidad de Salamanca. Ha sido coordinadora del Programa de Doctorado en Ingeniería Informática de la misma Universidad desde 2013 hasta 2020. Sus intereses de investigación se centran en el área del aprendizaje automático y su aplicación en varios dominios, especialmente en medicina, medios sociales y fundamentalmente en el campo de los sistemas de recomendación. En estos ámbitos, ha colaborado con grupos de investigación punteros, habiendo realizado estancias de investigación en el Intelligent System Lab de la Universidad de Bristol (Reino Unido) y en el College of Computing and Digital Media de la Universidad DePaul de Chicago, Estados Unidos. Ha organizado diferentes ediciones de sesiones especiales sobre "Web mining and recommender systems" sobre "Web and Social Media Mining" en conferencias internacionales. Es miembro del consejo editorial de revistas internacionales, así como editora invitada de varios números especiales y revisora de revistas indexadas en posiciones relevantes en WoS Journal Citation reports. También es autora de numerosos artículos publicados en revistas y conferencias de reconocido prestigio.



Vidal Moreno Rodilla, es Licenciado en Ciencias Físicas en la rama de Electrónica y Doctor en Ciencias en 1996. Actualmente es Profesor Titular de Universidad del área de Ingeniería de Sistemas y Automática de la Universidad de Salamanca. Ha impartido docencia sobre Robótica, Inteligencia Artificial y Control en niveles de Grado, Máster y Doctorado tanto en Salamanca como en el extranjero. Miembro fundador del Grupo de Investigación Reconocido "Robótica y Sociedad", es autor de varias decenas de publicaciones en revistas internacionales fruto de la dirección de una centena trabajos de investigación que incluye tesis doctorales, Trabajos de Fin de Master, etc.

Belén Pérez Lancho, es Licenciada y Doctora en Ciencias Físicas por la Universidad de Salamanca. Realizó su formación postdoctoral en la Universidad Paris VI (curso 1996-97) y desde 1998 es Profesora Titular en el Área de Ingeniería de Sistemas y Automática de la Universidad de Salamanca. Imparte o ha impartido docencia en las titulaciones de Ingeniería Informática, Física, Estadística y en el Máster en Sistemas Inteligentes. Pertenece al grupo de investigación BISITE y ha participado en más de 20 proyectos de investigación, principalmente en aplicaciones de sistemas de control y sistemas multiagente, colaborando en varias publicaciones científicas, en la dirección de tres tesis y de numerosos proyectos de fin de carrera. Ha ocupado cargos de gestión en la Facultad de Ciencias como Secretaria Académica (años 2000 a 2004) y como Vicedecana (años 2008 a 2016).



Roberto Therón Sánchez, es un investigador multidisciplinar con una sólida formación en Informática, Comunicación Audiovisual y Humanidades. Licenciado en Informática por las Universidades de Salamanca (diplomatura) y La Coruña, y Doctor en Informática por la Universidad de Salamanca, donde obtuvo el Premio Extraordinario de Doctorado por su trabajo sobre robótica.



Su trayectoria investigadora se centra en la Visualización de la Información y la Analítica Visual, combinando herramientas y técnicas de la informática, estadística, diseño gráfico y humanidades para comprender datos complejos. Actualmente dirige el grupo VisUsal (parte del Grupo de Investigación GRIAL), dedicado al desarrollo de herramientas avanzadas para visualizar datos multidimensionales, como los genéticos o paleoclimáticos. Ha colaborado en proyectos de relevancia internacional con instituciones como el Laboratorio de Ciencias del Clima y del Medio Ambiente (París), la Real Academia Española y la Academia de Ciencias de Austria. Además de su labor investigadora, ha sido coordinador del Programa de Doctorado en Informática y Automática de la USAL (2012-2017) y actualmente es el Director del Máster Universitario en Sistemas Inteligentes (desde 2012).



Diego Uribe Esquivel, es un profesionalista graduado de la Licenciatura Ingeniería en Mecatrónica y Producción por la Universidad Iberoamericana en la Ciudad de México y del máster en Sistemas Inteligentes de la Universidad de Salamanca. Sus áreas de investigación de interés son

la automática, informática, el control inteligente, la ingeniería electrónica, y la aplicación de la Inteligencia Artificial en estos campos.



Gabriel Villarrubia González, Doctor de la Universidad de Salamanca en el Departamento de Informática y Automática. Ingeniero Informático por la Universidad Pontificia de Salamanca (2011), habiendo estudiado el Máster en Sistemas Inteligentes en la Universidad de Salamanca (2012). Además, ha realizado el Máster Universitario en Seguridad e Internet (2014) y el Máster en Dirección de Sistemas de Información (2015). A lo largo de su formación ha seguido una línea de investigación bien definida, centrada en la aplicación de los sistemas multi-agente a los entornos de inteligencia ambiental, con especial atención a la definición de arquitecturas inteligentes y a la fusión de información. Ha participado en más de 30 proyectos de investigación, colaborando en la obtención de 15 propiedades intelectuales, habiendo realizado estancias internacionales. En el ámbito científico, es autor de más de 40 publicaciones científicas y codirector de 50 proyectos de fin de grado. Es miembro del comité científico de diferentes congresos, siendo miembro del comité organizador de ISAMI, PACBB, MIS4TEL, PAAMS y DCAI. A nivel profesional fue nombrado Microsoft Student Partner en el año 2010. Ha ganado diferentes premios en el desarrollo de aplicativos móviles. Actualmente centra sus esfuerzos en diseñar una arquitectura multi-agente especialmente diseñada para la fusión dinámica de información en entornos inteligentes caracterizados por estar dotados de sensores heterogéneos que capturan información del entorno. Actualmente es personal docente investigador en el Departamento de Informática.

