



**VNiVERSIDAD
D SALAMANCA**

CAMPUS DE EXCELENCIA INTERNACIONAL

UNIVERSIDAD DE SALAMANCA

Departamento de Informática y Automática

Facultad de Ciencias

TESIS DOCTORAL

**Machine learning and econometric applications
for increasing profitability and efficiency: A case
study on sustainable production and trade in
agro-based industries**

Autora:

Dña. María Eugenia Pérez Pons

Directores:

Dr. D. Juan Manuel Corchado Rodríguez

Dr. D. Javier Parra Domínguez

Diciembre de 2021



**VNiVERSIDAD
D SALAMANCA**

CAMPUS DE EXCELENCIA INTERNACIONAL

UNIVERSITY OF SALAMANCA

Department of Computer Science and Automation

Faculty of Sciences

DOCTORAL THESIS

**Machine learning and econometric applications
for increasing profitability and efficiency: A case
study on sustainable production and trade in
agro-based industries**

Author:

Dña. María Eugenia Pérez Pons

Advisors:

Dr. D. Juan Manuel Corchado Rodríguez

Dr. D. Javier Parra Domínguez

December, 2021



**VNiVERSiDAD
D SALAMANCA**

CAMPUS DE EXCELENCIA INTERNACIONAL

UNIVERSIDAD DE SALAMANCA

Departamento de Informática y Automática

Facultad de Ciencias

**Machine learning and econometric applications
for increasing profitability and efficiency: A case
study on sustainable production and trade in
agro-based industries**

Autor:

Dña. María Eugenia Pérez Pons

Directores:

Dr. D. Juan Manuel Corchado Rodríguez

Dr. D. Javier Parra Domínguez

TRIBUNAL

Presidente:

Vocal:

Secretario:

Suplentes:

FECHA DE LECTURA:

CALIFICACIÓN:

Solicitud de Presentación de Tesis Doctoral

Estimado Coordinador del Programa de Doctorado en Ingeniería Informática:

Dña. María Eugenia Pérez Pons, con DNI 47910780-D y alumna del programa de DOCTORADO EN INGENIERÍA INFORMÁTICA, matriculada en el plan de estudios D015 - INGENIERÍA INFORMÁTICA (R.D.99/2011):

Solicita que se tenga en consideración la información aportada en este documento con el objeto de poder presentar la Tesis Doctoral con título *Machine learning and econometric applications for increasing profitability and efficiency: A case study on sustainable production and trade in agro-based industries* mediante el formato de compendio de artículos/publicaciones. La información aportada se corresponde con lo establecido en el Procedimiento para la presentación de la Tesis Doctoral en la Universidad de Salamanca en el Formato de Compendio de Artículos/Publicaciones.

A continuación se detallan los documentos adjuntos en esta solicitud:

- Autorización de los directores para la presentación de la Tesis Doctoral mediante el formato de compendio de artículos/publicaciones.
- Introducción y resumen de la Tesis Doctoral presentada.
 - Introducción.
 - Metodología de investigación.
 - Objetivos de la Tesis Doctoral.
 - Estado del arte.
 - Contribuciones.
 - Publicaciones.
 - Proyectos.
 - Conclusiones.
 - Trabajo futuro.
- Se incluye un resumen de cada publicación, una introducción con los antecedentes del tema objeto de estudio, la hipótesis de trabajo y los objetivos de la investigación, así como los principales resultados y conclusiones finales de cada uno de ellos.

- En el anexo, se incorpora una copia completa de las publicaciones originales que conformarán la Tesis Doctoral.

En Salamanca, a 10 de enero de 2022

La doctoranda

Dña. María Eugenia Pérez Pons

Autorización de los Directores

En Salamanca, a 15 de diciembre de 2021,

HACEMOS CONSTAR:

Que, como directores de la Tesis Doctoral de **María Eugenia Pérez Pons**, con DNI 47910780D, autorizamos a presentar la Tesis Doctoral **“Machine learning and econometric applications for increasing profitability and efficiency: A case study on sustainable production and trade in agro-based industries”** mediante la modalidad de compendio de artículos, al disponer de los siguientes artículos publicados:

1. Pérez-Pons, M. E., Parra-Domínguez, J., Sigeru Omatu, Herrera-Viedma, E & Corchado, JM. (2021). Machine Learning and Traditional Econometric Models: A Systematic Mapping Study. JAISCR, 12(2), 79-100
2. Pérez-Pons, M. E., Plaza-Hernández, M., Alonso, R. S., Parra-Domínguez, J., & Prieto, J. (2021). Increasing Profitability and Monitoring Environmental Performance: A Case Study in the Agri-Food Industry through an Edge-IoT Platform. Sustainability, 13(1), 283 (14), 5706.
3. Pérez-Pons, M. E., Alonso, R. S., García, O., Marreiros, G., & Corchado, J. M. (2021). Deep Q-Learning and preference based multi-agent system for sustainable agricultural market. Sensors, 21(16), 5276.
4. Pérez-Pons, M. E., Parra-Dominguez, J., Hernández G., Herrera-Viedma., E & Corchado, J.M. (2022). Evaluation metrics and dimensional reduction for binary classification algorithms: A case study on bankruptcy prediction. The Knowledge Engineering Review (accepted, pending for publication)

Los directores:

CORCHADO
RODRIGUEZ JUAN
MANUEL -
70978310B

Firmado digitalmente por
CORCHADO RODRIGUEZ
JUAN MANUEL - 70978310B
Fecha: 2021.12.05 18:34:58
+01'00'

Dr. Juan Manuel Corchado Rodríguez

PARRA
DOMINGUEZ
JAVIER - 70935333K

Firmado digitalmente
por PARRA DOMINGUEZ
JAVIER - 70935333K
Fecha: 2021.12.03
16:59:36 +01'00'

Dr. Javier Parra-Dominguez



D. /Dª. Juan Manuel Corchado Rodriguez

HAGO CONSTAR:

Que soy COAUTOR/A de los siguientes trabajos:

Pérez-Pons, M. E., Alonso, R. S., García, O., Marreiros, G., & Corchado, J. M. (2021). Deep Q-Learning and preference based multi-agent system for sustainable agricultural market. *Sensors*, 21(16), 5276.

Pérez-Pons, M. E., Parra-Domínguez, J., Sigeru Omatu, Herrera-Viedma, E & Corchado, JM. (2021). Machine Learning and Traditional Econometric Models: A Systematic Mapping Study. *JAISCR*, 12(2), 79-100

Pérez-Pons, M. E., Parra-Domínguez, J., Hernández G., Herrera-Viedma., E & Corchado, J.M. (2022). Evaluation metrics and dimensional reduction for binary classification algorithms: A case study on bankruptcy prediction. *The Knowledge Engineering Review* (accepted, pending for publication)

Y MANIFIESTO QUE:

Como COAUTOR/A NO DOCTOR/A del trabajo del doctorando _____ expreso mi RENUNCIA a presentar el artículo como parte de otra Tesis Doctoral.

Como COAUTOR/A del trabajo del doctorando María Eugenia Pérez Pons acepto que dicho trabajo sea presentado como parte de su Tesis Doctoral y declaro que el doctorando es el autor principal de la investigación recogida en estos trabajos.

Salamanca a de de

CORCHADO
RODRIGUEZ JUAN
MANUEL - 70978310B

Firmado digitalmente por
CORCHADO RODRIGUEZ JUAN
MANUEL - 70978310B
Fecha: 2021.12.05 18:35:48
+01'00'

Fdo: _____

COMISIÓN ACADÉMICA DEL PROGRAMA DE DOCTORADO

D. /Dña. Javier Parra Domínguez

HAGO CONSTAR:

Que soy COAUTOR/A de los siguientes trabajos:

Pérez-Pons, M. E., Plaza-Hernández, M., Alonso, R. S., Parra-Domínguez, J., & Prieto, J. (2021). Increasing Profitability and Monitoring Environmental Performance: A Case Study in the Agri-Food Industry through an Edge-IoT Platform. Sustainability, 13(1), 283 (14), 5706.

Pérez-Pons, M. E., Parra-Domínguez, J., Sigeru Omatu, Herrera-Viedma, E & Corchado, JM. (2021). Machine Learning and Traditional Econometric Models: A Systematic Mapping Study. JAISCR, 12(2), 79-100

Pérez-Pons, M. E., Parra-Domínguez, J., Hernández G., Herrera-Viedma, E & Corchado, J.M. (2022). Evaluation metrics and dimensional reduction for binary classification algorithms: A case study on bankruptcy prediction. The Knowledge Engineering Review (accepted, pending for publication)

Y MANIFIESTO QUE:

Como COAUTOR/A NO DOCTOR/A del trabajo del doctorando _____ expreso mi RENUNCIA a presentar el artículo como parte de otra Tesis Doctoral.

Como COAUTOR/A del trabajo del doctorando María Eugenia Pérez Pons acepto que dicho trabajo sea presentado como parte de su Tesis Doctoral y declaro que el doctorando es el autor principal de la investigación recogida en estos trabajos.

Salamanca a 3 de diciembre de 2021

PARRA DOMINGUEZ
JAVIER - 70935333K

Firmado digitalmente por PARRA
DOMINGUEZ JAVIER - 70935333K
Fecha: 2021.12.03 17:01:14
+01'00'

Fdo: Javier Parra Domínguez

COMISIÓN ACADÉMICA DEL PROGRAMA DE DOCTORADO



D. /Dª. Sigeru Omatu

HAGO CONSTAR:

Que soy COAUTOR/A de los siguientes trabajos:

Pérez-Pons, M. E., Parra-Domínguez, J., Sigeru Omatu, Herrera-Viedma, E & Corchado, JM. (2021). Machine Learning and Traditional Econometric Models: A Systematic Mapping Study. JAISCR, 12(2), 79-100

Y MANIFIESTO QUE:

Como COAUTOR/A NO DOCTOR/A del trabajo del doctorando _____ expreso mi RENUNCIA a presentar el artículo como parte de otra Tesis Doctoral.

Como COAUTOR/A del trabajo del doctorando María Eugenia Pérez Pons acepto que dicho trabajo sea presentado como parte de su Tesis Doctoral y declaro que el doctorando es el autor principal de la investigación recogida en estos trabajos.

Hiroshima a 4 de diciembre de 2021

Fdo: _____

COMISIÓN ACADÉMICA DEL PROGRAMA DE DOCTORADO

D. /Dª. Oscar García

HAGO CONSTAR:

Que soy COAUTOR/A de los siguientes trabajos:

Pérez-Pons, M. E., Alonso, R. S., García, O., Marreiros, G., & Corchado, J. M. (2021). Deep Q-Learning and preference based multi-agent system for sustainable agricultural market. Sensors, 21(16), 5276.

Y MANIFIESTO QUE:

Como COAUTOR/A NO DOCTOR/A del trabajo del doctorando _____ expreso mi RENUNCIA a presentar el artículo como parte de otra Tesis Doctoral.

Como COAUTOR/A del trabajo del doctorando María Eugenia Pérez Pons acepto que dicho trabajo sea presentado como parte de su Tesis Doctoral y declaro que el doctorando es el autor principal de la investigación recogida en estos trabajos.

Salamanca a 3 de diciembre de 2021



Digitally signed by GARCIA
GARCIA OSCAR - 07982656T
Date: 2021.12.03 19:09:48
+01'00'

Fdo: Óscar García García

COMISIÓN ACADÉMICA DEL PROGRAMA DE DOCTORADO

D. /Dña. Enrique Herrera-Viedma

HAGO CONSTAR:

Que soy COAUTOR/A de los siguientes trabajos:

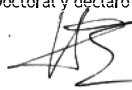
Pérez-Pons, M. E., Parra-Domínguez, J., Sigeru Omatu, Herrera-Viedma, E & Corchado, JM. (2021). Machine Learning and Traditional Econometric Models: A Systematic Mapping Study. JAISCR, 12(2), 79-100

Pérez-Pons, M. E., Parra-Domínguez, J., Hernández G., Herrera-Viedma., E & Corchado, J.M. (2022). Evaluation metrics and dimensional reduction for binary classification algorithms: A case study on bankruptcy prediction. The Knowledge Engineering Review (accepted, pending for publication)

Y MANIFIESTO QUE:

Como COAUTOR/A NO DOCTOR/A del trabajo del doctorando _____ expreso mi RENUNCIA a presentar el artículo como parte de otra Tesis Doctoral.

Como COAUTOR/A del trabajo del doctorando María Eugenia Pérez Pons acepto que dicho trabajo sea presentado como parte de su Tesis Doctoral y declaro que el doctorando es el autor principal de la investigación recogida en estos trabajos.



Salamanca a 6 de diciembre de 2021

Fdo: Enrique Herrera Viedma

COMISIÓN ACADÉMICA DEL PROGRAMA DE DOCTORADO



D. /Dña. Marta Plaza-Hernández

HAGO CONSTAR:

Que soy COAUTOR/A de los siguientes trabajos:

Pérez-Pons, M. E., Plaza-Hernández, M., Alonso, R. S., Parra-Domínguez, J., & Prieto, J. (2021). Increasing Profitability and Monitoring Environmental Performance: A Case Study in the Agri-Food Industry through an Edge-IoT Platform. Sustainability, 13(1), 283 (14), 5706.

Y MANIFIESTO QUE:

Como COAUTOR/A NO DOCTOR/A del trabajo del doctorando María Eugenia Pérez Pons
expreso mi RENUNCIA a presentar el artículo como parte de otra Tesis Doctoral.

Como COAUTOR/A del trabajo del doctorando _____
acepto que dicho trabajo sea presentado como parte de su Tesis Doctoral y declaro que el doctorando es el autor principal de la investigación recogida en estos trabajos.

Salamanca a 29 de noviembre de 2021

PLAZA HERNANDEZ
MARTA - 70903717F

Firmado digitalmente por
PLAZA HERNANDEZ MARTA
- 70903717F
Fecha: 2021.11.29 13:29:04
+01'00'

Fdo: Marta Plaza Hernández

COMISIÓN ACADÉMICA DEL PROGRAMA DE DOCTORADO



D. /Dª. Ricardo Serafín Alonso Rincón

HAGO CONSTAR:

Que soy COAUTOR/A de los siguientes trabajos:

Pérez-Pons, M. E., Plaza-Hernández, M., Alonso, R. S., Parra-Domínguez, J., & Prieto, J. (2021). Increasing Profitability and Monitoring Environmental Performance: A Case Study in the Agri-Food Industry through an Edge-IoT Platform. Sustainability, 13(1), 283 (14), 5706.

Pérez-Pons, M. E., Alonso, R. S., García, O., Marreiros, G., & Corchado, J. M. (2021). Deep Q-Learning and preference based multi-agent system for sustainable agricultural market. Sensors, 21(16), 5276.

Y MANIFIESTO QUE:

Como COAUTOR/A NO DOCTOR/A del trabajo del doctorando _____ expreso mi RENUNCIA a presentar el artículo como parte de otra Tesis Doctoral.

Como COAUTOR/A del trabajo del doctorando María Eugenia Pérez Pons acepto que dicho trabajo sea presentado como parte de su Tesis Doctoral y declaro que el doctorando es el autor principal de la investigación recogida en estos trabajos.

Salamanca a 28 de noviembre de 2021

Firmado digitalmente
por ALONSO RINCON
RICARDO SERAFIN -
45685315R
Fecha: 2021.11.28
21:46:03 +01'00'

Fdo: Ricardo Serafín Alonso Rincón

COMISIÓN ACADÉMICA DEL PROGRAMA DE DOCTORADO

D. /Dña. Goreti Marreiros

HAGO CONSTAR:

Que soy COAUTOR/A de los siguientes trabajos:


Pérez-Pons, M. E., Alonso, R. S., García, O., Marreiros, G., & Corchado, J. M. (2021). Deep Q-Learning and preference based multi-agent system for sustainable agricultural market. Sensors, 21(16), 5276.

Y MANIFIESTO QUE:

Como COAUTOR/A NO DOCTOR/A del trabajo del doctorando _____ expreso mi RENUNCIA a presentar el artículo como parte de otra Tesis Doctoral.

Como COAUTOR/A del trabajo del doctorando María Eugenia Pérez Pons acepto que dicho trabajo sea presentado como parte de su Tesis Doctoral y declaro que el doctorando es el autor principal de la investigación recogida en estos trabajos.

Salamanca a 9 de diciembre de 2021

Fdo: 

COMISIÓN ACADÉMICA DEL PROGRAMA DE DOCTORADO

D. /D^a. Javier Prieto Tejedor

HAGO CONSTAR:

Que soy COAUTOR/A de los siguientes trabajos:

Pérez-Pons, M. E., Plaza-Hernández, M., Alonso, R. S., Parra-Domínguez, J., & Prieto, J. (2021). Increasing Profitability and Monitoring Environmental Performance: A Case Study in the Agri-Food Industry through an Edge-IoT Platform. Sustainability, 13(1), 283 (14), 5706.

Y MANIFIESTO QUE:

Como COAUTOR/A NO DOCTOR/A del trabajo del doctorando _____ expreso mi RENUNCIA a presentar el artículo como parte de otra Tesis Doctoral.

Como COAUTOR/A del trabajo del doctorando María Eugenia Pérez Pons acepto que dicho trabajo sea presentado como parte de su Tesis Doctoral y declaro que el doctorando es el autor principal de la investigación recogida en estos trabajos.

Salamanca a 10 de diciembre de 2021

**PRIETO TEJEDOR
JAVIER -
11969812Z**

Firmado digitalmente por PRIETO
TEJEDOR JAVIER - 11969812Z
Fecha: 2021.12.10 10:02:20 +01'00'

Fdo: Javier Prieto Tejedor

COMISIÓN ACADÉMICA DEL PROGRAMA DE DOCTORADO

D. /Dª. Guillermo Hernández

HAGO CONSTAR:

Que soy COAUTOR/A de los siguientes trabajos:

Pérez-Pons, M. E., Parra-Dominguez, J., Hernández G., Herrera-Viedma., E & Corchado, J.M. (2022). Evaluation metrics and dimensional reduction for binary classification algorithms: A case study on bankruptcy prediction. The Knowledge Engineering Review (accepted, pending for publication)

Y MANIFIESTO QUE:

Como COAUTOR/A NO DOCTOR/A del trabajo del doctorando María Eugenia Pérez Pons
expreso mi RENUNCIA a presentar el artículo como parte de otra Tesis Doctoral.

Como COAUTOR/A del trabajo del doctorando María Eugenia Pérez Pons
acepto que dicho trabajo sea presentado como parte de su Tesis Doctoral y declaro que el doctorando es el autor principal de la investigación recogida en estos trabajos.

Salamanca a 6 de diciembre de 2021

Firmado por HERNANDEZ.GONZALEZ
GUILLERMO - 52414240J el día 06/12/2021 con
un certificado emitido por AC FNMT Usuarios

Fdo: Guillermo Hernández González

COMISIÓN ACADÉMICA DEL PROGRAMA DE DOCTORADO

“ Much to learn, you still have.”

Yoda

Resumen

Se espera que en 2050 la población mundial haya alcanzado los 9 000 millones de personas. El crecimiento de la población a nivel mundial va a dar lugar a un crecimiento en la demanda de productos de consumo derivados de la agricultura. Es por ello que se espera que la producción agrícola incremente alrededor de un 70 % a nivel mundial. Las aplicaciones tecnológicas en agricultura en los últimos años han permitido aumentar la producción y contribuir a la reducción de la contaminación, utilizar de forma sostenible los recursos naturales, prever costes, mitigar riesgos e incluso poder anticiparse a una potencial quiebra. En este sentido, uno de los retos de los próximos años a nivel global es diseñar metodologías que permitan a las empresas del sector agropecuario ser más eficientes y sostenibles independientemente del índice de desarrollo tecnológico de la región en la que se encuentren.

En esta investigación, tomando como caso de estudio el sector agropecuario, se ha diseñado y validado una metodología modular que combina algoritmos de aprendizaje automático y modelos econométricos orientada a la mejora de la gestión de los recursos, permitir a las empresas más competitivas y sostenibles para fomentar su inversión. Para desarrollar la metodología se han llevado a cabo tres experimentos. El primer experimento se ha orientado a medir la eficiencia de recursos a partir de un método no paramétrico para la estimación de las fronteras de producción en el que se han incorporado los costes derivados de las transferencias de datos, permitiendo identificar las fronteras de producción más óptimas teniendo en cuenta los costes tecnológicos. Para el segundo experimento se ha diseñado un sistema multi-agente para predecir oscilaciones de los precios en los mercados de futuros en productos derivados del sector agropecuario. El sistema multi-agente está diseñado como un sistema de apoyo a la toma de decisiones en el que los potenciales compradores o vendedores pueden incorporar parámetros de impacto medioambiental. Finalmente, el último experimento consiste en el diseño de una metodología de razonamiento basado en casos para la recomendación de inversión en una empresa. El último experimento permite incorporar capital a las empresas del sector agropecuario para invertir en tecnología. Además, se ha implementado una mejora sobre tercer experimento en el que se ha podido incrementar el rendimiento del sistema de recomendación de inversiones, en el que en la parte de clasificación se comparan distintas métricas de evaluación en las situaciones en los que los datos no están balanceados.

Abstract

By 2050, the world population is expected to reach 9 billion people. Global population growth will lead to an increase in consumer demand for products derived from agriculture. As a result, agricultural production is expected to increase by approximately 70% worldwide. Over the last years, technological applications have made it possible to increase agricultural production and contribute to reduce pollution, the sustainable use of natural resources, cost forecast, risk mitigation and even potential bankruptcy anticipation. In this regard, one of the global, near-future challenges is designing methodologies that enable companies in the agricultural sector to be more efficient and sustainable regardless of the technological development index of the region in which they are located.

In this research, taking the agricultural sector as a case study, has been designed and validated a modular methodology that combines machine learning algorithms and econometric models aimed at improving the management of resources, allowing companies to be more competitive and sustainable in order to encourage their investment. To develop the methodology, three experiments were carried out. The first experiment was aimed at measuring resource efficiency based on a non-parametric method for estimating production frontiers in which the costs derived from data transfers were incorporated, making it possible to identify the most optimal production frontiers taking into account technological costs. For the second experiment, a multi-agent system has been designed to predict price variations in futures markets for agricultural products. The multi-agent system is been designed as a decision support system in which potential buyers or sellers can incorporate environmental impact parameters. Finally, the last experiment consists of the design of a case-based reasoning methodology for the recommendation of investment in a company. The last experiment enables the incorporation of investment of capital to companies in the agricultural sector. In addition, to increase the performance of the investment recommender system an improvement has been implemented in the third experiment. This improvement has made it possible for the system's classification element to compare different evaluation metrics in situations where the data labels are not balanced.

Agradecimientos

En primer lugar, agradecer a mi director de tesis Juan Manuel Corchado su confianza depositada desde el inicio y la oportunidad de realizar la Tesis Doctoral en el grupo de investigación BISITE y en la Universidad de Salamanca.

A Parra, mi codirector de tesis con el que me he sentido muy acompañada. A Marta, por estar siempre para ayudarme con la tesis. A Sergio y Ludeiro, mis primeros amigos y compañeros de doctorado en Salamanca. A mis amigos del doctorado y equipo titular de los buenos momentos: Kike, Guille y Richi, gracias por toda vuestra ayuda siempre. Por tantos buenos momentos y los que vendrán. A mis compañeros de BISITE.

A mis padres y a mi hermano. A mi abuelita Ana y abuelito Francisco, y avi Miquel, porque estarían orgullosos de mí, y a mi abuelita Geni por su seguimiento activo en todos mis avances a lo largo de esta Tesis Doctoral.

Índice general

1. Introducción	1
1.1. Descripción del problema	6
1.2. Hipótesis y objetivos	8
1.3. Metodología de investigación	9
1.4. Estructura de la Tesis Doctoral	11
1. Introduction	13
1.1. Problem description	17
1.2. Hypothesis and objectives	19
1.3. Research Methodology	20
1.4. Doctoral Thesis structure	22
2. Estado del Arte	25
2.1. Introducción	29
2.2. Los modelos econométricos y el aprendizaje automático	29
2.2.1. Revisión sistemática de la literatura	33
2.2.2. Identificación de casos de uso	35
2.3. AgTech: Nuevas tecnologías aplicadas a la industria agropecuaria	36
2.3.1. Sostenibilidad y gestión de recursos	37
2.3.2. El internet de las cosas y el internet de las cosas industrial	38
2.3.3. Edge Computing	39
2.4. Metodologías para optimización de la producción en AgTech	40
2.4.1. Medidas de producción e implicaciones tecnológicas	40
2.4.2. Fronteras de producción	42
2.5. Sistemas de apoyo a la toma de decisiones	43
2.5.1. Sistemas de razonamiento basado en casos	44
2.5.2. Optimización multi-objetivo y multi-atributo	45
2.5.3. Procesos de decisiones secuenciales	46
2.6. Conclusiones	50
3. Contribuciones	53
3.1. Aprendizaje automático y modelos econométricos tradicionales: Un estudio de mapeo sistemático	55
3.2. Un caso práctico en la industria agroalimentaria a través de una plataforma Edge-IoT	58
3.2.1. Tecnología en el sector agropecuario	61
3.3. Sistemas de apoyo a la toma de decisiones basados en preferencias	63

3.3.1. Sistema multi-agente basado en preferencias y aprendizaje profundo para un mercado agrícola sostenible	64
3.3.2. Modelo híbrido de apoyo a la toma de decisiones en escenarios de inversión a partir de razonamiento basado en casos	66
3.4. Conclusiones	73
4. Evidencias y Resultados	75
4.1. Publicaciones	77
4.1.1. Publicaciones en revistas científicas internacionales	77
4.1.2. Publicaciones en congresos internacionales y workshops	79
4.2. Proyectos de investigación	79
4.3. Estancias internacionales	80
5. Conclusiones y Trabajo Futuro	81
5.1. Conclusiones	83
5.2. Futuras Líneas de Investigación	85
5. Conclusions and Future Work	87
5.1. Conclusions	89
5.2. Future Lines of Research	91
Bibliografía	93

Índice de figuras

1.1. Diagrama con los objetivos que se cubren en la presente Tesis Doctoral.	5
1.1. Diagram illustrating the objectives of this Doctoral Thesis.	17
2.1. Diagrama de clasificación general de inteligencia artificial.	31
2.2. Diagrama de clasificación general de aprendizaje automático.	32
2.3. Ejemplos de tipos y funcionamientos de aprendizaje automático.	34
2.4. Aplicaciones econométricas y de aprendizaje automático.	36
2.5. Diseño de una metodología basada en casos.	45
3.1. Tipos de contribución por industria.	57
3.2. Diagrama seguimiento eficiencia ambiental en explotaciones agrícolas.	59
3.3. Diagrama agricultura de precisión.	63
3.4. Sistema multi-agente propuesto	65

Siglas y acrónimos

AI	Artificial Intelligence
ANN	Artificial Neural Network
ANN	Agricultura Precisión
API	Application Programming Interface
ASIC	Application-Specific Integrated Circuit
CBR	Case Base Reasoning
CC	Cloud Computing
CS	Cloud Services
CV	Computer Vision
DDQN	Double Deep Q-Network
DL	Deep Learning
DLT	Distributed Ledger Technologies
DQN	Deep Q-Network
DRL	Deep Reinforcement Learning
DRNN	Deep Recurrent Neural Network
DRQN	Deep Recurrent Q-Network
DSS	Decision Support Systems
FMDP	Finitne Markov Decision Process
GECA	Global Edge Computing Architecture
GPU	Graphics Processing Unit
EC	Edge Computing
EN	Edge Nodes
IA	Inteligencia Artificial
IIC	Industrial Internet Consortium
IIoT	Industrial Internet of Things

IoT	I nternet o f T hings
JCR	J ournal C itation R eports
JSON	J ava S cript O bject N otation
LoRa	L ong R ange
LQI	L ink Q uality I ndicator
LSTM	L ong S hort- T erm M emory
MARL	M ulti- A gent R einforcement L earning
MAS	M ulti- A gent S ystem
MDP	M arkov D ecision P rocess
MEC	M obile E dge C omputing
MEMS	M icro- E lectro- M echanical S ystem
ML	M achine L earning
MLP	M ulti- L ayer P erceptron
NLP	N atural L anguage P rocessing
PaaS	P latform a s a S ervice
PA	P recision A griculture
REST	R Epresentational S tate T ransfer
RA	R eference A rchitecture
RF	R andom F orest
ReLU	R ectified L inear U nit
RFID	R adio- F requency I Dentification
RL	R einforcement L earning
RN	R edes N euronales (también <i>RRNN</i>)
RNA	R edes N euronales A rtificiales
RNN	R ecurrent N eural N etworks
SaaS	S oftware a s a S ervice
SARSA	S tate- A ction- R eward- S tate- A ction
SDN	S oftware- D efined N etwork
SE	S mart E nergy
SF	S mart F arming
SFC	S ervice F unction C haining
SL	S upervised L earning
SMP	S ystematic M apping S tudy

SN	S ensor N etwork
SVM	S upport V ector M achine
SVR	S upport V ector R egression
TPU	T ensor P rocessing U nit
USB	U niversal S erial B us
vCPU	v irtual C entral P rocessing U nit
w Wi-Fi	W ireless F idelity

Capítulo 1

Introducción



**VNiVERSiDAD
D SALAMANCA**

CAMPUS DE EXCELENCIA INTERNACIONAL

Introducción

La Organización de las Naciones Unidas para la Agricultura y la Alimentación estima que cada año se pierde entre el 20% y el 40% de la producción mundial de cultivos debido a plagas y enfermedades (King, 2017). Una agricultura óptima es fundamental para que la civilización humana siga prosperando. Esto se debe a que la agricultura debe satisfacer las necesidades básicas de la vida cotidiana. La industria agropecuaria debe adaptarse al mercado actual, utilizando los recursos de forma eficiente y respetando el medio ambiente. Las tecnologías de la información pueden contribuir a la transición de la sostenibilidad agroalimentaria aumentando la productividad de los recursos, reduciendo las ineficiencias, reduciendo los costes de gestión y mejorando la coordinación de la cadena alimentaria (Skvortsov, Skvortsova, Sandu, & Iovlev, 2018). Además, garantizar la seguridad alimentaria mundial, debido a las limitaciones de recursos naturales en muchos países (Bellmann & Hepburn, 2017).

En los procesos de producción propios de la industria agropecuaria intervienen múltiples factores tanto internos como externos. Los factores internos son aquellos que pueden ser gestionados por la propia explotación o empresa como, por ejemplo la optimización de los recursos propios para la producción. Por otro lado, un ejemplo de factores externos son los precios de venta en el mercado futuros o los fuertes cambios meteorológicos como consecuencia al cambio climático. La globalización y la posibilidad de exportar e importar, así como de competir en un mercado mayor, han llevado a un paradigma en el que los productos derivados del sector agropecuario son también un indicador financiero de la economía y participan en los precios de los mercados futuros (De Gorter, Drabik, & Just, 2015).

La intensificación sostenible de los sistemas agrícolas ofrece oportunidades sinérgicas para la coproducción de resultados agrícolas y del capital natural. La eficiencia y la

digitalización son pasos hacia la intensificación sostenible, pero el rediseño del sistema es esencial para obtener resultados óptimos a medida que cambian las condiciones ecológicas y económicas (Burliai, Nesterchuk, Nepochatenko, & Naherniuk, 2020). Gracias a los avances tecnológicos, las explotaciones agrícolas de los países desarrollados y en desarrollo pueden beneficiarse de la aplicación de tecnologías de bajo coste. En este sentido, el Internet de las Cosas (IoT) y, más concretamente, el Internet Industrial de las Cosas (IIoT), se presenta como una tecnología habilitadora clave para implementar y monitorizar soluciones de gestión de recursos en diversos escenarios de la Industria 4.0, incluyendo los entornos de agricultura inteligente (Sisinni, Saifullah, Han, Jennehag, & Gidlund, 2018). La monitorización de todos estos valores puede llevarse a cabo con sensores, utilizando nuevos paradigmas como el edge computing (EC) que permiten la monitorización a la vez que reducen el coste del análisis de datos en la nube (Alonso, Sittón-Candanedo, García, Prieto, & Rodríguez-González, 2020; Sittón-Candanedo, 2019) aumentando la eficiencia de los procesos.

Tradicionalmente se han utilizado análisis econométricos para realizar predicciones y optimizaciones de los recursos, así como algoritmos de inteligencia artificial para predicciones meteorológicas o de gestión de los recursos (Hatirli, Ozkan, & Fert, 2005). Sin embargo, hasta donde llega el conocimiento de la autora, no existe una metodología que desde un punto de vista de tecnología, inversión y optimización de recursos permita a las empresas o explotaciones del sector agropecuario ser más eficientes y sostenibles en términos económicos y de medioambiente.

En esta investigación, tomando como caso de estudio el sector agropecuario, se pretende diseñar una metodología híbrida que combine algoritmos de aprendizaje automático y modelos econométricos que esté orientada a mejorar la sostenibilidad, reducir los costes de la cadena de valor, prevenir la quiebra y fomentar la inversión en las empresas.

El objetivo de esta Tesis Doctoral es investigar una metodología que aúne técnicas de aprendizaje automático y modelos econométricos con el objeto de predecir el éxito de una explotación agropecuaria, maximizar su eficiencia energética y reducir los costes asociados al uso de tecnologías de información, así como recomendar precios de venta de los productos, permitiendo, así, que las pequeñas empresas agrícolas y ganaderas sean más sostenibles y competitivas.

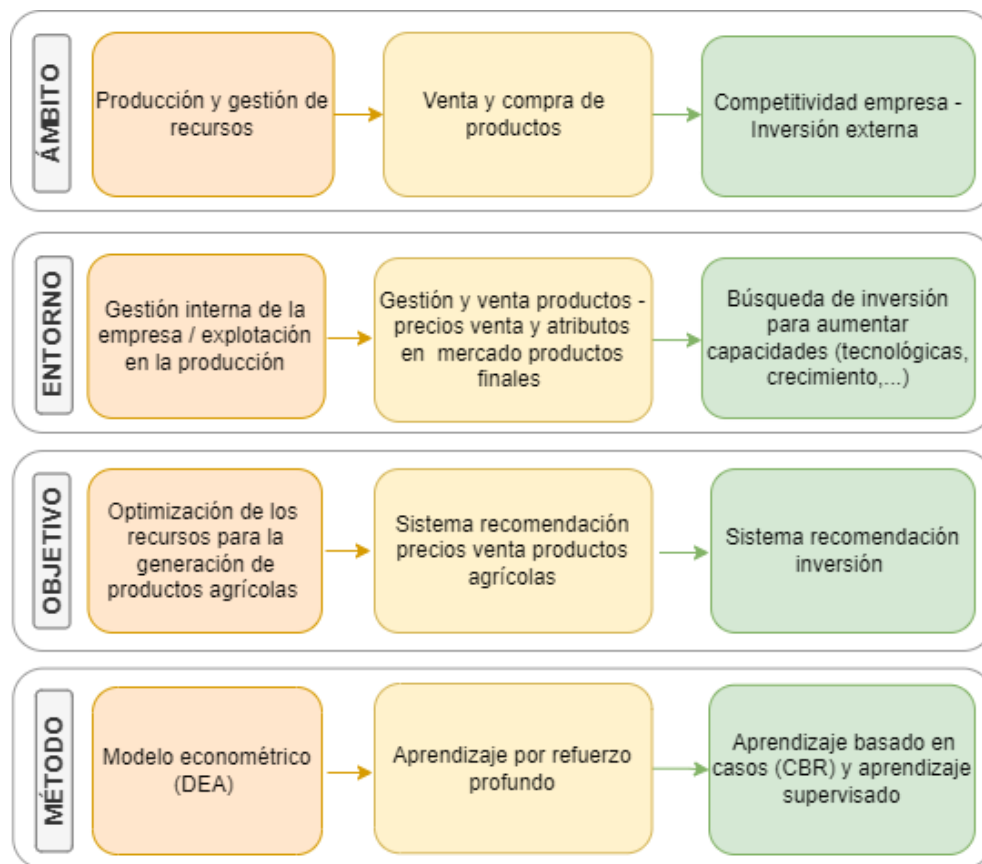


FIGURA 1.1: Diagrama con los objetivos que se cubren en la presente Tesis Doctoral.

Para desarrollar la metodología se han llevado a cabo tres experimentos consecutivos que se describen en la Figura 1.1. Tal y como se muestra en la Figura 1.1, el primer experimento se ha orientado a medir la eficiencia de recursos a partir de un método no paramétrico para la estimación de las fronteras de producción. Para el segundo experimento se ha diseñado un sistema multi-agente con aprendizaje por refuerzo profundo para predecir oscilaciones de los precios en los mercados bursátiles en productos agrícolas. Finalmente, el último experimento consiste en la aplicación de una metodología basada en casos para la recomendación de inversión en una empresa con base en unos criterios fijados por un potencial inversor. Además, se han ampliado los resultados de los experimentos y han contrastado los métodos tradicionalmente econométricos con los de aprendizaje supervisado para la predicción de bancarrota en diferentes industrias.

El presente Capítulo introductorio está estructurado de la siguiente forma. En la Sección 1.1 se ha realizado una introducción a los antecedentes del tema objeto de estudio y se describe el problema a tratar. La Sección 1.2 expone la hipótesis de trabajo, así como los objetivos de los distintos trabajos de investigación que han llevado a la elaboración

de esta Tesis Doctoral. La Sección 1.3 detalla la metodología seguida durante dichos trabajos de investigación. Finalmente, la Sección 1.4 describe la estructura de esta Tesis Doctoral.

1.1. Descripción del problema

La globalización ha provocado un crecimiento sin precedentes del flujo de bienes y capitales en las últimas décadas, que también ha afectado a la agricultura (Kastratović, 2019). Las empresas multinacionales, tanto de los países en desarrollo como de los países que carecen de tierras cultivables de los países desarrollados, se están convirtiendo en fuentes de inversión cada vez más importantes. Estas empresas, al invertir en la agricultura de los países en desarrollo, tienen como principal objetivo estabilizar la oferta de productos agrícolas en un entorno económico incierto de precios volátiles y creciente demanda de productos agrícolas (Hallam, 2011).

En períodos de crisis financiera, la volatilidad del mercado agrícola ha sido tradicionalmente elevada (Du, Cindy, & Hayes, 2011; Kurupparachchi & Premachandra, 2016). Esas fluctuaciones en los mercados de futuros de productos básicos (Kang, McIver, & Yoon, 2017) contribuyen a las oscilaciones de los precios tanto de compra como venta para las empresas del sector agropecuario. Sin embargo, no todos los productos del sector agrario se ven afectados por los mismos factores. El mercado agrícola es muy diverso en cuanto a los atributos y particularidades de los productos, ya que están sujetos a factores externos. Por ejemplo, los precios de los futuros del maíz pueden estar determinados por ciertos factores (como la ubicación, los costes de transporte, la contaminación, los almacenes, las condiciones meteorológicas, etc.) en comparación con los precios del trigo, que son menos sensibles a la meteorología. Los cambios en el clima son algo que se ha acentuado en los últimos años debido al efecto invernadero (Agovino, Casaccia, Ciommi, Ferrara, & Marchesano, 2019). El análisis de las emisiones de efecto invernadero, así como el gasto de recursos limitados o el consumo de energía, se está generalizando en la agricultura de precisión (AP) (Eshel, Shepon, Makov, & Milo, 2014) y la agricultura inteligente (Sisinni et al., 2018). Concretamente, la AP se refiere al concepto de utilizar las nuevas tecnologías para aumentar el rendimiento y la rentabilidad de los cultivos, reduciendo al mismo tiempo

los recursos necesarios para el cultivo (Skobelev, Budaev, Gusev, & Voschuk, 2018; N. Zhang, Wang, & Wang, 2002).

La agricultura mundial alimenta a más de 7 000 millones de personas, pero también es una de las principales causas de la degradación del medio ambiente (Clark & Tilman, 2017). Los estudios realizados hasta la fecha sugieren que el cambio climático ha reducido el crecimiento del rendimiento de los cultivos entre un 1 % y un 2 % por década durante el último siglo, y se prevé que los impactos adversos aumenten en el futuro (Change, 2014). Sin embargo, entender la magnitud de estos impactos es complicado por la interacción de numerosos factores biofísicos y socioeconómicos.

Estudios anteriores han combinado modelos climáticos, de cultivos y económicos para examinar el impacto del cambio climático en la producción agrícola y la seguridad alimentaria, pero los resultados han variado mucho debido a las diferencias en los modelos, los escenarios y los datos de entrada (Hunter, Smith, Schipanski, Atwood, & Mortensen, 2017). Comprender el cambio climático, especialmente los efectos de la temperatura, es fundamental para que se pueda garantizar la seguridad alimentaria mundial (Zhao et al., 2017).

El crecimiento de la agricultura y su productividad se consideran esenciales para lograr un crecimiento sostenible y una reducción significativa de la pobreza en los países en desarrollo (Beddington et al., 2012). El crecimiento de la productividad en el sector agrícola es fundamental para que la producción agrícola aumente a un ritmo lo suficientemente rápido como para hacer frente a la pobreza (Conway & Barbier, 2013). Teniendo en cuenta la disminución de la tierra cultivable per cápita, los elevados costes de producción, combinados con el rápido crecimiento de la población y la consiguiente necesidad de asentamientos humanos, así como la creciente urbanización, se requieren mejoras significativas en el crecimiento de la productividad en la agricultura para aumentar la producción agrícola mediante innovaciones tecnológicas.

Los productos derivados del sector agropecuario son una de las fuentes principales de subsistencia para el ser humano. Es por ello necesario que se desarrollen metodologías para que las empresas o explotaciones del sector agropecuario sean más eficientes, más sostenibles y también puedan fomentar inversiones de capital externas para ampliar sus posibilidades.

1.2. Hipótesis y objetivos

En esta investigación, tomando como caso de estudio el sector agropecuario, se pretende validar una metodología híbrida que combine algoritmos de aprendizaje automático y modelos econométricos y que esté orientada a mejorar la sostenibilidad, reducir los costes de la cadena de valor y recomendar inversiones tanto a nivel producto como a nivel empresa.

Así, el **objetivo principal** de esta Tesis Doctoral es investigar, diseñar e implementar una metodología basada en técnicas de inteligencia artificial y modelos econométricos que contribuya al crecimiento de la productividad en la agricultura mediante innovaciones tecnológicas orientada a la mejora de la gestión de los recursos y hacer las empresas más competitivas y sostenibles para fomentar su inversión.

Para alcanzar el objetivo principal, es necesario definir un listado de **objetivos específicos**, que se describen a continuación:

- (OB1) Realizar un estudio del estado del arte en la combinación en el uso de técnicas de aprendizaje automático con modelos econométricos para comprobar las aplicaciones en diferentes industrias y objetivos en los que se han planteado diferentes casos de estudio.
- (OB2) Identificar los requerimientos y limitaciones para optimizar los recursos en empresas del sector agropecuario con el uso de nuevas tecnologías. Además de analizar las posibilidades y retos que sucederán en los próximos años.
- (OB3) Aplicar técnicas que permitan reducir los costes asociados al uso de nuevas tecnologías de información en las empresas / explotaciones del sector agropecuario.
- (OB4) Analizar los retos, motivaciones y problemas abiertos que surgen en la compra y venta de productos agrícolas.
- (OB5) Diseñar y validar una metodología que ayude a reducir los riesgos para las empresas del sector agropecuario a la hora de definir precios de venta de productos agrícolas. Entre ellos incluir elementos que fomenten una agricultura

sostenible y permitan vender productos en base a criterios de medición de sostenibilidad.

(OB6) Implementar y evaluar un sistema de apoyo a la toma de decisiones que fomente la recomendación e inversión de empresas en múltiples sectores.

(OB7) Confeccionar un plan de trabajo futuro de cara a facilitar que las pequeñas empresas agrícolas y ganaderas sean más sostenibles y competitivas.

1.3. Metodología de investigación

La metodología *action-research* (traducida al castellano habitualmente como *investigación-acción*) ha ganado importancia en los proyectos de Ingeniería de Software, donde, por sus características, su aplicación resulta beneficiosa para la investigación de campo, ya que simultáneamente permite la realización de investigaciones y acciones. La *acción* generalmente se asocia con una transformación en una determinada comunidad, organización o proyecto, mientras que la *investigación* se distingue por una amplia comprensión de un fenómeno de cambio por parte del investigador (comunidad investigadora), persona (cliente) o ambos. Baskerville (1999) ha indicado que la metodología *action-research* es el tipo de estudio donde se encuentra el escenario de mayor realismo al involucrar un contexto real para investigar los resultados de acciones concretas.

De acuerdo con Baskerville (1999), citado por Tüzün, Tekinerdogan, Macit, & İnce (2019), los estudios desarrollados bajo esta metodología comparten cuatro características comunes:

1. Una orientación a la acción y al cambio;
2. Un enfoque del problema;
3. Un proceso orgánico que implica etapas sistemáticas y, a veces, iterativas;
4. Así como la colaboración entre los participantes.

En general, la investigación-acción consta de cinco fases (Eden & Ackermann, 2018; Tüzün et al., 2019) que, para el desarrollo de esta Tesis Doctoral, se describen a continuación:

1. **Diagnóstico:** Esta fase se corresponde con la identificación y planteamiento del problema expuesto en esta Tesis Doctoral. Es necesario delimitar el alcance, proponer una hipótesis y definir los objetivos que se pretenden alcanzar. A este respecto, se identificarán y analizarán los diferentes requisitos existentes en entornos agropecuarios y el potencial beneficio de la incorporación de aprendizaje automático en combinación con modelos econométricos. Esto incluirá el análisis de los retos, motivaciones y problemas abiertos a la hora de aplicar técnicas de aprendizaje automático supervisado, aprendizaje por refuerzo y aprendizaje por refuerzo profundo en escenarios de la industria agropecuaria.
2. **Revisión de la literatura:** Se realiza un estudio del estado del arte con el objeto de identificar las soluciones existentes al problema planteado y sentar las bases para el diseño de una solución innovadora. En este trabajo de investigación se incluye una revisión del estado del arte en la aplicación de técnicas econométricas y de aprendizaje automático para diferentes escenarios en los que se comparan los resultados alcanzados por los algoritmos, los diferentes sectores de aplicación y los métodos y modelos utilizados.
3. **Solución:** A partir de los resultados obtenidos en las dos fases anteriores, se propone una solución que responda a los objetivos establecidos. En esta Tesis Doctoral se plantean diferentes escenarios a partir de los cuales las empresas del sector agropecuario pueden utilizar de forma más óptima sus recursos, pueden vender sus productos basándose en predicciones realizadas con aprendizaje profundo y teniendo en cuenta precios de mercado reales y preferencias de los compradores, así como un sistema para recomendar inversiones con una metodología basada en casos.
4. **Evaluación:** En esta fase se lleva a cabo la evaluación de los procesos y metodologías propuestas. Además de la incorporación de mejoras en los sistemas propuestos en la solución.
5. **Resultados:** Comprende la última fase del ciclo investigación-acción. Por medio de las actividades de esta etapa, la comunidad científica valora los resultados obtenidos en base a un conjunto de publicaciones que se someten a evaluación en revistas de alto impacto, incluyendo *Sensors*, *Sustainability*, *Journal of Artificial Intelligence and Soft Computing Research* , *Knowledge Engineering Review*,

así como congresos y conferencias internacionales o participación en doctoral consortium como la International Conference on Case-Based Reasoning (ICCBR) o la International conference on Decision Economics (DECON), entre otras.

1.4. Estructura de la Tesis Doctoral

La estructura de esta Tesis Doctoral por compendio de artículos/publicaciones incluye la presentación de las evidencias y resultados originados en el trabajo de investigación llevado a cabo durante la elaboración de esta Tesis Doctoral, así como las publicaciones acreditativas de la investigación realizada. Finalmente, se exponen las conclusiones y las líneas de trabajo futuras.

De este modo, y tal como se ha adelantado, con el fin de facilitar el seguimiento de la investigación, se estructura la memoria de la presente Tesis Doctoral a través de cinco Capítulos. El primero de ellos se corresponde con la actual Introducción (Capítulo 1), en la que se describe el problema a resolver planteado, se formula la hipótesis de trabajo, se detallan los diferentes objetivos y se describe la metodología de investigación seguida para la culminación de esta Tesis Doctoral.

El Capítulo 2 presenta una revisión del estado del arte en el que se analizan progresivamente los enfoques existentes de aprendizaje automático y modelos econométricos, el concepto de AgTech, así como los mecanismos inteligentes de apoyo a la toma de decisiones para la gestión de recursos.

El Capítulo 3 describe las contribuciones resultantes de la investigación llevada a cabo y que se compendia en esta Tesis Doctoral. Para desarrollar la metodología se han llevado a cabo tres experimentos. El primer experimento se ha orientado a medir la eficiencia de recursos a partir de un método no paramétrico para la estimación de las fronteras de producción. En segundo lugar se ha diseñado un sistema multi-agente con aprendizaje por refuerzo para predecir oscilaciones de los precios en los mercados bursátiles en productos agrícolas. Finalmente, el último experimento consiste en la recomendación de inversión en una empresa con base en unos criterios fijados por un potencial inversor.

El Capítulo 4 presenta los resultados alcanzados mediante el desarrollo de la investigación que culmina en esta Tesis Doctoral y que se validan a través de un conjunto de

publicaciones en revistas científicas, Capítulos de libro, conferencias, congresos y workshops internacionales en los cuales se ha contribuido, así como los proyectos de investigación en los que se ha participado.

Finalmente, el Capítulo 5 detalla las principales conclusiones que se han obtenido a partir del trabajo de investigación desarrollado en esta Tesis Doctoral y las aportaciones más relevantes realizadas. Además, se definen las líneas para el desarrollo de trabajo futuro que se abren a partir de los resultados de la presente Tesis Doctoral. Para concluir, en el anexo se incluye una copia de las publicaciones acreditativas y en el último lugar se incluye el listado de todas las fuentes bibliográficas que se han citado en esta memoria con el propósito de respaldar las afirmaciones y conceptos que se presentan.

Chapter 1

Introduction



**VNiVERSiDAD
D SALAMANCA**

CAMPUS DE EXCELENCIA INTERNACIONAL

Introduction

The Food and Agriculture Organisation of the United Nations estimates that between 20-40% of the world's crop production is lost each year due to pests and diseases. Optimal agriculture is essential for human civilisation to continue to thrive. This is because agriculture must meet the basic needs of daily life. The agricultural industry must adapt to today's market, using resources efficiently and respecting the environment. Information technologies can help transition to agri-food sustainability by increasing resource productivity, reducing inefficiencies, reducing management costs and improving the coordination of the food chain (Skvortsov et al., 2018), as well as ensuring global food security in the face of natural resource constraints in many countries (Bellmann & Hepburn, 2017).

Multiple internal and external factors are involved in the processes of the agricultural industry. Internal factors are those ones that can be managed by the farm or company itself, such as optimising the use of its resources in production. An example of external factors would be the selling prices on the futures market or strong meteorological changes as a consequence of climate change. Globalisation and the possibility of exporting and importing, as well as competing on a larger market, have led to a paradigm in which products derived from the agricultural sector are a financial indicator of the economy and participate in futures market prices.

The sustainable intensification of agricultural systems offers synergistic opportunities for the co-production of agricultural outputs and natural capital. Efficiency and digitisation are the steps towards sustainable intensification, however, system redesign is essential for optimal outcomes as ecological and economic conditions change (Burliai et al., 2020). Thanks to technological advances, farms in developed and developing countries can benefit from the application of low-cost technologies. In this regard, the Internet of

Things (IoT) and, more specifically, the Industrial Internet of Things (IIoT), is a key enabling technology for the implementation and monitoring of resource management solutions in various Industry 4.0 scenarios, including smart farming environments. Parameter values can be monitored through sensors, using new paradigms such as edge computing (EC) that enable monitoring while reducing the cost of data analysis in the cloud (Alonso, Sittón-Candanedo, García, et al., 2020; Sittón-Candanedo, 2019), and increasing the efficiency of processes.

Traditionally, econometric analyses have been used for resource prediction and optimisation while artificial intelligence algorithms have been used for weather forecasting or resource management. However, there is no methodology that, from an investment and resource optimisation point of view, allows agricultural companies or farms to become more efficient and sustainable in what regards economic and environmental aspects.

Taking the agricultural sector as a case study, this research aims to design a hybrid methodology that combines machine learning algorithms and econometric models to improve sustainability, reduce costs in the value chain, prevent bankruptcy and encourage investment in them.

The aim of this Doctoral Thesis is to investigate a methodology that combines machine learning techniques and econometric models for the prediction of the success of a farm, maximising its energy efficiency and reducing the costs associated with the use of information technologies. Moreover, this methodology is to be capable of recommending the prices at which to sell products, thus enabling small agricultural and livestock enterprises to become more sustainable and competitive.

To develop the methodology, three consecutive experiments have been carried out, as described in Figure 1.1. The first experiment is aimed at measuring resource efficiency on the basis of a non-parametric method for estimating production frontiers. In the second experiment, a multi-agent system has been designed that incorporates deep reinforcement learning for the prediction of price movements on agricultural commodity stock markets. Finally, the last experiment involved the application of a case-based methodology for the recommendation of investment in a company based on criteria set by a potential investor. In addition, the results have been extended, and the bankruptcy

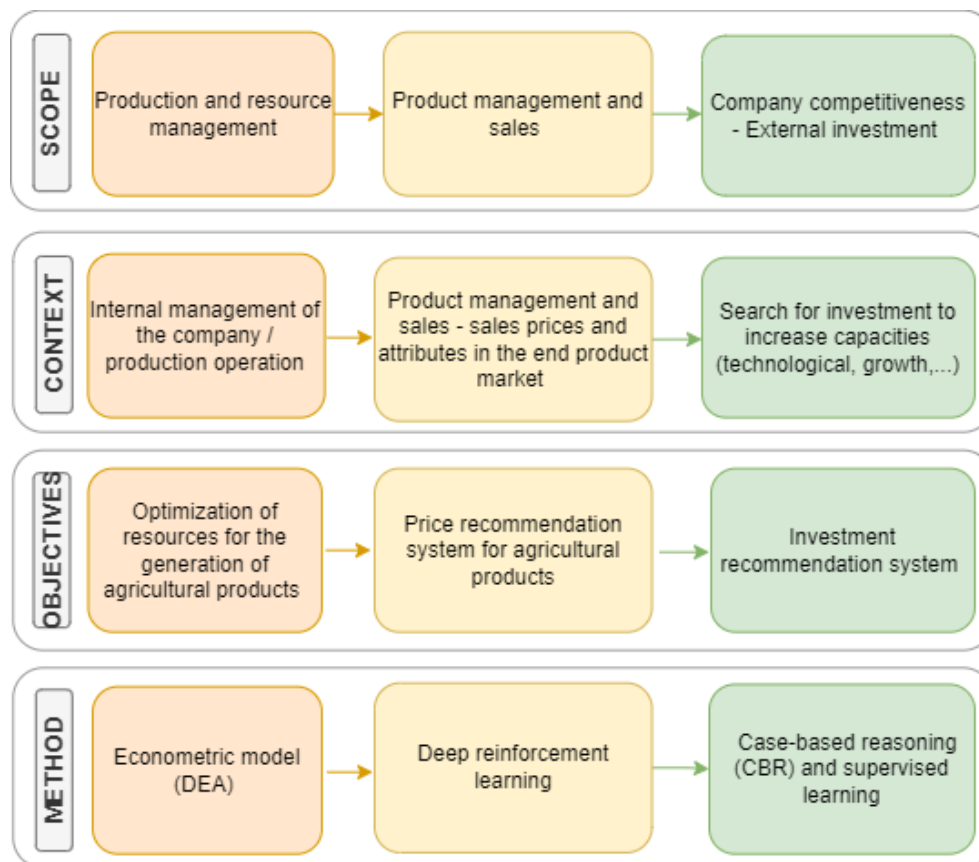


FIGURE 1.1: Diagram illustrating the objectives of this Doctoral Thesis.

prediction capabilities of traditional econometric methods have been contrasted with those of supervised learning methods.

This introductory chapter is structured as follows. Section 1.1 provides background on the subject under study and describes the problem to be addressed. Section 1.2 states the working hypothesis, as well as the objectives of the different research works that have led to the elaboration of this Doctoral Thesis. Section 1.3 details the methodology followed during the research work. Finally, Section 1.4 describes the structure of this Doctoral Thesis.

1.1. Problem description

Globalization has led to an unprecedented growth in the flow of goods and capital in recent decades, which has also affected agriculture (Kastratović, 2019). Multinational companies from developing countries (especially China and India), countries lacking arable land (such as the Gulf states) and developed countries are becoming increasingly

important sources of investment. By investing in agriculture in developing countries, the main objective of these companies is to stabilize the supply of agricultural products in an uncertain economic environment of volatile prices and growing demand for agricultural products (Hallam, 2011).

In periods of financial crisis, agricultural market volatility has been very high (Du et al., 2011; Kuruppuarachchi & Premachandra, 2016). These fluctuations in the commodity futures markets (Kang et al., 2017) contribute to swings in both buying and selling prices for companies in the agricultural sector. However, not all agricultural commodities are affected by the same factors; the agricultural market is very diverse in terms of attributes and particularities of products that are subject to external factors. For example, corn futures prices can be determined by certain factors (such as location, transportation costs, contamination, warehousing, weather conditions, etc.) compared to wheat prices, which are less sensitive to weather. Changes in weather have become more visible in recent years due to the greenhouse effect (Agovino et al., 2019). The analysis of greenhouse emissions, as well as the expenditure of limited resources or energy consumption, is becoming widespread in precision agriculture (PA) (Eshel et al., 2014) and intelligent agriculture (Sisinni et al., 2018). Specifically, PA refers to the concept of using new technologies to increase crop yields and profitability while reducing the resources required for cultivation (Skobelev et al., 2018; N. Zhang et al., 2002).

Global agriculture feeds more than 7 billion people. By 2050, the world's population is expected to have reached 9 billion people. Global population growth is going to lead to a growth in demand for consumer products derived from agriculture. As a result, the production of agricultural products is expected to increase by around 70% globally.

However, agriculture is also identified as a major cause of environmental degradation (Clark & Tilman, 2017). Studies to date suggest that climate change has reduced crop yield growth by 1 to 2 percent per decade over the last century, and adverse impacts are expected to increase in the future (Change, 2014). However, understanding the magnitude of these impacts is complicated because numerous biophysical and socioeconomic factors intertwine.

Previous studies have combined climate, crop, and economic models to examine the impact of climate change on agricultural production and food security, but results have varied widely due to differences in models, scenarios, and input data (Hunter et al.,

2017). Understanding climate change, especially its effects on temperature, is critical if global food security is to be assured (Zhao et al., 2017).

Growth in agriculture and its productivity are considered essential to achieving sustainable growth and reducing poverty in developing countries (Beddington et al., 2012). The increase in the productivity of the agricultural sector is critical for agricultural yield to grow at a fast enough pace to address poverty (Conway & Barbier, 2013). Given the decline in arable land per capita, high production costs, combined with rapid population growth and the consequent need for human settlement and urbanization, significant improvements in the growth of agricultural productivity are required to increase agricultural production through technological innovations.

Products derived from the agricultural sector are one of the main sources of human livelihoods, and it is therefore necessary to develop methodologies to make agricultural enterprises and farms more efficient, more sustainable and also to enable them to have external capital investments that will expand their possibilities.

1.2. Hypothesis and objectives

In this research, taking the agricultural sector as a case study, the aim is to design and validate a methodology that combines supervised learning algorithms and econometric models and that is oriented towards improving sustainability, reducing value chain costs and recommending investments at both product and company levels.

Thus, the main objective of this Doctoral Thesis is to investigate, design and implement a methodology based on artificial intelligence techniques and econometric models that contributes to productivity growth in agriculture through technological innovations aimed at improving resource management and making companies more competitive and sustainable to encourage their investment.

To achieve the main goal of this Doctoral Thesis, it is necessary to define a list of specific objectives, which are described below:

(OB1) To conduct a review of the state of the art on the combined use of machine learning techniques with econometric models and to perform a comparative

study of the industries and the objectives with which these case studies has been proposed.

- (OB2) To identify the requirements and limitations associated with the optimization of resources, through the use of new technologies, in companies operating within the agricultural sector.
- (OB3) To apply techniques for the reduction of the costs associated with the use of information technologies in agricultural companies or farms.
- (OB4) To analyze the challenges, motivations and unresolved problems that arise in the purchase and sale of agricultural products.
- (OB5) To design and validate a methodology for companies in the agricultural sector that will reduce the risk involved in defining the selling prices of their products. These include elements that promote sustainable agriculture and allow products to be sold on the basis of sustainability measurement criteria.
- (OB6) To implement and evaluate a decision support system that recommends investment in companies across multiple sectors.
- (OB7) To create a roadmap for the future to enable small agricultural and livestock enterprises to become more sustainable and competitive.

1.3. Research Methodology

The *action-research* methodology has gained importance in Software Engineering projects, where, due to its characteristics, its application is beneficial for field research, as it simultaneously enables research and action to be carried out. Action is generally associated with a transformation in a given community, organization or project, while research is distinguished by a broad understanding of a change phenomenon by the researcher (research community), person (client) or both. Baskerville (1999) has indicated that the *action-research* methodology is the type of study where the most realistic scenario is found by involving a real context in which to investigate the results of concrete actions.

According to Baskerville (1999), cited by Tüzün et al. (2019), studies developed under this methodology share four common characteristics:

1. An orientation to action and change;
2. A problem focus;
3. An organic process involving systematic and sometimes iterative stages;
4. As well as collaboration between participants.

In general, action research consists of five phases, which, for the purposes of this dissertation, are described below:

1. **Diagnosis:** This phase corresponds to the identification and statement of the problem set out in this Doctoral Thesis. It is necessary to delimit the scope, propose a hypothesis and define the objectives to be achieved. In this regard, the different requirements in agricultural environments and the potential benefit of incorporating machine learning in combination with econometric models will be identified and analysed. This will include the analysis of the challenges, motivations and open problems in applying supervised machine learning, reinforcement learning and deep reinforcement learning techniques in agricultural industry scenarios.
2. **Literature review:** A state-of-the-art study is carried out in order to identify existing solutions to the problem and to lay the foundations for the design of an innovative solution. This research work includes a review of the state of the art in the application of econometric and machine learning techniques for different scenarios in which the results and the different sectors and uses of econometric and machine learning techniques are compared.
3. **Solution:** On the basis of the results obtained in the two previous phases, a solution is proposed that responds to the established objectives. In this Doctoral Thesis, different scenarios are proposed from which companies in the agricultural sector can use their resources more optimally, can sell their products on the basis of the predictions made with deep learning, which take into account real market prices and buyer preferences, as well as a system for recommending investments with a case-based methodology.

4. **Evaluation:** In this phase, the evaluation of the proposed experiments and methodologies is carried out. In addition to the incorporation of improvements in the systems proposed in the solution.
5. **Results:** This is the last phase of the action research cycle. Through the activities carried out at this stage, the scientific community evaluates the obtained results on the basis of a set of publications that are submitted for evaluation in high-impact journals, including *Sensors*, *Sustainability*, *Journal of Artificial Intelligence and Soft Computing Research*, *Knowledge Engineering Review*, as well as international congresses and conferences or participation in doctoral consortium such as the International Conference on Case-Based Reasoning (ICCBR) or the International conference on Decision Economics (DECON), among others.

1.4. Doctoral Thesis structure

The structure of this Doctoral Thesis by compendium of articles/publications includes the presentation of the evidence and results originating from the research work carried out during the preparation of this Doctoral Thesis, as well as the publications accrediting the research that has been carried out. Finally, the conclusions and future lines of research are presented.

In this way, and as has already been mentioned, in order to facilitate the monitoring of the research included in this Doctoral Thesis easier to follow, it is structured into five chapters. The first one corresponds to the present Introduction (Chapter 1), in which the problem to be solved has been described, the working hypothesis has been formulated, the different objectives have been set out and the research methodology followed throughout this Doctoral Thesis has been outlined.

Chapter 2 presents a review of the state of the art of approaches to machine learning and econometric models, the concept of AgTech, as well as intelligent decision support mechanisms for resource management

Chapter 3 describes the contributions resulting from the conducted research, which are summarised in this Doctoral Thesis. To develop the methodology, three experiments have been carried out. The first experiment was aimed at measuring resource efficiency

using a non-parametric method for the estimation of production frontiers. Secondly, a multi-agent system with reinforcement learning has been designed to predict price oscillations of agricultural products on stock markets. Finally, the last experiment consists of recommending an investor a company to invest in on the basis of the criteria set by them or using the criteria of other investors with a similar profile.

Chapter 4 presents the results obtained from the researches that culminate in this doctoral thesis and which are validated through a set of publications in scientific journals, book chapters, conferences, congresses and international workshops, as well as the research projects in which they have been included.

Finally, Chapter 5 draws conclusions from the research carried out in this Doctoral Thesis and details its most relevant contributions. In addition, future lines of research are defined on the basis of the results presented in this Doctoral Thesis. To conclude, the appendix includes a copy of the accredited publications and, lastly, a list of all the bibliographical sources that have been cited throughout this report to support statements and concepts, has been provided.

Capítulo 2

Estado del Arte



**VNiVERSiDAD
D SALAMANCA**

CAMPUS DE EXCELENCIA INTERNACIONAL

Estado del Arte

La globalización ha llevado a un nuevo paradigma en el que las industrias tradicionales, como la agropecuaria, emplean tecnologías de vanguardia para ampliar sus posibilidades en lo que se conoce como agricultura inteligente e industria agroalimentaria 4.0. Esta industria debe adaptarse al mercado actual mediante un uso eficiente de los recursos y respetuoso con el medio ambiente.

Para los países de la Unión Europea (UE), la Política Agrícola Común (PAC) (Gardner, 1996) gestiona y financia, a nivel europeo, los recursos del presupuesto de la UE relativos al apoyo a las rentas de los agricultores, la orientación del mercado y el medio ambiente. La PAC también regula la evolución de las cuotas de las industrias agrícolas. Por ejemplo, en 2015 con la supresión de la cuota láctea y las políticas correspondientes (“European Commission, The End of Milk Quotas”, 2015), o más recientemente, el fin de la cuota del azúcar en 2017 (Wimmer & Sauer, 2020), así como la evaluación del entorno comercial post-cuota. A pesar de que las cuotas han llegado a su fin, hay importantes preocupaciones entre los lobbies agrícolas y los efectos reales en el cambio de tiempo de esas cuotas, así como el impacto económico de años con un mercado en aparente competencia perfecta (Cathagne, Guyomard, & Levert, 2006; Philippidis & Waschik, 2019).

Los productores agrícolas del siglo XXI se enfrentarán al reto de lograr la seguridad y la eficiencia alimentaria. Esto debe lograrse al tiempo que se garantizan sistemas agrícolas sostenibles y se superan los problemas que plantean el cambio climático, el agotamiento de los recursos hídricos y la posibilidad de que aumente la erosión y la pérdida de productividad debido a condiciones meteorológicas extremas. Estas consecuencias medioambientales afectan directamente tanto al proceso de fijación de precios de venta como a los costes derivados de las actividades principales.

Es por ello que dentro de los costes de las actividades principales y en el análisis de las eficiencias en la producción se deben incluir los costes de las tecnologías implementadas.

Gracias a la interdependencia económica mundial, hoy en día se dispone de mucha más información para realizar predicciones. Estas grandes cantidades de datos requieren diferentes tipos de métodos para el procesamiento óptimo de la información. Los modelos y aplicaciones económicas tradicionales requieren herramientas con mayor capacidad de cálculo y nuevas metodologías de predicción que proporcionen pronósticos más precisos, como los que se pueden realizar con aprendizaje automático.

El uso de las nuevas tecnologías y grandes cantidades de datos que se pueden recoger en el sector agropecuario permiten que las empresas sean más competitivas, sin embargo, los costes derivados de las tecnologías que se implementan para poder monitorizar y realizar predicciones es un coste que cada vez es más importante. Es por ello necesario empezar a realizar estudios y metodologías para la optimización de recursos que incluyan paradigmas como la computación en el borde o EC que permitan reducir los costes de las transacciones de los datos a la nube.

Así, en el resto de este Capítulo se describen progresivamente los enfoques existentes en las aplicaciones tecnológicas en agricultura, los modelos tradicionalmente econométricos y los algoritmos de aprendizaje automático que se emplean para mejorar eficiencias en agricultura, así como los diferentes métodos y sistemas de recomendación que permiten a las empresas del sector agropecuario ser más eficientes y competitivas.

La Sección 2.1 expone un primer acercamiento a las tecnologías bajo estudio y analiza las revisiones existentes sobre el aprendizaje automático, los modelos econométricos y el escenario del sector agropecuario. A continuación, la Sección 2.3 analiza los enfoques existentes en el campo de las aplicaciones tecnológicas en agricultura desde el punto de vista de sostenibilidad y la posibilidad de parametrizar información para una adecuada gestión de los recursos gracias al Internet de las Cosas y el Internet de las Cosas Industrial. La Sección 2.4, por su parte, describe diferentes medidas para optimización de producción y métodos para determinar fronteras de producción. Finalmente, la Sección 2.5 recoge diferentes ejemplos de sistemas de recomendación que permiten a las empresas del sector agropecuario ser más competitivas en fijación de precios y potenciales inversores.

2.1. Introducción

La intensificación sostenible de los sistemas agrícolas ofrece oportunidades sinérgicas para la coproducción de resultados agrícolas y del capital natural. La eficiencia y la sustitución son pasos hacia la intensificación sostenible, pero el rediseño del sistema es esencial para obtener resultados óptimos a medida que cambian las condiciones ecológicas y económicas. La digitalización es uno de los procesos de transformación en curso más importantes en la agricultura y las cadenas alimentarias mundiales.

Las nuevas tecnologías y los incentivos para la producción sostenible han tenido un impacto positivo en términos económicos en las empresas de la industria agropecuaria (Balafoutis et al., 2017). En los últimos años, las empresas se están centrande en utilizar proveedores del sector agropecuario que cumplan los requisitos de sostenibilidad y no solo en considerar la eficiencia del producto en sí (Kamble, Gunasekaran, & Gawankar, 2020).

Por tanto, ser una explotación sostenible se ha convertido en un valor añadido, no solo en términos de reducción de gastos, sino también por ser apto para nuevos socios. El último proyecto europeo Horizonte, lanzado por la PAC, tiene como objetivo constituir en legislación la ambición política de ser el primer continente neutro desde el punto de vista climático para el año 2050 (“European Commission Horizon 2050”, 2019).

La adición de sistemas de apoyo a la decisión o (Decision Support Systems - DSS) a la AP representa una contribución más a una óptima toma de decisiones (Zhai, Martínez, Beltran, & Martínez, 2020). Mientras que los niveles de producción y la efectividad en las fronteras de producción incorporan variables más alineadas con el entorno y la situación macroeconómica, cuando se trata de medir la rentabilidad, cualquier inestabilidad económica puede ser un factor determinante en la Ecuación (Machek & Špička, 2014).

2.2. Los modelos econométricos y el aprendizaje automático

En términos generales, la econometría engloba el conjunto de métodos estadísticos aplicados a los datos económicos y financieros para proporcionar apoyo empírico a las

teorías económicas. En la práctica, sin embargo, este conjunto de métodos estadísticos se centra tradicionalmente en el modelo de regresión lineal multivariante. En los últimos 100 años, los modelos de regresión lineal multivariante han sido populares por varias razones: los conjuntos de datos económicos han tendido a ser numéricos, cortos, pequeños y con una baja relación señal-ruido. Las limitaciones de los datos justificaron a menudo el uso de especificaciones relativamente limitadas.

La modelización de la regresión lineal multivariante no es una tecnología nueva. Su historia se remonta al menos a 1795, cuando Carl Friedrich Gauss aplicó los mínimos cuadrados ordinarios (MCO) a conjuntos de datos geodésicos y astronómicos (Stigler, 1981). El eugenista británico Sir Francis Galton acuñó el término regresión.^{en} 1886, al estimar ecuaciones lineales para argumentar que los rasgos físicos y morales humanos hereditarios muestran una regresión hacia la media. En los años siguientes, Ronald Fisher estudió y demostró las propiedades matemáticas del análisis de regresión y popularizó la estimación de máxima verosimilitud. Estas ideas dieron origen a gran parte de lo que hoy estudiamos con el nombre de econometría. El concepto de “Econometría”, fue introducido por primera vez por Pawel Ciompi a principios del siglo XX. Después de él, Jan Tinbergen, fue uno de los primeros investigadores en aplicar las matemáticas en la comprobación de las hipótesis económicas (Tinbergen, 1962). Su canon se ejemplifica mejor en el contenido de los libros de texto estándar de econometría (Wooldridge, 2010).

Los nuevos conjuntos de datos disponibles incluyen las redes sociales, los metadatos extraídos de los sitios web, las imágenes de satélite, los datos de los sensores o el análisis de sentimiento extraído del texto. La Inteligencia Artificial (IA), y concretamente los algoritmos de aprendizaje automático son capaces de procesar estos datos tan variados y aprender patrones complejos en un espacio de alta dimensión con poca especificación del modelo.

La inteligencia artificial pretende desarrollar y utilizar sistemas informáticos que intenten replicar los procesos de la inteligencia humana. Estos procesos de la inteligencia humana son necesarios para aprender, comprender, resolver problemas y tomar decisiones (Panesar, 2019). Como se muestra en la Figura 2.1, la IA es, por tanto, una amplia disciplina que reúne varios campos como sistemas expertos (Dorri, Kanhere, & Jurdak, 2018), los sistemas recomendadores (García, Prieto, Alonso, & Corchado, 2017), los sistemas multi-agente (Tapia, Fraile, Rodríguez, Alonso, & Corchado, 2013),

el procesamiento del lenguaje natural (Chowdhury, 2003), el análisis de voz (Partila et al., 2018), la visión por ordenador (Seo, Han, Lee, & Kim, 2015), la denominada computación evolutiva (Salleh et al., 2015).

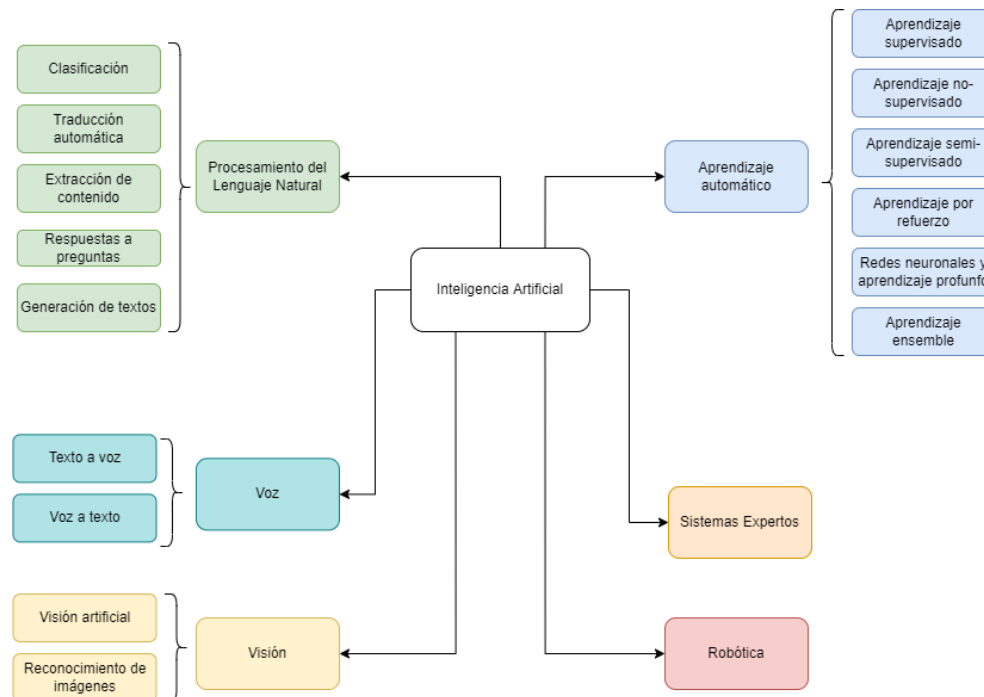


FIGURA 2.1: Diagrama de clasificación general de inteligencia artificial.

Tal y como se puede apreciar en la Figura 2.2, el aprendizaje automático es una rama de la IA que se refiere a la construcción de programas de ordenador que mejoran automáticamente su rendimiento en una determinada tarea con la experiencia. De acuerdo con las diferentes definiciones de IA de Russell & Norvig (2016), el aprendizaje automático se encuadra dentro de las definiciones relacionadas con el pensamiento humano (relacionadas con los sistemas cognitivos). En el aprendizaje automático, se utilizan algoritmos para analizar y aprender de los datos. Después de ello, los algoritmos realizan predicciones y toman decisiones sobre eventos en el mundo real. Y luego estos algoritmos son evaluados a partir de diferentes métricas que resultan determinantes para medir la bondad de ajuste un algoritmo.

Dentro del aprendizaje automático, el aprendizaje automático clásico, centrado en datos simples y características claras, se puede dividir en aprendizaje supervisado, aprendizaje no supervisado y aprendizaje semi-supervisado. En el aprendizaje supervisado, los datos de entrenamiento utilizados por el algoritmo para la construcción del modelo incluyen casos en los que la clase o el resultado que se espera es numérico o ha sido

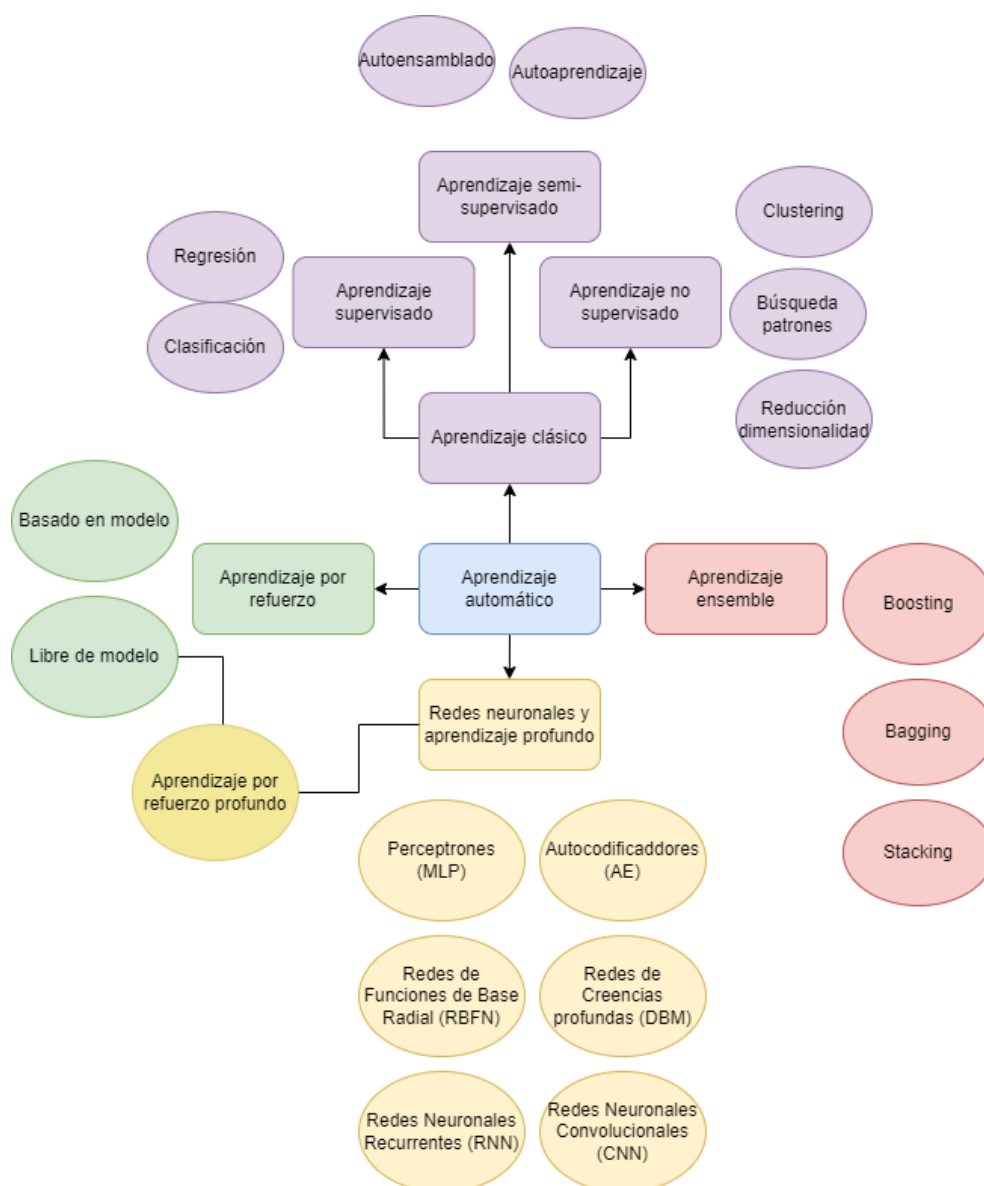


FIGURA 2.2: Diagrama de clasificación general de aprendizaje automático.

previamente etiquetado (normalmente por un humano), como los problemas de regresión y clasificación, respectivamente (Sen, Hajra, & Ghosh, 2020). En el aprendizaje no supervisado, los datos de entrenamiento no están etiquetados de antemano y no se conoce el resultado esperado. Por ejemplo, algunos algoritmos de aprendizaje no supervisado son el agrupamiento, la reducción de dimensionalidad o el aprendizaje basado en reglas de asociación (Usama et al., 2019).

En el aprendizaje semi-supervisado, algunos de los datos de entrenamiento de entrada están etiquetados y otros no lo están (van Engelen & Hoos, 2019). El aprendizaje automático clásico, que surgió en los años 50 a partir de los métodos estadísticos,

se sigue utilizando ampliamente en la actualidad y es útil para muchos problemas emergentes. Con el tiempo han surgido nuevos paradigmas como el aprendizaje ensemble, también conocido como aprendizaje integrado (Krawczyk, Minku, Gama, Stefanowski, & Woźniak, 2017), cuando la calidad de los resultados obtenidos resulta un problema; las redes neuronales y el aprendizaje profundo, son una alternativa en los casos en que se presentan datos complejos con características poco claras. El aprendizaje por refuerzo, por otro lado, es un caso de uso cuando no existen datos de entrenamiento de entrada, pero existe un entorno con el que se puede interactuar (Leike et al., 2018). La aplicación de redes neuronales profundas a tareas de aprendizaje por refuerzo dio lugar al aprendizaje por refuerzo profundo.

El resto de esta Sección está estructurada de la siguiente forma. La Sección 2.2.1 expone la revisión de la literatura en la que convergen los dos conceptos descritos anteriormente, y la Sección 2.2.2 describe casos de uso de modelos econométricos y de aprendizaje automático para resolver problemas de predicción en diferentes industrias.

2.2.1. Revisión sistemática de la literatura

En los últimos años, diversos autores han presentado revisiones de las aplicaciones del aprendizaje automático en los campos econométricos (López de Prado, 2019). Las posibilidades que ofrece el aprendizaje automático han llevado a varios investigadores en econometría financiera a realizar estudios comparativos sobre el rendimiento de las aplicaciones de aprendizaje automático frente a los modelos tradicionales, tal y como se sugiere López de Prado (2019). Sin embargo, ambos enfoques no siempre han estado enfrentados. Por ejemplo, si solo un subconjunto de variables de control es predictivo, un enfoque de selección de modelos de aprendizaje automático podría ayudar a apuntar al más relevante. Los métodos de minería de datos también pueden ser útiles si hay efectos de interacción significativos, por lo que uno se centra en la predicción de los efectos para ciertos individuos en lugar de un impacto global para toda la población (Belloni, Chen, Chernozhukov, & Hansen, 2012).

El objetivo del aprendizaje automático es resolver problemas de la vida real aprendiendo y mejorando automáticamente a partir de la experiencia sin estar programado explícitamente para un problema específico, sino para un tipo de problema genérico.

En la Figura 2.3 se describe brevemente el funcionamiento de los grandes grupos de aprendizaje automático.

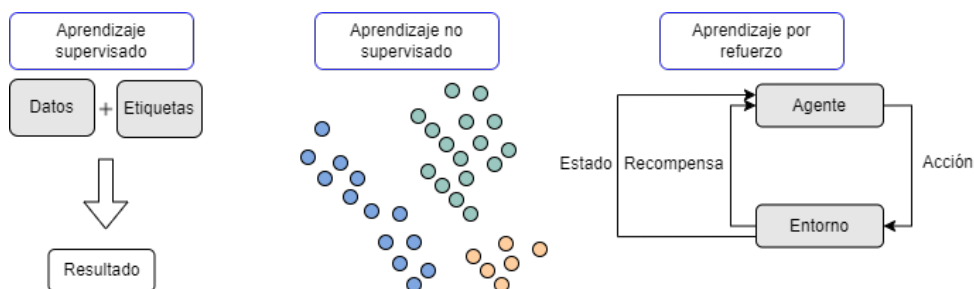


FIGURA 2.3: Ejemplos de tipos y funcionamientos de aprendizaje automático.

Mullainathan & Spiess (2017), y Athey (2018) publicaron resultados teóricos que enfatizaban el rango de posibilidades de predicción que ofrece el aprendizaje automático en el campo de la economía.

Los algoritmos supervisados buscan funciones que predigan bien fuera de la muestra. Los economistas lo denominarían como la variable dependiente; la que cambiará en función de los cambios en otras variables. Si el atributo etiquetado no existe, es necesario un algoritmo no supervisado para la exploración de los datos más que para la predicción de los resultados. Por ejemplo, se podría intentar predecir el valor “y” de una empresa a partir de sus características observadas “x”. Dentro del campo de los algoritmos supervisados hay muchas aplicaciones diferentes que se pueden emplear Caruana & Niculescu-Mizil (2006), ya sea para la clasificación o la regresión. Supongamos que nos proponemos medir si tener un gimnasio en el lugar de trabajo mejoraría la eficiencia de los empleados (donde la productividad se midió como los proyectos realizados al mes en una semana laboral de 40 horas). Los economistas buscarían un experimento lógico que pudiera incitar a ciertos trabajadores a utilizar las instalaciones del gimnasio por razones no relacionadas con su productividad actual (es decir, un gimnasio temporal en el lugar de trabajo). Y se podría estimar un modelo mediante una regresión lineal, como se muestra en la Ecuación 2.1:

$$Y_i = \alpha + \beta_0 + \beta_1 X_i + \varepsilon_i \quad (2.1)$$

o una regresión lineal múltiple como en la Ecuación 2.2:

$$Y_i = \alpha + \beta_0 + \sum \beta_p X_{pi} + \varepsilon_i \quad (2.2)$$

donde Y_i es el resultado (la productividad de los individuos dentro de un año), X_i es la política de interés (en caso de que el trabajador haya ido al gimnasio), β es el parámetro clave de interés (el efecto de ir al gimnasio dentro de las horas de trabajo), α denota los otros parámetros, y ε_i es un término de error. Utilizando los mismos datos, un enfoque de aprendizaje automático consistiría en identificar las variables que están estrictamente asociadas a la productividad. Esto se debe a la amplia gama de indicadores potenciales en los datos, y a la probabilidad de construir un modelo que sea capaz de predecir bien la rentabilidad, ya sea dentro o fuera de los datos de la muestra.

Dado que la econometría incluye muchas aplicaciones, algunos autores han reducido el alcance de sus estudios. Por ejemplo, Athey & Imbens (2019) destaca los principales modelos de aprendizaje automático que se aplican en los modelos econométricos, incluyendo los métodos basados en la regresión, la clasificación, los métodos de aprendizaje no supervisado y los métodos de completación de matrices, otros autores exploran áreas concretas de aplicaciones de aprendizaje automático en econometría, como los métodos para la estimación de la demanda (Bajari, Nekipelov, Ryan, & Yang, 2015). Además, otros realizaron un examen en profundidad de los principales modelos de aprendizaje automático para la previsión de series temporales (Ahmed, Atiya, Gayar, & El-Shishiny, 2010), (M.-W. Hsu, Lessmann, Sung, Ma, & Johnson, 2016) realizaron estudios comparativos y conclusiones derivadas de las aplicaciones tradicionales y de aprendizaje automático en la previsión de mercados financieros. Otros estudios comparativos que han señalado el buen desempeño del aprendizaje automático trabajaron en la detección de patrones irregulares y la realización de pronósticos a corto plazo utilizando datos heterogéneos (Y. Liu & Xie, 2019).

2.2.2. Identificación de casos de uso

A partir de la revisión sistemática realizada y que se describe en la Sección 3.1 y que forma parte de esta tesis doctoral, se observó que en los diferentes campos las aplicaciones utilizadas eran muy diferentes así como los campos de estudio.

En la Figura 2.4, y que es una adaptación del diagrama presentado en la publicación: "Machine Learning and Traditional Econometric Models: A Systematic Mapping Study", y que forma parte de esta Tesis Doctoral, se comparan los algoritmos y sus aplicaciones en econometría y aprendizaje automático. En algunos casos se contrastan sus rendimientos, en otros se utilizan conjuntamente. Los métodos más investigados son redes neuronales artificiales y bosques de decisiones frente a ARIMA y los diferentes tipos de regresión (y las diferentes variantes que puede tener).



FIGURA 2.4: Aplicaciones econométricas y de aprendizaje automático.

Adaptación del diagrama presentado en la publicación: "Machine Learning and Traditional Econometric Models: A Systematic Mapping Study" que forma parte de esta Tesis Doctoral

2.3. AgTech: Nuevas tecnologías aplicadas a la industria agropecuaria

Gracias a los avances tecnológicos, las explotaciones agrícolas de los países desarrollados y en desarrollo pueden beneficiarse de la aplicación de tecnologías de bajo coste. En este sentido, el internet de las cosas (Internet of Things- IoT) y, más concretamente, el internet industrial de las cosas (Industrial Internet of Things - IIoT), se presenta como una tecnología habilitadora clave para implementar y monitorizar soluciones de gestión

de recursos en diversos escenarios de la industria 4.0, incluyendo entornos de agricultura inteligente (Sisinni et al., 2018).

El IoT puede utilizarse en combinación con otras tecnologías como computación en la nube, big data, AI, o tecnologías como por ejemplo blockchain para implementar soluciones que mejoren la trazabilidad y la productividad de los procesos industriales (Yu et al., 2017). Sin embargo, cuando se trata de transmitir datos a la nube, surgen varios retos relacionados con la privacidad de los datos, el consumo de energía o los costes asociados al uso de los servicios en la nube (Alonso, Sittón-Candanedo, Rodríguez-González, García, & Prieto, 2019). En este sentido, los proveedores de servicios cobran tasas en función de la cantidad de datos que se transfieren, almacenan y procesan en la nube (Wu, Toosi, Buyya, & Ramamohanarao, 2018). Mediante el uso de tecnologías EC, es posible reducir el volumen de tráfico transferido entre la capa IoT y la nube (Ai, Peng, & Zhang, 2018). La computación de borde ofrece un modelo potencialmente manejable para la integración de la agricultura inteligente.

El resto de esta Sección está estructurada de la siguiente forma. La Sección 2.3.1 describe los diferentes conceptos de sostenibilidad y gestión de los recursos gracias al AgTech. La Sección 2.3.2 expone las características principales del IoT y finalmente la Sección 2.3.3, detalla el estado del arte en relación con el Edge Computing en escenarios IoT.

2.3.1. Sostenibilidad y gestión de recursos

A mediados de los años noventa, autores como Schmidheiny & Timberlake (1992) comenzaron a examinar los términos y conceptos de la ecoeficiencia. La ecoeficiencia es la capacidad de producir más bienes y servicios con menos impacto ambiental y menos consumo de recursos naturales (Ichimura et al., 2009). La ecoeficiencia se mide normalmente como la relación entre el valor añadido de lo producido (por ejemplo, el PIB) y los impactos ambientales añadidos del producto o servicio producido (normalmente utilizando las emisiones de CO_2). En el contexto de la ecoeficiencia, la OCDE establece diferentes indicadores para evaluar el progreso medioambiental a diferentes niveles. Esos indicadores ambientales conforman tipos de políticas e indicadores que pueden ser conceptualizados como un ratio, expresado como un indicador de valor económico dividido por un indicador de impacto ambiental. Para evaluar esas medidas, los sensores del IoT son un instrumento clave para monitorear las

diferentes condiciones ambientales y su evolución. En los últimos 20 años, y alineados con el concepto de ecoeficiencia, diferentes autores como Tiwari, Loof, & Paudyal (1999) desarrollaron un sistema utilizando técnicas de toma de decisiones multicriterio para lograr la eficiencia económico-ambiental incluyendo criterios de sostenibilidad y económicos ambientales. Asimismo, Simar & Wilson (2007) demostró que la eficiencia medioambiental estaba relacionada negativamente con el tamaño de la explotación, la edad de los agricultores y las subvenciones a los cultivos, y positivamente con la rotación de cultivos. También introdujeron la variabilidad de las condiciones meteorológicas (niveles de temperatura y precipitación) para captar el efecto de la incertidumbre de la producción en la eficiencia medioambiental, comprobando que tiene un impacto significativo. Lansink & Wall (2014) presentó una visión general sobre la evaluación de los modelos de frontera y su eficiencia ambiental. La intensificación sostenible de los sistemas agrícolas ofrece oportunidades sinérgicas para la coproducción de resultados agrícolas y del capital natural. La eficiencia y la sustitución son pasos hacia la intensificación sostenible, pero el rediseño del sistema es esencial para obtener resultados óptimos a medida que cambian las condiciones ecológicas y económicas. Las tecnologías de la información pueden contribuir a la transición de la sostenibilidad agroalimentaria aumentando la productividad de los recursos, reduciendo las ineficiencias, reduciendo los costes de gestión y mejorando la coordinación de la cadena alimentaria (El Bilali & Allahyari, 2018).

2.3.2. El internet de las cosas y el internet de las cosas industrial

El interés en el IoT creció a medida que las grandes empresas y los gobiernos la vieron como una tecnología clave. En este sentido, Google comenzó a almacenar datos relacionados con las redes Wi-Fi de los usuarios de sus servicios en 2010. El concepto de IoT fue acuñado por un miembro de la comunidad de desarrollo de la identificación por radiofrecuencia (RFID) en el año 2000, y recientemente ha cobrado más importancia en el mundo real. Esto se debe en gran medida al aumento de los dispositivos móviles, la comunicación integrada y ubicua, la computación en la nube y el análisis de datos (Patel, Patel, et al., 2016).

Internet de las cosas (IoT) se refiere a la idea general de las cosas conectadas, especialmente los objetos cotidianos que son legibles, reconocibles localizable,

direccionable por un dispositivo de detección y / o controlable a través de Internet, independientemente de los medios de comunicación. En la literatura existen múltiples modelos de IoT orientados a la monitorización de las condiciones ambientales de una finca a través de una serie de sensores desplegados. Todos estos modelos requieren, para su correcto funcionamiento, de una transmisión robusta sobre la que posteriormente poder desarrollar modelos de inteligencia artificial predictivos.

El IoT ofrece muchas soluciones para cada una de sus áreas de aplicación. Algunas de las funciones más importantes son: la gestión de múltiples protocolos de comunicación, el procesamiento de datos, la información y la respuesta en tiempo real, el almacenamiento de big data, la seguridad y privacidad de los datos (Sisinni et al., 2018).

Sin embargo, la implementación de estas funciones conlleva una serie de retos que deben ser resueltos: la heterogeneidad de las fuentes de datos, la seguridad, la privacidad, la latencia, la respuesta en tiempo real y el uso de recursos informáticos compartidos. Aunque el uso de capas de ingestión de datos de IoT puede resolver el problema de la heterogeneidad, es necesario abordar otras cuestiones. Uno de ellos es el elevado volumen de datos que pueden transmitir a la plataforma IoT cientos, miles o incluso millones de dispositivos. En este sentido, han surgido soluciones como el paradigma de la computación de borde o EC en respuesta a la necesidad de reducir la cantidad de tráfico de datos entre la capa de IoT y la nube.

2.3.3. Edge Computing

El EC permite la ejecución de modelos de aprendizaje automático en el borde de la red, reduciendo el tiempo de respuesta y proporcionando un cierto nivel de servicio aunque se interrumpa la comunicación con la nube. Esto es habitual en escenarios en los que la conectividad a Internet es limitada (por ejemplo, entornos agrícolas rurales) (Alonso, Sittón-Candanedo, García, et al., 2020).

Sittón-Candanedo (2019) realizaron un estudio de las diferentes aplicaciones de EC y se concluyó que muchas aplicaciones tenían un potencial muy alto en la industria agrícola. Además, Alonso, Sittón-Candanedo, García, et al. (2020) presentaron una plataforma orientada a la aplicación de técnicas de IoT, EC, IA y blockchain en entornos agrícolas inteligentes.

2.4. Metodologías para optimización de la producción en AgTech

Mientras que los niveles de producción y la efectividad en las fronteras de producción incorporan variables más alineadas con el entorno y la situación macroeconómica, cuando se trata de medir la rentabilidad, cualquier inestabilidad económica es un factor determinante (Machek & Špička, 2014).

En la actualidad, como se ha detallado en los apartados anteriores, la agricultura también se está beneficiando de las soluciones de IoT y del aprendizaje automático, este hecho puede considerarse como un híbrido entre las mejoras de la industria y las posibilidades de globalización.

Es por ello que las metodologías y métricas utilizadas para analizar las fronteras de producción se han de adaptar a los usos de las nuevas tecnologías así como sus costes derivados, como por ejemplo el almacenamiento de la información en la nube.

El resto de esta Sección está estructurada de la siguiente forma. La Sección 2.4.1 describe los diferentes metodologías referentes al análisis de la producción. La Sección 2.4.2 expone fórmulas relacionadas con las fronteras de producción.

2.4.1. Medidas de producción e implicaciones tecnológicas

Según diferentes métricas econométricas, la medida de la producción óptima puede traducirse en una función de producción como la Ecuación 2.3, donde K es el capital como la maquinaria (incluyendo los dispositivos de IoT), L es la tierra y mientras W representa la mano de obra. Con los años y los avances tecnológicos, la variable K contiene cada vez más variables, que se desglosarán más adelante en esta Sección. Donde Q es la variable de producción dependiente, como se muestra en la Ecuación 2.3.

$$Q = (K, L, W) \tag{2.3}$$

Durante muchos años, el crecimiento de los rendimientos a escala (que es la función que describe lo que ocurre con los rendimientos a largo plazo a medida que aumenta

la escala de producción) Shephard (2015) era una prueba de cómo se realizaban las explotaciones en el sector agrícola, con pocos cambios en la industria o la maquinaria Takeshima (2017).

Otro enfoque que se puede considerar para entender la productividad y la eficiencia en la producción en el siglo XXI debido a la proliferación de las nuevas tecnologías es la Productividad Total de los Factores (Productivity Total Factor - PTF). La PTF es la parte de la producción que no se explica por el número de insumos utilizados en la producción. Mide el crecimiento residual de la producción total de una empresa, industria o economía nacional que no puede explicarse por la acumulación de insumos tradicionales, como el trabajo y el capital. Como tal, su nivel viene determinado por la eficiencia y la intensidad con la que se utilizan los insumos en la producción. El crecimiento de la PTF suele medirse mediante el residuo de Solow, como en la Ecuación 2.4, pero la variable de productividad suele estar unida a la variable del trabajo en el modelo Solow-Swan presentado por (Solow, 1957; Swan, 1956; Van Beveren, 2012)

$$Q(t) = K(t)^\alpha (A(t) W(t))^{1-\alpha} \quad (2.4)$$

donde t denota el tiempo, α en $[0, 1]$ es la elasticidad de la producción con respecto al capital, y $Q(t)$ representa la producción total. A se refiere a la tecnología que aumenta el trabajo o el conocimiento, por lo que A y W representan el trabajo efectivo. Todos los factores de producción están totalmente empleados, y se dan los valores iniciales $A(0)$, $K(0)$ y $W(0)$. El número de trabajadores (labor - L) así como el nivel de tecnología crecen exógenamente a tasas n y g , respectivamente:

$$W(t) = W(0) e^{nt} \quad (2.5)$$

$$A(t) = A(0) e^{gt} \quad (2.6)$$

La función Cobb-Douglas puede estimarse como una relación lineal utilizando la Ecuación (2.7), donde el I representa los Insumos y el Q es la producción y los a son los coeficientes del modelo:

$$\ln(Q) = a_0 + \sum_i a_i \ln(I_i) \quad (2.7)$$

Una vez que se alcanza la optimización de la producción, la rentabilidad financiera puede utilizarse para medir la capacidad de la empresa para seguir produciendo beneficios. La rentabilidad financiera de una empresa o una explotación puede calcularse con el retorno de los activos (Return on Assets - ROA), y donde los activos totales representan todos los activos que la empresa utiliza para producir los productos.

2.4.2. Fronteras de producción

El El análisis de frontera estocástica (Stochastic Frontier Analysis-SFA) y el análisis envolvente de datos (Data Envelopment Analysis -DEA) se han comparado como herramientas de medición para la economía agrícola durante muchos años Hjalmarsson, Kumbhakar, & Heshmati (1996). Por ejemplo, Theodoridis & Anwar (2011) realizó un experimento en explotaciones agrícolas de Bangladesh en el que los resultados del SFA se apoyan en los resultados del DEA.

El DEA es un método de programación lineal no paramétrico, que ha obtenido una amplia gama de aplicaciones que miden la eficiencia comparativa de múltiples entradas y salidas de un conjunto homogéneo de unidades de decisión (Decision Making Unit - DMU). El DEA es un método de frontera que trata de optimizar la medida de eficiencia de cada unidad analizada. Charnes, Cooper, & Rhodes (1978) definió, en su modelo básico de DEA, la función objetivo DMU. El DEA permite que cada DMU especifique sus propias ponderaciones para obtener su máxima puntuación de eficiencia, lo que puede dar lugar a un número relativamente alto de DMUs eficientes y evitar que DEA aparezca como un enfoque robusto para determinar la unidad más eficiente Doyle & Green (1994). Por lo tanto, se puede utilizar una eficiencia minimax para discriminar entre las unidades que parecen relativamente eficientes al aplicar el modelo DEA clásico.

La función objetivo de un modelo DEA contiene un parámetro que debe seleccionarse mediante un método de prueba y error para alcanzar una de eficiencia relativa de la DMU. También contiene una restricción relativa a una DMU específica, lo que significa que tiene que resolver n veces, una programación lineal para cada DMU (n es el número de DMUs).

2.5. Sistemas de apoyo a la toma de decisiones

Con los recientes avances en el campo de la inteligencia artificial, cada vez son más las tareas de toma de decisiones que se delegan en sistemas informáticos. Un requisito fundamental para el éxito y la adopción de estos sistemas es que los usuarios confíen en las elecciones del sistema o incluso en las decisiones totalmente automatizadas. Con los algoritmos de aprendizaje automático cada vez más sofisticados, surgen constantemente nuevos retos en el contexto de las explicaciones, la responsabilidad y la confianza hacia estos sistemas. Un requisito clave para el éxito y la adopción práctica de estos sistemas en muchos casos es que los usuarios confíen en las recomendaciones y decisiones automatizadas de los sistemas informáticos o, al menos, que confíen en la imparcialidad de los consejos. Dado un ejemplo concreto en el que a unos usuarios dados se les ofrecen diferentes artículos, el objetivo principal de un sistema de recomendación es sugerir artículos relevantes a los usuarios, aunque a menudo se tienen en cuenta otras dimensiones de utilidad, como la diversidad, la novedad, la confianza o la posibilidad de dar explicaciones.

Entre los diferentes sistemas de apoyo a la toma de decisiones se encuentran los sistemas expertos, los propiamente denominados de apoyo a la toma de decisiones y los sistemas de recomendación. Un sistema experto (Expert Systems - SE) es un sistema que utiliza el conocimiento humano capturado en un ordenador para resolver problemas que normalmente requieren experiencia humana. Los SE imitan los procesos de razonamiento de los expertos para resolver problemas específicos. El objetivo principal de un SE es proporcionar conocimientos a un usuario nuevo, que puede mostrar un rendimiento de nivel experto en su toma de decisiones. Los SE propagan los escasos recursos de conocimiento para obtener resultados mejorados y consistentes. Estos sistemas también pueden ser utilizados por los expertos como conocimientos. A medida que los conocimientos de un SE mejoran y son más precisos, el sistema puede llegar a funcionar a un nivel superior al de cualquier experto humano a la hora de emitir juicios en un área específica de conocimientos (denominada dominio).

Los sistemas de recomendación se clasifican generalmente en filtrado (Colaborative Filtering - CF) y en filtrado basado en el contenido. En general, el FC utiliza una técnica de filtrado de información basada en la evaluación previa de los artículos por

parte del usuario o en el historial de compras anteriores. Para mejorar el rendimiento de los sistemas de recomendación se usan técnicas de minería de datos. En consecuencia, en ocasiones también tiene sentido clasificar los sistemas de recomendación según las técnicas de minería de datos aplicadas. En general, las técnicas de minería de datos se definen como la extracción o conocimiento de los datos. Estas técnicas se utilizan para la exploración y el análisis de grandes cantidades de datos con el fin de descubrir patrones y reglas significativas. También pueden utilizarse para dirigir la toma de decisiones y para predecir el efecto de las decisiones.

2.5.1. Sistemas de razonamiento basado en casos

El razonamiento basado en casos (Case Base Reasoning - CBR) pretende obtener el caso fuente en la memoria a través de la sugerencia del caso objetivo y luego guiar el caso objetivo utilizando el caso origen. La metodología CBR funciona sobre la base del conocimiento previo; estos sistemas pueden ayudar en el proceso de toma de decisiones, mediante la reutilización de casos previamente recuperados (Aamodt & Plaza, 1994). En cualquier sistema de razonamiento basado en casos, cada caso está representado por dos vectores: El vector problema y el vector solución (Corchado & Lees, 1996). Los sistemas CBR revisan cada nueva solución y, una vez ejecutada, el caso anterior se conserva para reutilizarlo en una posible solución futura. La metodología CBR tradicional está compuesta por 4 fases:

- Recuperación: Consiste en la recuperación de casos de similares (a partir de medidas de distancia).
- Re-uso: Las soluciones de los casos previamente recuperados se analizan y ordenan. Estos casos se comparan con el caso actual para encontrar la solución más similar.
- Revisión: Revisión de los casos reutilizados. Esta parte puede incluir también el conocimiento de un experto.
- Retener: El sistema de razonamiento basado en casos almacena los casos cuyas soluciones son consideradas como válidas o correctas.

El uso de CBR conlleva muchas ventajas; aplica el conocimiento de los expertos, proporciona sugerencias y decisiones rápidas para un tipo de problema concreto y almacena las soluciones para futuras aplicaciones (aprendizaje incremental).

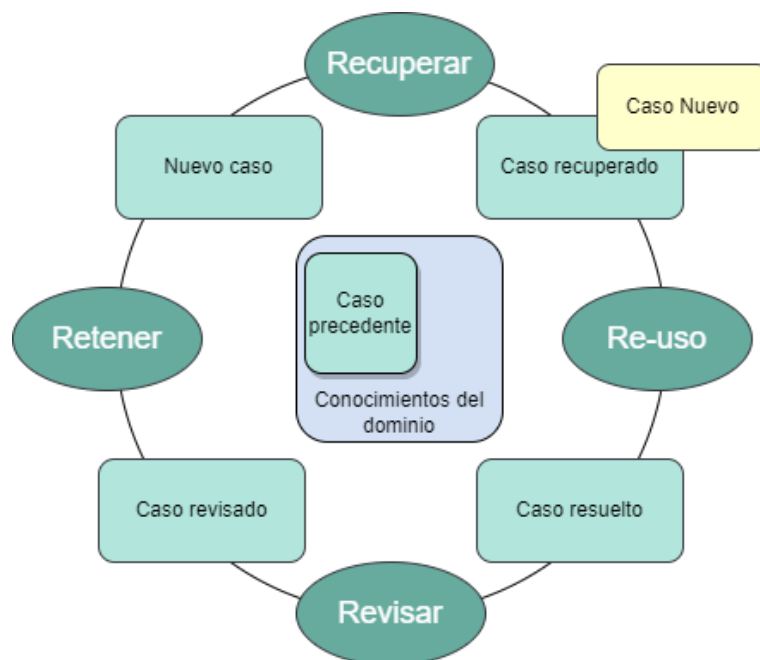


FIGURA 2.5: Diseño de una metodología basada en casos.

2.5.2. Optimización multi-objetivo y multi-atributo

Se han hecho varias revisiones sobre los métodos y la aplicación de la optimización multiobjetivo (Multi objective optimization - MOO). Hay dos métodos de MOO que no requieren complicadas ecuaciones matemáticas, por lo que el problema se vuelve sencillo. Estos dos métodos son el de Pareto y el de escalarización. En el método de Pareto, hay una solución dominada y una solución no dominada que se obtiene mediante un algoritmo de actualización continua. Por su parte, el método de escalarización crea funciones multiobjetivo convertidas en una única solución mediante pesos. Hay tres tipos de ponderaciones en la escalarización, que son las ponderaciones iguales, las ponderaciones del centro del rango y las ponderaciones de la suma del rango.

Un problema de optimización multiobjetivo puede definirse como la siguiente Ecuación 2.8, donde $f(x)$ es el vector objetivo k -dimensional.

$$f(x) = f_1(x), f_2(x), \dots, f_k(x) \quad (2.8)$$

Por otro lado, la función de utilidad multiatributo se utiliza para representar las preferencias de un usuario sobre paquetes de bienes, en condiciones de certeza sobre los resultados de cualquier elección potencial. Van Calker, Berentsen, Romero, Giesen, & Huirne (2006) presentaron un modelo de función multiatributo de sostenibilidad para evaluar la sostenibilidad en diferentes sistemas agrícolas.

Las preferencias pueden caracterizarse mediante funciones de utilidad, en las que la información relativa a la preferencia está implícita en la función, lo que permite clasificar las soluciones. Las funciones de utilidad asignan diferentes pesos a determinados atributos. En un caso hipotético en la que hay un comprador en el sector agropecuario se podría dar la siguiente situación: la función de utilidad para el agente comprador en este caso sería la siguiente, en la que el producto A (que podría ser Maíz vendido a X precio y Contaminación de invernadero de Y puntos), es preferido sobre el producto B solo si la expectativa de la función U es mayor bajo A que bajo B , como se muestra en la Ecuación 2.9.

$$E_A[u(x_1, \dots, x_n)] > E_B[u(x_1, \dots, x_n)] \quad (2.9)$$

2.5.3. Procesos de decisiones secuenciales

Un proceso de decisión de markov (Markov Decision Process - MDP) es un proceso que transcurre a través de diferentes estados y en el que existe una probabilidad específica de transitar entre un estado i y un estado j . Así, existe un conjunto de estados finitos en los que el agente se puede encontrar en el laberinto. Asimismo, existe un conjunto de acciones (avance, retroceso, giro, etc.) que el agente puede llevar a cabo en cada uno de los estados. También existe una recompensa (positiva o negativa) asociada a cada transición. Por otro lado, se considera que existe un *factor de descuento* o *discount factor* γ , un valor entre 0 y 1 que cuantifica la importancia entre las recompensas inmediatas y las recompensas futuras. Por ejemplo, un $\gamma = 0.95$ implica que una recompensa de 10 puntos que se obtiene después de 5 pasos ofrecerá un valor actual como recompensa de $0.95^5 \cdot 10 \approx 7.73$ puntos.

$$\sum_{t=0}^{t=\infty} \gamma^t r(s(t), a(t)) \quad (2.10)$$

Finalmente, en el MDP existe lo que se conoce como falta de memoria, es decir, el agente no necesita conocer los estados anteriores al actual, considera que el futuro solo depende del presente. Así, el objetivo del agente es maximizar la suma de las recompensas a largo plazo desde el momento presente hasta su horizonte en el tiempo futuro: donde $r(s, a)$ es una función de recompensa que depende del estado s y de la acción a .

El aprendizaje por refuerzo (Reinforcement Learning - RL) consiste en que un agente que interactúa con el entorno, aprende una política óptima, por ensayo y error, para los problemas de toma de decisiones secuenciales Sutton & Barto (2018). El RL estándar consiste en un agente que interactúa con un entorno, que puede ser modelado como un MDP.

2.5.3.1. Aprendizaje por refuerzo

El RL y el DRL conforman una de las tendencias más prometedoras de los últimos años en el ámbito del aprendizaje automático (Leike et al., 2018).

El algoritmo Q-learning (Watkins & Dayan, 1992) es una de las técnicas libres de modelo de aprendizaje por refuerzo. Para cualquier proceso de decisión finito de Markov (Finite Markov Decision Process FMDP), el Q-learning descubre una política óptima, es decir, se maximiza el valor esperado de la recompensa total en cada paso siguiente al actual (Melo, Meyn, & Ribeiro, 2008).

El aprendizaje Q, es un método de RL sin modelo y sin política, que consiste en agentes cuyo objetivo es alcanzar el estado-acción-valor de una función $Q = (s, a)$ interactuando en un entorno determinado. A medida que el agente explora el entorno, Q devuelve una aproximación cada vez más precisa del valor esperado de una acción a , dado un estado s del valor esperado de una acción a , dado un estado s . Por tanto, la función Q se actualiza progresivamente.

Se formula como un MDP que puede definirse mediante la 5-tupla (s, a, p, r, γ) , donde s es el estado, a es la acción, p es la probabilidad de transición, r es la función de recompensa, y γ es el factor de descuento.

2.5.3.2. Aprendizaje por refuerzo profundo

El aprendizaje profundo (Deep Learning - DL) ha acelerado el progreso en la RL, con el uso de algoritmos de aprendizaje profundo dentro de la RL definiendo el campo del aprendizaje de refuerzo profundo (Deep Reinforcement Learning - DRL) Arulkumaran, Deisenroth, Brundage, & Bharath (2017). El DL permite extender el RL a problemas de toma de decisiones previamente intratables, es decir, a entornos con un elevado número de estados y espacios de acción dimensionales. Como una red neuronal es una aproximación funcional universal, puede utilizarse como sustituto de la tabla Q. En el proceso de aprendizaje, DL optimiza los pesos, θ , para minimizar el error estimado por la función de pérdida. El error o pérdida se mide como la diferencia entre el resultado predicho y el resultado real. La red Q profunda (Deep Q Network - DQN) fue introducida por primera vez por Mnih et al. (2013) y luego Mnih et al. (2015) introdujo técnicas adicionales, como (Deep Q Learning DQL). El algoritmo base para DQN es el RL basado en valores, que es un método que aproxima un valor de acción (es decir, un valor Q) en cada estado. Un algoritmo basado en el aprendizaje Q que aproxima la función Q utilizando DNN es la base de DQN Mnih et al. (2015). En el aprendizaje por refuerzo, la función objetivo de la diferencia temporal es siempre desconocida. Antes de que un agente tome una acción, el valor Q puede definirse como $Q(s, a)$, y después de que la acción se toma el nuevo estado es $R(s, a) + \gamma \max_{a'} Q(s', a')$, por lo que la diferencia temporal se define en la Ecuación 2.11.

$$T(a, s) = R(s, a) + \gamma \cdot \max_{a'} Q(s', a') - Q_{t-1}(s, a) \quad (2.11)$$

La función es aproximada por una red neuronal $Q(s, a; \theta)$ parámetro, θ donde el parámetro se aprende minimizando la pérdida de la diferencia temporal (Temporal Difference - TD). Así, la función de pérdida resulta ser

$$Loss = Q^*(s_t, a_t) - Q(s_t, a_t) \quad (2.12)$$

La idea clave de DQN es aprender una aproximación de la función de valor óptimo Q , que se ajusta a la Ecuación de optimalidad de Bellman Bradtke & Duff (1995). En el algoritmo DQN, la Q corresponde a la función que representa las recompensas esperadas

para una acción dada en un estado determinado. DQN refina la política con respecto a los valores de la acción mediante el operador \max Sutton (1988). Una forma de minimizar la función de pérdida es mediante el método de descenso de gradiente Hasselt (2010). En este método, la política $Q(s, a)$ se actualiza en base a la recompensa actual y al valor máximo de las recompensas futuras esperadas. En el DQN, la función de aprendizaje puede describirse como en el Algoritmo 1, donde ϵ es la tasa de aprendizaje y π es la política óptima.

Algoritmo 1: Algoritmo adaptado de la investigación realizada por Sutton (1988).

Resultado: Q

inicialización; $Q : XA \rightarrow R$

mientras Q *no converge* **hacer**

 Empieza en estado sX

mientras s *no es terminal* **hacer**

 calcula π acorde a Q y estrategia de exploración

$a \leftarrow \pi(s)$

$r \leftarrow R(s, a)$

$s' \leftarrow T(s, a)$

$Q(s', a) \leftarrow (1 - \alpha) \cdot Q(s, a) + \alpha \cdot (r + \gamma \cdot \max_{a'} Q(s', a'))$

$s \leftarrow s'$

fin

fin

2.6. Conclusiones

La investigación presentada en este Capítulo es el resultado de un análisis pormenorizado del estado del arte de las diferentes tecnologías y propuestas existentes aplicadas a entornos del sector agropecuario.

En ese sentido, el estudio del estado del arte llevado a cabo en este Capítulo permite extraer algunas conclusiones importantes asociadas a los retos que este tipo de entornos presentan, así como plantear una metodología de cara a abordar dichos desafíos.

- Las posibilidades que ofrece el aprendizaje automático han llevado a varios investigadores en econometría financiera a realizar estudios comparativos sobre el rendimiento de las aplicaciones de aprendizaje automático frente a los modelos tradicionales. La Sección 2.2 describe de forma resumida las características principales de los modelos tradicionalmente econométricos y contextualizando los de inteligencia artificial, más concretamente los de aprendizaje automático. Como punto de partida, en el próximo Capítulo 3, la Sección 3.1 describe el planteamiento inicial en el que se llevó a cabo una revisión sistemática para contrastar los usos de los modelos econométricos y de aprendizaje automático en diferentes campos de estudio, concretamente evaluando únicamente los artículos que comparaban los rendimientos de unos modelos u otros o bien en que se aplicaban en conjunto.
- En el sector agropecuario, para que las empresas de dicho sector sean más competitivas y eficientes necesitan tener datos sobre los que aplicar fórmulas que permitan ver y mejorar su rentabilidad. El Internet de las Cosas y, en particular, el Internet Industrial de las Cosas, se presenta como una tecnología clave para implementar soluciones de monitorización y gestión de recursos en diversos escenarios. Es necesario encontrar soluciones que reduzcan el coste y el consumo de energía de los dispositivos IoT para facilitar efectivamente la digitalización de los diferentes escenarios. Además, las tecnologías de EC permiten reducir el volumen de tráfico transferido entre los nodos IoT y la nube. En este sentido, la Sección 2.3 describe las diferentes tecnologías aplicadas como punto de partida, y en la Sección 2.4 se presentan diferentes métodos para optimizar los recursos. Por tanto, en el próximo Capítulo, en la Sección 3.2.1 se desarrolla un caso de uso en el que a partir de datos recolectados con sensores IoT y computando el coste del almacenamiento

de los datos y haciendo uso del paradigma edge computing permiten medir las fronteras de decisión y optimización para un caso de uso real de una explotación agropecuaria.

- La Sección 2.5 resume diferentes sistemas para apoyo a la toma de decisiones, entrando en detalle en aquellos sobre los que se ha realizado dos experimentos y que forman parte de la metodología propuesta y que se presentan en el Capítulo 3, en las Secciones 3.3 y 3.3.2.

Capítulo 3

Contribuciones



**VNiVERSIDAD
D SALAMANCA**

CAMPUS DE EXCELENCIA INTERNACIONAL

Contribuciones

La agricultura de precisión es un término que se ha acuñado en los últimos años. Se refiere al concepto de utilizar las nuevas tecnologías para aumentar el rendimiento y la rentabilidad de los cultivos, reduciendo al mismo tiempo los recursos necesarios para el cultivo (Skobelev et al., 2018; N. Zhang et al., 2002).

En períodos de crisis financiera, la volatilidad del mercado agrícola ha sido muy elevada (Du et al., 2011; Kuruppuarachchi & Premachandra, 2016). Esas fluctuaciones en los mercados de futuros de productos básicos (Kang et al., 2017) contribuyen a las oscilaciones de los precios agrícolas para los agricultores. Sin embargo, no todos los productos agrícolas se ven afectados por los mismos factores, el mercado agrícola es muy diverso en cuanto a los atributos y particularidades de los productos que están sujetos a factores externos.

El presente Capítulo ha sido estructurado como se describe a continuación. En primer lugar, en la Sección 3.1 se describe la publicación en la que se realizó la revisión sistemática en aprendizaje automático y modelos econométricos. A continuación, en la Sección 3.2 se describe el caso de estudio que se llevó a cabo en para la optimización de los recursos en una empresa del sector agropecuario. Finalmente en la Sección 3.3 se presentan las dos investigaciones basadas en sistemas de apoyo a la toma de decisiones, tanto para productos del sector agropecuario como para inversión.

3.1. Aprendizaje automático y modelos econométricos tradicionales: Un estudio de mapeo sistemático

Los principales modelos de aprendizaje automático que se aplican en los modelos econométricos, incluyendo los métodos basados en la regresión, la clasificación y

los métodos de aprendizaje no supervisado. Por otro lado, autores como (Bajari et al., 2015) exploraron áreas concretas de aplicaciones de aprendizaje automático en econometría, como los métodos para la estimación de la demanda. Ahmed et al. (2010) realizaron un examen en profundidad de los principales modelos de aprendizaje automático para la previsión en series temporales, y M.-W. Hsu et al. (2016) realizaron estudios comparativos y de las aplicaciones tradicionales y de aprendizaje automático en la previsión de mercados financieros. El estudio comparativo realizado por Y. Liu & Xie (2019), señaló el buen desempeño del aprendizaje automático en la detección de patrones irregulares y la realización de pronósticos a corto plazo utilizando datos heterogéneos.

Dadas estas casuísticas, se realizó una revisión sistemática con el objetivo de identificar y hacer un estudio comparativo del rendimiento de los modelos econométricos y los de aprendizaje automático. La revisión sistemática se realizó con la finalidad de dar respuesta a evidencias empíricas de si los algoritmos de aprendizaje automático obtienen mejores resultados que los modelos econométricos cuando se comparan en el mismo problema, o bien si cuando se utilizan conjuntamente modelos de aprendizaje automático y econométricos era posible conseguir mejores resultados.

El objetivo principal de la investigación era identificar las últimas aplicaciones publicadas en las que se utilizaba y se comparaba el resultado obtenido a partir de predicciones realizadas con aprendizaje automático o modelos econométricos. Más concretamente el objetivo residía en encontrar evidencias empíricas de si los algoritmos de aprendizaje automático obtienen mejores resultados que los modelos econométricos cuando se comparan con el mismo problema, o por el contrario, si existían evidencias de que cuando se utilizan conjuntamente modelos de aprendizaje automático y econométricos es posible conseguir mejores resultados.

Los tres objetivos de investigación fueron los siguientes:

- **RQ1** - ¿Cómo y en qué campos se ha implementado el aprendizaje automático en modelos de aplicación tradicionalmente econométricos?
- **RQ2** - ¿Cómo complementa el aprendizaje automático supervisado a los modelos econométricos tradicionales?

- **RQ3** - Comparando el aprendizaje automático y los modelos econométricos ¿cuáles son los métodos más frecuentemente aplicados, y en qué casos de estudio?

Dados los objetivos de investigación y las cuestiones planteadas, las conclusiones de esta investigación derivaron a partir de las siguientes preguntas de investigación:

- El análisis del rendimiento futuro se utiliza en varias industrias, mientras que el análisis de las series temporales se utiliza sobre todo en el mercado de valores y en aplicaciones de inversión. Una de las cuestiones en las que es prácticamente un denominador común en todos los casos es la previsión de los precios futuros.
- En la mayoría de los casos en los que se han comparado los modelos de aprendizaje automático y los econométricos tradicionales, el aprendizaje automático ha obtenido mejores resultados a la hora de realizar predicciones. Sin embargo, en algunas situaciones en las que la mayor precisión se consigue utilizando los algoritmos conjuntamente
- Los métodos más investigados son RNA y RF frente a ARIMA y diferentes tipos de regresión (y las diferentes variantes que puede tener).

Como se observa en la Figura 3.1, el análisis del rendimiento futuro está orientado a todos los sectores de actividad, mientras que el análisis de las series temporales se utiliza sobre todo en las aplicaciones bursátiles y de inversión. Una de las cuestiones en las que es prácticamente un denominador común en todos los casos es la previsión de los precios futuros.

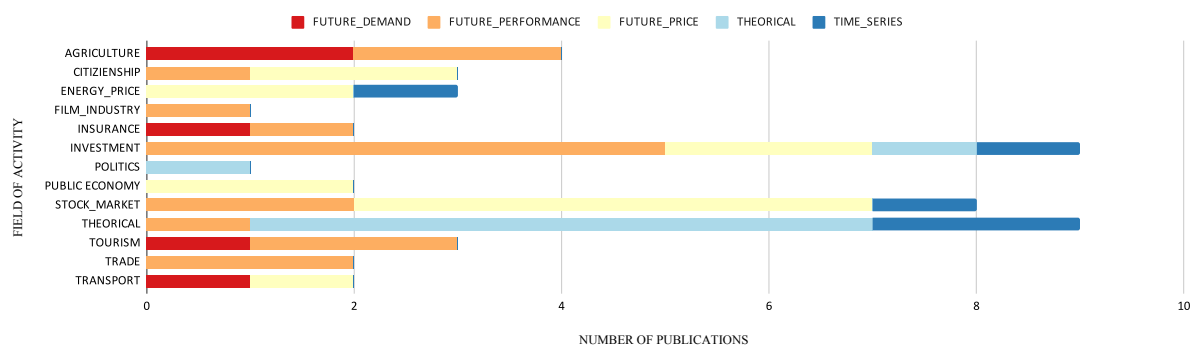


FIGURA 3.1: Tipos de contribución por industria.

Adaptación del diagrama presentado en la publicación: "Machine Learning and Traditional Econometric Models: A Systematic Mapping Study" que forma parte de esta Tesis Doctoral

Dadas las conclusiones obtenidas en la investigación resultó interesante realizar más investigaciones en los campos de la inteligencia artificial y econometría en el sector agropecuario siendo uno de los sectores en los que había menos investigaciones y casos de uso.

El contenido de este Capítulo ha sido adaptado del artículo *Machine Learning and Traditional Econometric Models: A Systematic Mapping Study* de la revista *Journal of Artificial Intelligence and Soft Computing Research*.

3.2. Un caso práctico en la industria agroalimentaria a través de una plataforma Edge-IoT

El internet de las cosas (IoT) y, más concretamente, la Internet industrial de las cosas (IIoT), se presentan como una tecnología habilitadora clave para la implementación de soluciones de monitorización y gestión de recursos en una variedad de escenarios de la Industria 4.0, incluyendo entornos agrícolas inteligentes (Sisinni et al., 2018). A la hora de transmitir datos a la nube, surgen varios retos en cuanto a la privacidad de los datos, el consumo de energía o los costes asociados al uso de los servicios en la nube (Alonso, Sittón-Candanedo, Casado-Vara, Prieto, & Corchado, 2020). En este sentido, los proveedores de servicios cobran a los usuarios en función de la cantidad de datos que se transfieren, almacenan y procesan en la nube (Wu et al., 2018). Mediante la utilización de las tecnologías de EC, es posible reducir el volumen de tráfico transferido entre la capa de IoT y la nube (Ai et al., 2018). Además, esta tecnología permite la ejecución de modelos de aprendizaje automático en el borde de la red, reduciendo el tiempo de respuesta y proporcionando un cierto nivel de servicio aunque se interrumpa la comunicación con la nube; algo frecuente en escenarios donde la conectividad a Internet es limitada (por ejemplo, entornos agrícolas en zonas rurales) (Alonso, Sittón-Candanedo, García, et al., 2020).

Se llevó a cabo una investigación en la que se pretendió desarrollar una estrategia sobre cómo las explotaciones agrícolas pueden hacer un seguimiento de su eficiencia ambiental y evaluar sus posibilidades de aumentar su rentabilidad, tal y como se muestra en la Figura 3.2.

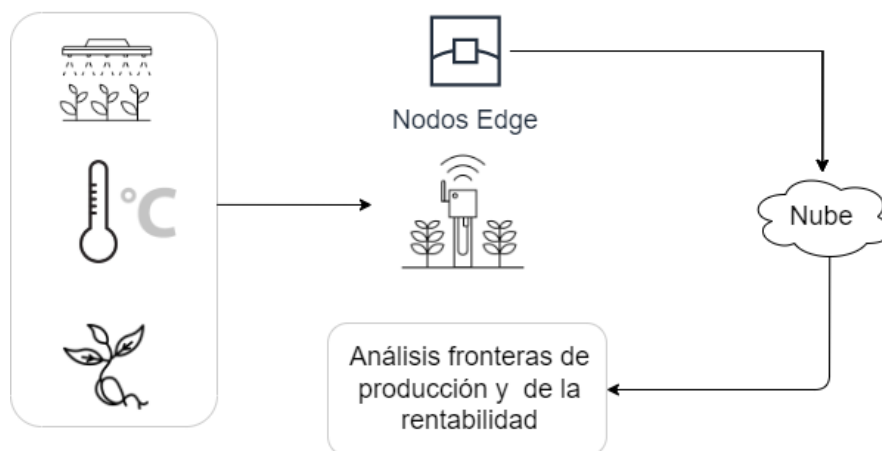


FIGURA 3.2: Diagrama seguimiento eficiencia ambiental en explotaciones agrícolas.

Diagrama adaptado de la publicación (Pérez-Pons, Parra-Domínguez, Chamoso, Plaza, & Alonso, 2020).

Debido a todas las regulaciones y niveles de producción permitidos y sus implicaciones, esta investigación presenta un estudio de caso orientado a la eficiencia en una granja mixta en España. En este sentido, existen diferentes enfoques para medir la rentabilidad y la eficiencia ambiental, siendo el SFA y el DEA. El SFA y el DEA, son dos de los más practicados, como recogió Lansink & Wall (2014). El DEA es un enfoque basado en datos para evaluar el rendimiento de diferentes unidades, conocidas como Unidades de Toma de Decisiones (Decision Making Units - DMU) (Charnes, Cooper, Lewin, & Seiford, 1997). Las DMU representan la eficacia con la que los inputs se convierten en outputs. Igualmente, las entradas y salidas pueden medirse en diferentes unidades, siendo utilizadas en muchos estudios en el campo de la eficiencia ambiental y ecológica (Dyckhoff & Allen, 2001; Theodoridis & Anwar, 2011) y también, lo que hace que sea un enfoque interesante para el actual análisis de eficiencia.

Hace casi 20 años, en 2002, Nuthall (2004) realizó un experimento para conocer cómo la tecnología afectaba a la rentabilidad y rendimiento de las explotaciones; y, en ese caso, tener acceso a un ordenador ya suponía un aumento de su beneficio. Varios años después, Piedra-Muñoz, Galdeano-Gómez, & Pérez-Mesa (2016) tomaron diferentes experimentos y revisaron años de mejoras en la aplicación de la tecnología y su compatibilidad con la sostenibilidad. Teniendo en cuenta que la tecnología es ya una realidad en el sector agrícola, en los próximos años los datos transferidos a la nube son un input que se utilizará porque afecta a la eficiencia del output. El estudio realizado muestra que la aplicación de técnicas de vanguardia,

como IoT y EC, a largo plazo puede representar una ventaja competitiva a la hora de medir la eficiencia de las Unidades de Decisión. Con este fin, se ha realizado un estudio que incorpora el tráfico de datos a la Nube como input, reflejando la importancia de la tecnología a la hora de analizar la producción del DEA. Se han presentado los resultados tras analizar el DEA, mostrando los valores más eficientes en términos de eficiencia de las DMU cuando existen variaciones en el tráfico de datos a la Nube, siendo un activo representativo.

El paradigma agro-tecnológico está llevando a escenarios de gran escala (es decir, explotaciones con millones de hectáreas con un gran número de sensores) lo que se traduce en un aumento del tráfico de datos a la Nube.

El objetivo de esta investigación era permitir la monitorización de las condiciones meteorológicas a través de estaciones agro-meteorológicas inalámbricas para evitar situaciones inciertas y medir la eficiencia de la granja. Para ello se dividieron los experimentos realizados en dos partes. La primera parte está relacionada con la monitorización de las condiciones meteorológicas, realizada a través de la plataforma IoT y Edge Computing. La segunda se refiere a la medición de la eficiencia a través del DEA. Para cumplir con la primera parte, se han considerado tres magnitudes diferentes (lluvia, temperatura y humedad del aire) entre las recogidas por el conjunto de estaciones meteorológicas inalámbricas agro-meteorológicas instaladas en la finca (Alonso, Sittón-Candanedo, García, et al., 2020). Los datos recogidos de estas estaciones se transfieren a la nube, teniendo en cuenta las medidas medias diarias.

Los datos transferidos a la nube se midieron tanto en la versión anterior del sistema (es decir, sin utilizar la plataforma Edge Computing) como en la nueva versión del sistema de monitorización (es decir, basada en la plataforma GECA Edge-IoT) para comparar el efecto en la eficiencia; considerando las variaciones de costes como inputs para la producción.

Alonso, Sittón-Candanedo, García, et al. (2020) demostraron una reducción en los costes de transferencia de datos al aplicar Edge Computing para filtrar y preprocesar datos utilizando la nueva plataforma basada en GECA. Otros autores como Chen, Shi, Yang, & Xu (2018) y Guillén et al. (2020), también demostraron una reducción en la transferencia de datos a la nube en AP.

La solución propuesta tiene en cuenta algunas de las variables más importantes del índice de rendimiento ambiental (Index, 2018) y cómo reducir su consumo con y la plataforma Edge-IoT. En la investigación se consideran tres efectos directos con la metodología propuesta. Se han colocado diferentes sensores en la granja, y otras variables se han utilizado como entradas para el seguimiento de la información.

Las futuras líneas de investigación que surgieron de este artículo se describen en el siguiente apartado y que contemplaban el incluir aprendizaje automático para predecir los niveles futuros de consumo de recursos. Se investigó la aplicación de tecnologías que permitan la monitorización de los cultivos o del ganado sin requerir excelentes conexiones a la red, así como el análisis de la eficiencia de la forma más inclusiva a nivel tecnológico y de recursos.

El contenido de este Capítulo ha sido adaptado del artículo *Increasing Profitability and Monitoring Environmental Performance: A Case Study in the Agri-Food Industry through an Edge-IoT Platform* de la revista *Sustainability*.

3.2.1. Tecnología en el sector agropecuario

En general, las nuevas tecnologías y aplicaciones, como el aprendizaje automático, ayudan a aumentar la precisión de los algoritmos de predicción. Una revisión sistemática sirve para tener una visión general del estado del arte en el campo, la revisión llevada a cabo y descrita al inicio de esta Sección muestra que en términos de número de publicaciones, este campo está creciendo y cada vez empiezan a surgir modelos más innovadores y conjuntos que ofrecen mejores capacidades de predicción. Teniendo en cuenta los criterios de búsqueda de este artículo, la tendencia y el mayor grado de innovación, así como el número de artículos publicados, la bolsa es donde hay más aplicaciones de aprendizaje automático en econometría. Teniendo en cuenta los resultados de esta revisión sistemática, existen muchas posibilidades para futuras líneas de investigación, donde se podrían combinar modelos econométricos con modelos de aprendizaje automático supervisado, y abrir un nuevo paradigma para la creación e implementación de modelos híbridos. Además, otras posibles líneas de investigación futuras serían analizar con más detalle algunos sectores de actividad específicos en los que la mayoría de las publicaciones están creciendo, en los que convergen varios tipos de

aplicaciones de aprendizaje automático y modelos econométricos. Un ejemplo de estos tipos de actividad podría ser el sector agropecuario.

Tradicionalmente, el manejo del campo consiste en inspeccionar visualmente el desarrollo de los cultivos para llegar a un diagnóstico con el que los agricultores toman decisiones y actúan dando diferentes tratamientos a sus cultivos. En los lugares donde la tecnología aún no ha llegado, se sigue utilizando este enfoque que se basa en la experiencia de los agricultores en el campo y en sus observaciones.

Hoy en día, hay un número creciente de sistemas y arquitecturas en la industria agrícola que aprovechan la tecnología para aumentar la eficiencia de los recursos de las explotaciones agrícolas (Alonso, Sittón-Candanedo, Casado-Vara, et al., 2020; Pérez-Pons, Plaza-Hernández, Alonso, Parra-Domínguez, & Prieto, 2021). La aparición de los sistemas de apoyo a la decisión (DSS) en la agricultura de precisión (AP) ha sido posible gracias al continuo progreso de las tecnologías de la información.

Uno de los principales objetivos de la AP es mejorar la sostenibilidad agrícola y optimizar las decisiones de cultivo, teniendo en cuenta la variabilidad del campo y los valores de los parámetros específicos del lugar. Las mediciones en bruto de los parámetros clave de los cultivos deben procesarse de forma eficiente para que los números o las imágenes se conviertan en información inequívocamente valiosa. En ocasiones, las limitaciones técnicas y de costes dificultan el despliegue de infraestructuras de AP como sistemas de apoyo a la toma de decisiones en entornos de pequeños agricultores. Por ello, han surgido aplicaciones, sistemas y arquitecturas que pretenden reducir los costes.

Uno de los aspectos fundamentales de la AP es el seguimiento de las mediciones de los sensores (Kpienbaareh, Kansanga, & Luginaah, 2019), así como los sistemas basados en la nube para la recogida, el procesamiento y el almacenamiento de datos. A partir de estos datos, se pueden realizar predicciones. Últimamente, la implementación de arquitecturas de borde se ha convertido en una solución popular en el campo de la agricultura de precisión (Alonso, Sittón-Candanedo, Casado-Vara, et al., 2020; Sittón-Candanedo, Alonso, Corchado, Rodríguez-González, & Casado-Vara, 2019).

Los avances actuales en la gestión de datos están haciendo que la AP crezca exponencialmente, ya que los datos se han convertido en el elemento clave de

la agricultura moderna, ayudando a los agricultores a tomar decisiones críticas. En la figura 3.3, se muestran algunos de los aspectos que se suelen rastrear en la AP.

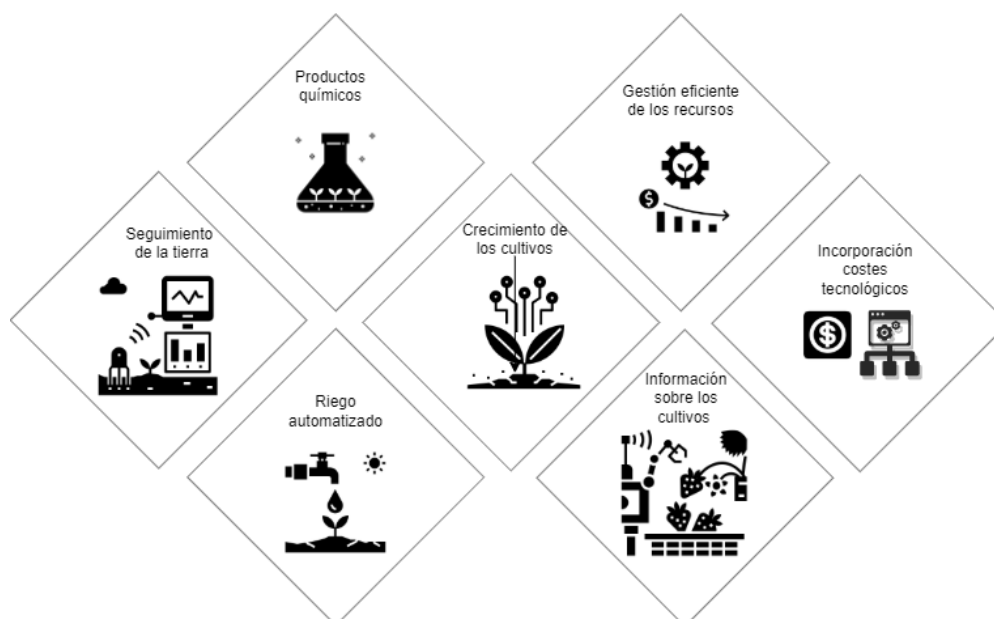


FIGURA 3.3: Diagrama agricultura de precisión.

3.3. Sistemas de apoyo a la toma de decisiones basados en preferencias

El mercado de la inversión es un sector competitivo con barreras de entrada relativamente bajas, pero con un alto riesgo asociado (Kaplan & Schoar, 2005). Siempre hay una serie de riesgos asociados a una inversión (Panousi & Papanikolaou, 2012), por ejemplo, en términos de finanzas, se pueden identificar dos riesgos principales; (a) el riesgo sistémico (Hill & Stone, 1980) es el que se deriva directamente del mercado mientras que (b) el riesgo no sistémico se deriva directamente de la empresa que se está evaluando (Peterson & Arun, 2018) y es aquel cuya predicción es más factible.

A la hora de tomar una decisión tanto dentro de una empresa, e intentar utilizar la que parece más óptima, también hay otros factores que son determinantes como las compras de productos o su venta, aún pudiendo hacer los procesos los más óptimos posibles en cuanto al funcionamiento de la empresa, lo generado o lo comprado por la misma tienen que tener en cuenta terceras partes, y por ello,

resulta indispensable el poder contar con sistemas de apoyo a la toma de decisiones (tanto para los inputs como para los outputs).

3.3.1. Sistema multi-agente basado en preferencias y aprendizaje profundo para un mercado agrícola sostenible

El uso de nuevas tecnologías, así como la reducción del consumo de recursos naturales o la generación de impactos de efecto invernadero, es un elemento cada vez más importante a la hora de elegir un proveedor (Gurel, Acar, Onden, & Gumus, 2015). Para mitigar la dificultad de elegir un proveedor según parámetros de sostenibilidad, así como disponer de información transparente del mercado se diseñó un sistema multi-agente. El sistema presentado por (Pérez-Pons, Alonso, García, Marreiros, & Corchado, 2021) y también anexado en esta Tesis Doctoral ayuda a los potenciales compradores a identificar un proveedor según sus preferencias en términos de sostenibilidad, así como de uso eficiente de los recursos. Por tanto, el sistema diseñado tiene dos resultados principales que se presentan a continuación. El primer resultado está relacionado con el ranking de selección de proveedores, y el segundo con la previsión de precios. Para desarrollar el sistema se ha utilizado una arquitectura Multi-Agente - MAS ya que un MAS puede definirse como una colección de entidades computacionales, posiblemente heterogéneas, que tienen sus propias capacidades de resolución de problemas y que son capaces de interactuar para alcanzar un objetivo global. En el caso de esta investigación, cada agente tiene un rol (Oliveira, Fischer, & Stepankova, 1999).

Como se muestra en la figura 3.4, hay tres agentes que se comunican para proporcionar la mejor opción según las preferencias del usuario; se trata de un sistema de agentes que colaboran estrechamente y en el que cada agente tiene sus propias capacidades y conocimientos especializados, y ningún agente tiene un conocimiento completo del mundo. Las funciones principales incluyen la recuperación de información del mercado de futuros agrícolas, el aprendizaje por refuerzo profundo para la identificación del precio más bajo y el sistema de ponderación basado en las preferencias de los usuarios o potenciales compradores. La arquitectura diseñada para el MAS se ha implementado con la biblioteca SPADE (Gregori, Cámara, & Bada, 2006).



FIGURA 3.4: Sistema multi-agente propuesto

Adaptación de la publicación Pérez-Pons, Alonso, et al. (2021).

Según H. Liu, Zhang, Zhou, & Ru (2018), una política de venta óptima puede aumentar el beneficio esperado. Por lo tanto, poder combinar y obtener información de precios en tiempo real permite ajustar los precios de compra en previsión del problema de tomar decisiones estratégicas de venta (Shi, Zhao, Kiwanuka, & Chang, 2019). En el caso que nos ocupa, se propone un paradigma en el que cada parte interesada (es decir, las cadenas de gran consumo directas, los mayoristas, etc.) actúa de forma individual.

Para probar el MAS, un agente recoge información de un periodo concreto de un producto en el mercado de futuros. A continuación, las funciones de utilidad basadas en las preferencias recuperan la información de la base de datos histórica y también las preferencias del usuario en términos de pesos para cada atributo. En el caso de este estudio, como el enfoque está orientado a las preferencias y hay unos pesos según cada atributo, la ecuación se representa como en la Ecuación 3.1. Para los valores de los demás atributos, la equivalencia se muestra en la Ecuación 3.2. A continuación, el agente de aprendizaje profundo se encarga de modelar las tendencias no lineales de las series temporales de precios de las acciones, mediante la predicción de valores y la identificación de los precios más bajos.

$$(w_2, \dots, w_n)(w_2, \dots, w_n) \quad (3.1)$$

$$(y_1, w) \sim < (x_1, w) : (z_1, w) > \quad (3.2)$$

El resultado y futura línea de investigación resultaba poder evaluar el modelo con un conjunto de proveedores con el fin de incluir la ubicación como otro factor interesante a la hora de determinar los precios y crear una realidad entre los requisitos de los posibles compradores y proveedores. Para realizar esas recomendaciones también a la hora de recomendar potenciales proveedores o empresas en las que invertir.

El contenido de este Capítulo ha sido adaptado del artículo *Deep Q-Learning and Preference Based Multi-Agent System for Sustainable Agricultural Market* de la revista *Sensors*.

3.3.2. Modelo híbrido de apoyo a la toma de decisiones en escenarios de inversión a partir de razonamiento basado en casos

Dada la necesidad de un método válido y universal que ayude a los inversores a tomar decisiones óptimas sobre una posible inversión, se ha desarrollado un modelo híbrido basado en un sistema de Razonamiento Basado en Casos (Aamodt & Plaza, 1994). La metodología propuesta tiene como objetivo maximizar la eficiencia y facilitar el proceso de toma de decisiones. El modelo propuesto tiene en cuenta los factores financieros que normalmente intervienen en una decisión de inversión (Cronqvist, Siegel, & Yu, 2015) pero también los activos intangibles que son analizados subjetivamente por el potencial inversor. Esta propuesta es muy enriquecedora porque se ha demostrado que los inversores que invierten en campos más diversos e innovadores consiguen mayores rendimientos en sus inversiones (Haro-de Rosario, del Carmen Caba-Pérez, & Cazorla-Papis, 2014) sin embargo, los inversores tienen tendencia a invertir en sectores con los que están familiarizados.

A lo largo de los años, autores como Fried & Hisrich (1994), Tyebjee & Bruno (1984), Sirower & O'Byrne (1998) evaluaron diferentes criterios de inversión que se han considerado fundamentales para cualquier proceso de adquisición de empresas, sin embargo, ninguno de ellos aportó aún un método concluyente. La inversión en pequeñas y medianas empresas (PYME) que no cotizan en bolsa y en empresas de nueva creación es clave para el crecimiento de la economía, independientemente

del país en el que opere la empresa (Jeng & Wells, 2000). En los últimos años, han surgido nuevos modelos de negocio y las empresas que necesitan una inyección de capital externo para seguir creciendo tienen que ser capaces de atraer a los inversores. Como es lógico, este tema ha recibido mucha atención por parte de la comunidad investigadora, en particular las inversiones realizadas por el Capital Riesgo (VC) (Granz, Henn, & Lutz, 2020) y también por los Business Angels (BA) (White & Dumay, 2017). Asimismo, han surgido nuevas teorías de optimización y eficiencia a partir del análisis de los criterios evaluados por un conjunto diverso de tipos de inversores (Granz et al., 2020) (Block, Fisch, Vismara, & Andres, 2019). El sistema OCI-CBR pretende maximizar las posibilidades de inversión pero también reducir la incertidumbre. En el modelo de inversión empresarial, al igual que en el modelo bancario descrito por Bontempi (2016), los inversores se enfrentan a un cierto grado de incertidumbre. La incertidumbre es un riesgo y un factor que puede provocar un retraso en la toma de decisiones (Y. Li & Mahoney, 2011), donde las inversiones potencialmente óptimas se vieron afectadas negativamente por la incertidumbre.

El modelo de recomendación de inversiones que se ha creado tiene los elementos fundamentales de un sistema de razonamiento basado en casos. Para que un sistema de ayuda a la toma de decisiones funcione correctamente es necesario establecer un mecanismo que garantice la presencia de información adecuada en la base de casos y para ello se utiliza un sistema de clasificación.

Durante el ciclo de trabajo de OCI-CBR, se mide la similitud entre empresas y entre inversores. La similitud se determina mediante un conjunto de atributos seleccionados en cada caso, a los que se aplica una fórmula común. Para ello se utiliza una metodología basada en el CBR, empleando un conjunto de datos con parejas inversor-empresa.

Para un inversor que busca una empresa en la que invertir, el proceso es el siguiente:

1. Las bases de datos se filtran para encontrar las empresas que cumplen los requisitos del inversor.
2. Se recuperan los perfiles de inversores más similares, utilizando métricas de similitud basadas en el cuantil de la distribución de atributos en el conjunto de datos.

3. Se extrae una lista de empresas en las que han participado los inversores más afines.
4. Dentro de la lista de empresas que han sido seleccionadas como válidas en el paso 1 definido como P_Csim, buscamos aquellas que son más similares a las encontradas en el paso 3, definido como P_Isim

Para garantizar un buen rendimiento en la fase **recuperación**, la escala de todos los atributos debe ser la misma. Para ello, los valores de los atributos se han distribuido mediante el método de transformación robusta

$$s_A(x, y) = 1 - |Q_A(x) - Q_A(y)|, \quad (3.3)$$

donde A es un atributo aleatorio y Q_A es la función cuantil de ese atributo. Para tener en cuenta varios atributos simultáneamente, se utiliza una media ponderada de los mismos, es decir.

$$S(X, Y) = \frac{\sum_{A \in \mathcal{A}} w_A \cdot s_A(X(A), Y(A))}{\sum_{A \in \mathcal{A}} w_A}, \quad (3.4)$$

donde \mathcal{A} es el conjunto de atributos, $X(A), Y(A)$ son los valores del atributo A para dos instancias X, Y y w_A es el peso dado a cada atributo A , que define su importancia.

A falta de un estudio posterior que considere la distribución de pesos más adecuada para la noción de similitud anterior, en los resultados de este trabajo se utiliza un peso uniforme ($w_A = 1$).

En la fase de **reutilización**, el sistema obtiene una lista de empresas a las que se asignan dos puntuaciones, cada una de ellas en el intervalo $[0, 1]$. A saber, una puntuación asignada por el modelo de recomendación en base a la probabilidad de similitud en cuanto a los atributos de la empresa, P_{Csim} , y una puntuación combinada calculada en base a las similitudes entre los inversores y las empresas que han dado lugar a la propuesta de recomendación, P_{Isim} , obtenida por

$$P_{Isim} = \max_{I', C'} S(I, I') \cdot S(C, C'), \quad (3.5)$$

donde I es el inversor considerado, C es la empresa recomendada y I', C' es un inversor y una empresa similares para los que se hace la recomendación. El máximo en la fórmula indica que en los casos en que más de un inversor o empresa proporciona la misma recomendación, se toma el valor más alto de las similitudes. La **fase de revisión** implica el conocimiento de los expertos; es donde los inversores analizan las posibles inversiones. Así, las empresas que son recomendadas por el sistema como oportunidades de inversión son clasificadas por un inversor experto, que indica las empresas en las que consideraría invertir. No obstante, en este caso, para evaluar la fase de revisión, el CBR se ha evaluado teniendo en cuenta las empresas en las que los inversores ya estaban invirtiendo. Posteriormente, en la fase **retain**, se genera una tabla independiente en la base de datos con información adicional sobre los inversores y las empresas elegidas por ellos, esto permite al CBR almacenar las empresas clasificadas por puntuación, similitud y afinidad en su base de casos. Como resultado del proceso CBR, se obtiene una lista de empresas, a cada empresa se le asignan tres puntuaciones en el intervalo $[0, 1]$. Una puntuación combinada de las similitudes entre los inversores y las empresas que llevaron a la propuesta de la recomendación, P_{Csim} y P_{Isim} . Finalmente, la tercera puntuación, denominada $P_{Affinity}$ se obtiene mediante P_{Csim} y P_{Isim} .

El modelo propuesto ha coordinado la aplicación del aprendizaje automático y RF, conduciendo a la obtención de resultados óptimos en forma de recomendaciones de inversión en empresas que tienen buena solvencia y buenas perspectivas de futuro, debido al proceso de poda, haciendo poco probable que quiebren.

Una de las futuras líneas de investigación que surgieron fue analizar las diferentes métricas de evaluación que se utilizaban en el proceso de poda en la clasificación binaria. Es por ello que se investigó y se realizó un artículo que se describe en la próxima Sección en el que se comparaban diferentes métricas de evaluación para ver que algoritmo es mejor.

El contenido de este Capítulo ha sido adaptado del artículo: [Hybrid model for decision support in preference-aware investment scenarios](#) que se encuentra en estado *en revisión* en la revista *Expert Systems with applications*.

3.3.2.1. Evolución del modelo híbrido basado en razonamiento basado en casos: Sistema para la identificación de métricas de evaluación y reducción dimensional para algoritmos de clasificación binaria

En los últimos años, varios investigadores han demostrado que los algoritmos de aprendizaje automático obtienen mejores resultados que los métodos tradicionales en la predicción de quiebras cuando se consideran los mismos atributos (Barboza, Kimura, & Altman, 2017; Hosaka, 2019; Kim, Cho, & Ryu, 2020). La investigación que se describe a continuación se enmarca en las predicciones de clasificación binaria utilizando el mínimo número de atributos. En este caso, el éxito de la predicción binaria viene dado no por la precisión sino por la métrica de evaluación. Los algoritmos se evalúan según la certeza de las métricas de clasificación de las clases y no sólo según la precisión del algoritmo.

A lo largo de los años, ha habido un interés continuado por los modelos y metodologías de predicción de quiebras con diferentes fines, como prevenir situaciones de quiebra inesperadas o hacer estudios de viabilidad financiera de empresas que puedan ser de interés para los inversores (Bellovary, Giacomino, & Akers, 2007). Existen muchas variantes en cuanto a las diferencias en la estructura financiera, las variaciones pueden ir desde tener mayores niveles de liquidez hasta mayores ratios de solvencia. Las diferencias en las estructuras financieras entre industrias son indicadores importantes y relevantes (Huang, Liu, Bai, & Zhang, 2020; L. Li & Islam, 2019). Gracias a las múltiples oportunidades que ofrece la inteligencia artificial, se han desarrollado nuevos enfoques para tratar las situaciones de quiebra (Kim et al., 2020). En los últimos años, varios autores han informado sobre las aplicaciones de la inteligencia artificial para la predicción de quiebras. Diferentes revisiones sistemáticas han recogido las nuevas técnicas y combinaciones de atributos, (Wang et al., 2017), (W. Zhang et al., 2017), (Devi & Radhika, 2018), (W. Zhang et al., 2017), (Qu, Quan, Lei, & Shi, 2019), y (Alaka et al., 2018), comparando el rendimiento de diferentes modelos de aprendizaje automático en la predicción de quiebras.

Debido a las incertidumbres en cuanto a qué método y métrica de evaluación para la predicción binaria identificadas en el estado del arte, no existe ningún método que pueda dar una solución integral a todas ellas. Para los fines de este estudio, se ha considerado un problema de clasificación binaria en bancarrota con datos del

mundo real. Utilizando métodos estadísticos tradicionales y técnicas de aprendizaje automático, el objetivo de esta investigación es proponer una metodología para los siguientes tres objetivos de investigación:

- La posibilidad de predecir la quiebra sin considerar numerosos atributos financieros, y sin sesgar los datos para los futuros algoritmos de clasificación de aprendizaje automático.
- Evaluar cómo se comportan los métodos de clasificación binaria en cada industria y valorar la relevancia del peso de cada atributo en función de una etiqueta en condiciones similares (en este caso atributos financieros).
- Evitar los sesgos de resultado proporcionando una comparación de las métricas de evaluación en la clasificación binaria.

La selección de atributos consiste en reducir el número de atributos mediante el análisis de la varianza acumulada explicada (VAE). El VAE es un análisis a menudo utilizado para analizar los datos antes de hacer un análisis de componentes principales (PCA), sin embargo, en este caso también se ha utilizado para reforzar la extracción de características. Teniendo en cuenta el tipo de datos y el hecho de que el problema a resolver implica la predicción de una etiqueta binaria, se han utilizado algoritmos de clasificación supervisada de aprendizaje automático. Por tanto, se han comparado seis algoritmos de clasificación de aprendizaje automático supervisado para predecir la probabilidad de quiebra en las diferentes muestras de datos. Por último, se ha evaluado la clasificación de clase de los algoritmos, utilizando las siguientes métricas de evaluación; Coeficiente de Correlación de Matthew (MCC), Área bajo la curva de la característica operativa del receptor (ROC-AUC) y Área bajo la curva de precisión-recuperación (AUC-PR).

La lista final de los algoritmos que se han considerado en este estudio son: Gradient Boosting (GB) que construye el modelo de forma escalonada y generaliza el modelo permitiendo la optimización arbitraria de una función de pérdida diferenciable (Zieba, Tomczak, & Tomczak, 2016), Gaussian Naive Bayes (NB) que se basa en el teorema de Bayes (Eirola et al., 2015) (Sharma & Mavani, 2011), Decision Tree Classifier (DTC) que se basa en un árbol de decisión, es un modelo predictivo que mapea las observaciones sobre un ítem a conclusiones sobre el valor objetivo del ítem (Foroghi, Monadjemi, et al., 2011), Random Forest (RF) que consiste en una

combinación de árboles predictores tal que cada árbol depende de los valores de un vector aleatorio probado independientemente y con la misma distribución para cada árbol predictor, K-Nearest Neighbors (KNN) que busca las observaciones más cercanas que se intentan predecir y clasifica el punto de interés en función de la mayor cantidad de datos que lo rodean, (Imandoust & Bolandraftar, 2013) y Support Vector Machine Clasificador (SVMC) que se basa en el concepto de hiperplano (C.-W. Hsu, Chang, Lin, et al., 2003). Para superar el posible sesgo en el entrenamiento de los algoritmos, se utilizó la validación cruzada de 5 veces. La técnica de validación cruzada ha permitido reducir los problemas de sobreajuste y también evaluar los resultados del análisis asegurando que son independientes de la partición entre los datos de entrenamiento y los de prueba. La técnica de Grid Search se ha implementado para encontrar los hiperparámetros óptimos para cada uno de los modelos, lo que resulta en un modelo con mayor precisión de predicción. Las tres métricas que se han tenido en cuenta son: el Área Bajo la Curva de Característica Operativa del Receptor (ROC-AUC), el Coeficiente de Correlación de Matthew (MCC) y el Área Bajo la Curva de Precisión-Recalificación (AUC-PR). Cuando se aplica un algoritmo de clasificación a un conjunto de datos para la predicción de clases, uno de los factores más cruciales que hay que evaluar es el rendimiento del algoritmo. Existen diferentes opciones para determinar la precisión del algoritmo. Por ejemplo, se podría utilizar cualquier métrica conocida, pero sólo podría indicar el rendimiento global, pero no podría identificar los falsos positivos o los falsos negativos. Esto ocurre cuando un algoritmo identifica una clase como otra, incurriendo en errores estadísticos de tipo I y II, lo que da lugar a falsos negativos y falsos positivos. La Característica Operativa del Receptor (ROC) es un enfoque para evaluar la precisión de los algoritmos. El ROC ofrece una representación gráfica de la sensibilidad a la especificidad para un sistema de clasificación binaria a medida que se varía el umbral de discriminación. En cuanto a la variedad de datos, en el caso de los resultados obtenidos al separar los datos por industria, algunos de los lotes de datos de la industria estaban desequilibrados en cuanto a la clase. En el ámbito de la predicción de quiebras, se ha estudiado que las características de un conjunto de datos desequilibrado sesgan el rendimiento del algoritmo (Veganzones & Séverin, 2018). Por ello, se deben emplear métricas de evaluación (He & Garcia, 2009), (Davis & Goadrich,

2006). Por ejemplo, las referenciadas anteriormente, se comportan mejor que ROC en esta situación porque no tienen en cuenta las clases que hacen que el conjunto de datos esté desequilibrado. Para contrastar el ROC-AUC y el MCC (Chicco & Jurman, 2020), se ha considerado como alternativa el Área Bajo la Curva de la Precisión-Recuperación (Saito & Rehmsmeier, 2015).

Los resultados llevan a la conclusión de que, al aplicar algoritmos de clasificación de aprendizaje automático, la reducción de características generada por el modelo univariante funciona tan bien como los atributos tradicionales elegidos por los especialistas en la materia, como los atributos seleccionados por Altman en el modelo de puntuación Z. Además, en un intento de identificar un algoritmo que pudiera tener un mejor rendimiento en el análisis de empresas de un determinado sector, los resultados no son concluyentes. Por otro lado, uno de los resultados más importantes obtenidos en la investigación es que las métricas de evaluación son decisivas a la hora de elegir un algoritmo para predecir las clasificaciones. En este trabajo se ha demostrado que a la hora de evaluar los diferentes algoritmos hay que tener en cuenta todas las métricas posibles para evitar conclusiones sesgadas. El uso de una sola métrica de evaluación puede conducir a errores en las aplicaciones posteriores. En este trabajo se ha comparado y concluido el progreso de los modelos de aprendizaje automático de clasificación y la importancia de las métricas de evaluación con respecto a la predicción de quiebras.

[El contenido de este Capítulo ha sido adaptado del artículo *Evaluation metrics and dimensional reduction for binary classification algorithms: A case study in Bankruptcy prediction* de la revista *Knowledge Engineering Review*.](#)

3.4. Conclusiones

El presente Capítulo 3 recoge los diferentes experimentos llevados a cabo y que forman esta Tesis Doctoral por compendio de artículos.

La Sección 3.1 presenta el estado del arte en aplicaciones econométricas y de aprendizaje automático, así como los diferentes campos de aplicación y sus derivados casos de uso. A continuación, la Sección 3.1 presenta el experimento realizado en el que se evaluaron las diferentes fronteras de producción incorporando los costes de tecnología en términos de transferencia de datos. Finalmente, en la

Sección 3.3 se explican los últimos dos experimentos basados en sistemas de apoyo a la toma de decisiones. En primer lugar, en la sección 3.3.1 se presenta el sistema multi-agente para la compra sostenible de productos agrícolas, y a continuación, en la Sección 3.3.2 se presenta un modelo híbrido para la toma de decisiones en inversiones. Además, la Sección 3.3.2.1 incorpora las mejoras que se realizaron sobre la metodología en el experimento presentado en la Sección continuación, en la Sección 3.3.2.

Capítulo 4

Evidencias y Resultados



**VNiVERSIDAD
D SALAMANCA**

CAMPUS DE EXCELENCIA INTERNACIONAL

Evidencias y Resultados

En este Capítulo se presenta el conjunto de publicaciones en revistas científicas y congresos internacionales en los cuales se ha contribuido, reflejando el desarrollo y los resultados de las diferentes líneas de investigación que han culminado en esta Tesis Doctoral. En este sentido, la estructura de este Capítulo es la siguiente. La Sección 4.1 detalla las publicaciones realizadas en revistas científicas, conferencias y workshops internacionales, expuestas en orden cronológico inverso. La Sección 4.2 enumera los proyectos en los que se ha participado y que han sustentado las diferentes líneas de investigación desarrolladas que han llevado a esta Tesis Doctoral. La Sección 4.3 describe las estancias llevadas a cabo en organismos de investigación internacionales relacionadas con las líneas de investigación de esta Tesis Doctoral.

4.1. Publicaciones

En primer lugar, se detallan las diferentes publicaciones llevadas a cabo en revistas científicas internacionales, capítulos de libro, así como congresos internacionales y workshops, relacionadas con las líneas de investigación que han llevado a la elaboración de esta Tesis Doctoral. En cada bloque, se presentan las entradas en orden cronológico inverso según la fecha de publicación. En el caso de las revistas científicas internacionales, se señalan en **negrita** aquéllas que forman parte de la Tesis Doctoral como compendio de artículos/publicaciones.

4.1.1. Publicaciones en revistas científicas internacionales

- **Pérez-Pons, M. E., Parra-Domínguez, J., Sigeru Omatu, Herrera-Viedma, E Corchado, JM. (2022). Machine Learning and**

Traditional Econometric Models: A Systematic Mapping Study.
JAISCR, 12(2), 79-100 (Q3 - JCR 2.50)

- **Pérez-Pons, M. E., Parra-Domínguez, J., Hernández G, Herrera-Viedma, E Corchado, JM. (2021). Evaluation metrics and dimensional reduction for binary classification algorithms: A case study in Bankruptcy prediction. KER, Knowledge Engineering Review (Q4 - JCR 1.09)**
- **Pérez-Pons, M. E., Alonso, R. S., García, O., Marreiros, G., Corchado, J. M. (2021). Deep Q-Learning and preference based multi-agent system for sustainable agricultural market. Sensors, 21(16), 5276(Q1 - JCR 3.57)**
- **Pérez-Pons, M. E., Plaza-Hernández, M., Alonso, R. S., Parra-Domínguez, J., Prieto, J. (2021). Increasing Profitability and Monitoring Environmental Performance: A Case Study in the Agri-Food Industry through an Edge-IoT Platform. Sustainability, 13(1), 283(14), 5706(Q2 - JCR 3.25)**
- **Queiroz, J., Leitão, P., Pontes, J., Chaves, A., Parra, J., Perez-Pons, M. E. (2020). A Quality Innovation Strategy for an Inter-regional Digital Innovation Hub. ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal, 9(4), 31-45.**
- **Pérez-Pons, M. E., Parra-Domínguez, J., Chamoso, P., Plaza, M., Alonso, R. (2020). Efficiency, profitability and productivity: Technological applications in the agricultural sector. ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal, 9(4).**

4.1.2. Publicaciones en congresos internacionales y workshops

- Pérez-Pons, M. E., Alonso, R. S., Parra-Domínguez, J., Plaza-Hernández, M., Trabelsi, S. (2021, April). An Edge-IoT Architecture and Regression Techniques Applied to an Agriculture Industry Scenario. In Sustainable Smart Cities and Territories International Conference (pp. 92-102). Springer, Cham.
- Pérez-Pons, M. E., Parra, J., Hernández, G., González, J., Corchado, J. M. (2020, June). Machine Learning and Financial Ratios as an Alternative to Altman's Z-Score Bankruptcy Model in Spanish Companies. In The International Conference on Decision Economics (pp. 130-139). Springer, Cham.
- Mezquita, Y., Parra, J., Perez, E., Prieto, J., Corchado, J. M. (2019, May). Blockchain-Based Systems in Land Registry, A Survey of Their Use and Economic Implications. In Computational Intelligence in Security for Information Systems Conference (pp. 13-22). Springer, Cham.
- Pérez-Pons, M. E. Methodology for efficient capital investments in private sector companies. Doctoral Consortium Proceedings, 17.
- Javier Parra, María E Pérez-Pons, Jorge González. The Impact and Correlation of the Digital Transformation on GDP Growth in Different Regions Worldwide. International Symposium on Distributed Computing and Artificial Intelligence (2020).

4.2. Proyectos de investigación

A continuación se describen los proyectos en los que se ha participado y que han sustentado las diferentes líneas de investigación desarrolladas que han llevado a esta Tesis Doctoral:

- ÑNTELFIN: Artificial Intelligence for investment and value creation in SMEs through competitive analysis and business environment”, Reference: RTC-2017-6536-7, funded by the Ministry of Science, Innovation and Universities (Challenges-Collaboration 2017), the State Agency for Research (AEI) and the European Regional Development Fund (ERDF).
- Interreg Spain-Portugal V-A Program (PocTep) under grant 0677 DISRUPTIVE-2-E (Intensifying the activity of Digital Innovation Hubs within the PocTep region to boost the development of disruptive and last generation ICTs through cross-border cooperation).
- European Regional Development Fund (ERDF) through the Interreg Spain-Portugal V-A Program (POCTEP) under grant 0631-DIGITEC-3-E (Smart growth through the specialization of the cross-border business fabric in advanced digital technologies and blockchain).

4.3. Estancias internacionales

A continuación se detalla la estancia internacional llevada a cabo fuera de España en una institución de enseñanza superior de prestigio, realizando trabajos de investigación relacionados con esta Tesis Doctoral. A continuación se desglosan los detalles de dicha estancia:

- Instituto Superior de Engenharia do Porto / Instituto Politécnico do Porto (ISEP/IPP).
- Ciudad de la entidad: Porto, Região Norte, Portugal.
- Fecha de inicio – Fecha de fin: 01/05/2021 – 31/07/2021

Como resultado de dicha estancia se originaron publicaciones en conjunto con los investigadores del ISEP/IPP, como Pérez-Pons, Alonso, et al. (2021).

Capítulo 5

Conclusiones y Trabajo Futuro



**VNiVERSIDAD
D SALAMANCA**

CAMPUS DE EXCELENCIA INTERNACIONAL

Conclusiones y Trabajo Futuro

El caso de estudio presentado en esta Tesis Doctoral se centra en el sector agropecuario. En esta Tesis Doctoral se ha presentado una metodología que permite a los empresarios del sector agropecuario optimizar sus costes en la monitorización de la producción y consumos de recursos, así como ajustar los precios de venta de sus productos a partir de un sistema de apoyo a la toma de decisiones. Finalmente, como último paso dentro de la metodología, se ha incorporado un sistema de recomendación para fomentar la inversión en múltiples sectores.

Este Capítulo sirve, por tanto, para concluir esta Tesis Doctoral y presentar posibles líneas de trabajo futuro. La Sección 5.1 a continuación enumera las conclusiones obtenidas durante el trabajo de investigación llevado a cabo en esta Tesis Doctoral, así como su alineamiento con los objetivos planteados al inicio de la misma. Tras ello, la Sección 5.2 propone diferentes líneas de investigación futura posibles que surgen a raíz de los resultados y conclusiones obtenidos por esta Tesis Doctoral.

5.1. Conclusiones

Tal y como se ha expuesto en los Capítulos anteriores, existen diferentes investigaciones en metodologías de inteligencia artificial para fomentar la sostenibilidad en las empresas del sector agropecuario. También existen investigaciones centradas en analizar la rentabilidad de las empresas del sector agropecuario a partir de modelos econométricos o de inteligencia artificial. Sin embargo, hasta donde alcanza el conocimiento de la autora de esta Tesis Doctoral, todavía no existía una propuesta de metodología orientada a mejorar

la rentabilidad en las empresas del sector agropecuario, realizar predicciones sobre los precios de venta de los productos y poder incorporar dichas empresas a sistemas de recomendación en el que las mismas pudieran captar capital externo.

De este modo, los resultados obtenidos permiten afirmar que se ha cumplido tanto con el objetivo principal, lo que permite comprobar tanto la hipótesis planteada en el Capítulo 1, como con los objetivos específicos que fueron definidos en el mismo. A continuación se describen los resultados alcanzados para cada uno de los objetivos planteados inicialmente:

- (R1) Para dar respuesta al primer objetivo (OB1) de la Tesis Doctoral planteado en la Sección 1.2 del Capítulo 1, se ha realizado una revisión del estado del arte en predicciones realizadas a partir de métodos econométricos y de aprendizaje automático tal y como se detalla en la Sección 3.1. Del mismo modo, los experimentos realizados en la Sección anteriormente mencionada en esta Tesis Doctoral comparan directamente los modelos econométricos con los de aprendizaje automático.
- (R2) A continuación, a partir del segundo objetivo (OB2) y tercer objetivo (OB3) planteados al inicio de la investigación, se ha demostrado que la utilización de nuevas tecnologías como el EC permite a las empresas del sector agropecuario reducir sus costes e incrementar su rentabilidad. Para ello se ha realizado un experimento en el que a partir del DEA se ha comparado cómo afecta la incorporación de los costes de enviar los datos al cloud utilizado EC. El mencionado experimento se describe en la Sección 3.2.
- (R3) Con respecto al cuarto objetivo (OB4) y quinto objetivo (OB5) planteados al inicio de la investigación, para mitigar la dificultad de elegir un proveedor según parámetros de sostenibilidad, así como disponer de información transparente del mercado, se ha diseñado un sistema multi-agente (Pérez-Pons, Alonso, et al., 2021). Dicho sistema se ha descrito en la Sección 3.3.1.
- (R4) En cuanto al sexto objetivo (OB6) planteado al inicio de la investigación, los sistemas de recomendación permiten mejorar la toma de decisiones y permiten realizar planificaciones, inversiones personalizadas, así como

fomentar las inversiones sostenibles. Para ello se ha desarrollado con base en una metodología CBR un sistema recomendación de inversiones. El sistema mencionado se ha descrito en la Sección 3.3.2. Además, para mejorar el sistema de recomendación se ha implementado una mejora que permite identificar las métricas de evaluación más adecuadas en los casos de aprendizaje supervisado con clases no balanceadas 3.3.2.1.

5.2. Futuras Líneas de Investigación

Esta Sección incorpora nuevas líneas de investigación derivadas del trabajo realizado en esta Tesis Doctoral. A continuación se enumeran dichas líneas:

- (L1) Debido al diseño modular y escalable de la metodología, la aplicación de los diferentes experimentos es un proceso sencillo. En las próximas líneas de investigación deberá existir una adaptación constante más allá de la técnica, implicando recursos propios y la implementación de la metodología en regiones con limitado acceso a internet y conexiones inestables.
- (L2) Otra línea de investigación orientada la sostenibilidad y a una economía sostenible será el diseño de un sistema para evitar el exceso de producción. La propuesta consistirá en el diseño de un sistema sobre el que se realizarán predicciones de stock de productos, oferta y demanda con el fin de reducir los sobrantes de cualquier producto. También será determinante la agregación de compras a partir de la identificación de potenciales compradores (y una distancia) ahorrando contaminación y costes de transporte.
- (L3) Tomando como ejemplo el mercado de valores, a partir del concepto de teoría de juegos, se creará un sistema multi-agente en el que se llevará a cabo la compra y venta de productos, en los que habrá unos agentes demandantes y otros ofertantes que seguirán las leyes de la oferta y la demanda propias de un mercado en equilibrio. Todo ello incorporando limitaciones computacionales y, por tanto, sobre los que existirán diferentes restricciones.

Chapter 5

Conclusions and Future Work



**VNiVERSIDAD
D SALAMANCA**

CAMPUS DE EXCELENCIA INTERNACIONAL

Conclusions and Future Work

The case study presented in this doctoral thesis focuses on the agricultural sector. In this Doctoral Thesis, a methodology has been presented that allows entrepreneurs in the agricultural sector to optimise their costs in the monitoring and consumption of resources, as well as to adjust the selling prices of their products based on a decision support system. Finally, as a further step in the methodology, a recommendation system is incorporated in which investment in different sectors is encouraged. In the agricultural sector capital injections are crucial for the incorporation of new technologies that enable such methodologies to be carried out.

This Chapter therefore serves to conclude this Doctoral Thesis and to present possible lines of research for the future. The following section lists the conclusions drawn from the research conducted as part of this Doctoral Thesis, as well as their alignment with the objectives set out at the beginning of the thesis. After that, Section 5.2 proposes a series of lines of research for the future that arise from the results and conclusions obtained in this Doctoral Thesis.

5.1. Conclusiones

As presented in the previous chapters, there is a variety of research on artificial intelligence methodologies to promote sustainability in agribusiness. There are different researches that explore the use of artificial intelligence methodologies to create more sustainable, profitable and efficient companies, however, to the best of the author's knowledge, a methodology of such characteristics as described in this Doctoral Thesis, has not been proposed yet. It improves profitability and makes predictions on the selling prices of products. Moreover, it incorporates these

companies into recommendation systems in which companies could raise external capital.

Thus, it is possible to draw the conclusion that the main objective of this Doctoral Thesis has been met, which has made it possible to verify both, the hypothesis set out in Chapter 1, and the specific objectives that were defined therein. The results achieved for each of the initially stated objectives are described below:

- (R1) To respond to the first objective (OB1) of the Doctoral Thesis set out in Section 1.2 of Chapter 1, a review of the state of the art has been carried out as detailed in Section 3.1, where the prediction capabilities of different machine learning methods were assessed. In addition, as part of the experiments carried out in this Doctoral Thesis, a comparative study has been performed of econometric models and machine learning models.
- (R2) Then, to meet the second (OB2) and third (OB3) objectives set out in the introduction, it has been evidenced that the use of new technologies, such as edge computing, makes it possible for companies in the agricultural sector to reduce their costs and increase their profitability. To this end, an experiment was carried out (Pérez-Pons et al., 2020). And the experiment is described in Section 3.2.
- (R3) Regarding the fourth (OB4) and fifth (OB5) objectives, a multi-agent system has been designed which uses sustainability parameters to mitigate the difficulty of choosing a supplier and which provides transparent market information. This proposal is presented in (Pérez-Pons, Alonso, et al., 2021) and described in Section 3.3.1.
- (R4) Regarding the sixth objective (OB6), recommender systems enable better decision-making and allow for planning, customised and sustainable investments. For this purpose, an investment recommendation system has been developed on the basis of a CBR methodology and is described in Section 3.3.2. Furthermore, to improve the recommender system, an investigation of the evaluation metrics used in the pruning phase of the recommender system was carried out, which demonstrates the decisive role of algorithm evaluation metrics in supervised learning cases with unbalanced classes, as described in Section 3.3.2.1.

5.2. Future Lines of Research

This Section outlines the new lines of research that have opened up thanks to the results obtained in this Doctoral Thesis. These lines of research are described below:

- (L1) Due to the modular and scalable design of the methodology, the implementation of the different experiments is a straightforward process. In future lines of research, beyond the development of new techniques, there will have to be constant adaptation involving equity and the implementation of the methodology in regions with limited internet access and unstable connections.
- (L2) Another line of research oriented towards sustainability and a sustainable economy will be the design of a system for overproduction prevention. The proposal will consist of the design of a system on which product stock, supply and demand predictions will be made in order to reduce surpluses of any product. The aggregation of purchases based on the identification of potential buyers (and the distance between them) will also be decisive, reducing pollution and transport costs.
- (L3) Using the stock market as an example, on the basis of the concept of game theory a multi-agent system could be created in which products are bought and sold, in which there are agents that follow the laws of supply and demand. All this would incorporate computational constraints which would imply certain restrictions.

Bibliografía

- Aamodt, A., & Plaza, E. (1994). Case-based reasoning: Foundational issues, methodological variations, and system approaches. *AI communications*, 7(1), 39–59.
- Agovino, M., Casaccia, M., Ciommi, M., Ferrara, M., & Marchesano, K. (2019). Agriculture, climate change and sustainability: The case of eu-28. *Ecological Indicators*, 105, 525–543.
- Ahmed, N. K., Atiya, A. F., Gayar, N. E., & El-Shishiny, H. (2010). An empirical comparison of machine learning models for time series forecasting. *Econometric Reviews*, 29(5-6), 594–621.
- Ai, Y., Peng, M., & Zhang, K. (2018, apr). Edge computing technologies for Internet of Things: a primer. *Digital Communications and Networks*, 4(2), 77–86. Descargado de <https://www.sciencedirect.com/science/article/pii/S2352864817301335> doi: 10.1016/J.DCAN.2017.07.001
- Alaka, H. A., Oyedele, L. O., Owolabi, H. A., Kumar, V., Ajayi, S. O., Akinade, O. O., & Bilal, M. (2018). Systematic review of bankruptcy prediction models: Towards a framework for tool selection. *Expert Systems with Applications*, 94, 164–184.
- Alonso, R. S., Sittón-Candanedo, I., Casado-Vara, R., Prieto, J., & Corchado, J. M. (2020). Deep reinforcement learning for the management of software-defined networks and network function virtualization in an edge-iot architecture. *Sustainability*, 12(14), 5706.
- Alonso, R. S., Sittón-Candanedo, I., García, Ó., Prieto, J., & Rodríguez-González, S. (2020). An intelligent edge-iot platform for monitoring livestock and crops in a dairy farming scenario. *Ad Hoc Networks*, 98, 102047.
- Alonso, R. S., Sittón-Candanedo, I., Rodríguez-González, S., García, Ó., & Prieto, J. (2019). A survey on software-defined networks and edge computing over

- IoT. En F. de la Prieta et al. (Eds.), *Highlights of Practical Applications of Survivable Agents and Multi-Agent Systems. The PAAMS Collection* (pp. 289–301). Cham: Springer International Publishing.
- Arulkumaran, K., Deisenroth, M. P., Brundage, M., & Bharath, A. A. (2017). Deep reinforcement learning: A brief survey. *IEEE Signal Processing Magazine*, 34(6), 26–38.
- Athey, S. (2018). The impact of machine learning on economics. En *The economics of artificial intelligence: An agenda* (pp. 507–547). University of Chicago Press.
- Athey, S., & Imbens, G. W. (2019). Machine learning methods that economists should know about. *Annual Review of Economics*, 11, 685–725.
- Bajari, P., Nekipelov, D., Ryan, S. P., & Yang, M. (2015). Machine learning methods for demand estimation. *American Economic Review*, 105(5), 481–85.
- Balafoutis, A. T., Beck, B., Fountas, S., Tsiropoulos, Z., Vangeyte, J., van der Wal, T., . . . Pedersen, S. M. (2017). Smart farming technologies—description, taxonomy and economic impact. En *Precision agriculture: Technology and economic perspectives* (pp. 21–77). Springer.
- Barboza, F., Kimura, H., & Altman, E. (2017). Machine learning models and bankruptcy prediction. *Expert Systems with Applications*, 83, 405–417.
- Baskerville, R. L. (1999). Investigating information systems with action research. *Communications of the Association for Information Systems*, 2(1), 19.
- Beddington, J. R., Asaduzzaman, M., Bremauntz, F. A., Clark, M. E., Guillou, M., Jahn, M. M., . . . others (2012). Achieving food security in the face of climate change: Final report from the commission on sustainable agriculture and climate change.
- Bellmann, C., & Hepburn, J. (2017). The decline of commodity prices and global agricultural trade negotiations: A game changer? *International Development Policy—Revue internationale de politique de développement*(8.1).
- Belloni, A., Chen, D., Chernozhukov, V., & Hansen, C. (2012). Sparse models and methods for optimal instruments with an application to eminent domain. *Econometrica*, 80(6), 2369–2429.
- Bellovary, J. L., Giacomino, D. E., & Akers, M. D. (2007). A review of bankruptcy prediction studies: 1930 to present. *Journal of Financial education*, 1–42.

- Block, J., Fisch, C., Vismara, S., & Andres, R. (2019). Private equity investment criteria: An experimental conjoint analysis of venture capital, business angels, and family offices. *Journal of Corporate Finance*.
- Bontempi, M. E. (2016). Investment–uncertainty relationship: differences between intangible and physical capital. *Economics of Innovation and New Technology*, 25(3), 240–268.
- Bradtke, S. J., & Duff, M. O. (1995). Reinforcement learning methods for continuous-time markov decision problems. *Advances in neural information processing systems*, 393–400.
- Burliai, A., Nesterchuk, Y., Nepochatenko, O., & Naherniuk, D. (2020). Ecological consequences of the digitization of agriculture.
- Caruana, R., & Niculescu-Mizil, A. (2006). An empirical comparison of supervised learning algorithms. En *Proceedings of the 23rd international conference on machine learning* (pp. 161–168).
- Cathagne, A., Guyomard, H., & Levert, F. (2006). Milk quotas in the european union: Distribution of marginal costs and quota rents. *European Dairy Industry Model Working Paper*, 1(2006), 23.
- Change, I. C. (2014). Synthesis report. contribution of working groups i. *II and III to the fifth assessment report of the intergovernmental panel on climate change*, 151(10.1017).
- Charnes, A., Cooper, W., Lewin, A. Y., & Seiford, L. M. (1997). Data envelopment analysis theory, methodology and applications. *Journal of the Operational Research society*, 48(3), 332–333.
- Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European journal of operational research*, 2(6), 429–444.
- Chen, X., Shi, Q., Yang, L., & Xu, J. (2018). Thriftyedge: Resource-efficient edge computing for intelligent iot applications. *IEEE network*, 32(1), 61–65.
- Chicco, D., & Jurman, G. (2020). The advantages of the matthews correlation coefficient (mcc) over f1 score and accuracy in binary classification evaluation. *BMC genomics*, 21(1), 1–13.
- Chowdhury, G. G. (2003). Natural language processing. *Annual review of information science and technology*, 37(1), 51–89.
- Clark, M., & Tilman, D. (2017). Comparative analysis of environmental impacts

- of agricultural production systems, agricultural input efficiency, and food choice. *Environmental Research Letters*, 12(6), 064016.
- Conway, G. R., & Barbier, E. B. (2013). *After the green revolution: sustainable agriculture for development*. Routledge.
- Corchado, J., & Lees, B. (1996). Case-base reasoning recommendation system. En *Ieee colloquium on knowledge discovery. london, uk*.
- Cronqvist, H., Siegel, S., & Yu, F. (2015). Value versus growth investing: Why do different investors have different styles? *Journal of Financial Economics*, 117(2), 333–349.
- Davis, J., & Goadrich, M. (2006). The relationship between precision-recall and roc curves. En *Proceedings of the 23rd international conference on machine learning* (pp. 233–240).
- De Gorter, H., Drabik, D., & Just, D. R. (2015). *The economics of biofuel policies: impacts on price volatility in grain and oilseed markets*. Springer.
- Devi, S. S., & Radhika, Y. (2018). A survey on machine learning and statistical techniques in bankruptcy prediction. *International Journal of Machine Learning and Computing*, 8(2), 133–139.
- Dorri, A., Kanhere, S. S., & Jurdak, R. (2018). Multi-agent systems: A survey. *Ieee Access*, 6, 28573–28593.
- Doyle, J., & Green, R. (1994). Efficiency and cross-efficiency in dea: Derivations, meanings and uses. *Journal of the operational research society*, 45(5), 567–578.
- Du, X., Cindy, L. Y., & Hayes, D. J. (2011). Speculation and volatility spillover in the crude oil and agricultural commodity markets: A bayesian analysis. *Energy Economics*, 33(3), 497–503.
- Dyckhoff, H., & Allen, K. (2001). Measuring ecological efficiency with data envelopment analysis (dea). *European Journal of Operational Research*, 132(2), 312–325.
- Eden, C., & Ackermann, F. (2018). Theory into practice, practice to theory: Action research in method development. *European Journal of Operational Research*, 271(3), 1145–1155.
- Eirola, E., Gritsenko, A., Akusok, A., Björk, K.-M., Miche, Y., Sovilj, D., ... Lendasse, A. (2015). Extreme learning machines for multiclass classification: refining predictions with gaussian mixture models. En *International*

work-conference on artificial neural networks (pp. 153–164).

- El Bilali, H., & Allahyari, M. S. (2018). Transition towards sustainability in agriculture and food systems: Role of information and communication technologies. *Information Processing in Agriculture*, 5(4), 456–464.
- Eshel, G., Shepon, A., Makov, T., & Milo, R. (2014). Land, irrigation water, greenhouse gas, and reactive nitrogen burdens of meat, eggs, and dairy production in the united states. *Proceedings of the National Academy of Sciences*, 111(33), 11996–12001.
- European commission horizon 2050. (2019). Descargado de https://ec.europa.eu/commission/presscorner/detail/en/IP_19_6691
- European commission, the end of milk quotas. (2015). Descargado de http://ec.europa.eu/agriculture/milk-quota-end/index_en.htm
- Foroghi, D., Monadjemi, A., y cols. (2011). Applying decision tree to predict bankruptcy. En *2011 ieee international conference on computer science and automation engineering* (Vol. 4, pp. 165–169).
- Fried, V. H., & Hisrich, R. D. (1994). Toward a model of venture capital investment decision making. *Financial management*, 28–37.
- García, Ó., Prieto, J., Alonso, R., & Corchado, J. M. (2017). A framework to improve energy efficient behaviour at home through activity and context monitoring. *Sensors*, 17(8), 1749.
- Gardner, B. (1996). *European agriculture: policies, production, and trade*. Psychology Press.
- Granz, C., Henn, M., & Lutz, E. (2020). Research on venture capitalists' and business angels' investment criteria: A systematic literature review. En *Contemporary developments in entrepreneurial finance* (pp. 105–136). Springer.
- Gregori, M. E., Cámara, J. P., & Bada, G. A. (2006). A jabber-based multi-agent system platform. En *Proceedings of the fifth international joint conference on autonomous agents and multiagent systems* (pp. 1282–1284).
- Guillén, M. A., Llanes, A., Imbernón, B., Martínez-España, R., Bueno-Crespo, A., Cano, J.-C., & Cecilia, J. M. (2020). Performance evaluation of edge-computing platforms for the prediction of low temperatures in agriculture using deep learning. *JOURNAL OF SUPERCOMPUTING*.

- Gurel, O., Acar, A. Z., Onden, I., & Gumus, I. (2015). Determinants of the green supplier selection. *Procedia-Social and Behavioral Sciences*, 181, 131–139.
- Hallam, D. (2011). International investment in developing country agriculture—issues and challenges. *Food Security*, 3(1), 91–98.
- Haro-de Rosario, A., del Carmen Caba-Pérez, M., & Cazorla-Papis, L. (2014). Efficiency of venture capital firms: evidence from Spain. *Small Business Economics*, 43(1), 229–243.
- Hasselt, H. (2010). Double q-learning. *Advances in neural information processing systems*, 23, 2613–2621.
- Hatirli, S. A., Ozkan, B., & Fert, C. (2005). An econometric analysis of energy input–output in Turkish agriculture. *Renewable and Sustainable Energy Reviews*, 9(6), 608–623.
- He, H., & Garcia, E. A. (2009). Learning from imbalanced data. *IEEE Transactions on Knowledge and Data Engineering*, 21(9), 1263–1284.
- Hill, N. C., & Stone, B. K. (1980). Accounting betas, systematic operating risk, and financial leverage: A risk-composition approach to the determinants of systematic risk. *Journal of Financial and Quantitative Analysis*, 15(3), 595–637.
- Hjalmarsson, L., Kumbhakar, S. C., & Heshmati, A. (1996). Dea, dfa and sfa: a comparison. *Journal of Productivity Analysis*, 7(2-3), 303–327.
- Hosaka, T. (2019). Bankruptcy prediction using imaged financial ratios and convolutional neural networks. *Expert Systems with Applications*, 117, 287–299.
- Hsu, C.-W., Chang, C.-C., Lin, C.-J., y cols. (2003). *A practical guide to support vector classification*. Taipei, Taiwan.
- Hsu, M.-W., Lessmann, S., Sung, M.-C., Ma, T., & Johnson, J. E. (2016). Bridging the divide in financial market forecasting: machine learners vs. financial economists. *Expert Systems with Applications*, 61, 215–234.
- Huang, W.-B., Liu, J., Bai, H., & Zhang, P. (2020). Value assessment of companies by using an enterprise value assessment system based on their public transfer specification. *Information Processing & Management*, 57(5), 102254.
- Hunter, M. C., Smith, R. G., Schipanski, M. E., Atwood, L. W., & Mortensen, D. A. (2017). Agriculture in 2050: recalibrating targets for sustainable intensification. *Bioscience*, 67(4), 386–391.

- Ichimura, M., Nam, S., Bonjour, S., Rankine, H., Carisma, B., Qiu, Y., & Khrueachotikul, R. (2009). Eco-efficiency indicators: Measuring resource-use efficiency and the impact of economic activities on the environment. *ESCAP: Bangkok, Thailand*.
- Imandoust, S. B., & Bolandraftar, M. (2013). Application of k-nearest neighbor (knn) approach for predicting economic events: Theoretical background. *International Journal of Engineering Research and Applications*, 3(5), 605–610.
- Index, E. P. (2018). Environmental performance index. *Yale University and Columbia University: New Haven, CT, USA*.
- Jeng, L. A., & Wells, P. C. (2000). The determinants of venture capital funding: evidence across countries. *Journal of corporate Finance*, 6(3), 241–289.
- Kamble, S. S., Gunasekaran, A., & Gawankar, S. A. (2020). Achieving sustainable performance in a data-driven agriculture supply chain: A review for research and applications. *International Journal of Production Economics*, 219, 179–194.
- Kang, S. H., McIver, R., & Yoon, S.-M. (2017). Dynamic spillover effects among crude oil, precious metal, and agricultural commodity futures markets. *Energy Economics*, 62, 19–32.
- Kaplan, S. N., & Schoar, A. (2005). Private equity performance: Returns, persistence, and capital flows. *The journal of finance*, 60(4), 1791–1823.
- Kastratović, R. (2019). Impact of foreign direct investment on greenhouse gas emissions in agriculture of developing countries. *Australian Journal of Agricultural and Resource Economics*, 63(3), 620–642.
- Kim, H., Cho, H., & Ryu, D. (2020). Corporate default predictions using machine learning: Literature review. *Sustainability*, 12(16), 6325.
- King, A. (2017). Technology: The future of agriculture. *Nature*, 544(7651), S21–S23.
- Kpienbaareh, D., Kansanga, M., & Luginaah, I. (2019). Examining the potential of open source remote sensing for building effective decision support systems for precision agriculture in resource-poor settings. *GeoJournal*, 84(6), 1481–1497.
- Krawczyk, B., Minku, L. L., Gama, J., Stefanowski, J., & Woźniak, M. (2017). Ensemble learning for data stream analysis: A survey. *Information Fusion*,

- 37, 132–156.
- Kurupparachchi, D., & Premachandra, I. (2016). Information spillover dynamics of the energy futures market sector: A novel common factor approach. *Energy Economics*, 57, 277–294.
- Lansink, A. O., & Wall, A. (2014). Frontier models for evaluating environmental efficiency: an overview. *Economics and Business Letters*, 3(1), 43–50.
- Leike, J., Krueger, D., Everitt, T., Martic, M., Maini, V., & Legg, S. (2018). Scalable agent alignment via reward modeling: A research direction. *arXiv preprint arXiv:1811.07871*.
- Li, L., & Islam, S. Z. (2019). Firm and industry specific determinants of capital structure: Evidence from the Australian market. *International Review of Economics & Finance*, 59, 425–437.
- Li, Y., & Mahoney, J. T. (2011). When are venture capital projects initiated? *Journal of Business Venturing*, 26(2), 239–254.
- Liu, H., Zhang, J., Zhou, C., & Ru, Y. (2018). Optimal purchase and inventory retrieval policies for perishable seasonal agricultural products. *Omega*, 79, 133–145.
- Liu, Y., & Xie, T. (2019). Machine learning versus econometrics: prediction of box office. *Applied Economics Letters*, 26(2), 124–130.
- López de Prado, M. (2019). Beyond econometrics: A roadmap towards financial machine learning. *Available at SSRN 3365282*.
- Machek, O., & Špička, J. (2014). Productivity and profitability of the Czech agricultural sector after the economic crisis. *WSEAS Transactions on Business and Economics*, 11, 700–706.
- Melo, F. S., Meyn, S. P., & Ribeiro, M. I. (2008). An analysis of reinforcement learning with function approximation. En *Proceedings of the 25th international conference on machine learning* (pp. 664–671).
- Mnih, V., Kavukcuoglu, K., Silver, D., Graves, A., Antonoglou, I., Wierstra, D., & Riedmiller, M. (2013). Playing atari with deep reinforcement learning. *arXiv preprint arXiv:1312.5602*.
- Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A. A., Veness, J., Bellemare, M. G., ... others (2015). Human-level control through deep reinforcement learning. *nature*, 518(7540), 529–533.

- Mullainathan, S., & Spiess, J. (2017). Machine learning: an applied econometric approach. *Journal of Economic Perspectives*, 31(2), 87–106.
- Nuthall, P. L. (2004). Case studies of the interactions between farm profitability and the use of a farm computer. *Computers and Electronics in Agriculture*, 42(1), 19–30.
- Oliveira, E., Fischer, K., & Stepankova, O. (1999). Multi-agent systems: which research for which applications. *Robotics and Autonomous Systems*, 27(1-2), 91–106.
- Panesar, A. (2019). What is machine learning? En *Machine Learning and AI for Healthcare* (pp. 75–118). Springer.
- Panousi, V., & Papanikolaou, D. (2012). Investment, idiosyncratic risk, and ownership. *The Journal of Finance*, 67(3), 1113–1148.
- Partila, P., Tovarek, J., Voznak, M., Rozhon, J., Sevcik, L., & Baran, R. (2018). Multi-classifier speech emotion recognition system. En *2018 26th Telecommunications Forum (TELFOR)* (pp. 1–4).
- Patel, K. K., Patel, S. M., y cols. (2016). Internet of things-iot: definition, characteristics, architecture, enabling technologies, application & future challenges. *International journal of engineering science and computing*, 6(5).
- Pérez-Pons, M. E., Alonso, R. S., García, O., Marreiros, G., & Corchado, J. M. (2021). Deep q-learning and preference based multi-agent system for sustainable agricultural market. *Sensors*, 21(16), 5276.
- Pérez-Pons, M. E., Parra-Domínguez, J., Chamoso, P., Plaza, M., & Alonso, R. (2020). Efficiency, profitability and productivity: Technological applications in the agricultural sector. *ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal*, 9(4).
- Pérez-Pons, M. E., Plaza-Hernández, M., Alonso, R. S., Parra-Domínguez, J., & Prieto, J. (2021). Increasing profitability and monitoring environmental performance: A case study in the agri-food industry through an edge-iot platform. *Sustainability*, 13(1), 283.
- Peterson, O. K., & Arun, T. G. (2018). Income smoothing among european systemic and non-systemic banks. *The British Accounting Review*, 50(5), 539–558.

- Philippidis, G., & Waschik, R. (2019). Melitz meets milk: The impact of quota abolition on eu dairy export competitiveness. *Journal of agricultural economics*, *70*(1), 44–61.
- Piedra-Muñoz, L., Galdeano-Gómez, E., & Pérez-Mesa, J. C. (2016). Is sustainability compatible with profitability? an empirical analysis on family farming activity. *Sustainability*, *8*(9), 893.
- Qu, Y., Quan, P., Lei, M., & Shi, Y. (2019). Review of bankruptcy prediction using machine learning and deep learning techniques. *Procedia Computer Science*, *162*, 895–899.
- Russell, S. J., & Norvig, P. (2016). *Artificial intelligence: a modern approach*. Malaysia; Pearson Education Limited,.
- Saito, T., & Rehmsmeier, M. (2015). The precision-recall plot is more informative than the roc plot when evaluating binary classifiers on imbalanced datasets. *PloS one*, *10*(3).
- Salleh, A. H. M., Mohamad, M. S., Deris, S., Omatu, S., Fdez-Riverola, F., & Corchado, J. M. (2015). Gene knockout identification for metabolite production improvement using a hybrid of genetic ant colony optimization and flux balance analysis. *Biotechnology and Bioprocess Engineering*, *20*(4), 685–693.
- Schmidheiny, S., & Timberlake, L. (1992). *Changing course: A global business perspective on development and the environment* (Vol. 1). MIT press.
- Sen, P. C., Hajra, M., & Ghosh, M. (2020). Supervised classification algorithms in machine learning: A survey and review. En *Emerging Technology in Modelling and Graphics* (pp. 99–111). Springer.
- Seo, J., Han, S., Lee, S., & Kim, H. (2015). Computer vision techniques for construction safety and health monitoring. *Advanced Engineering Informatics*, *29*(2), 239–251.
- Sharma, M., & Mavani, M. (2011). Development of predictive model in education system: using naïve bayes classifier. En *Proceedings of the international conference & workshop on emerging trends in technology* (pp. 185–186).
- Shephard, R. W. (2015). *Theory of cost and production functions*. Princeton University Press.
- Shi, J., Zhao, Y., Kiwanuka, R. B. K., & Chang, J. (2019). Optimal selling policies for farmer cooperatives. *Production and Operations Management*, *28*(12),

3060–3080.

- Simar, L., & Wilson, P. W. (2007). Estimation and inference in two-stage, semi-parametric models of production processes. *Journal of econometrics*, *136*(1), 31–64.
- Sirower, M. L., & O’Byrne, S. F. (1998). The measurement of post-acquisition performance: Toward a value-based benchmarking methodology. *Journal of Applied Corporate Finance*, *11*(2), 107–121.
- Sisinni, E., Saifullah, A., Han, S., Jennehag, U., & Gidlund, M. (2018). Industrial internet of things: Challenges, opportunities, and directions. *IEEE Transactions on Industrial Informatics*, *14*(11), 4724–4734.
- Sittón-Candanedo, I. (2019). Edge computing: A review of application scenarios. En *International symposium on distributed computing and artificial intelligence* (pp. 197–200).
- Sittón-Candanedo, I., Alonso, R. S., Corchado, J. M., Rodríguez-González, S., & Casado-Vara, R. (2019). A review of edge computing reference architectures and a new global edge proposal. *Future Generation Computer Systems*, *99*, 278–294.
- Skobelev, P., Budaev, D., Gusev, N., & Voschuk, G. (2018). Designing multi-agent swarm of uav for precise agriculture. En *International conference on practical applications of agents and multi-agent systems* (pp. 47–59).
- Skvortsov, E., Skvortsova, E., Sandu, I., & Iovlev, G. (2018). Transition of agriculture to digital, intellectual and robotics technologies. *Ekonomika regiona*(3), 1014.
- Solow, R. M. (1957). Technical change and the aggregate production function. *The review of Economics and Statistics*, 312–320.
- Stigler, S. M. (1981). Gauss and the invention of least squares. *the Annals of Statistics*, 465–474.
- Sutton, R. S. (1988). Learning to predict by the methods of temporal differences. *Machine learning*, *3*(1), 9–44.
- Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction*. MIT press.
- Swan, T. W. (1956). Economic growth and capital accumulation. *Economic record*, *32*(2), 334–361.

- Takehima, H. (2017). Custom-hired tractor services and returns to scale in smallholder agriculture: a production function approach. *Agricultural Economics*, 48(3), 363–372.
- Tapia, D. I., Fraile, J. A., Rodríguez, S., Alonso, R. S., & Corchado, J. M. (2013). Integrating hardware agents into an enhanced multi-agent architecture for ambient intelligence systems. *Information Sciences*, 222, 47–65.
- Theodoridis, A. M., & Anwar, M. M. (2011). A comparison of dea and sfa methods: a case study of farm households in bangladesh. *The Journal of Developing Areas*, 95–110.
- Tinbergen, J. (1962). Shaping the world economy; suggestions for an international economic policy. *NA*.
- Tiwari, D., Loof, R., & Paudyal, G. (1999). Environmental-economic decision-making in lowland irrigated agriculture using multi-criteria analysis techniques. *Agricultural systems*, 60(2), 99–112.
- Tüzün, E., Tekinerdogan, B., Macit, Y., & İnce, K. (2019). Adopting integrated application lifecycle management within a large-scale software company: An action research approach. *Journal of Systems and Software*, 149, 63–82.
- Tyebjee, T. T., & Bruno, A. V. (1984). A model of venture capitalist investment activity. *Management science*, 30(9), 1051–1066.
- Usama, M., Qadir, J., Raza, A., Arif, H., Yau, K.-L. A., Elkhatib, Y., ... Al-Fuqaha, A. (2019). Unsupervised machine learning for networking: Techniques, applications and research challenges. *IEEE Access*, 7, 65579–65615.
- Van Beveren, I. (2012). Total factor productivity estimation: A practical review. *Journal of economic surveys*, 26(1), 98–128.
- Van Calker, K., Berentsen, P., Romero, C., Giesen, G., & Huirne, R. (2006). Development and application of a multi-attribute sustainability function for dutch dairy farming systems. *Ecological Economics*, 57(4), 640–658.
- van Engelen, J. E., & Hoos, H. H. (2019). A survey on semi-supervised learning. *Machine Learning*, 1–68.
- Veganzones, D., & Séverin, E. (2018). An investigation of bankruptcy prediction in imbalanced datasets. *Decision Support Systems*, 112, 111–124.
- Wang, N., y cols. (2017). Bankruptcy prediction using machine learning. *Journal of Mathematical Finance*, 7(04), 908.

- Watkins, C. J., & Dayan, P. (1992). Q-learning. *Machine learning*, 8(3-4), 279–292.
- White, B. A., & Dumay, J. (2017). Business angels: a research review and new agenda. *Venture Capital*, 19(3), 183–216.
- Wimmer, S., & Sauer, J. (2020). Profitability development and resource reallocation: The case of sugar beet farming in germany. *Journal of Agricultural Economics*.
- Wooldridge, J. M. (2010). *Econometric analysis of cross section and panel data*. MIT press.
- Wu, C., Toosi, A. N., Buyya, R., & Ramamohanarao, K. (2018). Hedonic pricing of cloud computing services. *IEEE Transactions on Cloud Computing*, 1-1.
- Yu, W., Liang, F., He, X., Hatcher, W. G., Lu, C., Lin, J., & Yang, X. (2017). A Survey on the Edge Computing for the Internet of Things. *IEEE Access*, 6, 6900–6919. doi: 10.1109/ACCESS.2017.2778504
- Zhai, Z., Martínez, J. F., Beltran, V., & Martínez, N. L. (2020). Decision support systems for agriculture 4.0: Survey and challenges. *Computers and Electronics in Agriculture*, 170, 105256.
- Zhang, N., Wang, M., & Wang, N. (2002). Precision agriculture—a worldwide overview. *Computers and electronics in agriculture*, 36(2-3), 113–132.
- Zhang, W., y cols. (2017). Machine learning approaches to predicting company bankruptcy. *Journal of Financial Risk Management*, 6(04), 364.
- Zhao, C., Liu, B., Piao, S., Wang, X., Lobell, D. B., Huang, Y., ... others (2017). Temperature increase reduces global yields of major crops in four independent estimates. *Proceedings of the National Academy of Sciences*, 114(35), 9326–9331.
- Zieba, M., Tomczak, S. K., & Tomczak, J. M. (2016). Ensemble boosted trees with synthetic features generation in application to bankruptcy prediction. *Expert systems with applications*, 58, 93–101.

