



# A Data Mining Approach to Detecting Bias and Favoritism in Public Procurement

Yeferson Torres-Berru<sup>1,2,\*</sup>, Vivian F. Lopez-Batista<sup>1</sup> and Lorena Conde Zhingre<sup>3</sup>

<sup>1</sup>University of Salamanca, Salamanca, 37008, Spain

<sup>2</sup>Instituto Tecnológico Superior Sudamericano, Loja, 1101608, Ecuador

<sup>3</sup>Universidad Internacional del Ecuador, Loja, 1101608, Ecuador

\*Corresponding Author: Yeferson Torres-Berru. Email: yeferson.torres11@gmail.com

Received: 18 August 2022; Accepted: 29 November 2022

**Abstract:** In a public procurement process, corruption can occur at each stage, favoring a participant with a previous agreement, which can result in over-pricing and purchases of substandard products, as well as gender discrimination. This paper's aim is to detect biased purchases using a Spanish Language corpus, analyzing text from the questions and answers registry platform by applicants in a public procurement process in Ecuador. Additionally, gender bias is detected, promoting both men and women to participate under the same conditions. In order to detect gender bias and favoritism towards certain providers by contracting entities, the study proposes a unique hybrid model that combines Artificial Intelligence algorithms and Natural Language Processing (NLP). In the experimental work, 303,076 public procurement processes have been analyzed over 10 years (since 2010) with 1,009,739 questions and answers to suppliers and public institutions in each process. Gender bias and favoritism were analyzed using a Word2vec model with word embedding, as well as sentiment analysis of the questions and answers using the VADER algorithm. In 32% of cases (96,984 answers), there was favoritism or gender bias as evidenced by responses from contracting entities. The proposed model provides accuracy rates of 88% for detecting favoritism, and 90% for detecting gender bias. Consequently one-third of the procurement processes carried out by the state have indications of corruption and bias. In Latin America, government corruption is one of the most significant challenges, making the resulting classifier useful for detecting bias and favoritism in public procurement processes.

**Keywords:** Favoritism; bias; natural language processing; Word2vec; sentiment analysis; word embeddings

## 1 Introduction

The Ecuadorian government has been working with new technologies to integrate its processes, one of which is public procurement. According to Thi Nguyen et al. [1] several reports propose that the modernization of the state be undertaken as indicated by organizations such as the United Nations and the



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Organization for Economic Cooperation and Development (OECD). In order to simplify processes, improve productivity, gain transparency, reinforce good governance, and strengthen coordination and communication in institutions, the development and implementation of electronic initiatives were promoted. For community participation in government to increase, public services should be improved, policies should be effective, and citizen confidence must be increased [2]. A public procurement system is used by institutions to procure goods, services, and works for the community; therefore, it should meet three essential objectives: human development, efficiency, and transparency.

The 2008 constitution of Ecuador and the National Public Procurement System organic law created a set of operational and regulatory conditions for government purchases, modifying the scenarios for government purchases. Monitoring and controlling processes are key to achieving this productivity improvement. The National Public Procurement Service [3] promotes transparency and publicity of contracting acts in 2018. For example, it is prohibited to benefit from purchase orders generated through the “inclusive dynamic catalogue”. Public authorities decide to purchase goods, works, and services, but it is possible at this point that the decision does not follow a policy rationale or an existing need but rather the desire to channel benefits to an individual or/and organization.

It has been shown that the promotion of transparency and publicity of acts in contracting, should focus on an analysis to identify risks of corruption in public purchases, which requires advanced management of substantial amounts of information with typical techniques of big data [4], sentiment analysis [5], advanced analytics and Artificial Intelligence (AI) [6]. In previous research, we have highlighted the need to use different machine learning techniques to assess favoritism and therefore corruption in public procurement. The literature on this topic has been reviewed, and we have proposed a model which combining supervised and unsupervised learning, that allowed the detection of corruption in the allocation of public procurement contracts [7–10]. By analyzing text generated by each public process, we want to extend their study to finding other defined terms, such as gender bias and favoritism, in questions and answers.

As a result of corruption in public procurement, Transparency International estimates costs ranging from 20%–25%, and sometimes up to 40%–50% [11]. Some issues encountered include various kinds of corruption like bribery, collusion, embezzlement, misappropriation, fraud, abuse of discretion, and favoritism. Based on our previous theoretical discussions, we chose to focus on favoritism in this paper because only eight of the 147 reviewed articles address this type of corruption [10].

Favoritism is defined as the natural human propensity to favor friends, relatives and any close or reliable person involved in a public process [12] once a tender process is started the tender provider can still dissuade competitive bidders by keeping the contracting process non-transparent and by circulating private information to favor a particular supplier. For this reason, the aim of machine learning systems is to analyze the relationship between the behavior of providers and contracting entities. The unilateral modification consists of the authority to modify administrative contracts for reasons of public interest [13]. Conflict of interest is a clear red flag for corruption, this can be due to family, business or political ties. On multiple occasions, the legal instrument of the modification is used to articulate new contracts with fraud at the beginning of the bidding process [14]. On other occasions, the modifications double or even triple the initial price of the contract. Also, during the “*questions, answers, and clarifications*” stage in a public process, the entities knowing the deadline for making clarifications do them at the last minute; leaving all the participants without the necessary time to generate an offer with the requested changes automatically benefiting the bidder who was given the information beforehand, to direct the qualification towards conditions that only one proponent will present. The entity after holding conversations with the bidder which it wants to be the winner, modifies certain parameters in the “*questions, answers and clarifications*” stage, so that the offer of this supplier obtains maximum scores and this way to ensure that he is the person who wins the procedure. There is a risk that evaluation criteria aren’t clearly stated in

tender documents, leaving no grounds to justify awarding the tender to a corrupt supplier. Therefore, red flags are accumulations of clues that suggest corruption.

Section 2 shows the related works to highlight the importance of the current study, Section 3 details the techniques and methods used also describes the dataset and the technological tools such as programming languages. Section 4 shows the results obtained from the experimentation, the following section discusses the results and conclusions and finally presents future work of the paper.

## 2 Related Works

Natural language processing (NLP) is a branch of Artificial Intelligence (AI) that enables machines to understand the human language and is used in many fields. Pant et al. [15] in their research address NLP to detect subjective biases through models focused on a classifier based on Bidirectional Encoder for Transformers (BERT) which uses a bidirectional analysis model. This means that words that are both to the right and the left of a keyword are being analyzed. The method takes advantage of automatic denoising auto-encoders and a token-weight loss function considering bias when associations between gender and certain concepts are captured in processes, with the use of word embedding or in the parameters of the model. Bias is introduced into natural language through specific words and phrases. It can also be defined as the inclination or prejudice of a decision made by an AI system, unfairly for or against a person or group. For example, gender bias is the preference of one gender over another [16], which manifests itself in particular contexts or between social roles mainly, having prejudices against the less dominant gender. This occurs in [17] when there is a correlation between the representation of a neutral word concerning a word with gender (male or female) through a calculation called indirect inequality and the degree of bias with any gender [18].

Modrusan et al. [6] utilize AI to develop a model capable of identifying and extracting terms that determine the prediction of suspicious offers based on the quality and volume of the data under analysis, with a classification accuracy of 0,76%. Only 1,500 documents provided by the competent authority to the authors were reviewed. Hamishu et al. [19] use a model based on Bag-of-Words (BoW) designed to detect and classify fraud in tax advances, through the identification of frequently used words used by scammers, when processing substantial amounts of data. Also, it is important to note that these two parsers are based on English grammar and syntax. The NLP allows the implementation of systems for the detection of fraud taking text classification, information retrieval and information extraction as a reference.

The public procurement marketplace is an important instrument for achieving the inclusiveness of women and improving their purchasing power because it has the potential to promote their empowerment as well as its inclusion in non-traditional sectors such as science and technology by providing funds for women's companies to maintain and diversify employment [20]. For this reason, it is extremely necessary to detect if there are gender biases within this sector, as the complexity of natural language constructs makes this task even more challenging [21]. The work of Akhter et al. [22] indicate that more research is required on the evaluation of abstraction techniques for text summaries in Spanish especially novel techniques based on transformers. The results obtained by the authors suggest that Word2vec word embeddings achieved the best results based on the ROUGE-1 [23] and BLEU [24] metrics. As Tunyan et al. [25] describe, NLP can be used to detect subjective biases at the level of sentences, sentiments, opinions, as well as bias mitigation, but the vast majority of recent work on bias is focused on identifying hate speech on social media. In effect, as a result of the selection bias, these studies fail to differentiate those indicators that can effectively distinguish corrupt public procurements from the rest. However, reducing bias is a new challenge for the NLP and AI research community.

Traditionally, the information generated by suppliers and public entities oriented to the identification of suspicious activities of corruption in a public procurement process is not used, given that the complaints

generated by suppliers or suspicious responses made by financial institutions are not reviewed or taken into account by any controlling entity. This causes uneasiness on the part of suppliers who feel disadvantaged by the state, believing that there is a predilection (favoritism) towards a certain supplier or group of suppliers. Consequently, this research focuses on developing AI based model, which uses NLP techniques for detecting suspicious processes in the Spanish Language. The research objective is to develop a machine-learning model that recognizes gender biases and favoritism in the public procurement process for Spanish speakers.

The main novelty of the work is the evaluation of the validity of the evidence extracted from the questions and answers made in a public purchase process to detect biased (fraudulent) purchases in the Spanish Language, based on an evaluation of equal conditions, in terms of gender and chances of winning the process.

### 3 Materials and Methods

In this section, the detection of favoritism and gender bias in public procurement for Spanish speakers is described. This model complements the methodology for detecting corruption in public purchases proposed by the authors in previous work [10]. For the evaluation, the following indicators are considered: the publication date of the call for vendors' proposals, the selection and contract decisions, the prices, the number of bids submitted, the kind of institutions' responsibilities, and the origin of the bidders.

In addition, red flags related to the bidding process, such as a really short time between the call for tenders and the submission of the bid, or requirements that are excessively difficult to achieve by the participant in the selection processes. Moreover, suspicious indicators related to the results of the selection processes, such as the high participation of a supplier or contractor in the public contracts of a given institution. On the other hand, the model runs an analysis of the opinions of the questions and answers made by the contracting entities to complement the evaluation of favoritism.

#### 3.1 Dataset Description

The experimental work 303,076 of public procurement processes executed from 2010 to 2020 were evaluated. Data collection is explained in our previous works [7,8]. The dataset in each process includes questions or clarifications made and answered by the contracting entity and vendors. In total 1,009,739 questions and answers are evaluated with the following characteristics:

- Questions issued by the contracting party where necessary aspects for the bidding are consulted or it is intended to report any novelty in the process.
- Responses issued by the entity that conducts the contracting, which serve to clarify the doubts of the contracting parties. Furthermore, it can be used to issue clarifications or modifications to the contract. It is understood that the answers and clarifications must be partial and include the references. Therefore, there should not be biases or sentiments for or against the provider.

In Fig. 1 we show an example of the questions and answers made in each process. The texts are written in Spanish Language and within each purchase process additional information is added, such as: code, town, amount, date, time, web address of the process, the status of the process, payment method, type of purchase and the type of hiring. The questions, answers, dates, and attached files can be seen in the questions section.

#### 3.2 Proposed Methodology

Fig. 2 shows a proposed model for detecting favoritism and gender bias in public procurement processes. The model is trained based on the questions and answers made by the suppliers and contracting entities. As an initial step, the pre-processing of the data is conducted as explained in Section 3.3, for its treatment using Word2vec [26] based on word embeddings [27] (Section 3.4) What is

more, sentiment analysis is performed to detect favoritism, with a pre-trained classifier Balance Aware Dictionary for Sentiment Reasoning (VADER) at the same time. It is evaluated if there is any bias in the answers given by the institutions, to benefit a specific type of contracting party.

```

"CODIGO": "SIE-002-DPLD-2020",
"CANTON": "QUITO",
"MONTO": -1,
"FECHA_HORA": "5/8/20 13:00",
"LINK": "https://www.compraspublicas.gob.ec/ProcesoContratacion/compras/PC/informacionProce
"Estado del Proceso:": "Desierta\nRaz\u00f3n: Art. 1.- DECLARAR DESIERTO El Proceso de CONTRATA
"Forma de Pago:": "Anticipo: 0% Saldo: Pago contra entrega de bienes obras o servicio 100.0
"Tipo Compra:": "Servicio",
"Tipo de Adjudicaci\u00f3n:": "Total",
"Tipo de Contrataci\u00f3n:": "Subasta Inversa Electr\u00f3nica",
"preguntas": [
  {
    "pregunta": " Se\u00f1ores considerar plazo de entrega 60 d\u00edas",
    "respuesta": "Estimado oferente NO es posible porque el tiempo estimado es de 30 d\u00edas
    "fechaPre": " 2020-05-11 00:09:09",
    "archivos": "archivos"
  },
  {
    "pregunta": " Se\u00f1ores por favor indicar si se puede realizar una visita t\u00e9cnica",
    "respuesta": "Estimado oferente si es posible para que realice una inspecci\u00f3n t\u00e9cnica y
    "fechaPre": " 2020-05-11 00:10:17",
    "archivos": "archivos"
  }
]

```

Figure 1: An example of the questions and answers that are taken as input data

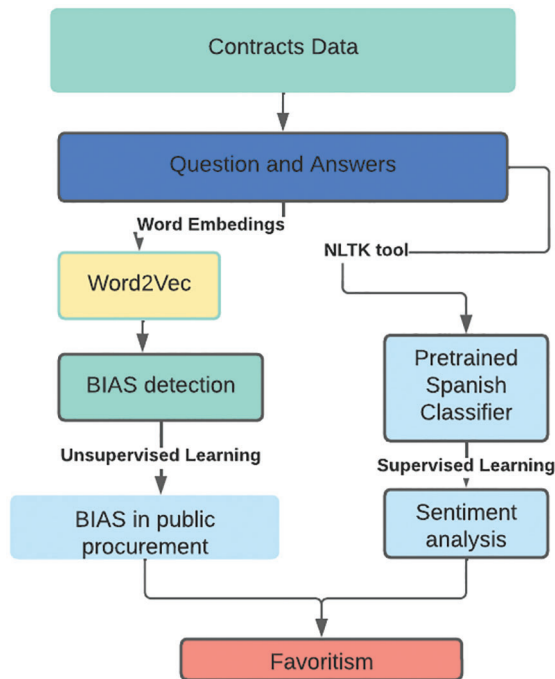


Figure 2: Proposed methodology

### 3.3 Data Pre-Processing

To precisely label each of the terms and normalize the text, stop words are eliminated, the text is converted to lowercase, punctuation marks, question marks, extra spaces, and tabulations are eliminated. Despite the fact that term root extraction (stemming) is common in NLP, some results were observed that showed how it affected sensitivity and precision, so we decided not to apply it. For this phase, the NLTK [28] and PANDAS [29] Python libraries are used.

### 3.4 Data Processing

Word2vec [26] word embeddings were used to analyze gender bias and favoritism. With Word2vec, a large amount of text corpus is taken as input and used to generate word embeddings. The word embeddings are represented in a high-dimensional vector space and the semantic information of the words is represented. As a result, similar words sharing the same context are grouped together in this vector space. For word embedding generation, the FastText algorithm [27] was also evaluated. In both algorithms, the questions and answers of each public procurement process can be projected from the original space to the new multidimensional space. In this way, word embeddings learn relationships derived from concurrency statistics. We also test Dipper throated optimization (DTO) algorithm's with its suitability for solving complex real-world issue [30].

The evaluation of connected words is obtained through the conditional probability that a word, see Eq. (1), (represented by  $w$ ) is part of the context ( $w_i$ ) and also  $w_j$  (next word) is given by (1) according to the matrix  $W$  and  $W'$  (the inverse) and its matrix of transformation ( $T$ ) so that the equation allows to obtain the similarity between the words denominated as  $w_i, w_j, w_k$  thus substantively.

$$p(w_j|w_i) = \frac{\exp(v_{w_i} T w_j)}{\exp(v'_k T v_{w_i})} \quad (1)$$

#### 3.4.1 Sentiment Analysis

The questions and answers are analyzed as individual sentences to extract the opinion contained in each of them (positive, negative, or neutral). A value is assigned to each opinion, whose sum between the three sentiments must be equal to 1, for example (neutral: 0.1, positive: 0.0, and negative: 0.9). The algorithm VADER combines lexical features with the consideration of five general rules, which incorporate grammatical and syntactic conventions to express and emphasize sentiment and tension of the same. This algorithm is available through the NLTK *Sentiment Intensity Analyzer function* and is implemented in Google Colab.

VADER is used because it allows for obtaining a better accuracy than other techniques analyzed [31]. To evaluate the results, standard machine learning metrics [32] are used, such as precision, accuracy, F1 score, area under the receiver operating characteristics curve (AUC), and the loss function (loss) which establishes the penalty for not achieving the expected result. If the deviation from the expected value by our model is large, then the loss function returns the highest number and if the deviation is much smaller, it is moved closer to the expected value.

#### 3.4.2 Data Bias Detection

In order to detect gender bias and favoritism in the datasets, the generated word embedding models were used for the questions and answers respectively. As we mentioned these models are based on Word2vec because this technique can be used to treat a large amount of data in the corpus. It is based on the application of Boolean type logical operators using a smooth transformation matrix in a linear subspace, in which word embedding obtains the highest variance [33,34]. The bias detail assessed is summarized in Table 1.

**Table 1:** Description of bias type to be evaluated

Bias type	Verification reason	Where to evaluate
Gender	Male terms are used to address suppliers and to request staff for each contract.	<ul style="list-style-type: none"> <li>• Questions</li> <li>• Answers</li> </ul>
Favoritism	The responses of the contracting entities must be the same for all providers, without directing or favoritism.	<ul style="list-style-type: none"> <li>• Answers</li> </ul>

For implementation, the Keras and TensorFlow tools were used provided by Google Colab. Due to the amount of data used, a convolutional neural network is trained in Tensorflow with the representation model “Continuous Skip-gram” [35], which predicts the words that are within a certain range before and after the current parsed word. In this phase, the following parameters are established for each sequence: *vocab\_size* = 4096 (range between words) and *sequence\_length* = 10 (sequence length). The source code and implementation are available in Google Colab. You can also download the *vector.tsv* (vectorized words) and *metadata.tsv* files (approximation of each word) to allow the visualization in the *TensorBoard Embedding Projector*.

#### 4 Results

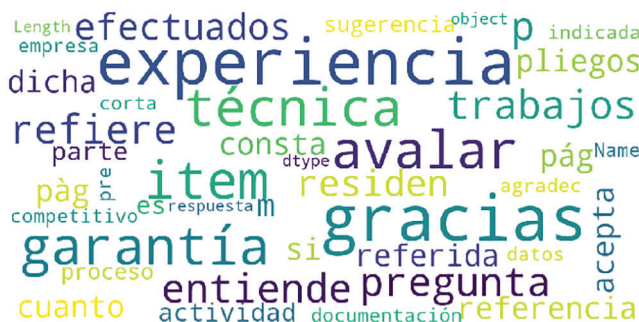
Fig. 3 shows the word cloud corresponding to the most used terms for the elaboration of questions by providers. Some of them in the image are in Spanish Language. Words such as: late, yes, confirm, validate, money (*plata* in colloquial Spanish), etc. In Fig. 3 is intended to get an idea of the terminology used by vendors.

**Figure 3:** Word cloud of vendor questions

Fig. 4 gives the word cloud corresponding to the terminology frequently used for preparing responses by the contracting entities. It should be noted that the answer provided by the state entities is expected will not be biased. However, frequently appearing biased words like experience (responses oriented to a provider with the institution previously worked), reference (response-oriented to the product must meet certain references to a particular model or brand), and gentlemen (answers that are aimed at a male audience).

##### 4.1 Bias Analysis

Algorithms such as Word2Vec and FastText were evaluated to analyze gender bias (in questions and answers) towards other providers.



**Figure 4:** Word cloud of public entity answers

#### 4.1.1 Bias in Questions

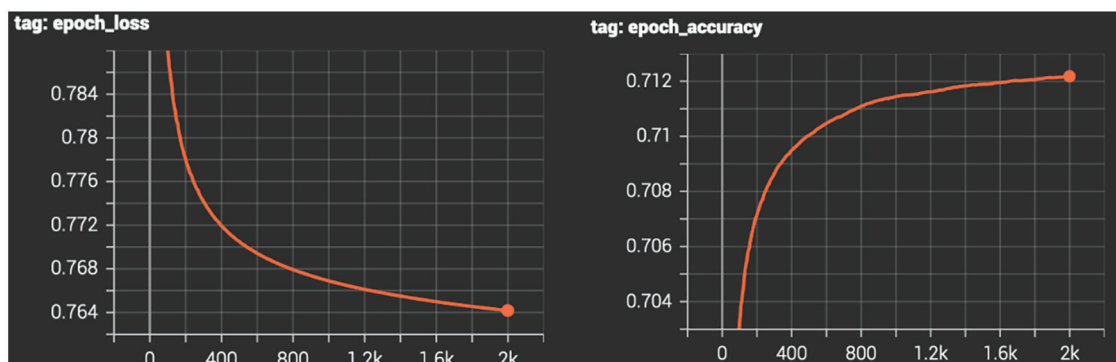
The process begins with the creation of the vocabulary of the obtained *tokens*, allowing the inverse vocabulary to be created to finally vectorize and obtain the transformation matrix of the words so that the equivalent vector is obtained in each word.

Table 2 analyzes the results obtained for the gender bias analysis through the comparison between Word2vec and FastText algorithms. Accuracy was used as a metric, which measures the percentage of cases in which the model was correct and also the function of loss (loss) which evaluates the discrepancy between the prediction and the expected result, showing Word2vec provides better results for the model.

**Table 2:** Gender bias results in questions

Algorithm	Metric	Bias in gender
FastText	Accuracy	0.71
Word2vec	Loss	0.78
	Accuracy	0.89
	Loss	0.59

Fig. 5 elaborates the results of evaluation SCORE with the ROC curve with 2000 iterations of one of the trained models (FastText) on the left side where the Specificity (0.71) is shown, and on the right, the loss function (0.78). It can be seen that the evaluated model cannot significantly improve its results, thus justifying the choice of Word2vec instead of FastText. The ROC curve and the loss function were used to measure the accuracy percentages of all the models. It is worth noting that these graphs were generated for all the biases in both questions and answers and the results are shown in Tables 2–4.



**Figure 5:** ROC curve with 2000 interactions, accuracy, and loss of the FastText algorithm



**Table 3:** Results of the response bias assessment

Technique	Metric	Gender bias	Favoritism
FastText	Accuracy	0.80	0.77
Word2vec	Loss	0.65	0.65
	Accuracy	0.90	0.88
	Loss	0.37	0.42

**Table 4:** Example of analysis elaborated by the model

Answer	Prediction	Probability
Dear Sir or Mrs., If the case of an interview or news is presented in an international media, the awarded supplier will be notified at least 24 h in advance so that the corresponding recording can be made, and the respective information can be sent to us immediately.	Without bias	0.9
Dear mrs bidder, it is not possible to present a woman professional with electronic engineering skills to replace the telecommunications professional, they must have a telecommunications degree, and the bidder must comply with what is requested in the specifications.	With bias (gender)	0,8
Dear supplier, form No. 7 will only be requested from the awarded supplier as a document prior to the execution of the contract. Form 7 is not part of the technical offer. At the time of including, it in it, it would be disqualified.	With bias (favoritism)	0,88

In the questions from suppliers, it was frequently mentioned that there are processes that are directed or limited to certain contractors or professionals which are previously chosen by the public institution. So, for example the word “*dirigido (biased)*” is related to the words: “*exist, defects, contractor, agree, specifications*”. It should be noted that the terms of conditions made by the contracting entity are called “specifications”, therefore it is understood that the providers in their questions directly indicate when there is favoritism on the part of the public institution, but these “complaints in many cases are not taken care of by the public institution”. An example of this is the following question asked by a supplier: “*In a biochemical place, can we deliver professional mechanized dilutors for a correct dilution of chemicals? since a biochemist, chemical engineer, and pharmaceutical chemist prepare or manufacture and indicate their technical datasheets. This is very biased, and even more so with the 30% advance payment, please check the specifications*”.

Fig. 6 analyzes the closeness of words, according to their similarity for bias detection using Word2vec, it is observed that in the responses made by the contracting entities, there are words such as: “*unfortunately*” which is directly related to the words “*vendors*” and “*number*” referring to the fact that the contracting entities use phrases such as “*unfortunately the process has such a number of invited vendors*” which suggests favoritism towards directly invited suppliers by the contracting entity.

In addition, words such as “*validate*” and “*require*” are common words with regard to the answers issued by the contracting party, referring to the number of suppliers invited to the public procurement processes, since the entity always asks if the supplier complies with the specific and on many occasions complex requirements.

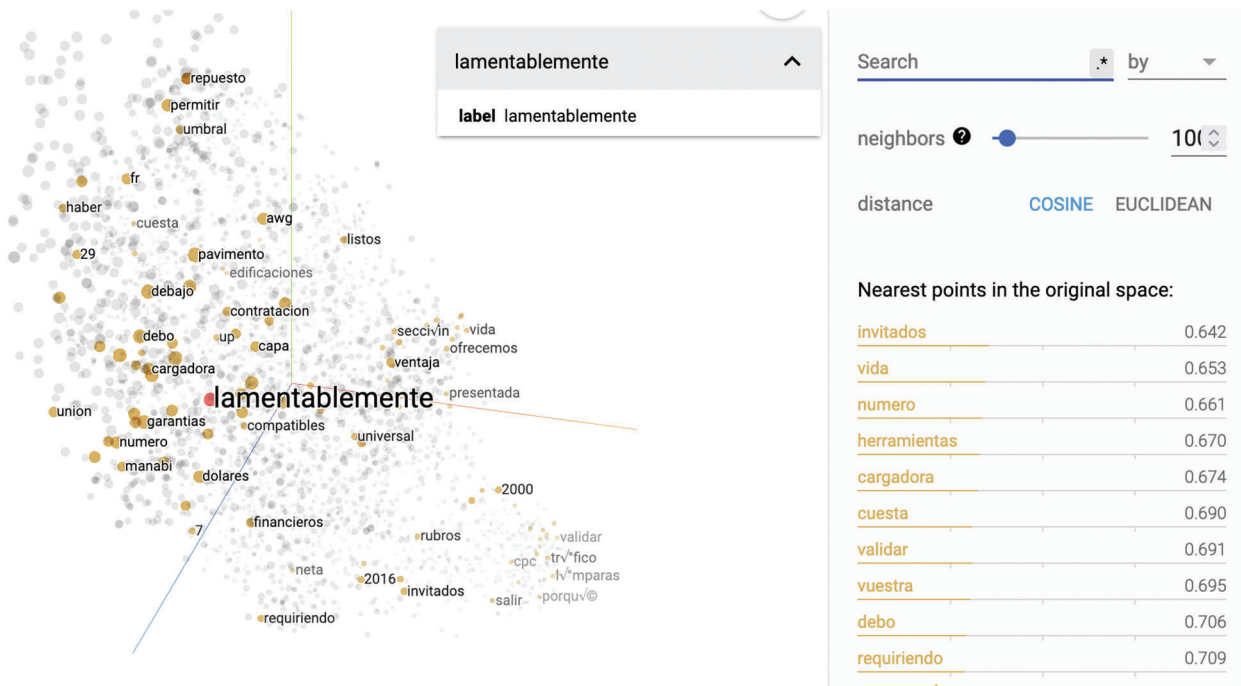


Figure 6: Word visualization questions

4.1.2 Bias in Answers

Table 3 shows the results of the evaluation of the metrics accuracy and loss function corresponding to gender bias and favoritism in the answers issued by the contracting entities in a public procurement process. It might be seen that the Word2vec technique presents better results, reaching 90% in the detection of gender bias and 88% in favoritism.

Fig. 7 shows the ROC curve, where the classificatory potentiality of the proposed model is indicated, showing the accumulation on the X-axis and the number of iterations for model training on the Y-axis.

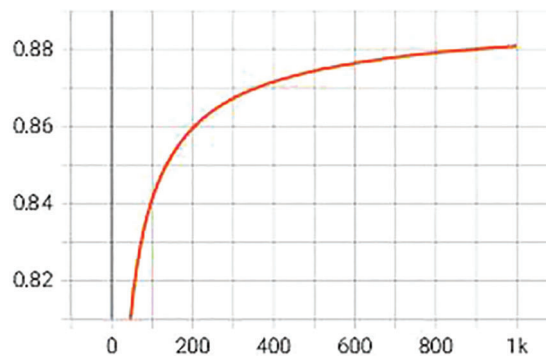
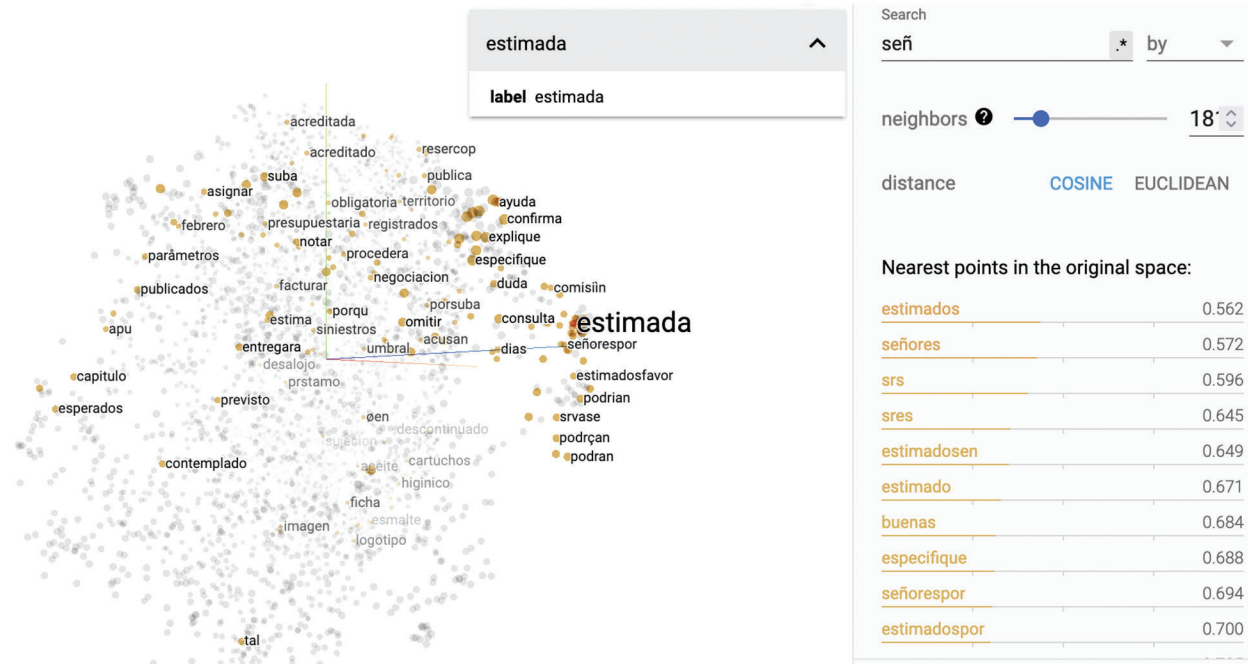


Figure 7: ROC accuracy answers

It can be observed in Fig. 8 that the visualization of words associated with gender bias, is predominantly oriented towards the masculine gender such as: gentlemen, Dear Sir/Mr., etc., with respect to equivalent words in the feminine gender or in neutral language. Both contractors and public institutions assume that the person with whom they are interacting is a male. The table further shows the similarity score of the

word “estimado (Dear Sir/Mister)” with other masculine gender-based expressions. It is also evident that there is no direct percentage score with the highlighted word “estimada (Dear Miss/Mrs/Ma’am)” which would be the feminine counterpart of “estimado”.



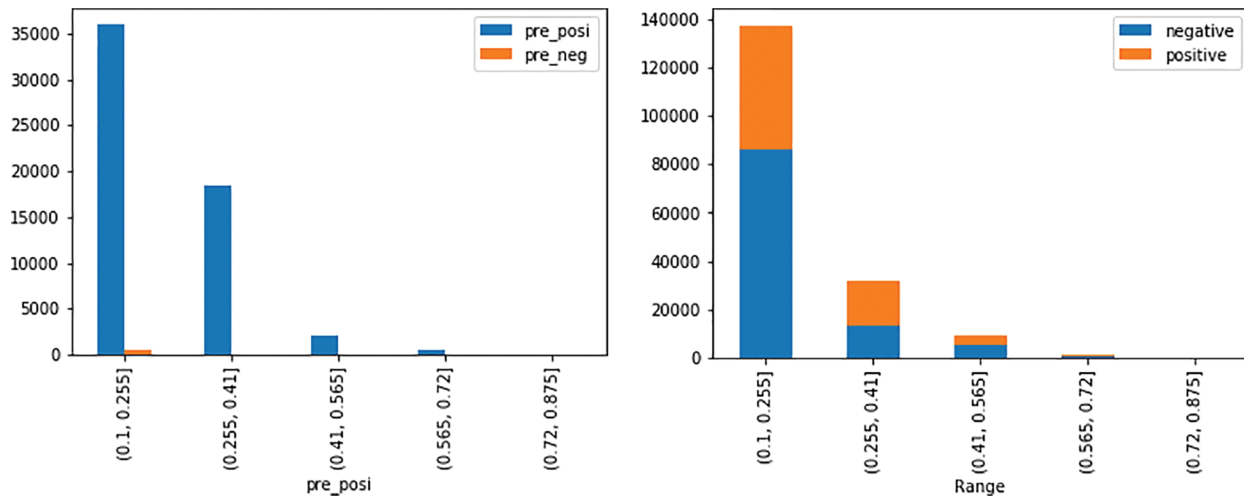
**Figure 8:** Word visualization in answers

Table 4 shows an example of the phrases used to detect bias in the responses made by the contracting institutions, including their evaluation, where 3 cases are observed: gender bias, favoritism (corruption), and no bias.

#### 4.2 Sentiment Analysis

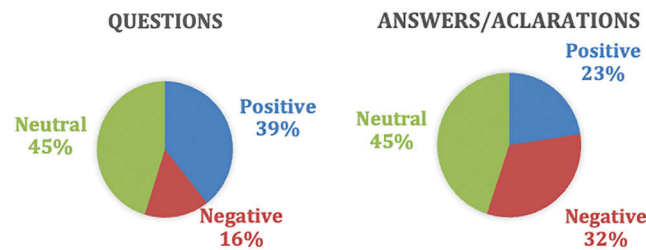
This section presents the sentiment analysis of questions and answers made by suppliers and contracting entities.

On the left side of Fig. 9, the sentiment evaluation scales are plotted at an interval of 0.15 (X-axis), while the number of questions is plotted on a Y-axis. In the questions asked by providers to contracting entities, we found a higher percentage of positive opinions. On the right side of Fig. 9, the classification of the sentiment analysis is also shown, in the responses of public institutions, highlighting only the positive and negative ones (discarding the neutral ones). There is evidence of a greater number of negative responses (105,547) compared to 73,665 positive responses, this being an indication of a possible bias. The negative sentiment was also evaluated in the responses of public institutions, highlighting that the majority are in the range of 0.1 to 0.41, these responses being “partially negative”, however, it is important to highlight the number of responses with a high percentage of negativism, which are those with scores greater than 0.41 to 0.875.



**Figure 9:** Distribution of sentiment in questions and responses

The result of the classification of the sentiment analysis in the questions asked by the providers is summarized in Fig. 10, which indicates that only 16% of the providers send negative questions to the contracting institution, while 39% are positive. In the case of the responses made by the contracting public entities, 32% of the institutions issue negative responses to the providers, while 45% are neutral and 23% are positive, so it can be inferred that practically 1 out of 3 responses from the contracting entities are negative.



**Figure 10:** Distribution of sentiment analysis in questions and answers

With this, it can be seen that there is a trend of between 30% and 35% of public procurement processes, where it is intended to favor a certain contractor, either through favoritism in questions and answers or with the modification of qualification parameters.

### 5 Discussion

The importance of this work is based on trying to achieve equality and sustainability in public procurement processes, applying the model proposed to detect bias and favoritism in public procurement. This will contribute to achieving the goal of sustainable public procurement, which finds staunch support in the Johannesburg Implementation Plan (PIJ), approved by all governments at the United Nations World Summit for Sustainable Development.

In the evaluation of the questions and answers of each process, it can be observed in the experiments that, far from a treatment based on equality, negative responses are detected by the contracting entities towards certain suppliers. This could mean that in 32% of cases there is bias and favoritism on the part of

the state entity and this is attested with previous work by other authors [7], wherein in 35% of the cases, there was a modification of the qualification parameters of the winner of the offer, with the aim of favoring a certain contractor.

Regarding public procurement policies, the inclusion of women in the entire process can affect the quality of life of this group and their economic autonomy [36]. In addition, there is evidence that women tolerate acts of corruption less than men according to Guerra et al. [37]. Following initiatives like the UK [38] that includes women in public processes in the European construction and industry. Therefore, our work aims to demonstrate the existing bias towards male suppliers, in the responses of the contracting entities, using a Word2vec-based gender bias detection model, which reaches an accuracy of 90%.

In the evaluation of the model of bias between contractors (favoritism), an accuracy of 88% was obtained. These results support the model proposed by the authors in their previous work [11], validated by using NLP to detect corruption. This agrees with the work of Modrusan et al. [6], which makes use of the BoW model for the payment of tax advances, obtaining an accuracy of 71%, but using only 15,000 records, a much smaller quantity than the one used in the present work. Our model makes use of word embeddings just like in Skorková [14], however, it also uses Word2vect as an alternative to BERT.

## 6 Conclusion

With the implemented experiments, the researchers demonstrated that the indications extracted from the questions and answers to the participants in a Spanish corpus public procurement allow biased purchases to be detected. Both gender bias and favoritism towards certain suppliers on the part of the contracting public entities are detected. Also, to complement the model, sentiment analysis of the questions and answers provided by the contracting entities was conducted. The model shows an accuracy of 88% for bias detection and 90% for vendors favoritism. For this reason, the implemented algorithms produced significant improvement in terms of accuracy.

The model shows that one-third of the procurement processes carried out by the state have indications of corruption and bias, which represents millions of dollars in damages suffered by the state, discrimination against suppliers (woman bias), and potential legal cases against those involved. This shows the potential of text analysis in the detection of corruption and bias since the documents attached to procurement processes contain large amounts of information that has not been exploited.

Therefore, the work presented contributes to public contracting being directed to acquiring sustainable, innovative, environmentally friendly design goods, with the criteria of inclusion, equality, and social equity contributing directly to achieving sustainable development goals.

## 7 Future Work

As future work, it is intended to evaluate the model, with other machine learning algorithms like Lexicalized Dependency Paths (LDPs), for usage with active learning methods [39] with noise removal techniques [40], decomposition methods a reduction technique [41,42], or feature selection techniques [43] and finally other techniques for sentiment analysis, in addition to the review of the current methods that are being applied in other countries for the detection of bias and favoritism in public purchases. The proposed method has been empirically verified with purchases from Ecuador and it would be ideal to analyze whether it can be replicated to other Spanish-speaking countries, with the purpose of improving their skills in different dialects and local terminologies.

**Funding Statement:** The authors received no specific funding for this study.

**Author Contributions:** Conceptualization, Y.T.-B. and V.F.L.B.; methodology, Y.T.-B., L.C.Z and V.F.L.B.; formal analysis, Y.T.-B.; investigation, Y.T.-B.; writing—original draft preparation, Y.T.-B. and L.C.Z; writing—review and editing, Y.T.-B. and V.F.L.B. All authors have read and agreed to the published version of the manuscript.

**Availability of Data and Materials:** In this paper the dataset is available in <https://drive.google.com/file/d/11W7hp5nVP09WHSYHBJQXoxi0h6Ov49sY/view?usp=sharing>,

Corpus de questions: <https://www.dropbox.com/s/tlfa5jfuhrvqxtyc/preguntas.txt?dl=0>.

Corpus de answers: <https://www.dropbox.com/s/ktmuj07mbw2jqew/respuesta.txt?dl=0>.

Source code. <https://colab.research.google.com/drive/1-WB6yywjyXKpWevbxwpukGtpxqa8JRse?usp=sharing>.

**Conflicts of Interest:** The authors declare that they have no conflicts of interest to report regarding the present study.

## References

- [1] M. Thi Nguyen and N. T. Bui, “Government expenditure and economic growth: Does the role of corruption control matter?,” *Heiyon*, vol. 8, no. 10, pp. 127–147, 2022.
- [2] E. Kehler, J. Paciello and J. Pane, “Anomaly detection in public procurements using the open contracting data standard,” in *Proc. 2020 Seventh Int. Conf. on eDemocracy & eGovernment (ICEDEG)*, Buenos Aires, Argentina, pp. 127–135, 2019.
- [3] Servicio Nacional de Contratación Pública, “Rendición de cuentas,” SERCOP, Quito, Ecuador, 2018.
- [4] M. S. Rad and A. Shahbahrami, “Detecting high risk taxpayers using data mining techniques,” in *Proc. 2016 2nd Int. Conf. of Signal Processing and Intelligent Systems (ICSPIS)*, Tehran, Iran, pp.14–15, 2017.
- [5] D. M. E. D. M. Hussein, “A survey on sentiment analysis challenges,” *Journal of King Saud University-Engineering Sciences*, vol. 30, no. 4, pp. 330–338, 2018.
- [6] N. Modrusan, K. Rabuzin and L. Mršić, “Improving public sector efficiency using advanced text mining in the procurement process,” in *Proc. of the 9th Int. Conf. on Data Science, Technology and Applications (DATA 2020)*, Paris, France, pp. 200–206, 2020.
- [7] Y. Torres-Berru and V. F. L. Batista, “Data mining to identify anomalies in public procurement rating parameters,” *Electronics*, vol. 10, no. 22, pp. 1–15, 2021.
- [8] E. Ortiz-Prado, R. Fernandez-Naranjo, Y. Torres-Berru, R. Lowe and I. Torres, “Exceptional prices of medical and other supplies during the COVID-19 pandemic in Ecuador,” *The American Journal of Tropical Medicine and Hygiene*, vol. 105, no. 1, pp. 81–87, 2021.
- [9] Y. Torres Berru, V. F. López Batista, P. Torres-Carrión and M. G. Jimenez, “Artificial intelligence techniques to detect and prevent corruption in procurement: A systematic literature review,” *Communications in Computer and Information Science*, vol. 1194, pp. 254–268, 2020.
- [10] Y. Torres-Berru, V. F. L. Batista and P. Torres-Carrión, “Data mining to detect and prevent corruption in contracts: Systematic mapping review,” *RISTI-Revista Iberica de Sistemas e Tecnologias de Informacao*, vol. 2020, no. E29, pp. 13–26, 2020.
- [11] Y. Torres-Berru and V. F. López Batista, “Data and text mining for the detection of fraud in public contracts: A case study of Ecuador’s official public procurement system,” in *Doctoral Symp. on Information and Communication Technologies-DSICT. Lecture Notes in Electrical Engineering*, vol. 846, pp. 116–127, 2022.
- [12] Y. Bramoullé and S. Goyal, “Favoritism,” *Journal of Development Economics*, vol. 122, no. 1, pp. 16–27, 2016.

- [13] N. W. Rustiarini, T. Sustrino, N. Nurkholis and W. Andayani, "Why people commit public procurement fraud? The fraud diamond view," *Journal of Public Procurement*, vol. 19, no. 4, pp. 345–362, 2019.
- [14] Z. Skorková, "Competency models in public sector," *Procedia-Social and Behavioral Sciences*, vol. 230, pp. 226–234, 2016.
- [15] K. Pant, T. Dadu and R. Mamidi, "Towards detection of subjective bias using contextualized word embeddings," in *Proc. Companion Proc. of the Web Conf. 2020*, Taipei, Taiwan, pp. 75–76, 2020.
- [16] C. A. Moss-Racusin, J. F. Dovidio, V. L. Brescoll, M. J. Graham and J. Handelsman, "Science faculty's subtle gender biases favor male students," *Proceedings of the National Academy of Sciences*, vol. 114, pp. 3–14, 2013.
- [17] T. Bolukbasi, K. W. Chang, J. Zou, V. Saligrama and A. Kalai, "Man is to computer programmer as woman is to homemaker? Debiasing word embeddings," *Advances in Neural Information Processing Systems*, pp. 4356–4364, 2016.
- [18] T. Sun, A. Gaut, S. Tang, Y. Huang, M. ElSherief *et al.*, "Mitigating gender bias in natural language processing: Literature review," in *Proc. ACL, 2019 Proc. of the 57th Annual Meeting of the Association for Computational Linguistics*, Florence, Italy, pp. 1630–1640, 2019.
- [19] M. Hamisu and A. Mansour, "Detecting advance fee fraud using NLP bag of word model," in *Proc. 2020 IEEE 2nd Int. Conf. on Cyberspace CYBER*, Niger, pp. 94–97, 2021.
- [20] A. J. Ruiz, *Inclusión de mujeres en las contrataciones públicas: la experiencia latinoamericana*, La Haya, Netherlands: Hivos, 2020.
- [21] C. Caparrós-Laiz, J. A. García-Díaz and R. Valencia-García, "Evaluating extractive automatic text summarization techniques in Spanish," *Communications in Computer and Information Science*, vol. 1460, pp. 79–92, 2021.
- [22] M. P. Akhter, Z. Jiangbin, I. R. Naqvi, M. Abdelmajeed and M. T. Sadiq, "Automatic detection of offensive language for Urdu and Roman Urdu," *IEEE Access*, vol. 8, pp. 91213–91226, 2020.
- [23] J. M. Conroy, J. D. Schlesinger and D. P. O'Leary, "Nouveau-ROUGE: A novelty metric for update summarization," *Association for Computational Linguistics*, vol. 37, no. 1, pp. 1–8, 2011.
- [24] E. Reiter, "A structured review of the validity of BLEU," *Computational Linguistics*, vol. 44, no. 3, pp. 393–401, 2018.
- [25] E. V. Tunyan, T. A. Cao and C. Y. Ock, "Improving subjective bias detection using bidirectional encoder representations from transformers and bidirectional long short-term memory," *International Journal of Cognitive and Language Sciences*, vol. 15, pp. 329–333, 2021.
- [26] K. W. Church, "Emerging Trends: Word2Vec," *Natural Language Engineering*, vol. 23, no. 1, pp. 155–162, 2017.
- [27] B. Athiwaratkun, A. G. Wilson and A. Anandkumar, "Probabilistic FastText for multi-sense word embeddings," in *Proc. 56th Annual Meeting of the Association for Computational Linguistics, ACL 2018*, Melbourne, Australia, pp. 1–11, 2018.
- [28] F. Millstein, "Natural language processing with python: Natural language processing using NLTK," in *Frank Millstein: North Charleston, USA*, 2020.
- [29] J. Reback, W. McKinney, J. Van den Bossche, T. Augspurger, P. Cloud *et al.*, "Pandas," in *Zenodo [code]*, 2020. <https://doi.org/10.5281/zenodo.4524629>
- [30] A. E. Takieldeem, E. M. El-kenawy, M. Hadwan and R. M. Zaki, "Dipper throated optimization algorithm for unconstrained function and feature selection," *Computers, Materials & Continua*, vol. 72, no. 1, pp. 1465–1481, 2022.
- [31] G. Veena, A. Vinayak and A. Nair, "Sentiment analysis using improved Vader and dependency parsing," in *Proc. 2021 2nd Global Conf. for Advancement in Technology (GCAT)*, Bangalore, India, pp. 1–6, 2021.
- [32] R. Bakshi, N. Kaur, R. Kaur and G. Kaur, "Opinion mining and sentiment analysis," in *2016 3rd Int. Conf. on Computing for Sustainable Global Development*, India, pp. 452–455, 2016.
- [33] K. Patel and P. Bhattacharyya, "Towards lower bounds on number of dimensions for Word Embeddings," in *Proc. of the 8th Int. Joint Conf. on Natural Language Processing*, Taipei, Taiwan, pp. 31–36, 2017.
- [34] R. Popović, F. Lemmerich and M. Strohmaier, "Joint multiclass debiasing of word embeddings," *Lecture Notes in Computer Science*, vol. 12117, pp. 79–89, 2020.

- [35] S. Sabra and V. Sabeeh, "A comparative study of N-gram and Skip-gram for clinical concepts extraction," in *Proc. Int. Conf. on Computational Science and Computational Intelligence (CSCI)*, Las Vegas, USA, pp. 807–812, 2020.
- [36] G. Pierri, M. J. Jarquin and R. De Michele, *Transparencia y género: el impacto de las compras electrónicas en el acceso a licitaciones públicas de las pymes lideradas por mujeres*, Banco Interamericano de Desarrollo, 2021.
- [37] A. Guerra and T. Zhuravleva, "Do women always behave as corruption cleaners?," *Public Choice*, vol. 191, no. 1–2, pp. 173–192, 2022.
- [38] T. Wright, "New development: Can 'social value' requirements on public authorities be used in procurement to increase women's participation in the UK construction industry?," *Public Money & Management*, vol. 35, no. 2, pp. 135–140, 2015.
- [39] H. Sun and R. Grishman, "Lexicalized dependency paths based supervised learning for relation extraction," *Computer Systems Science and Engineering*, vol. 43, no. 3, pp. 861–870, 2022.
- [40] M. T. Sadiq, X. Yu, Z. Yuan and M. Z. Aziz, "Motor imagery BCI classification based on novel two-dimensional modelling in empirical wavelet transform," *Electronic Letters*, vol. 56, pp. 1367–1369, 2020.
- [41] M. T. Sadiq, X. Yu, Z. Yuan, Z. Fan, A. U. Rehman *et al.*, "Motor imagery EEG signals classification based on mode amplitude and frequency components using empirical wavelet transform," *IEEE Access*, vol. 7, pp. 127678–127692, 2019.
- [42] M. T. Sadiq, X. Yu, Z. Yuan and M. Z. Aziz, "Exploiting dimensionality reduction and neural network techniques for the development of expert brain–computer interface," *Expert Systems with Applications*, vol. 164, pp. 114031, 2021.
- [43] E. -S. M. El-kenawy, A. Ibrahim, N. Bailek, B. Kada, M. Hassan *et al.*, "Sunshine duration measurements and predictions in Saharan Algeria region: An improved ensemble learning approach," *Theoretical and Applied Climatology*, vol. 147, no. 3–4, pp. 1015–1031, 2022.