CENTRO DE INVESTIGACIÓN Y DOCENCIA ECONÓMICAS, A.C.

CLASSIFICATION OF PUBLIC PROCUREMENT PROCEDURES BASED ON RED FLAGS WITH A MACHINE LEARNING APPROACH FOR THE DETECTION OF POSSIBLE RISK OF CORRUPTION

TESINA

QUE PARA OBTENER EL GRADO DE

MAESTRO EN MÉTODOS PARA EL ANÁLISIS DE POLÍTICAS PÚBLICAS

PRESENTA

EMILIO PERFECTO MARTÍNEZ ARÉVALO

DIRECTORA DE LA TESINA: DRA. DANIELA ALEJANDRA MOCTEZUMA OCHOA

AGUASCALIENTES, AGS. 2023

A mis hermosa familia:
María Dolores Arévalo Zenteno
Ricardo Martínez Martínez
Ricardo Israel Martínez Arévalo
Miguel Ángel Martínez Arévalo
Carlos Maximiliano Martínez Arévalo,
que con su ejemplo, soporte y acompañamiento en esta etapa, me motivaron a superar los momentos difíciles y a hacer más amena mi estancia en el CIDE.

Acknowledgments

Quiero agradecer a:

Dra. Daniela Alejandra Moctezuma Ochoa, por su invaluable experiencia y guía durante mi investigación y maestría,
mi familia, por alentarme a salir adelante y brindarme las oportunidades para trabajar,
mis profesores, por sus valiosas enseñanzas a lo largo de estos dos años,
mis amigos, por hacer mi estancia en el CIDE más placentera,
al Consejo Nacional de Ciencia y Tecnología (CONACYT), por financiar mis estudios de maestría,
y a todos los demás que me ayudaron a concluir esta etapa de forma satisfactoria.

Abstract

This study addresses corruption in public procurement by integrating machine learning and red flags. Objectives include evaluating red flags' impact on corruption detection using machine learning models and comparing different algorithms' performance.

Using data from Mexico's CompraNet platform, red flags—indicators of potential corruption—are incorporated. Supervised machine learning models (e.g., XGBoost, Random Forest, Logistic Regression) are trained and evaluated with using both inputs, with and without red flags variables. The key findings of this study underscore the significant positive influence of red flags on the accuracy of corruption detection across various machine learning models. The integration of red flags consistently improves precision, recall, and F1-scores, reaffirming their effectiveness as valuable corruption risk indicators. Furthermore, the comparative assessment of machine learning algorithms reveals variations in performance, emphasizing the critical nature of model selection.

In conclusion, red flags effectively help to improve the detection of potential corruption risks in public procurement. Machine learning's role in leveraging red flags shows promise for corruption detection. These insights have implications for public governance and policy-making, emphasizing the potential of data-driven approaches in mitigating corruption's adverse effects. Furthermore, this research highlights avenues for future exploration, such as tailored red flag frameworks, real-time detection, and cross-domain application, providing a comprehensive outlook for advancing corruption detection and prevention strategies.

Key words: Public Procurement, Corruption, Machine learning, Red Flags.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

All public policies are materialized in public contracts. It is a priority to know if what is implemented by the governments is achieving the expected impact. For instance, that public policies reach the beneficiaries of the desired economic sectors, more efficient processes are generated, greater competition and transparency are also generated. All of this can be achieved by being able to analyze the data when we have all the data from public policies and their respective public procurement contracts. In other words, having this data is a good starting point to have accountability effectively.

Public procurement is the process by which governments and other bodies governed by public law purchase products, services, and public works. Representing on average from 13% to 20% of Gross Domestic Product (GDP) (World Bank, 2023), public procurement is an essential area of any government, and following Magakwe, J, (2022), plays a pivotal role in driving economic activities. This area is particularly vulnerable to fostering corruption and illegality at any stage of the process. According to the United Nations Office of Drugs and Crime (UNODC) (2013), 10% to 25% of a public contract's overall value may be lost due to corruption. In the literature, numerous reasons contribute to the vulnerability of public procurement to corruption. Firstly, the disconnection between those securing contracts and those funding them provides an opportunity for corrupt practices (Decarolis & Giorgiantonio, 2022). Additionally, the unique

characteristics of public funding set public procurement apart from private acquisitions and related activities (Sun & Sales, 2018). Corrupt actors may exploit corruption techniques to give the appearance of legality and conceal their actions from the public eye (Fazekas, Tóth, & King, 2013b). Moreover, the lack of reliable indicators of corruption poses challenges (Fazekas & Tóth, 2016), and the involvement of various actors throughout the process, coupled with the handling of substantial sums of money, adds to the vulnerability.

Corruption is an ancient problem; it has always been with us because it is a social phenomenon with different incidences at different times at different places, with varying degrees of damaging consequences (Bardhan, 1997). There is no specific definition of corruption due to its multidimensional complexity; corrupt behaviors vary because the word "corruption" is used to mean different things in different contexts. Among the most widely accepted definitions in the literature, International Transparency, one of the most important anti-corruption agencies, defines corruption as the abuse of public power to obtain private benefits.

In the case of public procurement, corruption refers to using one's position or influence in the government to gain an unfair advantage in awarding contracts for public projects. According to (Fazekas, Tóth, & King, 2016), one of the authors cited in the field, public procurement corruption denotes the allocation and performance of public procurement contracts to benefit a closed network while denying access to all others. This can take many forms, such as accepting bribes in exchange for awarding contracts to specific companies, manipulating the bidding process to favor certain companies, or using insider information to secure contracts for oneself or one's associates. The desire for better understanding and the fight against corruption have become relevant topics. This motivation is mainly due to their different consequences and the realization among experts in the field and the public that development requires good public governance (Jain, 2001).

The consequences of corruption in public procurement are vast, and in countries with political and economic instability, the negative impacts are more significant (Mizoguchi & Van Quyen,

2014). As pointed out by Sun and Sales (2018), corruption not only results in a misuse of public money, but it also represents incomplete service or inferior quality. Additionally, corruption is widely recognized as one of the most significant obstacles to achieving efficient and sustainable economic and social development (Anderson, Kovacic, & Müller, 2011).

Understanding and identifying corruption in public procurement has received extensive academic and policy attention due to its central role in effective public policy. However, this is a social complex phenomenon that, by its nature, is difficult to detect, analyze, and measure (Mufutau & Mojisola, 2016) because it is illicit and secretive (Shleifer & Vishny, 1993).

Different approaches and indicators exist for measuring and detecting corruption, such as surveys of corruption perceptions and attitudes, reviews of institutional and legal frameworks, and detailed analysis and audits of individual cases (Fazekas, Tóth, & King, 2013a). Nonetheless, some field research indicates a need to use new approaches such as red flags (Ferwerda, Deleanu, & Unger, 2017). This recommendation is because of the deficiencies of the above traditional indicators. For instance, some key arguments include that perceptions may not be related to experience (Rose & Peiffer, 2015). Since these indicators are typically produced from non-representative surveys, representative bias will likely occur. Regarding surveys of corruption experience, the main problem is the insufficient data source since only a tiny fraction of the population has direct experience with corruption.

It is essential to consider using red flags as an efficient measurement; having quality data of public procurement records is necessary. Furthermore, the legislation of each country differs, and the data and the red flags implementation may be specific to certain countries. Nevertheless, as Modrusan and their colleagues (2021) said with some efforts and specific techniques, scientists can use that data to analyze the public procurement process and find adequate corruption indicators.

In the case of Mexico, public procurement records derive from the electronic Mexican system of public governmental information on public procurement, a.k.a as CompraNet. This e-

government system was created in 1996 to increase the efficiency and transparency of public procurement. As IT technology advances and the Public Procurement Process (PPP) undergoes digitalization, a growing volume of data becomes accessible, enabling new tools and methods like machine learning.

Machine learning can be used to detect corruption in public procurement by analysing patterns in the data related to the awarding of contracts. For example, a machine learning model could be trained to identify unusual patterns in the bidding process, such as a sudden increase in the number of bids from a particular company or a sudden decrease overall. The model could also be trained to detect unusual patterns in the award of contracts, such as a disproportionate number of contracts awarded to a single company or a sudden increase in the number of contracts awarded to companies with connections to government officials. By identifying these and other potential corruption indicators, machine learning can help detect and prevent corruption in public procurement.

As we mentioned above, each data and the Red Flags implementation may be specific for each country. In the case of Mexico, it is important to implement research with this approach since corruption is one of the main problems in the country, only behind violence and insecurity Amparo, M. (2015). Between 2013 and 2020, spending in public procurement reached 10% of the approved federal spending and 2% of this money was lost in corrupt practices (Falcón-Cortés, Aldana, & Larralde, 2022). Furthermore, there are large amounts of administrative data coming from the main public procurement platform, CompraNet, which allows studying corruption with different perspectives, technologies, and new data-driven perspectives.

The purpose of this study is to implement different supervised machine learning models using red flags as additional input data to detect and classify public procurement procedures with indications of corruption and analyze the performance of this kind of disruptive methodology. The proposed methodology is based on the study of Aldana et al. (2022) and Falcón-Cortés et al. (2022) since are the most recent studies focused on the implementation of machine learning

tools to detect corruption in public procurement in Mexico with Red Flags.

In the work of Falcón-Cortés and colleagues (2022) the red flags are based only on the work of the Mexican Institute of Competitiveness (MIC). Hence, we propose the research and implementation of red flags that different organizations have developed based on available data.

The primary objective of this study is to test the capability of red flags established by different organizations in classifying public procurement processes into those with potential risks and those without corruption risks. This will be achieved by implementing various well-known machine learning models renowned for their effectiveness in solving classification problems. Given the main objective, the research question addressed in this work focuses on examining the new red flags proposed to improve the identification of potential risks of corruption in public procurement through different supervised machine learning algorithms.

Consequently, the research question guiding this study is:

- Does adding the proposed red flags improve the performance of machine learning algorithms classifying public procurement processes into those with potential risk and those without?

The nature of the question is evaluative and exploratory because it aims to assess the effectiveness of new red flags in the Mexican public procurement context in enhancing the identification of potential corruption risks in public procurement. Besides, it investigates the impact and value of new red flags, suggesting that there is a need to understand and measure the effectiveness of these red flags in the context of public procurement.

The research outline is divided into four sections. The first section is a literature review to present the knowledge and gaps about using red flags in the case of Mexico and machine learning to detect corruption in the public procurement process. The following section will cover data understanding, particularly how risk indicators or red flags are created. Then, the

fourth section presents the research methodology, including the data analysis method and the modeling process. Next, we present the results of the classification models, and finally, the concluding section summarizes the present work, highlights its contribution to the literature, and provides the direction for future work.

Chapter 2

Literature Review

This section briefly spells out the general direction of new opportunities presented for the detection of possible risks of corruption in public procurement with machine learning. This literature review aims to synthesize areas of conceptual knowledge that contribute to a better understanding of the issue. The sources for the literature review were obtained through a systematic search of academic databases and reputable organizations, ensuring the inclusion of credible and peer-reviewed articles.

2.1 Red Flags in Public Procurement

Public procurement's vulnerability and consequences have been a key motive for implementing efforts to monitor, measure, and fight corruption. Following Rakhel and Putera (2021), the publication trend has continued to rise together with initiatives worldwide. Different institutions, anti-corruption agencies, and researchers have proposed different frameworks for studying and combating corruption. Multilateral organizations like the United Nations, the World Bank, the World Trade Organization, and the Organization for Economic Cooperation and Development aim to fight the problem.

Tina Soroide (2002) wrote an article entitled 'Corruption in Public Procurement: Causes, Consequences, and Cures', where she discussed various strategies to reduce corruption in governmental acquisitions. Concurrently, the American Institute of Certified Public Accountants (AICPA) published 'Consideration of Fraud in a Financial Statement Audit', which included a list of fraud indicators. Subsequently, in 2004, the Organization for Economic Co-operation and Development (OECD) organized a global forum meeting on 'Fighting Corruption and Promoting Integrity in Public Procurement'. During this meeting, the OECD initiated further work to review the risks associated with public procurement and tried to enhance understanding of the methods and techniques involved in corruption cases. Building upon these efforts, in 2006, the Financial Action Task Force recommended using red flags to minimize the risk of financial institutions handling criminal money. The following year, 2007, the World Bank adopted a new anti-corruption strategy recommending red flags. Recognizing the need for global action, Transparency International, a global coalition against corruption, has implemented various strategies worldwide to advocate for policies tackling corruption.

The literature reflects the discussions and efforts to understand the corruption phenomenon in public procurement and the tendency to propose red flags as corruption measurements in the field. Before addressing the studies that have analyzed corruption in public procurement processes using red flags, it is important to define a red flag. Dorn and colleges (2008) consider that red flags indicate specific risks of various forms of economic misconduct, including corruption in public procurement. Ferwerda et al.(2017) state that the logic behind the red flags is that corrupt activities require specific forms of economic behavior, and this behavior leaves traces. Hence, red flags are an accumulation of traces that may point to the presence of corrupt activities. With the above definitions, it is of utmost importance to bear in mind that in this work, the presence of red flags does not prove that we can safely say these are corrupt competitions. It simply proves additional checks or investigations are warranted and must talk about corruption risk score.

The literature focusing on applying red flags specifically within the context of public procurement corruption is relatively new and has garnered significant academic interest. One of the first efforts is the work of Fazekas and colleagues (2016); in their paper titled 'An objective corruption risk index using public procurement data', they propose a new composite indicator of grand corruption based on a wide range of elementary indicators. Using administrative data from Hungarian public procurement, the composite indicator is constructed by red flags to restrict market access in each stage or process (submission, assessment, and delivery). They consider 14 input factors to capture key characteristics of the public procurement process from the beginning of the submission phase until the end of delivery. From their findings, it can be concluded that it is feasible and fruitful to construct a risk index and the micro-level based on objective behavioral data only. In addition, almost every corruption input displayed a relationship with corruption outcomes in line with prior expectations.

2.2 Corruption Detection in Public Procurement with Machine Learning

Numerous studies have started using various methods to detect and prevent corruption in public procurement. With the development of information technology and the digitalization of the public procurement process, the amount of data and the possibility of using new methods is increasing. For example, a World Bank study (2020) presents new technology trends to tackle corruption, such as big data, cloud computing platforms, artificial intelligence, and machine learning. The study proposes a modular set of approaches, entry points, and tools that can be drawn upon and adapted to their specific country context.

The use of new analytical methods, such as machine learning, has gained popularity in the last decade due to the increasing data availability in public procurement and the development of new technologies to handle large amounts of data. According to Modrušan et al. (2021),

advanced statistics and data mining techniques are used to develop models to analyze corruption in public procurement. However, they agree that detection methods and techniques largely depend on the input data. In other words, one of the main obstacles to creating efficient corruption detection models is insufficient quality and diverse data. The most common approach to try to characterize and prevent corruption in public procurement is building risk factors from contract data.

Corruption, as a complex phenomenon, needs tools according to this complexity, and machine learning algorithms help analyze and model complex relationships without having a theory and structure of the problem behind them. According to Decarolis et al. (2022), when assessing red flags, machine learning methods are adequate for two reasons: The first one is because this approach deals with the trade-off between the expressiveness of the model and the risk of over-fitting. The second reason, according to the author, is because few red flags have a close relationship with corruption. As these red flags are mere tools for corruption arrangement, they are easily substituted with others. In other words, the functionality of these tools depends on the modification of corruption practices.

The literature review shows that the application of machine learning for corruption analysis has been increasing worldwide in countries and anti-corruption agencies. In 2019, research focused on Croatia tested different machine learning algorithms with the tender documentation of particular public procurement procedures to determine whether they can be used to detect indications of corruption in public procurement (Rakhel & Putera, 2021). Sun and Sales (2018) develop a system that uses the characteristics of the bidding company in Brazil to predict the risk of public procurement irregularities such as contract default; they use data from 2011 to 2014. Another example is the case of Italia, Decarolis, and Giorgiantonio (2022), who analyzed how different procurement features are associated with risk corruption. In the case of Mexico, Aldana, A. et al. (2022) propose a machine learning model based on an ensemble of random forest classifiers to identify and predict corrupt contracts.

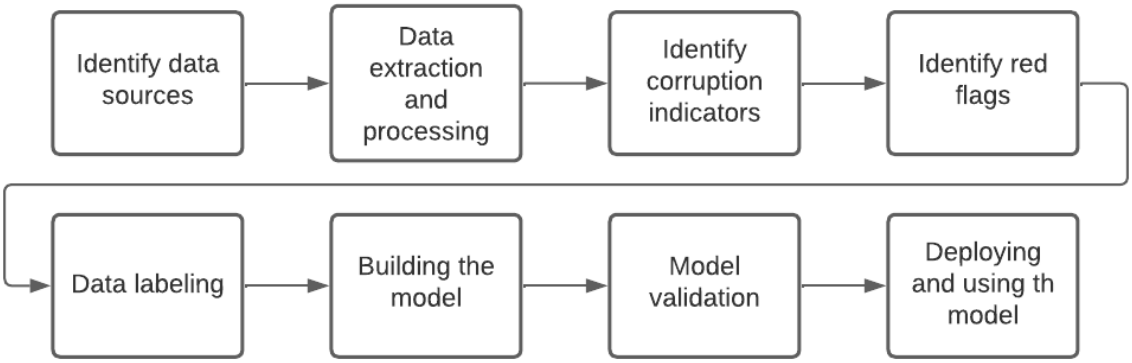
The work of Nikola and Modrusan (2021) aligns with the review of machine learning approaches in corruption detection because they address a comprehensive review of the emerging techniques and models used to detect suspicious or corrupted observations. After segmenting 23 scientific studies by analyzing scientific databases, the review shows that researchers used supervised and unsupervised machine learning. In addition, according to the authors, all studies generally show that the fraud detection model is divided into a set of steps:

- Identify Data Sources: Identify relevant data sources that might contain corruption-related information. This could include financial records, transaction logs, text documents, or any data that might provide insights into corrupt activities.
- Data Extraction and Processing: Extract data from the identified sources and preprocess it. This involves cleaning the data, handling missing values, and converting it into a format suitable for analysis.
- Identify Corruption Indicators: Determine potential features or attributes that might serve as indicators of corruption.
- Identify red flags: Define specific red flags or suspicious patterns that are often associated with corrupt activities
- Data Labeling: Manually label the data instances as either "corruption" or "non-corruption" based on the identified red flags. This labeled data will be used to train and evaluate the machine-learning model.
- Building the Model: Select an appropriate machine learning algorithm for the task (such as classification or anomaly detection). The labeled data is used to train the model, tuning its parameters for optimal performance.
- Model Validation: Assess the model's performance using validation techniques like cross-validation or a separate validation dataset. Measure metrics such as precision, recall, and F1-score to evaluate how well the model identifies corruption.

- Deploying and Using Model: Once satisfied with the model's performance, deploy it to a production environment. This could involve integrating the model into an existing system or creating an application that can take new data and predict whether corruption is likely present.

These can be seen in Figure 2.1.

Figure 2.1: Workflow to apply machine learning model to corruption detection task.



Source: Own elaboration.

The models are fitted on historical data and are used for several different purposes in the detection of corruption:

- Estimating the probability of corruption: Ferwerda and Deleanu (2013) estimate the direct costs of corruption in terms of economic loss for the public in public procurement procedures. The methodology is calculated by comparing effectiveness, efficiency, cost overruns, delays, and quality considerations.

- Predicting the number of bidding tenders: Mencia and colleges (2013) explore how data mining techniques such as discretization, processing of text fields, feature selection, and machine learning algorithms can be used on semantically linked data to estimate the number of bidders in public contracts.

- Predicting fraud risk in contracts and contractors: Wang (2016) employs game theory, machine learning, and statistical methods to detect fraud risk in Federal Procurement Contracts. Implementing a One-Class Support Vector Machine developed a classifier using historical data of contractors.

- Predictive models of fraud risk in contracts: Sales and Carvalho (2016) create a risk measurement model of companies that negotiate with the government using indicators grouped into four risk dimensions: operational capacity, history of penalties and findings, bidding profile and political ties. They use Bayesian Classifiers to contribute to selecting contracts to be audited. Another example is the other work of Sales (2013); the work aims to identify bidders likely to fail in the fulfillment of obligations under contracts with the government. He uses different statistical techniques.

- Anomaly detection: Domingos et al. (2016) investigates IT purchase anomalies in the Federal Government Procurement System by using a deep learning algorithm to generate a predictive model.

- Cartel detection: Ralha and Silva (2012) assess the problem of extracting useful information from the Brazilian federal procurement process databases used by government auditors in the process of corruption detection and prevention to identify cartel formation among applicants. They use clustering and association rules and a multi-agent approach to address the dynamic strategies of companies involved in cartel formation.

- Collusive behavior: Tas (2017) designed a method to identify and test for bid rigging in procurement auctions using limited information. He uses standard machine learning tools and statistical software using data from Turkish public procurement auctions, and he finds that collusion significantly increases procurement costs and decreases cost-effectiveness.

- Detection of fraudulent public procurement processes: The work of Arief et al. (2016) focuses on detecting potential fraud in the procurement process via the Indonesian E-

Procurement System (SPSE). They implement a fraud detection mechanism using data mining techniques based on supervised learning.

The most common methods in the studies are linear and logistic regression, neural networks, and Naive Bayes algorithms because they are the most used for classification and clustering. However, papers exist implementing neural networks, deep neural networks, bayesian networks, naive Bayes, support vector machines, discriminant function analysis, decision trees, lasso logistic regression, and more.

The plausibility of this approach lies mainly in that there exist different public procurement corruption detection techniques with emerging technologies. Governments worldwide have taken advantage of these new emerging technologies and the amount of data to study corruption. Compared to more advanced countries, the available data on the public procurement process in Mexico presents certain limitations; the data is unconnected with other electronic platforms that contain important information for analyzing corruption and is largely unstructured. Despite this, with some effort and specific techniques, other studies have succeeded in using the information of the public procurement process to find adequate corruption indicators. Additionally, machine learning methods experienced a boost thanks to data storage capability and the improvement of computing power. This allows these methods to be more innovative, faster, more intuitive, and structured more like the human brain (Sun & Sales, 2018).

There are two main machine learning categories: supervised learning and unsupervised learning. The main difference is that to build the classification model, we need the target variable, and in unsupervised learning, we do not. Implementing machine learning tools to analyze, understand, classify, and predict corruption in Mexico has gained popularity in the scientific community in recent years (Rabuzin & Modrusan, 2019). For example, Zumaya et al. (2021) implemented a deep neural network and a random forest to analyze electronic records of all taxable transactions since 2014 from the Mexican federal government. They trained each method with a portion of the test evader list, tested it with the rest, and showed evidence of a group

of highly suspicious contributors sorted by the amount of evaded tax. They conclude that it is possible to use tools from machine learning to identify patterns, but it should also be taken cautiously. Aldana and colleagues (2022), using data from CompraNet, implemented a machine learning model based on an ensemble random forest to detect corrupt contracts in México's public procurement data. Their model can detect corrupt contracts with an accuracy of 88% and non-corrupt contracts with 94% accuracy. An important conclusion reached in this study is that those variables relate directly to the relationship between buyer/supplier, and risk factors are more efficient predictors than those that only describe contract features. Another case is the study of Rabuzin and Modrusan (2019); through the use of the content of the tender documentation as a data source, they compare prediction models using text-mining techniques and machine learning methods to detect suspicious tenders in Croatia. They found that support vector machines and logistic regression are better at making predictions related to health and social work. At the same time, in almost all cases, the naive Bayes algorithm showed better results.

Chapter 3

Data

3.1 Datasets description

The two databases used for developing the quantitative indicators of the possible risk of corruption or red flags derive from Mexican public procurement records from the electronic Mexican system of public governmental information on public procurement, CompraNet. The first one is from the work of Falcón-Cortés and colleagues (2022). The second database derives from the Public Procurement Index of the Mexican Institute for Competitiveness;[1] this research center developed a tool to identify corruption risk practices in public procurement processes of federal institutions using data from CompraNet.

The data represent a complete database of all public procurement records conducted in Mexico. The principal motivators for using these two databases are because, on the one hand, the first one represents the data available for the public, and on the other hand, to create the second database represents better-quality data in public procurement. Thus, each database gives us two scenarios about the performance of the machine learning algorithms using two types of

[1] For further information, please visit: `https://imco.org.mx/indice-de-riesgos-de-corrupcion-2022/`.

databases.

For each normal database without the red flags proposed (Normal IMCO and Normal RPS), we create a new one, adding the red flags proposed in this work; therefore, they were named as:

- Normal IMCO: IMCO database without red flags proposed.
- IMCO with red flags: IMCO database with red flags proposed.
- Normal RPS: RPS database without red flags proposed.
- RPS with red flags: RPS database with red flags proposed

A comparison of the two primary datasets (Normal IMCO and Normal RPS) reveals a difference in size.

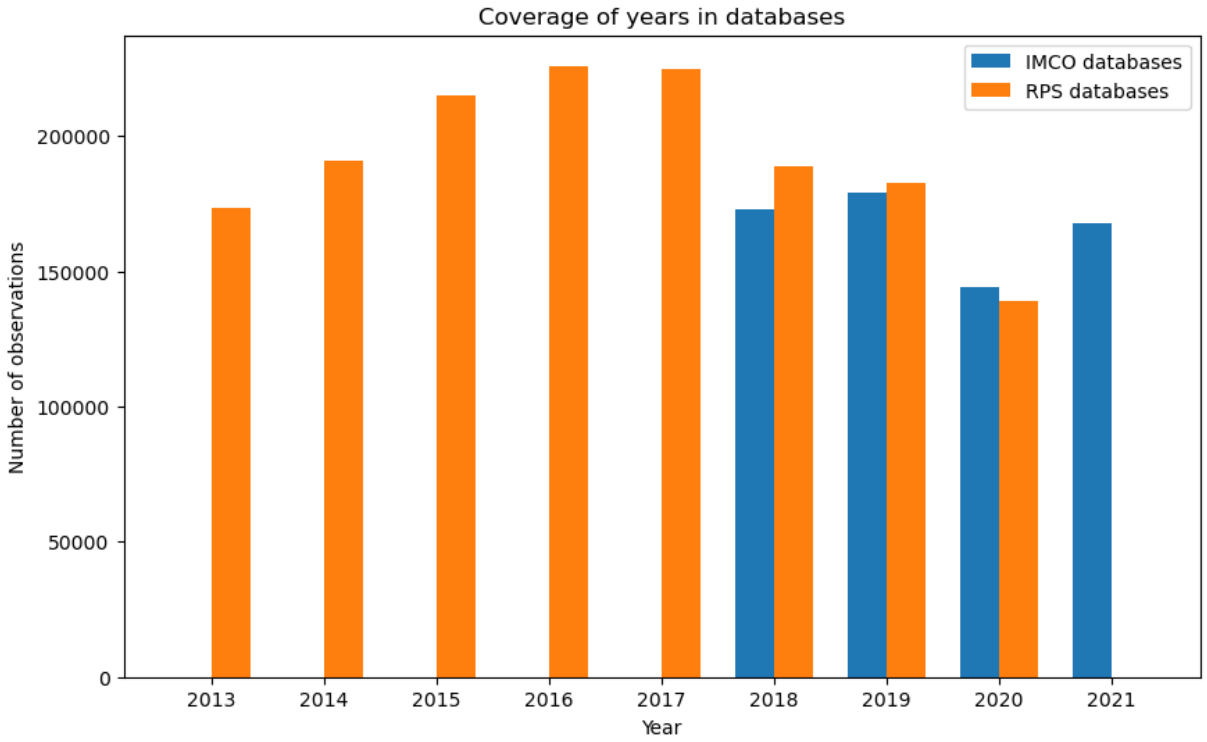
- The Normal IMCO dataset has 663,529 observations and 44 variables, whereas the second RPS dataset is relatively larger, with 1,540,386 observations and 29 variables. This discrepancy in size may have implications for data analysis and computational requirements. In the case of the augmented databases, IMCO with red flags has 663,529 observations and 62 variables, whereas RPS with Red Flags has 1,540,386 observations and 37 variables. This can be seen in Table 3.1.

Table 3.1: Shape of databases

Database	**Observations**	**Features**
Normal IMCO	663529	44
IMCO with Red Flags	663529	62
Normal RPS	1540386	29
RPS with Red Flags	1540386	37

- In terms of temporality, the IMCO dataset has public procurement procedures from 2018 to 2021, and RPS dataset has public contracts from 2013 to 2020. This can be seen in Figure 3.1, also the number of observations in each year.

Figure 3.1: Coverage of years in databases.



Source: Own elaboration.

- As you can see, there is a difference in the number of variables between databases. This is due to different reasons.
  - The Normal IMCO database contains more variables
  - Since Normal RPS covers more years, when one hot encoder is applied, each year from 2013 to 2020 is converted to a new variable.

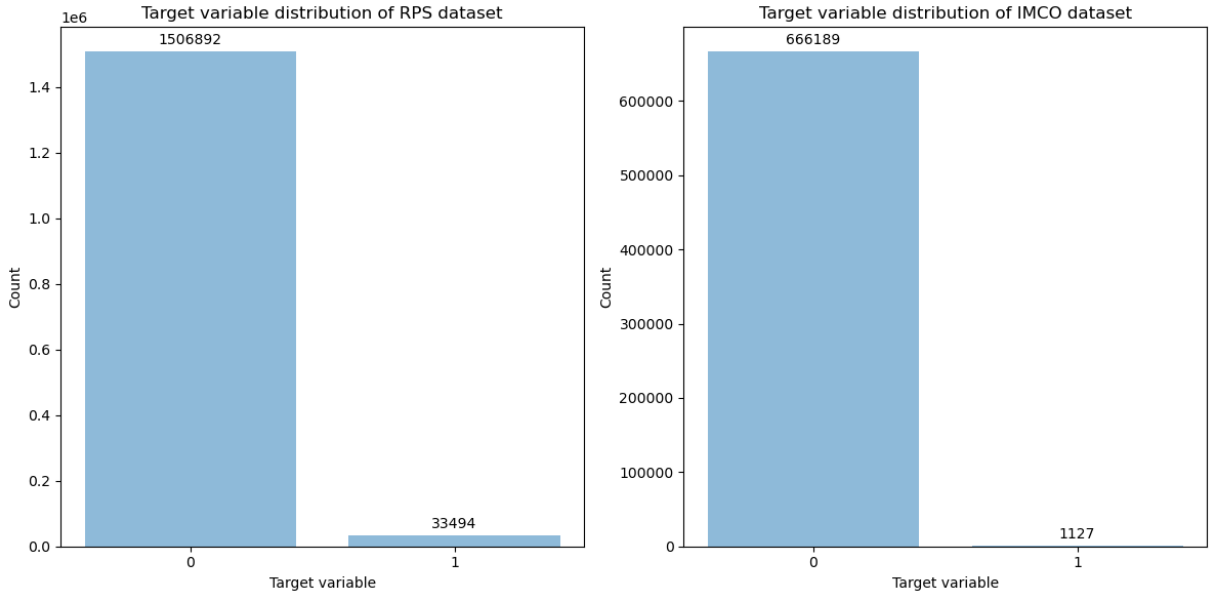
  In Appendix A of this work, two tables outline the variables used in this work. These tables offer a comprehensive overview of the variables employed in the study and their corresponding descriptions. The tables facilitate a better understanding of the data sources and enhance the transparency of the research methodology.

3.2 Imbalanced datasets

As fraud and corruption cases rarely occur and are difficult to detect because it is illicit and secretive, the number of positive labeled classes in public procurement records is minimal. Datasets in which one class is much more frequent than the other are often called imbalanced datasets. This phenomenon is very frequent in real-world problems.

Figure 4.1 shows the imbalanced distribution of the target variable, indicating if the procedure is corrupt or not.

Figure 3.2: Distribution of target variable across Normal IMCO and Normal RPS datasets



Source: Own elaboration.

There exist different approaches to handling imbalanced datasets. For instance, random undersampling, oversampling, class weight, changing the evaluation metric, and collecting more data (Müller & Guido, 2016).

In this work, we use a specific technique named the Synthetic Minority Over-sampling Technique (SMOTE), in which the minority class is over-sampled by creating "synthetic" examples

(Fernández, Garcia, Herrera, & Chawla, 2018). In other words, the minority class is over-sampled by taking each minority class sample and introducing synthetic examples along the line segments joining any/all of the k minority class nearest neighbors.

The main objective is to create new elements to solve the imbalance problem. We implement this technique only in the training stage because we want that the model learns effectively and test it with the real imbalance data in the test set. The SMOTE implementation is through scikit-learn library.[2]

[2] For further information, please visit: `https://imbalanced-learn.org/dev/references/generated/imblearn.over_sampling.SMOTE.html`.

Chapter 4

Methodology

The primary objective of this section is to provide a comprehensive description of the methods, techniques, and procedures employed during the study. By outlining the research design, data collection methods, and analysis techniques, the empirical strategy section allows readers to understand how the research was conducted and the results obtained.

4.1 Red Flags

To achieve our research objective, an extensive examination of the existing literature has been conducted to identify prevalent red flags suggested by various organizations, aiming to detect potential indications of corruption risk at the contract level. Moreover, two supplementary databases have been generated by integrating the comprehensive datasets compiled by Falcón-Cortés et al. (2022) and IMCO (2022). These databases incorporate novel red flags to facilitate a comparative analysis between the original database and the augmented databases. This comparative evaluation aims to ascertain the performance disparities exhibited by distinct machine learning models when utilized with and without incorporating the proposed variables as red flags.

This approach is mainly because red flags are predominantly employed in mitigating the potential risks associated with diverse types of misconduct, including corruption in public procurement. Moreover, following Ferwerda et al. (2017), numerous scholarly papers addressing the issue of corruption have advocated for using corruption indicators or red flags to distinguish between corrupt and non-corrupt public procurement processes.

To improve the performance of the machine learning model, we propose additional red flags based on the corruption risk factors literature in Mexico. To most adequately measure corruption risk, measurement is carried out on the level of individual contract awards. Besides, measurement is carried out at the organizational level to link procurement data and red flags to company or bureaucratic characteristics.

As mentioned earlier, we employ two database approaches: Normal IMCO, which contains a more significant number of variables related to the characteristics of the public procurement process. This abundance of data allows for identifying a higher number of red flags compared to RPS Normal. The primary objective of this approach is to demonstrate how the availability of more data can facilitate research, analysis, and measurement of potential corruption risks by generating additional red flags in the public procurement field. These proposed red flags are described as follows:

It is well-known that the lack of competition represents a risk signal in public procurement. This is because competition allows maximizing the value of money (Fazekas et al., 2016). Consequently, two red flags are created, one to measure the number of contracts without competition awarded by a government agency and the other to measure the number of contracts without competition awarded to a supplier.

- Fraction of single-bidder contracts by government agency per year:

$$NOPRGOV_{it} = \frac{\sum_m^M CPR_{it}}{\sum_m^M TNPG_{it}} \tag{4.1}$$

NOPRGOV refers to the proportion of non-open biddings overall procedures concluded by a government agency represented by *i* over a year represented by *t* (where *m* is the first procedure and *M* is the last procedure). This is calculated by dividing the number of procedures following the single-bidding procedure of the *ith* public government agency over period *t* (CPR) by the total number of procedures concluded by the *ith* public organization over period *t* (TNPG).

- Fraction of single-bidder contracts by supplier per year:

$$NOPRSUP_{it} = \frac{\sum_{m}^{M} CPR_{it}}{\sum_{m}^{M} TNPS_{it}} \tag{4.2}$$

NOPRSUP refers to the proportion of non-open biddings overall procedures concluded of a supplier represented by *i* over a period represented by *t* (where *m* is the first procedure and *M* is the last procedure). This is calculated by dividing the number of procedures following the single-bidding procedure of a supplier represented by *i* over a period represented by *t* (CPR) by the total number of procedures concluded by the supplier *i* over period *t* (TNPS).

- Percentage of split contracts per year:

There exist different ways to ensure the hiring of a specific company. One common way is to split a contract with a considerable amount of money into multiple contracts with a lower amount. According to the work of IMCO "Anexo Metodológico: Mapeando la Corrupción" (2019), this is due to construction and procurement laws allowing contracting through exception processes when the purchases to be made are small. Generally, the contracts resulting from splitting a big contract are awarded to the same company on close dates, even on the same day. Therefore, a contract with a risk of having been sliced would be one that has been executed through an exception process at the same time or very close to other contracts awarded to the same company. The calculation of the red flag is as follows:

$$PEROFSPLITC_{it} = \frac{\sum_{m}^{M} PER_{it}}{\sum_{m}^{M} TNPS_{it}} \tag{4.3}$$

Where PEROFSPLITC is calculated by dividing the number of biddings concluded and identified with the same supplier in the same week of the year with a supplier represented by *i* in a specific period represented with *t* (PER)(where *m* is the first procedure and *M* is the last procedure) by the total number of procedures concluded by the supplier *i* over the year *t* (TNPS).

- Frequency of Contracts Won:

  One of the critical factors in detecting government favoritism towards a supplier is the number of contracts the company has won within a specific time range. It is expected that the higher the number of contracts won, the greater the likelihood that a company is among the preferred ones by the government or agency. However, as reported by (*Anexo Metodológico: Mapeando la Corrupción*, 2019), it is no easy task because it is necessary to compare the success of the supplier in question with the other suppliers in the market. A government agency's favorite suppliers should substantially perform better than the rest to indicate favoritism. Thus, a proper measure would be to consider the number of contracts of the analyzed supplier based on the contracts won by the other suppliers.

  Therefore, IMCO proposes a way to calculate the frequency of the number of contracts of a supplier:

$$FCW_{itd} = \frac{\sum_{n}^{N} Cn_{itd}(100)}{Max(\sum_{n}^{N} Cn_{itd}))} \tag{4.4}$$

  Where FCW refers to the standardized contract's frequency of the supplier represented by *i* over a period represented by *t* awarded by the government agency *d* (where *m* is the first procedure and *M* is the last procedure), this is equal to the number of contracts that the supplier *i* had in that period *t* and with that government agency *d* (Cn) multiplied by 100,

divided by the number of contracts that the company *i* that won the most contracts in the specified period *t* with the government agency *d*.

- Amount contracted by the supplier

  To analyze favoritism, it is important take into account the number of contracts won by a supplier in function of the amount of money of the awarded contracts. (*Anexo Metodológico: Mapeando la Corrupción*, 2019). This is because the most contracted or successful company may not necessarily be the one that receives the highest amount of money.

  To address such situations, it is necessary to have a standardized measurement of the money received by companies within the same period and from the same agency. A higher amount received would imply a greater risk of being a favored company. This measurement should be standardized based on the performance of all other companies under the same conditions resulting in a variable between 0 and 100; therefore, to calculate this red flag, it is proposed:

$$MT_{itd} = \frac{\sum_{m}^{M} Cm_{itd}(100)}{Max(\sum_{n}^{N} Cm_{itd}))} \tag{4.5}$$

  Where MT refers to the standardized total amount of contracts for a company represented by *i* in a period represented by *t* awarded an agency represented by *d* (where *m* is the first procedure and *M* is the last procedure). It is calculated as the sum of the contract amounts for the company within the same period and awarded by the same agency (expressed as the summation of the amount m for contracts *c* of the company *i* at period *t* for agency *d*), multiplied by 100, divided by the sum of the contract amounts for the company that received the highest amount of money during the same period by the same agency (expressed as the maximum summation of the amount *m* for contracts *c* of all companies I at period *t* in agency *d*).

- Percentage of the amount allocated to exception processes.

  As we said earlier, corruption arises when there is a lack of competition. In public procurement, the most apparent signal of this is when a government agency awards contracts by single-bidder an invitation to only three companies. There exist cases when government agencies award through exceptional processes based on fixed reasons accompanied by a timely and legal justification. However, it remains a risk factor.

  Given the above, we calculate the percentage of the amount allocated to exception processes as follows:

$$PEREXCPRO_{it} = \frac{\sum_{m}^{M} SumExcPro_i}{Max(\sum_{n}^{N} TotSumProc_i))} \tag{4.6}$$

  PEREXCPRO is calculated as the summary of the amount allocated to exception processes (single-bidder and restricted invitation) for a company represented by *i* in a specific period represented by *t* (where *m* is the first procedure and *M* is the last procedure) divided by the total amount allocated to all types of procedures by the company *i* in a specific period *t*.

- Percentage of the amount allocated to sanctioned companies.

  In the case of Mexico, there exist two main organizations that can sanction companies dedicated to public procurement processes: The Tax Administration Services with the list of taxpayers allegedly non-existent transactions[3] and the list of sanctioned suppliers and contractors by the Secretariat of the Civil Service.[4] These two list shows the taxpayers who allegedly simulate transaction through the issuance of invoices of digital tax receipts and legal entities or individuals sanctioned by Internal Control Bodies and with the prohibition to submit proposals or enter into contracts with federal government agencies,

[3] For further information, please visit: `http://omawww.sat.gob.mx/cifras_sat/Paginas/datos/vinculo.html?page=ListCompleta69B.html`.

[4] For further information, please visit: `https://directoriosancionados.apps.funcionpublica.gob.mx/SanFicTec/jsp/Ficha_Tecnica/SancionadosN.htm`.

entities of the Federal Public Administration, and State Governments.

The fact that these contractors or suppliers are identified and have a history of being sanctioned results in all public procurement procedures related to these contractors or suppliers carrying a potential risk of corruption. Given the above, IMCO (2023) proposes in his Corruption Risk Index the Amount allocated to sanctioned companies as a way to measure the risk of corruption.

$$PSANCCOMP_{it} = \frac{\sum_{m}^{M} SumSancPro_i}{Max(\sum_{n}^{N} TotDepSpent_i))} \tag{4.7}$$

PSANCCOMP is calculated by dividing the sum of the amount of money spent in public procurement procedures by a government agency represented by *i* with sanctioned suppliers or contractors in a specific period represented by *t* (SumSancPro) (where *m* is the first procedure and *M* is the last procedure), divided by the total expenditure for each government agency *i* in specific period *t* (TotDepSpent).

4.2 Machine Learning

The literature shows different methods for detecting possible corruption risks; however, it is appropriate to implement machine learning as an innovative technique for this work because there is a favorable data environment with many input variables.

Within the machine learning field, there exist three main categories: supervised learning, unsupervised learning, and reinforcement learning. These categories provide a high-level overview of the different machine-learning approaches. However, we are going to focus on supervised learning. In this approach, the machine learning model is trained using labeled data in this category, where the input features and their corresponding target outputs are provided. Supervised machine learning models can be highly effective because we have labeled data of each public

procurement procedure.

We decided to use a set of supervised machine learning, which are among the most cited in the literature. Accordingly, with the literature review, authors in the field have used these models, demonstrating promising results.

According to Muller in his book 'Introduction to machine learning with Python: a guide for data scientists' (2016), the definitions of these models are:

- XGBoost.

  XGBoost stands for eXtreme Gradient Boosting, and the library implements the gradient boosting decision tree algorithm. Gradient-boosted decision trees are among the most powerful and widely used models for supervised learning due to execution speed and model performance (Brownlee, 2016). XGBoost uses boosting to learn from the errors committed in the preceding trees.

  The evidence shows that XGBoost dominates tabular datasets in classification, and this is one of the reasons to use it.

  Gradient Boosting is an approach where new models are created that predict the residuals of errors of prior models and then added together to make the final prediction. For instance, Velarde et al. (2023) evaluate XGBoost performance in fraud detection, examining the principles and performance with different percentages of positive samples (50, 45, 25, and 5 percent); they conclude that XGBoost recognition performance improves as more data is available, and deteriorates detection performance as the databases become more imbalanced. Some of the advantages and disadvantages of this model are:

  - Advantages:

    * High Performance: XGBoost is optimized for performance and efficiency. It can handle large datasets and process them faster compared to other gradient-boosting implementations.

    * Flexibility: It can be used for both classification and regression tasks
    * Imbalanced Data: XGBoost can handle imbalanced datasets by providing options to assign different weights to different classes, making it useful for tasks with rare classes.
  - Disadvantages:
    * Memory Usage: XGBoost can consume a significant amount of memory, especially when working with large datasets.
    * Black Box Nature: Like other ensemble methods, XGBoost's predictions can be hard to interpret due to its ensemble of decision trees.

- ExtraTreesClassifier.

  The ExtraTreesClassifier stands out as an Extremely Randomized Trees. This is a machine-learning algorithm that belongs to the family of decision tree ensembles. It is similar to the Random Forest classifier but with some key differences. The ExtraTreesClassifier, as a Random Forest algorithm, builds multiple decision trees using random subsets of features and training examples, a technique known as Bagging. These trees are constructed independently, and their predictions are combined through majority voting or averaging (Sarang, 2023).

  However, the Extremely Randomized Trees take the randomization one step further:

  - It not only selects a random subset of features but also chooses a random threshold for each feature, leading to an even higher level of randomness.
  - Also, from the splitting strategy approach, this algorithm uses random splits for all features under consideration. In other words, it does not evaluate different splits to find the best one.
  - The number of trees creates is fewer to achieve a similar performance to Random Forest.

- Since ExtraTreesClassifier takes the randomization one step further, often results in higher variance, which could lead to overfitting on smaller datasets.

Some of the advantages and disadvantages of this model are:

- Advantages:
  * Unlike traditional decision trees or even Random Forests, ExtraTreesClassifier selects a random subset of features for splitting at each node. This randomness can help prevent overfitting and improve generalization.
  * Reduced Overfitting: By employing both bootstrapping (random sampling with replacement) and random feature selection, ExtraTreesClassifier can reduce the risk of overfitting, making it more robust to noisy data.
- Disadvantages:
  * Black Box Nature: Like other ensemble methods and decision trees, the predictions of ExtraTreesClassifier can be challenging to interpret, especially when dealing with a large number of trees.
  * Hyperparameter Tuning: Similar to other ensemble algorithms, finding the optimal hyperparameters for ExtraTreesClassifier can be time-consuming and requires careful experimentation.

- K-Nearest Neighbors.

  K-Nearest Neighbors (KNN) consist of finding the closest data points in the "nearest neighbor". It is a non-parametric algorithm in the sense that it makes no assumptions about the underlying data. We can set arbitrarily the number of neighbors to take into account. This is where the name of the k-nearest neighbor algorithm comes from. Accordingly to Sarang(2023), some of the disadvantages and advantages of this algorithm are

  - Advantages:

    - Simple to implement
    - Robust to the noisy training data
    - Can be more effective for large datasets
  - Disadvantages:
    - An appropriate selection of K value can be tricky
    - Computation cost is high as you need to calculate the distance between the unknown point and all other points in the entire dataset

- Logistic Regression

  Despite the name, logistics is a classification algorithm because, unlike linear regression, where predictions are continuous, logistic regression is discrete. This algorithm is achieved by passing the output of the linear regression through an activation function that maps the real numbers to either 0 or 1, some of the disadvantages and advantages of this algorithm are :

  - Advantages
    - Logistic regression is easy to understand and interpret.
    - Logistic regression works well with small datasets and doesn't require a large amount of computational resources.
  - Disadvantages
    - Logistic regression assumes a linear relationship between the features and the log-odds of the target variable. This means it can only model linear decision boundaries. For more complex relationships, logistic regression might not perform well.
    - As mentioned earlier, logistic regression assumes a linear relationship between features and the target. It may struggle with capturing non-linear patterns present in the data, which can lead to suboptimal performance.

- Random Forest

  Random forests are an example of an ensemble method, meaning one that relies on aggregating the results of a set of simpler estimators (VanderPlas, 2016). Decision trees are extremely intuitive ways to classify or label objects: you simply ask a series of questions designed to zero in on the classification Random forests get their name from injecting randomness into the tree building to ensure each tree is different. There are two ways in which the trees in a random forest are randomized: by selecting the data points used to build a tree and by selecting the features in each split test (Müller & Guido, 2016):

  - Advantages
    - High Predictive Accuracy: Random Forest generally provides high predictive accuracy due to the aggregation of multiple decision trees.
    - Reduced Overfitting: By aggregating predictions from multiple trees and using techniques like bagging (bootstrap aggregating) and random feature selection, Random Forest is less prone to overfitting compared to single decision tree
  - Disadvantages
    - Complexity: The final model is an ensemble of multiple decision trees, which can be complex and challenging to interpret, especially when the number of trees is high
    - Scalability: Random Forest might not scale well to extremely large datasets, as constructing multiple trees can be computationally intensive.

4.3 Model Building Process

Remember that Machine Learning can be categorized into two main types: supervised and unsupervised learning, and in this work, we use supervised learning. This means that we have labels

associated with the public procurement procedures regarding identifying public procurement procedures with possible risks of corruption and otherwise.

As we mentioned earlier, in the case of Mexico, there exist two leading organizations that can sanction companies dedicated to public procurement processes:

- The Tax Administration Services a.k.a SAT with the list of taxpayers' allegedly non-existent transactions.
- The Secretariat of the Civil Service with the list of sanctioned suppliers and contractors by the internal control.

A procurement procedure is flagged with one if it meets at least one of these two conditions:

- The supplier is suspected of engaging in fictitious transactions by issuing invoices or tax receipts by the Tax Administration Service.
- The public institution entity or the supplier is found in the Directory of Bidders, Suppliers, and Contractors sanctioned with prohibition to submit proposals or enter into contracts with federal public administration departments, entities, and state governments.

Otherwise, it is flagged with 0, which means the procedure is not related to the risk of corruption.

This implies that the training set we feed each algorithm includes the desired solutions, called labels.

Once the nature of the problem has been addressed, it is crucial to delve into the series of steps constituting a workflow for the training and testing the selected machine learning models. Different approaches exist to explain the stages that compound this machine-learning workflow.

This work is based on specific books popular in data science (Müller Guido, 2016 ;VanderPlas, 2016 ; Géron, 2022 ; Sarang, 2023):[5]

- Data Preparation

  The Data from both datasets (Normal MCO and Normal RPS) comes with null values and duplicate observations. It is crucial to solve this issue because machine learning algorithms do not accept null values and because it may lead to the algorithm not learning correctly. After the primary processing, we focus on the categorical data. This is because it is crucial to ensure that the algorithm interprets the data accurately. One-hot encoding is a method where each variable is converted in as many 0/1 variables as there are different values, so we proceed with the implementation. Finally, some columns may have different maximum-minimum ranges, which can influence the model performance skewing the model's prediction. So to solve this, we delete the outliers outside the interquartile range.

- Exploratory Data Analysis

  One of the most critical steps is the Exploratory Data Analysis (EDA) because it enables the understanding of the data. To implement Data Understanding, scatter plots, histograms, and bar charts are used to understand the data and extract insights. In other words, it serves as a crucial preliminary step before training machine learning models because it sets the foundation for making informed decisions throughout the model development process, leading to more accurate, robust, and reliable models.

- Feature Engineering

  Once a proper understanding of the data is obtained, feature engineering is implemented. Feature engineering means excluding, creating, and selecting certain variables for the

[5] For the implementation of the aforementioned work, Python was employed alongside its relevant machine learning libraries. The utilization of Python allowed for the integration of various machine learning algorithms and techniques, enabling the analysis and classification of public procurement data using red flags. This approach facilitated the creation of predictive models and the assessment of potential corruption risks within the context of public procurement.

models. There are a variety of techniques at this stage, such as univariate selection or recursive feature elimination. For feature selection, Principal Component Analysis to implement dimensionality reduction in cases with excessive multicollinearity. In our case, we focus on the creation and validation of the red flags proposed.

- Deciding on Model Type

  It is crucial to understand and decide which algorithm is most suitable for the dataset and problem to solve. As previously indicated, we are implementing supervised learning classification because the algorithms learn from labeled training data to predict a new, unseen observation. In this approach, popular algorithms exist, including Decision Trees, Random Forests, and Support Vector Machines, among others. In this particular instance, we employ a famous library in Python named LazyPredict.[6] This library helps build a lot of basic models without much code and helps understand which models work better without any parameter tuning. Furthermore, the literature review of various implementations focused on the subject matter of this work revealed the machine learning models that were most commonly employed and yielded superior outcomes.

- Training and testing data.

  In this step, we assess the model's performance of each model. The training and testing process involves the following steps:

  - Data Splitting: In this step, each dataset is divided into two subsets: the training set and the testing set. The training set is used to train the model, while the testing set is used to evaluate its performance.
  - Model training: The algorithm is fed the training data along with the corresponding target values (labels). Then, the algorithm learns the relationship between the features and the target values, adjusting its internal parameters to minimize the prediction errors.

[6] For further information, please visit: `https://pypi.org/project/lazypredict/`.

- Model testing

  Once the model is trained, it is tested on a separate testing dataset that it has never seen before. The model predictions are compared to the actual target values to assess its performance.

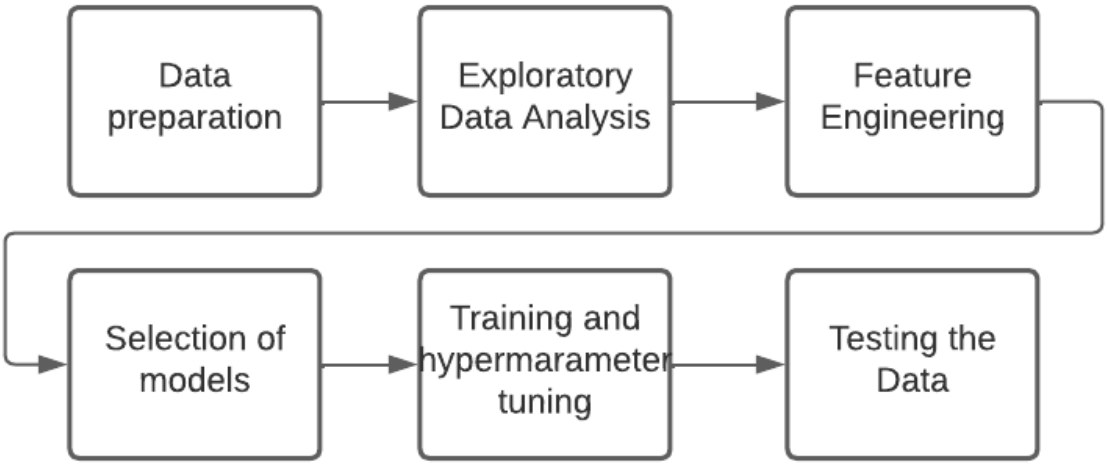
- Hyperparameter tuning.

  A hyperparameter is a parameter of a learning algorithm not off the model. This must be set prior to training and remains constant during training. Each algorithm has specific hyperparameters, and it is essential to know if they are a good fit to the data. To know this, there exist different approaches, including Grid Search, Random Search, and Bayesian optimization, to name a few. The hyperparameter tuning is implemented through a process of systematically searching for the best combination of hyperparameters that results in the optimal performance of a machine learning model.

  Since the main objective is to measure the performance of various models with and without the proposed red flags, rather than identifying the best-performing model, this approach was implemented using the GridSearchCV[7] library on the databases without the proposed red flags. The same set of hyperparameters was then applied to the augmented databases.

[7] For further information, please visit: `https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html`.

Figure 4.1: Methodology



Source: Own elaboration.

Chapter 5

Results

Before addressing the main findings and results of this work, it is important to understand how a machine-learning classification problem is evaluated. A good way to evaluate a model is to use a confusion matrix. The general idea is to count the number of times instances of class A are classified as class B for all A/B pairs. To compute the confusion matrix, we need to have a set of predictions to compare them to the actual targets. Table 5.1 shows that each row of the confusion matrix represents an actual class, while each column represents a predicted class.

Table 5.1: Confusion Matrix

	Predicted - 0	Predicted - 1
Actual - 0	True negatives	False positives
Actual - 1	False negatives	True positives

where:

- True Positives: It represents the number of instances correctly predicted as positive by the model. In other words, it is the number of positive cases correctly classified.
- True Negatives: It represents the number of instances correctly predicted as negative by the model. It is the number of negative cases correctly classified.

- False Positives: It represents the number of instances incorrectly predicted as positive by the model. It is the number of negative cases wrongly classified as positive.
- False Negatives: It represents the number of instances incorrectly predicted as negative by the model. It is the number of positive cases wrongly classified as negative.

The confusion matrix provides important information, and other parameters stem from it. An interesting one is the accuracy of the positive predictions; this is called the precision of the classifier. The precision tries to answer what proportion of positive instances was correct:

$$Precision = \frac{TruePositives}{TruePositives + FalsePositives} \tag{5.1}$$

Precision, commonly used in conjunction with the metric called recall or sensitivity, plays a vital role in evaluating classifiers. The recall represents the ratio of correctly identified positive instances by the classifier. Recall seeks to address the question of how accurately the classifier pinpointed positive instances:

$$Recall = \frac{TruePositives}{TruePositives + FalseNegative} \tag{5.2}$$

It is often convenient to combine precision and recall into a single metric called F1-Score. This metric is the harmonic mean of the precision and recall, and it is often used when we want to compare two or more classifiers:

$$F1 - Score = 2 * \frac{Precision * Recall}{Precision + Recall} \tag{5.3}$$

Since we have an imbalanced dataset, choosing the right evaluation metric is essential. Accuracy is a standard metric representing the degree of closeness to the acquired result and true value: however, classification accuracy is inappropriate for imbalanced classification. High accuracy is

achievable by a no-skill model that only predicts the majority class. For this reason, we do not consider this metric. For precision, recall, and F1-score, the formula for macro average can be expressed as:

$$MacroAveragePrecision = \frac{(PrecisionClass1 + PrecisionClass2 + ... + PrecisionClassN)}{N} \tag{5.4}$$

$$MacroAverageRecall = \frac{(RecallClass1 + RecallClass2 + ... + RecallClassN)}{N} \tag{5.5}$$

$$MacroAverageF1{-}Score = \frac{(F1 - ScoreClass1 + F1 - ScoreClass2 + ... + F1 - ScoreClassN)}{N} \tag{5.6}$$

Where N is the number of classes taken into account.

Macro Average refers to the average performance across all instances. This is an adequate metric because it calculates metrics such as precision, recall, and F1-score by treating each class equally, regardless of its size or prevalence in the dataset. Stated differently, it does not consider imbalance because the resulting performance is a simple average over the classes, so every class is given equal weight independently of their proportion.

Once the results foundations have been laid, we can proceed with the results and findings of this work. Since this work aims to explore the capacity of the Red Flags proposed to identify possible corruption risks in public procurement, we test the Red Flags with different supervised machine learning models and with four different databases.

The results of these tests are summarized in Table 5.2 and Table 5.3; each table shows the performance of each machine learning model through specific parameters with the four datasets.

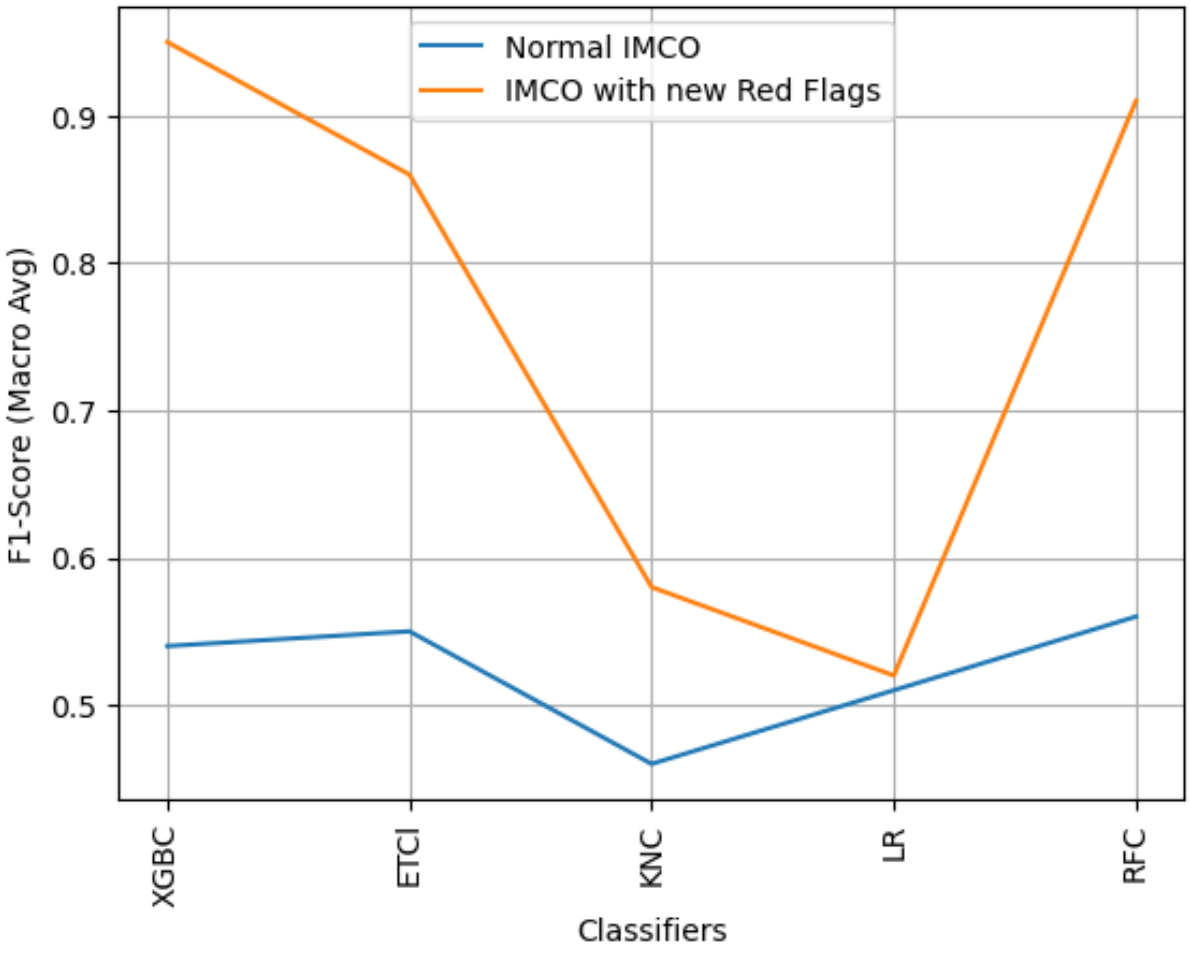
Table 5.2: Classification Report for different machine learning models with IMCO datasets

Model	Dataset	Parameter	Precision	Recall	F1-Score	Accuracy
XGBClassifier	Normal IMCO	0	1	1	1	1
		1	0.09	0.04	0.05	
		macro avg	0.55	0.52	0.53	
	IMCO with red flags	0	1	1	1	1
		1	0.94	0.69	0.80	
		macro avg	**0.97**	**0.85**	**0.90**	
ExtraTreesClassifier	Normal IMCO	0	1	1	1	1
		1	0.04	0.05	0.05	
		macro avg	0.52	0.53	0.52	
	IMCO with red flags	0	1	1	1	1
		1	0.91	0.50	0.64	
		macro avg	**0.95**	**0.75**	**0.82**	
K-Neasrest Classifier	Normal IMCO	0	1	0.78	0.88	0.78
		1	0	0.41	0.01	
		macro avg	0.5	0.59	0.44	
	IMCO with red flags	0	1	0.93	0.96	0.92
		1	0.00	0.21	0.01	
		macro avg	**0.50**	**0.57**	**0.49**	
Logistic Regression	Normal IMCO	0	1	0.98	0.99	0.98
		1	0.02	0.29	0.04	
		macro avg	0.51	0.63	0.51	
	IMCO with red flags	0	1	0.98	0.99	0.98
		1	0.03	0.36	0.05	
		macro avg	**0.51**	**0.67**	**0.52**	
RandomForestClassifier	Normal IMCO	0	1	1	1	1
		1	0.09	0.07	0.08	
		macro avg	0.55	0.53	0.54	
	IMCO with red flags	0	1	1	1	1
		1	0.96	0.57	0.72	
		macro avg	**0.98**	**0.79**	**0.86**	

Looking at Table 5.2, the following findings are observed:

- In terms of F1-Score in the macro average performance, all the models consistently demonstrate an improvement in their ability to classify instances adding the red flags proposed correctly:
  - The XGBClassifier model demonstrated a significant improvement of 0.37, increasing from an initial value of 0.53 to a final value of 0.90.
  - The ExtraTreesClassifier model demonstrated a significant improvement of 0.30, increasing from an initial value of 0.52 to a final value of 0.82.
  - The K-Nearest Classifier model showed an improvement of 0.05. increasing from an initial value of 0.44 to a final value of 0.49.
  - The Logistic Regression model experienced an improvement of 0.01, increasing from an initial value of 0.51 to a final 0.52. This is the model with the least improvement.
  - The RandomForestClassifier model demonstrated a significant improvement of 0.32, increasing from an initial value of 0.54 to a final value of 0.86
- The XGBClassifier, Random Forest, and Extra Trees Classifier models exhibit strong performance across most metrics in IMCO and RPS databases, adding the red flags proposed.
- Comparing the parameters between Normal IMCO and IMCO with the red flags for each supervised machine learning model, we can conclude that adding the red flags proposed can improve the ability to classify instances correctly. All of the above can be observed in Figure 5.1, where it can be seen how the line orange represents the IMCO database above line blue, which represents the Normal IMCO database.

Figure 5.1: F1-Score (Macro Average) Comparison of Classifiers with RPS datasets



Source: Own elaboration.

Now, moving on to the RPS datasets, the following findings are observed:

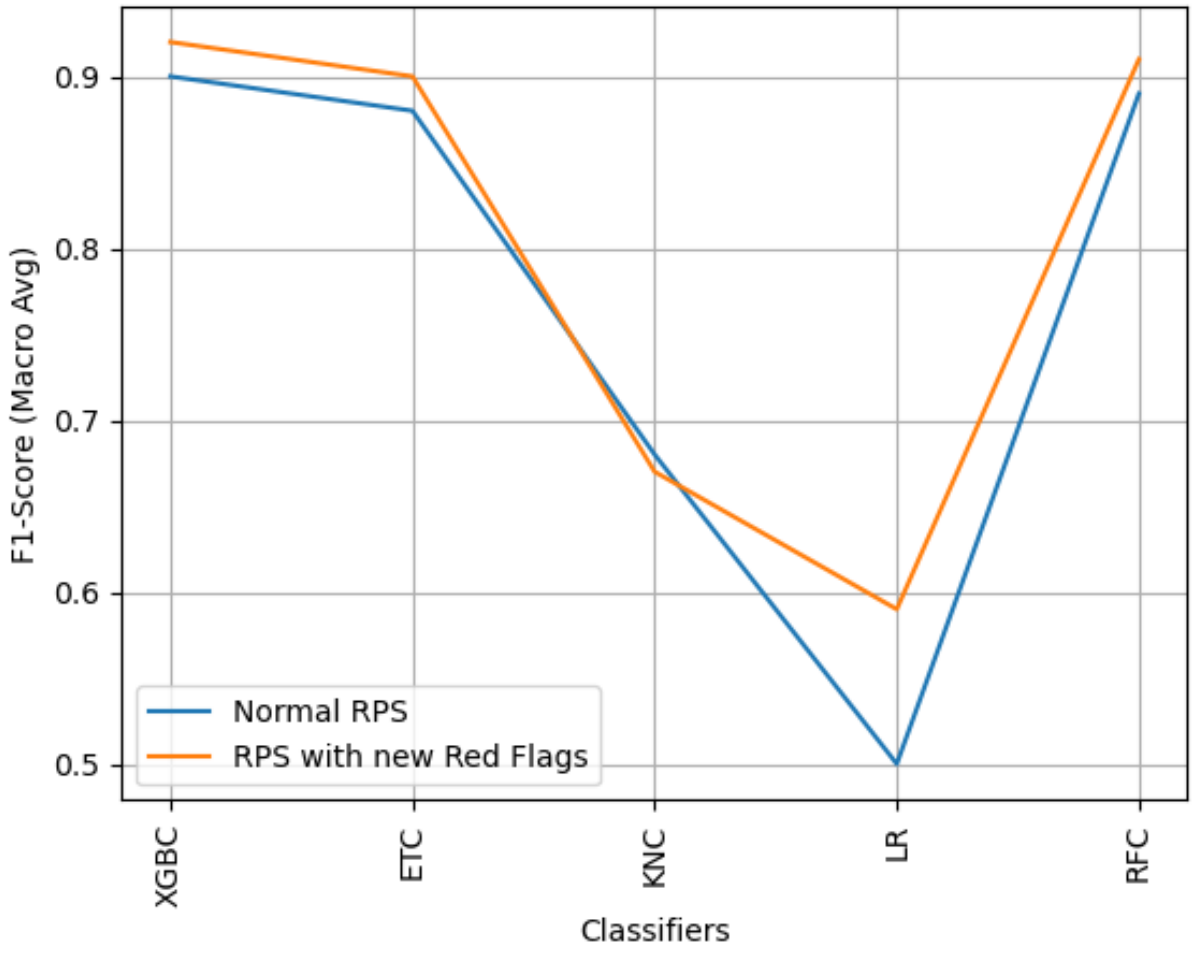
Table 5.3: Classification Report for different machine learning models with IMCO datasets

Model	Dataset	Parameter	Precision	Recall	F1-Score	Accuracy
XGBClassifier	Normal IMCO	0	0.98	1	0.99	0.98
		1	0.36	0.13	0.19	
		macro avg	0.67	0.56	0.59	
	IMCO with red flags	0	0.99	1	1	0.99
		1	0.98	0.74	0.85	
		macro avg	**0.99**	**0.87**	**0.92**	
ExtraTreesClassifier	Normal IMCO	0	0.98	0.98	0.98	0.99
		1	0.16	0.15	0.16	
		macro avg	0.57	0.57	0.57	
	IMCO with red flags	0	0.99	1	1	0.99
		1	0.97	0.67	0.79	
		macro avg	**0.98**	**0.84**	**0.9**	
K-Neasrest Classifier	Normal IMCO	0	0.98	0.84	0.90	0.83
		1	0.5	0.35	0.08	
		macro avg	0.51	0.60	0.49	
	IMCO with red flags	0	0.98	0.91	0.94	0.89
		1	0.05	0.21	0.08	
		macro avg	**0.51**	**0.56**	**0.51**	
Logistic Regression	Normal IMCO	0	0.98	1	0.99	0.98
		1	0.39	0.08	0.13	
		macro avg	0.68	0.54	0.56	
	IMCO with red flags	0	0.98	0.99	0.99	0.98
		1	0.40	0.25	0.31	
		macro avg	**0.69**	**0.62**	**0.65**	
RandomForestClassifier	Normal IMCO	0	0.98	0.99	0.99	0.97
		1	0.27	0.12	0.17	
		macro avg	0.62	0.56	0.58	
	IMCO with red flags	0	0.99	1	1	0.99
		1	0.98	0.70	0.82	
		macro avg	**0.99**	**0.85**	**0.91**	

Looking at Table 5.3, the following findings are observed:

- In terms of F1-Score in the macro average performance, all the models consistently demonstrate an improvement in their ability to classify instances adding the red flags proposed correctly:
  - The XGBClassifier model demonstrated an improvement of 0.33, increasing from an initial value of 0.59 to a final value of 0.92.
  - The ExtraTreesClassifier model demonstrated a significant improvement of 0.33, increasing from an initial value of 0.57 to a final value of 0.9.
  - The K-Nearest Classifier model showed a decline of 0.01, decreasing from an initial value of 0.68 to a final value of 0.67.
  - The Logistic Regression model experienced an improvement of 0.09, increasing from an initial value of 0.56 to a final 0.65.
  - The RandomForestClassifier model demonstrated an improvement of 0.33, increasing from an initial value of 0.58 to a final value of 0.91.
- The XGBClassifier, Random Forest, and Extra Trees Classifier models exhibit strong performance across most metrics in IMCO and RPS datasets, adding the red flags.
- Comparing the parameters between Normal RPS and RPS with red flags for each supervised machine learning model, we can conclude that adding the red flags proposed can improve the ability to classify instances correctly. All the above can be observed in Figure 5.2, where it can be seen how the line orange represents the IMCO database above line blue, which represents the Normal IMCO database.

Figure 5.2: F1-Score (Macro Average) Comparison of Classifiers for IMCO datasets



Source: Own elaboration.

Diving into the macro average parameters for the three best models between IMCO and RPS datasets, the following findings are observed:

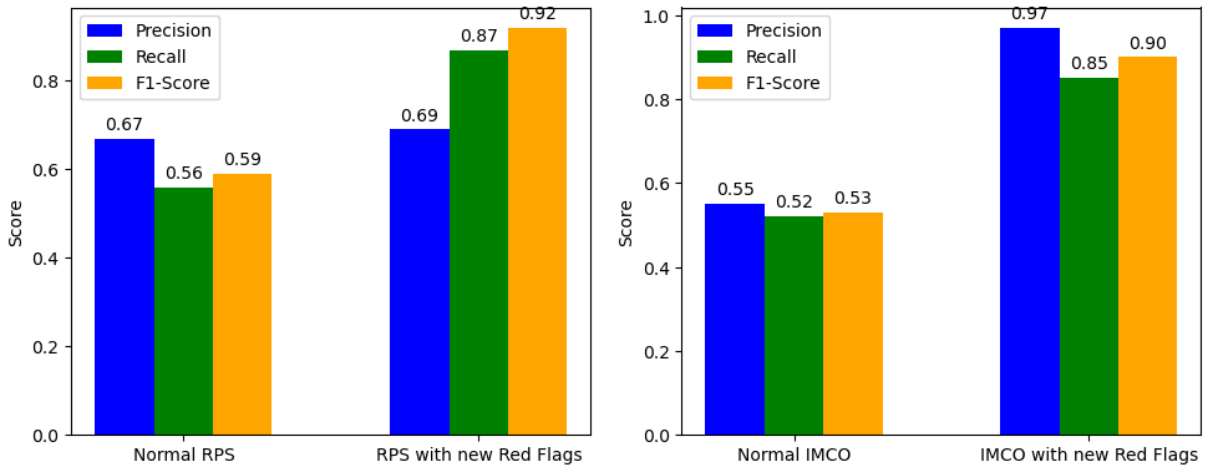
- We can observe that in both databases when comparing their normal version to their augmented version, there is an enhancement in performance metrics upon incorporating the proposed red flags.
- Taking into account the F1-score metric, we can observe that the improvement of the model is higher in the IMCO dataset than in the RPS dataset when adding the red flags.

The difference in F1-score is 0.33 in the RPS database, whereas, in the IMCO database, it is 0.37.

- Also, we can observe that the increase in precision is lower in the RPS databases compared to the IMCO databases.

The above can be observed in Figure 5.3, where the results of the best algorithm (XGBClassifier) are compared with the two datasets variations. :

Figure 5.3: XGBClassifier performance for macro average



Source: Own elaboration.

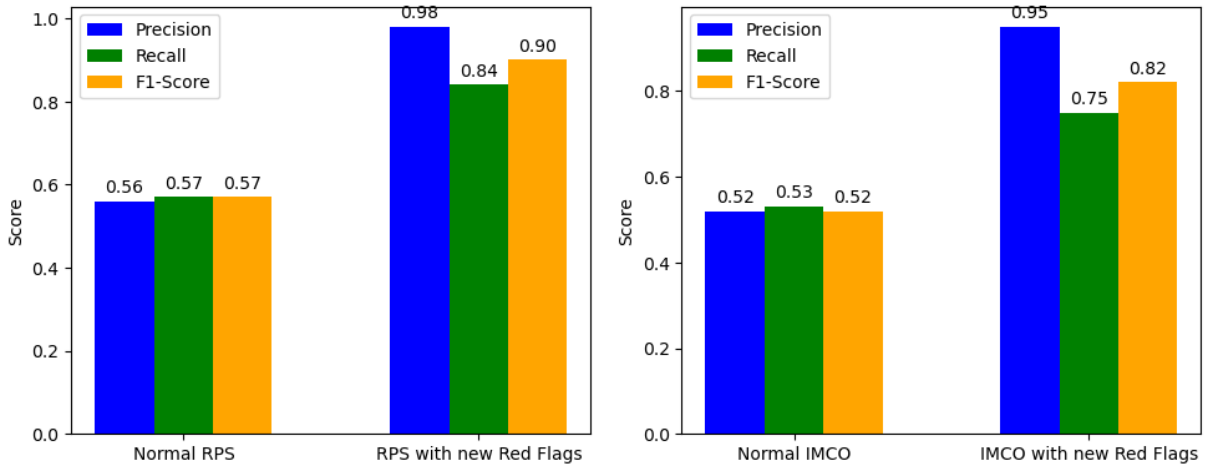
Secondly, moving on to the ExtraTreesClassifier model performance between IMCO and RPS datasets, the following findings are observed:

- We can observe that in both databases when comparing their normal version to their augmented version, there is an enhancement in performance metrics upon incorporating the proposed red flags.

- Taking into account the F1-score metric, we can observe that the improvement of the model is higher in the RPS dataset than in the IMCO dataset when adding the red flags. The difference in F1-score is 0.33 in the RPS database, whereas, in the IMCO database, it is 0.30.

- The model achieves higher Precision than Recall, with the highest value observed in IMCO dataset with an improvement of 0.46. The above can be observed in Figure 5.4:

Figure 5.4: ExtraTreesClassifier performance for macro average

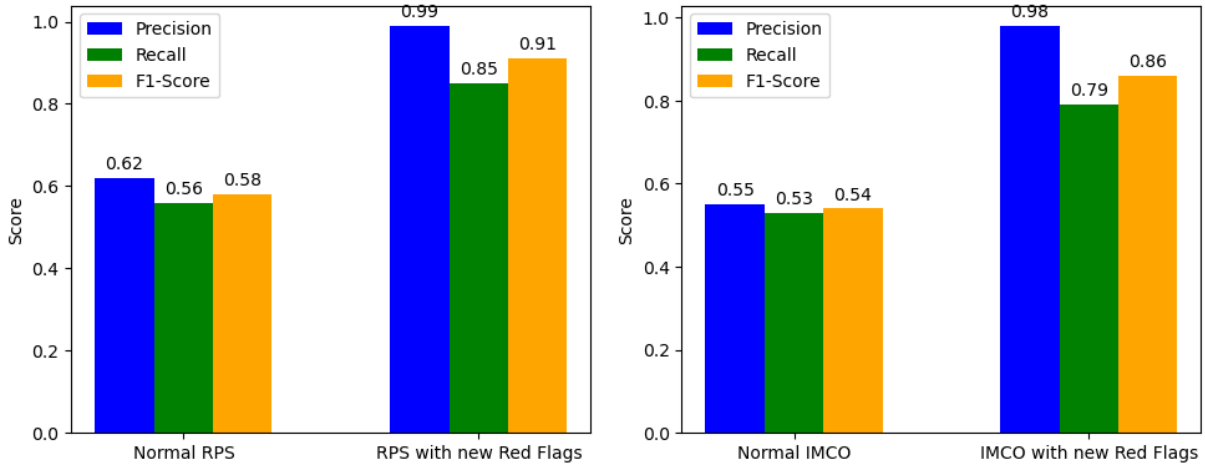


Source: Own elaboration.

Finally, moving on to the Random Forest Classifier model performance between IMCO and RPS datasets, the following findings are observed:

- We can observe that in both databases when comparing their normal version to their augmented version, there is an enhancement in performance metrics upon incorporating the proposed red flags.
- Taking into account the F1-score metric, we can observe that the improvement of the model is higher in the RPS dataset than in the IMCO dataset when adding the red flags. The difference in F1-score is 0.33 in the RPS database, whereas, in the IMCO database, it is 0.32.
- We can observe that the variation in the parameters is higher in IMCO dataset than RPS dataset when adding the red flags.
- The model achieves higher Precision than Recall, with the highest value observed in IMCO dataset with an improvement of 0.29. The above can be observed in figure 5.5:

Figure 5.5: Random Forest performance for macro average



Source: Own elaboration.

In general terms, when comparing the IMCO and RPS databases, we can draw the following conclusions:

- The difference in improvement by adding red flags is much smaller in the RPS database compared to the IMCO database.
- The models XGBClassifier, RandomForestClassifier, and ExtraTreesClassifier models exhibit the highest performance.

With the aim of obtaining a more reliable measure of the overall performance of the models, it was decided to train and measure the performance of each model ten times. This is a common practice in model evaluation to mitigate the impact of randomness on the results. This approach serves to facilitate a more accurate estimation of how the model will perform on future data and provides a solid foundation for the analysis of results, findings, and conclusions of this study.

Table 5.4: Mean of the macro average performance of ten machine learning models with RPS datasets

Model	**XGBClassifier**			**ExtraTreesClassifier**			**RandomForest**		
Dataset/Parameter	**Precision**	**Recall**	**F1-Score**	**Precision**	**Recall**	**F1-Score**	**Precision**	**Recall**	**F1-Score**
Normal RPS	0.9556	0.8618	0.9031	0.9529	0.8372	0.8863	0.9815	0.8332	0.8932
RPS with red flags	0.9846	0.8788	0.925	0.9751	0.8599	0.9091	0.9891	0.8536	0.9099

Upon observing Table 5.4 and comparing it with Table 5.3, it can be noted that the results of the macro average performance metrics of the RPS datasets remain consistent with minimal variations. The findings show a similarity between the two tables, indicating marginal changes in the performance metrics.

In the same manner, in the case of IMCO datasets and comparing Table 5.4 and Table 5.5, it is observed that the results of the macro average performance metrics remain consistent with minimal variations. The findings show a similarity between the two tables, indicating marginal changes in the performance metrics.

Table 5.5: Mean of the macro average performance of ten machine learning models with IMCO datasets

Model	**XGBClassifier**			**ExtraTreesClassifier**			**RandomForest**		
Dataset/Parameter	**Precision**	**Recall**	**F1-Score**	**Precision**	**Recall**	**F1-Score**	**Precision**	**Recall**	**F1-Score**
Normal IMCO	0.5352	0.5158	0.5217	0.5355	0.5404	0.5378	0.5467	0.5366	0.5410
IMCO with red flags	0.9661	0.8420	0.8945	0.9772	0.8204	0.8836	0.9856	0.7845	0.8591

Chapter 6

Discussion and future analyses

Analyzing the results and findings from a general perspective reveals important keys to address. One of these key findings is that some specific machine learning models outperform others in performance. To be precise, the models XGBClassifier, ExtraTreesClassifier, and Random Forest. This is mainly because these models are ensemble machine learning models that combine multiple base estimators to make decisions. Ensembles are methods that combine multiple machine-learning models to create robust models. These models have a more remarkable ability to capture complex relationships and patterns in the data, enabling them to better adapt to the characteristics of the datasets.

The application of machine learning models with red flags showcased promising results in detecting potential corruption risks within public procurement processes. The incorporation of red flags enhanced the models' ability to classify instances with corruption indicators accurately. This suggests that red flags offer valuable insights into irregularities and anomalies in procurement data, which machine learning models effectively leverage to make informed classifications.

The study's comparison of different machine learning algorithms provides a comprehensive perspective on their performance. Notably, certain algorithms exhibited higher precision, recall,

and F1-scores when coupled with red flags. This underscores the potential of specific algorithms to capitalize on the contextual information provided by red flags, enhancing their discriminatory power.

It's worth noting that the efficacy of red flags varied across different datasets. This variance could be attributed to the unique characteristics and contexts of each dataset, reinforcing the need for a tailored approach when implementing red flags for corruption detection. Furthermore, the comparison between datasets with and without red flags shed light on the discernible improvement achieved by incorporating this supplementary information.

From a public policy perspective, it is crucial to take advantage of the amount of information generated by the public administration to extract the most significant advantage of it. However, despite the efforts from different countries to implement rules and regulations to regulate public procurement, the question is how we can improve accountability and decrease discretion in the strategic use of public procurement records because corruption is a planned activity implemented in a way that is difficult to identify in the process because corrupt actors know how to use those rules and regulations o their advantage. To achieve the above, collective action between the government, the private sector, and the civil society is necessary. This is a critical issue to address because data can be used to lie and hide information about a public procurement process in spite of the existence or large quantities of data available as open data. This can be integrated with the idea of "garbage in, garbage out" used in data analysis to emphasize the importance of input quality in producing meaningful analysis and insights.

Conversely, another critical aspect of the public procurement open data policy in Mexico is identifying the main objective or purpose of the data being created. Therefore, it is important to understand which characteristics of the public procurement process must be gathered and in what format they should be well-planned and applied.

Governments aiming to implement artificial intelligence policies in public procurement face several critical considerations:

Firstly, it is imperative to establish policies and regulations governing the extent to which AI can be utilized in corruption detection. Decision-makers responsible for sanctioning or identifying risks in public procurement must possess a clear understanding of how these tools are employed without any legal or administrative ambiguities.

Secondly, personnel training is of paramount importance. Ensuring that staff members are adequately trained and educated is essential for effective decision-making using such methods.

Furthermore, one significant challenge for governments seeking to leverage AI in decision-making processes lies in the development of legal frameworks. Legislation against corruption should be enacted in a way that legitimizes the results of these efforts, making them admissible as evidence in administrative and judicial proceedings. This entails integrating AI technology into the laws, rules, and regulations governing the functions and responsibilities related to the identification and investigation of corruption.

The establishment of a legal framework is imperative for governments to incorporate AI technology into their efforts to combat corruption effectively. This framework will provide both an opportunity and a challenge in terms of integrating this technology into the legal, regulatory, and normative aspects of functions related to the identification and investigation of corruption.

Chapter 7

Conclusions

In conclusion, this study delved into the critical realm of public procurement and corruption detection using machine learning techniques. Public policies are implemented through public contracts, making it essential to ensure their efficacy and impact. By analyzing comprehensive data from public policies and procurement contracts, it becomes possible to enhance accountability and transparency.

Public procurement, constituting a substantial portion of a country's GDP, is susceptible to corruption, which can undermine economic activities and hinder development. Corruption within this domain can take various forms, such as bribery, biased bidding processes, or insider information utilization. This study recognized the need for innovative approaches to corruption detection, emphasizing the limitations of traditional indicators and the potential of red flags.

The utilization of red flags, along with machine learning, offers a promising avenue to detect and prevent corruption in public procurement. By identifying patterns and anomalies in procurement data, machine learning models can provide valuable insights into potential corrupt activities. This study acknowledges the complexity of the corruption phenomenon, driven by its secretive and illicit nature. However, the adoption of new methodologies, like the incorporation

of red flags, represents a proactive step toward addressing corruption.

In the context of Mexico, where corruption remains a significant challenge, this study's relevance is evident. Public procurement records from platforms like CompraNet provide a rich source of data for analysis. The research highlighted the potential benefits of implementing machine learning models with red flags, demonstrating their ability to classify procurement processes with potential corruption risks accurately.

While the study's primary objective was to assess the effectiveness of new red flags in enhancing machine learning algorithms' performance, it also emphasized the need for continued research and improvement in corruption detection methodologies. By embracing data-driven approaches, like those presented in this study, governments and organizations can make significant strides in combatting corruption and promoting transparency in public procurement processes.

In conclusion, this work presents a series of public policy recommendations for Latin American countries to harness this technology in the fight against corruption:

- Foster Collaborative Environments: It is essential to cultivate collaborative relationships among different levels of government and various public, private, and civil society entities. This collaboration should lead to the development of a unified policy that designs systems for sharing strategic anti-corruption data.
- Invest in Data Quality: Investing in data quality is crucial. Data should be standardized and made available in open data formats. The effectiveness of any artificial intelligence solution is heavily reliant on the quality and accessibility of information. This also entails investing in the interoperability of different systems.
- Long-term Perspective: Finally, it is essential to convey to decision-makers that efforts like these do not hinge on a single decision but on a multitude of public policy decisions at different levels and areas, all focused on a specific goal. Such projects should be viewed with a long-term perspective. Each public policy decision should contribute to and support

the long-term objectives. It is a coordinated endeavor that unfolds over time, requiring budget allocations, long-term policies, and the cultivation of capacity over the long run.

Appendix A

RPS and IMCO Databases

Table A.1: Normal RPS Variables

Variable	**Description**
Government Order · GO.APF · GO.GE · GO.GM	This variable indicates at which government level the public procurement procedure was implemented.
Procedure Character · PC.N · PC.I · PC.ITLC	Legal framework in which the public procurement procedure was implemented.

Continued on next page

Table A.1 – continued from previous page

Variable	**Description**
Contract Type · CT.OP · CT.S · CT.ADQ · CT.AR · CT.SLAOP	Type of services or commodities contracted.
Procedure Type · PT.AD · PT.I3P · PT.LP	Procedure by which the supplier won the contract.
Size · S.NOM · S.MED · S.PEQ · S.MIC · S.NA	Size of the supplier.
Year · 2013 · 2014 · 2015 · 2016 · 2017 · 2018 · 2019	Year in which the contract began.

Table A.2: Normal IMCO variables

Variable	**Description**
Fundamento.legal	Articles and clauses that, in accordance with the Law of Acquisitions, Leases, and Services, as well as the Law of Public Works and Related Services, contain the scenarios that government departments and entities must take into account when, under their responsibility, they choose not to proceed with the public bidding process and instead opt to enter into contracts through the methods of inviting at least three individuals or direct allocation.
Compra.Consolidada	Identify if the procedure's contract stems from a consolidated purchase. 1 = YES, it stems from a consolidated purchase. 0 = NO, it does not stem from a consolidated purchase.
Folio.en.el.RUPC	Identify whether the assigned reference number by the SFP (Federal Public Administration) exists for the individual or legal entity that was registered in the RUPC (Federal Register of Contractors and Suppliers) by a public entity with which they entered into a contract.
RFC.verificado.en.elSAT	A dichotomous variable that indicates 1 when the supplier's or contractor's Taxpayer Identification Number (RFC) is verified with the Tax Administration Service (SAT), or 0 when it is not verified.
exlusivo_mipymes	A dichotomous variable that indicates 1 when the procedure was exclusively conducted for SMEs (Small and Medium-sized Enterprises), and 0 when it was not exclusive.
testigo_social	A dichotomous variable that indicates 1 when the procedure had a social witness, and 0 when this figure was not present.

Continued on next page

Table A.2 – continued from previous page

Variable	**Description**
archivo_fallo	A dichotomous variable that identifies whether the purchasing process includes the document of the award decision report or not (1/0).
archivo_apertura	A dichotomous variable that identifies whether the purchasing process includes the document of the proposal opening minutes or not (1/0).
archivo_junta	A dichotomous variable that identifies whether the purchasing process includes the document of the clarification meeting or not (1/0).
archivo_convocatoria	A dichotomous variable that identifies whether the purchasing process includes the document of the invitation or not (1/0).
archivo_contrato	A dichotomous variable that identifies whether the purchasing process includes the document of relevant contract data or not (1/0).
missing_file	A dichotomous variable that identifies whether the contracting process is missing at least one document.
Spending	Contract amount excluding Value Added Tax in Mexican pesos.
Publicacion.EDCA	A dichotomous variable that identifies whether the purchasing process is published in Open Contracting Data Standard (1/0).
Sin.justificacion	A dichotomous variable that indicates 1 when direct awards lack legal basis, and 0 when it is included.
Publicacion.Tardia	Dichotomous variable that indicates 1 when the procedure's publication date is after the contract start date, and 0 when the publication date is before the start date.
Link.Funcional	Dichotomous variable that identifies whether the address of the announcement on Compranet works or not (1/0).

Continued on next page

Table A.2 – continued from previous page

Variable	**Description**
Carácter del procedimiento: · PC_I · PC_ITLC · PC_N · PC_OTHER	Legal framework in which the public procurement procedure was implemented.
Tipo de contratación: · CT_ADQ · CT_AR · CT_OP · CT_S · CT_SLAOP	Type of services or commodities contracted.
Tipo de procedimiento: · PT_AD · PT_CEEP · PT_I3P · PT_LP · PT_PC · PT_OTHER	Procedure by which the supplier won the contract.
Forma de Participación: · PF_ELE · PF_MIX · PF_PRE	Method of participation by the bidder.

Continued on next page

Table A.2 – continued from previous page

Variable	**Description**
Tamaño: · S_MED · S_MIC · S_NOMIPYME · S_PEQ	Size of supplier.
Año: · 2018 · 2019 · 2020 · 2021	Year in which the contract began.

References

Aldana, A., Falcón-Cortés, A., & Larralde, H. (2022). "A machine learning model to identify corruption in m\'exico's public procurement contracts." *arXiv preprint arXiv:2211.01478*.

Anderson, R. D., Kovacic, W. E., & Müller, A. C. (2011). "Ensuring integrity and competition in public procurement markets: a dual challenge for good governance." *The WTO Regime on Government Procurement: Challenge and Reform*, 681.

Anexo metodológico: Mapeando la corrupción. (2019). `https://cpcgto.com/wp-content/uploads/2020/11/Anexo-4-Metodologico-Mapeando-la-Corrupcion.pdf`. (Accessed on July 12, 2023)

Arief, H. A., Saptawati, G. P., & Asnar, Y. D. W. (2016). "Fraud detection based-on data mining on indonesian e-procurement system (spse)." In *2016 international conference on data and software engineering (icodse)* (pp. 1–6).

Bank, W. (2020). *Enhancing government effectiveness and transparency: The fight against corruption.* Author.

Bardhan, P. (1997). "Corruption and development: a review of issues." *Journal of economic literature*, *35*(3), 1320–1346.

Brownlee, J. (2016). *Xgboost with python: Gradient boosted trees with xgboost and scikit-learn*. Machine Learning Mastery.

Decarolis, F., & Giorgiantonio, C. (2022). "Corruption red flags in public procurement: new evidence from italian calls for tenders." *EPJ Data Science*, *11*(1), 16.

Domingos, S. L., Carvalho, R. N., Carvalho, R. S., & Ramos, G. N. (2016). "Identifying it purchases anomalies in the brazilian government procurement system using deep learning." In *2016 15th ieee international conference on machine learning and applications (icmla)* (pp. 722–727).

Dorn, N., Levi, M., & White, S. (2008). "Do european procurement rules generate or prevent crime?" *Journal of financial crime*, *15*(3), 243–260.

Falcón-Cortés, A., Aldana, A., & Larralde, H. (2022). "Practices of public procurement and the risk of corrupt behavior before and after the government transition in méxico." *EPJ Data Science*, *11*(1), 19.

Fazekas, M., & Tóth, I. J. (2016). "From corruption to state capture: A new analytical framework with empirical applications from hungary." *Political Research Quarterly*, *69*(2), 320–334.

Fazekas, M., Tóth, I. J., & King, L. P. (2013a). "Anatomy of grand corruption: A composite corruption risk index based on objective data." *Corruption Research Center Budapest Working Papers No. CRCB-WP/2013*, *2*.

Fazekas, M., Tóth, I. J., & King, L. P. (2013b). "Corruption manual for beginners:'corruption techniques' in public procurement with examples from hungary." *Corruption Research Center Budapest Working Paper no. CRCB-WP/2013*, *1*.

Fazekas, M., Tóth, I. J., & King, L. P. (2016). "An objective corruption risk index using public procurement data." *European Journal on Criminal Policy and Research*, *22*, 369–397.

Fernández, A., Garcia, S., Herrera, F., & Chawla, N. V. (2018). "Smote for learning from imbalanced data: progress and challenges, marking the 15-year anniversary." *Journal of artificial intelligence research*, *61*, 863–905.

Ferwerda, J., Deleanu, I., et al. (2013). "Identifying and reducing corruption in public procurement in the eu."

Ferwerda, J., Deleanu, I., & Unger, B. (2017). "Corruption in public procurement: finding the right indicators." *European Journal on Criminal Policy and Research*, *23*, 245–267.

Jain, A. K. (2001). “Corruption: A review.” *Journal of economic surveys*, *15*(1), 71–121.

Magakwe, J. (2022). “The root causes of corruption in public procurement: A global perspective.” In *Corruption-new insights.* IntechOpen.

Mencıa, E. L., Holthausen, S., Schulz, A., & Janssen, F. (2013). “Using data mining on linked open data for analyzing e-procurement information.” In *Proceedings of the first dmold: Data mining on linked data workshop at ecml/pkdd.*

Mizoguchi, T., & Van Quyen, N. (2014). “Corruption in public procurement market.” *Pacific Economic Review*, *19*(5), 577–591.

Modrušan, N., Rabuzin, K., & Mršic, L. (2021). “Review of public procurement fraud detection techniques powered by emerging technologies.” *International Journal of Advanced Computer Science and Applications*, *12*(2).

Mufutau, G. O., & Mojisola, O. V. (2016). “Detection and prevention of contract and procurement, fraud crtalyst to organization profitability.” *J Bus Manag*, *18*, 09–14.

Müller, A. C., & Guido, S. (2016). *Introduction to machine learning with python: a guide for data scientists*. " O’Reilly Media, Inc.".

on Drugs, U. U. N. O., & Crime). (2013). “Guidebook on anti-corruption in public procurement and the management of public finances.” *https://www.unodc. org/documents/corruption/Publications/2013/Guidebook_on_anti-corruption_in_public_procurement_and_the_management_of_public_finances. pdf*.

Rabuzin, K., & Modrusan, N. (2019). “Prediction of public procurement corruption indices using machine learning methods.” In *Kmis* (pp. 333–340).

Rakhel, T. M., & Putera, P. B. (2021). “Corruption in public procurement: A bibliometric analysis.” *COLLNET Journal of Scientometrics and Information Management*, *15*(2), 397–412.

Ralha, C. G., & Silva, C. V. S. (2012). “A multi-agent data mining system for cartel detection in brazilian government procurement.” *Expert Systems with Applications*, *39*(14), 11642–11656.

Riesgos de corrupción - IMCO. (2023). https://imco.org.mx/riesgosdecorrupcion/. (Accessed on July 12, 2023)

Rose, R., & Peiffer, C. (2015). *Paying bribes for public services: A global guide to grass-roots corruption*. Springer.

Sales, L., et al. (2013). *Risk prevention of public procurement in the brazilian government using credit scoring* (Tech. Rep.). OBEGEF-Observatório de Economia e Gestão de Fraude & OBEGEF Working Papers ....

Sales, L. J., & Carvalho, R. N. (2016). "Measuring the risk of public contracts using bayesian classifiers." In *Bma@ uai* (pp. 7–13).

Sarang, P. (2023). *Thinking data science: A data science practitioner's guide*. Springer Nature.

Shleifer, A., & Vishny, R. W. (1993). "Corruption." *The quarterly journal of economics*, *108*(3), 599–617.

Søreide, T. (2002). *Corruption in public procurement. causes, consequences and cures*. Chr. Michelsen Intitute.

Sun, T., & Sales, L. J. (2018). "Predicting public procurement irregularity: An application of neural networks." *Journal of Emerging Technologies in Accounting*, *15*(1), 141–154.

Tas, B. K. O. (2017). "Collusion detection in public procurement with limited information." *Available at SSRN 2929222*.

VanderPlas, J. (2016). *Python data science handbook: Essential tools for working with data*. " O'Reilly Media, Inc.".

Velarde, G., Sudhir, A., Deshmane, S., Deshmunkh, A., Sharma, K., & Joshi, V. (2023). "Evaluating xgboost for balanced and imbalanced data: Application to fraud detection." *arXiv preprint arXiv:2303.15218*.

Wang, Y. (2016). *Detecting fraud in public procurement* (Unpublished doctoral dissertation). State University of New York at Stony Brook.

World Bank. (2023). *Global public procurement database: Share, compare, improve.* https://www.worldbank.org/en/news/feature/2020/03/23/global

-public-procurement-database-share-compare-improve. (Accessed on July 12, 2023)

Zumaya, M., Guerrero, R., Islas, E., Pineda, O., Gershenson, C., Iñiguez, G., & Pineda, C. (2021). "Identifying tax evasion in mexico with tools from network science and machine learning." *Corruption Networks: Concepts and Applications*, 89–113.