Schneider dos Santos et al. *EPJ Data Science* (2025) 14:52
https://doi.org/10.1140/epjds/s13688-025-00569-3

EPJ Data Science
a SpringerOpen Journal

REVIEW

Open Access

Detection of fraud in public procurement using data-driven methods: a systematic mapping study

Everton Schneider dos Santos[1*], Matheus Machado dos Santos[2†], Márcio Castro[1†] and Jônata Tyska Carvalho[1†]

Handling Editor: Johannes Wachs

*Correspondence:
e.schneider.s@posgrad.ufsc.br
[1]Graduate Program in Computer Science, Federal University of Santa Catarina, Florianópolis, Santa Catarina, Brazil
Full list of author information is available at the end of the article
†Equal contributors

Abstract

The scientific literature dedicated to the detection of fraud in public procurement is vast, with several studies reporting the use of different methodologies to detect corruption. However, the literature still lacks a comprehensive study of the types of fraud being investigated and how data-driven techniques are being used to address this problem. This article aims to provide a better overview of how these techniques are used to detect corruption in public procurement. We systematically searched academic databases with the goal of finding papers that used data-driven techniques to predict or identify fraud in public procurement. We also performed a snowballing procedure to complement the database search with additional papers. 93 works were added to our study after screening and evaluation of more than 6000 papers. Relevant information was extracted from these papers to answer the research question defined during the planning phase. The results showed that most works use machine learning models to detect collusion and statistical analysis to detect instances of favoritism. Despite the promising results, there are some gaps that still need to be addressed. There is a lack of papers that employ the proposed methodologies in real-life systems to detect new cases of corruption. Another gap found is the lack of public available datasets, hindering the replication and dissemination of the proposed methodologies. The findings of our study contribute to a more comprehensive understanding of fraud detection in public procurement, pointing to areas for improvement and offering insights to researchers and institutions seeking to improve their processes.

Keywords: Survey; Corruption; Public sector; Government

1 Introduction

Public procurement is vulnerable to corruption and fraud, as in other types of spending processes in the public sector. The United Nations Office on Drugs and Crime indicates that, globally, the value of public contracts increases between 10% and 20% due to corruption [89]. Corruption can be defined as the misuse of power to obtain unlawful gains, including committing criminal acts such as bribery, embezzlement, and other forms of fraud [66]. Fraud, on the other hand, is a more broad term used to describe intentional deception by another party for economic gain [66]. These irregularities, even when unintentional, can also cause financial harm to public entities.

The adoption of new technologies by government agencies is creating new opportunities in areas such as the procurement process. The use of digital procurement systems makes it possible to attract a more significant number of participants, ensuring that the process is more transparent and available [82]. The adoption of these systems provides agencies with new tools to identify potential fraud. However, this process has several challenges. One such challenge is related to the analysis of vast amounts of data that can be easily accessed by digital procurement systems. The study of corruption using big data collected from public databases is a new research area with its own challenges [126].

Fraud and corruption detection has received widespread academic and political attention due to their central role in the quality of democracy, the provision of public goods, and economic growth [31]. The use of data-driven methods in the detection of fraud in public procurement has been extensively researched, from works that used Machine Learning (ML) models to detect cartels [128] to works that applied statistical tools to identify bidding anomalies [77].

This growth in academic production related to corruption in public procurement brings the need for research to evaluate the literature in the area and to identify the works being conducted and possible gaps to guide future research. This need comes from the fact that corruption occurrences can be hard to quantify and can take many forms, ranging from criminal activities to behavior that is inappropriate but not illegal [140]. The extensive research conducted on various forms of fraud and the techniques used to detect such fraudulent activities underscores the necessity for a comprehensive study that maps these different studies and identifies possible patterns and areas for improvement.

A systematic literature review (SLR) seeks to identify, evaluate, and interpret all available research related to a particular area or research question. This type of research is designed to summarize evidence on a particular technology, identify research gaps, and provide a framework for new research [63]. A systematic literature mapping (SLM) is a type of systematic study with a higher granularity level than an SLR. Therefore, an SLM is more appropriate for studies of broader topics [95], such as the one studied in this work.

Therefore, this work aims to perform an SLM to identify what types of fraud in public procurement are currently being studied in the literature, what types of data are being used in these studies, and how data-driven techniques are being used to solve this type of problem. The SLM described in this work is necessary because, to the best of our knowledge, there is no other secondary research whose purpose is to perform a systematic work with similar goals to those described in this document.

The remainder of the paper is organized as follows. Section 2 presents an overview of systematic studies related to data-driven methods and public procurement corruption, Sect. 3 describes the methodology used to conduct our research, Sect. 4 presents and discusses the results achieved by the proposed methodology. Finally, Sect. 5 concludes this paper.

2 Background

Given the complexity of fraud in public procurement and the different solutions that can be applied to each case, several systematic studies in this area can be found in the literature.

Table 1 Systematic studies found in the literature

Paper	Only public procurement	Data-driven methods
Artificial intelligence techniques to detect and prevent corruption in procurement: a systematic literature review [117]	•	•
Ready or not? A systematic review of case studies using data-driven approaches to detect real-world antitrust violations [3]	•	✓
Public Procurement Fraud Detection: A Review Using Network Analysis [76]	✓	•
Public procurement fraud detection and artificial intelligence techniques: a literature review [84]	✓	•
Fraud, corruption, and collusion in public procurement activities, a systematic literature review on data-driven methods [75]	•	✓

The goal of these reviews is to explore how different methods are used to detect fraud in the procurement of goods and services in the public and private sectors. Table 1 shows an overview of these studies.

Table 1 presents a list of the systematic studies identified in the literature, along with an analysis of the research methodologies employed in each study. This analysis differentiates between studies focused on public procurement auctions and works that also examined private auctions. Additionally, a distinction is made between papers that studied data-driven methods or more specific methods, such as Network Science (NS) or ML.

A systematic review was proposed by [117] to analyze the use of Artificial Intelligence (AI) techniques in detecting different types of corrupt practices. The study also investigated the characteristics of public and private organizations and the technological tools used. A systematic review of case studies using data-driven approaches to detect real-world antitrust violations was conducted by [3]. The review suggests that complex statistical analysis and ML models, such as leniency programs, can complement established tools to detect anti-competitive behavior.

A systematic review on the use of NS techniques for fraud detection was proposed by [76]. The study selected scientific papers over 10 years and found that cluster analysis and centrality measures were the most adopted approaches to detect public procurement fraud. A research conducted by [84] reviewed studies that used AI techniques such as ML and network analysis methods to detect fraud in processes of public organizations. The study also examined the challenges faced in this field.

The study most closely related to ours is presented in [75], in which the authors conducted a systematic literature review of data-driven approaches for detecting fraud in public procurement. However, their analysis primarily focused on bibliometric aspects, such as author affiliations and publication venues. The study addressed only one research question related to the techniques employed, while our study provides a more detailed analysis with five research questions. Their work also included primary studies that largely examined private auctions while our focus is the public sector. Additional limitations include the reliance on a single data source and the absence of strategies to mitigate the potential omission of relevant papers due to incomplete retrieval.

The present study aims to build upon the research carried out by [75]. We intend to achieve this objective with the definition of multiple research questions about other relevant aspects related to fraud detection in government procurement. We intend to solve the limited identification of relevant papers with the use of multiple data sources and with the implementation of additional search strategies such as snowballing.

As can be noticed, there exists a gap in the literature of systematic studies aimed at mapping what types of fraud in public procurement are being studied, the sources and formats of the data used in these studies, and how data-driven methods are being used to solve these problems. This research aims to fill this gap in the literature by performing a systematic analysis to map how data-driven techniques and frameworks are being applied in detecting, characterizing, and predicting corruption in public procurement.

3 Methodology

This section describes the planning and execution phases of our mapping study. We follow the guidelines for conducting systematic studies described in [63] and the protocol for identifying primary studies presented by [136]. We also used the guidelines described in the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) [94] to report the results of the search and selection of papers in our mapping study.

3.1 Planning phase

The first step in our systematic mapping concerns the definition of the objective of our research and our research questions. Our main goal is to *understand how data-driven techniques are being used in the prediction or identification of corruption and fraud in government procurement*. More specifically, we intend to discover the most used data-driven techniques for fraud detection in government procurement and the types of fraud detected by these techniques. This allowed us to define five research questions:

1. Q1: What data-driven techniques were proposed or used?
2. Q2: What types of fraud are considered by existing works?
3. Q3: What data are used by the primary works?
4. Q4: What features are used by the techniques applied by the primary works?
5. Q5: What are the results achieved by the primary works?

We considered the following data sources to carry out this research: Scopus,[1] IEEE Xplore,[2] ScienceDirect[3] and SBC OpenLib.[4] These sources were chosen because of their ample use in several systematic studies found in the literature. The SBC OpenLib was included because we were interested in the works executed by Brazilian researchers due to the unique nature of the public procurement process in Brazil.

The search string defined for each selected data source was designed to retrieve only papers whose titles or abstracts include terms related to "public procurement" and "fraud". The construction of the search string considers various synonyms and alternative terms for identifying fraud (e.g., "corruption" or "collusion") and public procurement (e.g., "public procurement" or "public tender"). The resulting search string is as follows: `(tender OR procurement) AND (corruption OR cartel OR collusion OR fraud OR "red flags" OR "bid rigging")`.

Since each database has unique characteristics, we had to adapt the search string accordingly. Table 2 shows the search string for all data sources.

The next step in the planning phase is to establish the criteria to select the primary works to be considered in our mapping. Based on the main objective and the research questions,

[1] https://www.scopus.com/home.uri.

[2] https://ieeexplore.ieee.org/Xplore/home.jsp.

[3] https://www.sciencedirect.com/.

[4] https://sol.sbc.org.br/index.php/indice.

Table 2 Search string adapted for each data source

Source	String
Scopus	(tender OR procurement) AND (corruption OR cartel OR collusion OR "red flag" OR fraud OR "bid rigging")
ScienceDirect	(tender OR procurement) AND (corruption OR cartel OR collusion OR "red flag" OR fraud OR "bid rigging")
IEEE Xplore	("Document Title":tender OR "Document Title":procurement) AND ("Document Title":corruption OR "Document Title":cartel OR "Document Title":collusion OR "Document Title":"red flag" OR "Document Title":fraud OR "Document Title":"bid rigging") OR (("Abstract":tender OR "Abstract":procurement) AND ("Abstract":corruption OR "Abstract":cartel OR "Abstract":collusion OR "Abstract":"red flag" OR "Abstract":fraud OR "Abstract":"bid rigging")) OR (("Author Keywords":tender OR "Author Keywords":procurement) AND ("Author Keywords": corruption OR "Author Keywords":cartel OR "Author Keywords":collusion OR "Author Keywords":"red flag" OR "Author Keywords":fraud OR "Author Keywords":"bid rigging"))
SBC OpenLib	(licitação OR licitações) AND (fraude OR fraudes OR corrupção OR conluio OR cartel OR cartéis OR "red flag" OR "bid rigging")

Table 3 Inclusion and exclusion criteria

Criteria	Inclusion	Exclusion
C1	Include only peer-reviewed primary research papers	Exclude non-peer reviewed, secondary works and surveys
C2	Include only works published in the last five years (2020 - 2024)	Exclude works published more than five years ago (before 2020)
C3	Include only works published in English or Portuguese	Exclude works published in languages other than English or Portuguese
C4	Include only works focused on government procurements	Exclude works that focused on private auctions or private procurements
C5	Include only works that used data-driven techniques	Exclude works that did not use data-driven techniques
C6	Include only works that detect, predict or identify fraud	Exclude works that do not detect, predict, or identify fraud

we defined six criteria for inclusion and exclusion of papers found in the database search. Table 3 shows a description of these criteria.

3.2 Execution phase

The execution phase of our research can be divided into three main steps: database search, selection of the primary works that will compose our mapping study, and extraction of data from the selected papers. The search step of the execution phase was performed in May 2024. This step consists of executing the adapted search strings, defined in Table 2, in all selected databases. We consolidated the works found in each database and excluded all duplicate papers. This process results in an initial set of papers.

Overall, 1777 papers were returned initially, and 566 papers remained after applying criteria C1, C2, and C3. This process was executed directly in each database using the available filtering tools. Table 4 shows the filter options used for each database.

Then, 75 duplicate papers were removed using Rayyan,[5] a web application used to support the execution of systematic reviews. After excluding all duplicates, 491 papers were left to apply the remaining inclusion and exclusion criteria (C4, C5, and C6). The Rayyan web tool was also used during this step of the selection procedure. While selecting primary works, we evaluated the title and abstract of all papers in the initial set, keeping those that

[5]https://new.rayyan.ai/.

Table 4 Filtering strategy for each database

Database	Criteria	Filter
Scopus	Search	Within title, abstract or author-specified keywords
	C1	Article and Conference paper selected in the Document type filter
	C2	2020, 2021, 2022, 2023, and 2024 selected in the Year filter
	C3	English and Portuguese selected in the Language filter
ScienceDirect	Search	Within title, abstract or author-specified keywords
	C1	Research articles selected in the Article type filter.
	C2	2020, 2021, 2022, 2023, and 2024 selected in the Years filter
IEEE Xplore	Search	Within title, abstract or keywords
	C1	Conference, Journals and Magazines filters selected
	C2	2020-2024 selected in the range of the Year filter
SBC OpenLib	Search	Within title, abstract or keywords
	C1	Event Proceedings and Journals filters selected
	C2	2020-2024 selected in the range of the Time period filter
	C3	English and Portuguese selected in the Language filter

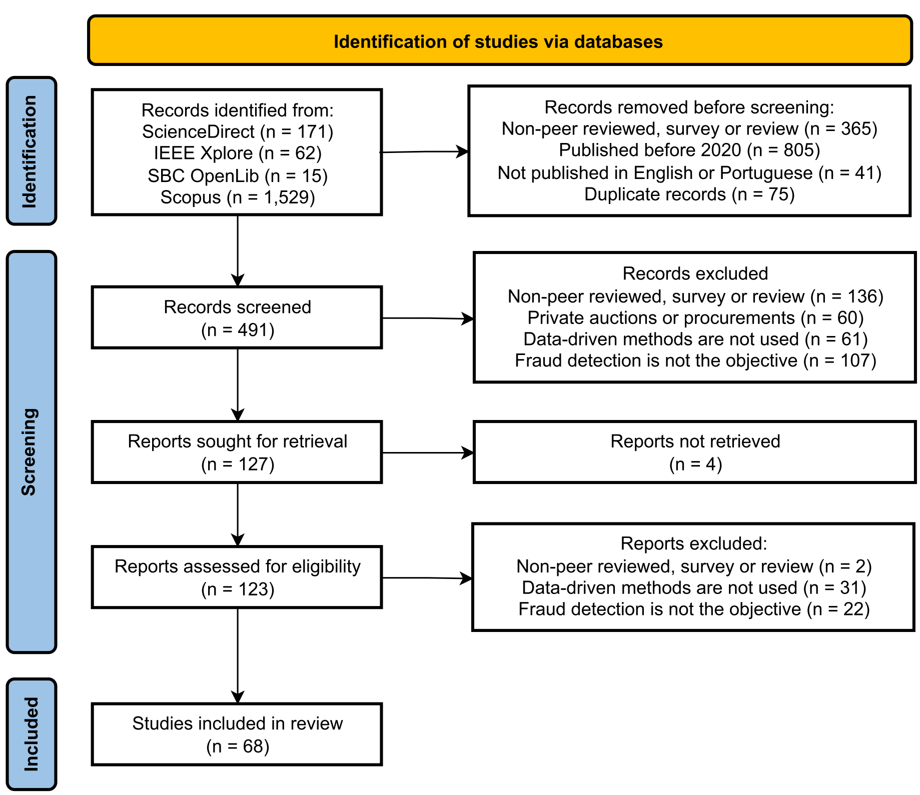

Figure 1 Primary works selection process

met the inclusion criteria and excluding those that met the exclusion criteria. Then, 68 papers remained after this evaluation. Figure 1 describes the execution of this process using the PRISMA guidelines.

3.3 Snowballing

Snowballing is a procedure used in systematic literature studies with the goal of identifying additional papers using a reference list of papers or citations [136]. Snowballing is a systematic search strategy comprising the following steps:

1. Define a start set of research papers
2. Application of inclusion and exclusion criteria in the start set
3. For each paper in the start set:
   (a) Backward Snowballing (BS): look at each reference in the paper and apply the inclusion and exclusion criteria
   (b) Forward Snowballing (FS): look at each work that referenced the paper and apply the inclusion and exclusion criteria
4. The backward and forward snowballing procedure must be executed for each new paper found in the current iteration of the snowballing procedure
5. If no new papers are found in the current iteration, the snowballing procedure is finished

We conducted five iterations of both backward and forward snowballing procedures in May, June and July of 2024. During this process, 5302 papers were retrieved and assessed for eligibility. The Google Scholar tool was used to find the studies that cited each paper in the start set. At the end of the snowballing process, 25 new papers were added to our mapping study. Figure 2 describes the execution of this process.

3.4 Threats to validity

There will always be threats to the validity of the results of a systematic study, even if the research has been carried out following strict quality criteria during its planning and execution phases. For this reason, it is essential for a systematic study to identify and present possible threats to its validity, providing a solid basis for the reader to decide whether or not the work will be useful to them [86].

A threat to the validity of our research is the lack of relevant papers in the study, caused by the use of incorrect or incomplete search terms during the database search. We mitigated this threat by trying to find as many relevant studies as possible. This was achieved

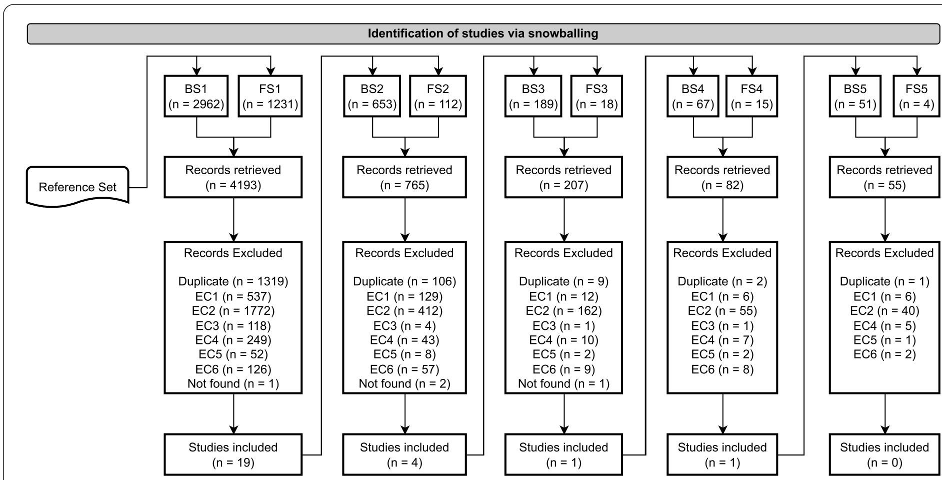

Figure 2 Selection of new studies using snowballing

Table 5 Data Extraction Form

Field	Description	RQ
ID	Identification number assigned to each paper	–
Year	Year of publication	–
Language	Language of publication	–
Publication Type	Type of publication (journal or conference)	–
Journal	Title of the publication	–
JIF	Journal Impact Factor	–
Title	Title of the paper	–
Method	Method used in the paper	RQ1
Method Group	Type of method used in the paper	RQ1
Field	Field of study of the paper	RQ1
Hybrid Strategy	Paper used methods from different fields	RQ1
Fraud Type	Type of fraud studied in the paper	RQ2
Fraud Group	Group of the fraud studied	RQ2
Data Type	Type of data used in the paper	RQ3
Country	Country of origin of the data	RQ3
Samples	Number of samples in the data	RQ3
Economic Sector	Economic sector of the data	RQ3
Available Data?	Authors published the data used in the paper	RQ3
Data Period	Range of dates of the data (start and end)	RQ3
Feature	Names of the features used in the paper	RQ4
Feature Group	Group of the feature used	RQ4
Red Flag?	Feature used is a red flag for corruption	RQ4
Target Type	Group of the target used	RQ4
Target	Names of the target used in the paper	RQ4
Measure Type	Type of measure used to describe the results	RQ5
Result	Value or textual synthesis of the results	RQ5

by searching in multiple scientific databases, and the execution of a snowballing procedure to complement the studies found during the database search.

3.5 Data extraction

We extracted information from the set of 93 papers included in the study with the goal of answering the research questions defined during the planning phase. A data extraction form was used to store the data extracted. Table 5 shows the fields in the form with a brief description and to which research question (RQ) each feature is related.

4 Results and discussion

This section presents the findings of the data extracted from the selected papers. Section 4.1 presents a summary of the main findings, Sect. 4.2 discusses the techniques used to detect fraud in public procurement, Sect. 4.3 describes the types of fraud studied, Sect. 4.4 presents the data used to build the models and Sect. 4.5 describes the types of features used. Finally, Sect. 4.6 discusses the results achieved with the proposed models.

4.1 Overview

We began by analyzing the number of publications per year and the types of publication venues. As shown in Fig. 3, there was a noticeable decline in the number of published works after 2020. This drop might be related to the COVID-19 pandemic, during which government procurement activity decreased, especially in sectors unrelated to healthcare. Conversely, a significant increase in publications was observed in 2023. This renewed interest may also be linked to the pandemic, as researchers began investigating how the temporary suspension of bureaucratic processes during the crisis may have influenced corruption levels in public procurement.

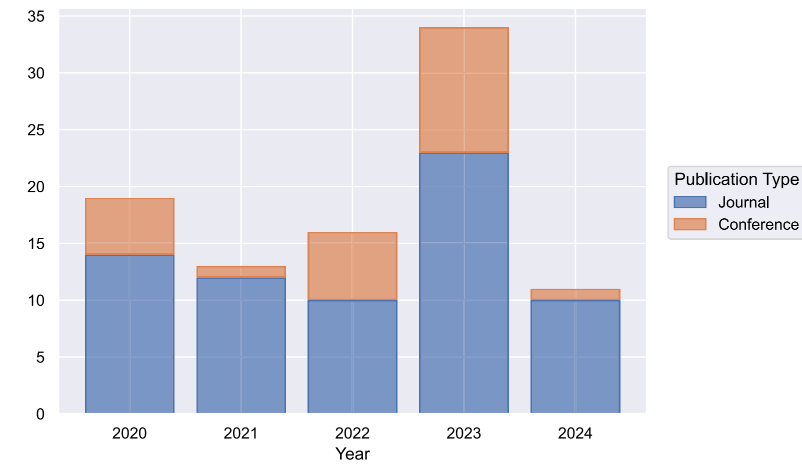

Figure 3 Type of publication by year

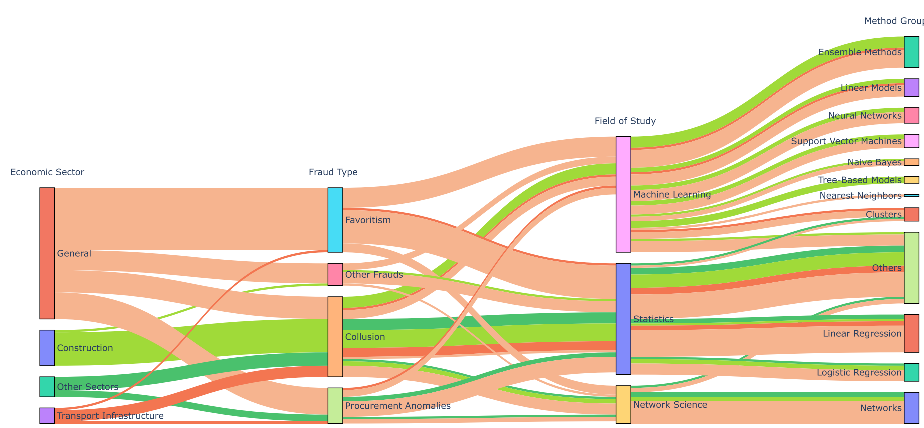

Figure 4 Overview of the studies by economic sector and type of fraud

Appendix A reveals that only 22.58% of the papers were published alongside their experimental data. This limited availability may stem from the confidential nature of the procurement process, making data sharing challenging for authors. To ensure the reliable reproduction of reported results, strategies should be developed to promote data sharing while safeguarding confidentiality.

To provide a brief overview of the results, we analyzed the economic sectors associated with the research presented in the primary set of papers. We identified four main classes of economic sectors: general, construction, transport infrastructure, and other sectors. The general sector includes papers that used data from public tenders spanning various economic domains. The transport infrastructure sector focuses on research involving data from highway projects, asphalt paving, and roadworks. Finally, the other sectors category encompasses diverse fields such as financial services, milk procurement, snow removal, and healthcare. Figure 4 shows these results.

The results suggest that most studies analyze data from multiple economic sectors to detect favoritism or other irregularities in public tenders. One possible explanation for this trend is that research on favoritism often uses the number of bids as a proxy for corruption, with tenders receiving only a single bid typically classified as fraudulent. Since single bidding can indicate fraud across various types of tenders, developing a more generalized approach may be more feasible, allowing for its application across different economic sectors.

The construction industry stands out most prominently in the results presented in Fig. 4. Research on this sector frequently highlights collusion as the predominant form of fraud. A key reason for this may be the substantial financial resources involved in construction tenders, which can attract corrupt individuals and organizations seeking to exploit the system through covert agreements for significant financial gain. A complete overview of these papers can be found in Appendix B.

4.2 RQ1 - data-driven techniques

This section presents the analysis of data extracted from the primary set of works used to address the first research question defined during the execution phase of our study: "Q1: What data-driven techniques are being studied in the literature?".

To answer this question, we collected the following information from the papers: the field of study of the method, the type of method used, and the specific method used in each study. Three fields of study were identified in our analysis: ML models, NS methods, and statistical tools. The following sections describe the methods employed in each of the fields previously described.

4.2.1 Machine learning models

The first identified field of study is ML, which refers to the ability of computer systems to extract meaningful patterns from raw data, enabling them to acquire knowledge autonomously [40]. Among the 93 papers in the primary set, 26 (27.95%) exclusively applied one or more ML methods. The majority of these studies leveraged ML models to detect instances of collusive behavior. Table 6 illustrates these findings.

Ensemble methods were utilized in 13 out of the 21 papers that exclusively applied ML models. These methods enhance predictive performance by combining multiple models into a single, more robust model. There are two main types of ensemble methods: averaging methods, which reduce variance by averaging predictions from multiple independent models, and boosting methods, where weak models are trained sequentially to minimize bias and improve overall accuracy [100].

The most commonly used ML model was Random Forest, an ensemble approach that aggregates predictions from multiple decision trees while employing bagging to mitigate overfitting and enhance out-of-sample accuracy [24]. Tree-based algorithms offer significant advantages, including higher explanatory power due to their flexible parametrization [30] and an easily implementable predictive framework. These characteristics make them particularly suitable for agencies aiming to develop fraud detection tools for public procurement [127]. However, we observed a distinction in studies that combined ML models with Natural Language Processing (NLP) techniques. Table 7 presents these findings.

NLP techniques enable computers to process and analyze human language, facilitating human-computer interactions and extracting valuable information from textual documents. These techniques typically involve large volumes of unstructured data and are widely used for text classification, information retrieval, and extraction tasks [44]. In our set of primary papers, studies that combined NLP with ML employed different models compared to those that did not incorporate NLP.

Table 6 Types of ML models used

Method group	Method	Papers (ID and reference)
Clusters	Gaussian Mixture Model	7 [111]
Clusters	Latent Markov Model	68 [25]
Clusters	PCA	17 [118]
Clusters	Soft-DTW k-means	72 [79]
Ensemble Methods	Ada Boost	9 [100]
Ensemble Methods	Extra Trees	9 [100]
Ensemble Methods	Gradient Boosting	9 [100], 12 [30], 14 [35], 84 [78]
Ensemble Methods	Isolation Forest	50 [17], 77 [89]
Ensemble Methods	Random Forest	2 [128], 3 [127], 8 [24], 9 [100], 10 [52], 11 [115], 12 [30], 13 [50], 14 [35], 22 [58]
Ensemble Methods	StackedEnsemble	84 [78]
Ensemble Methods	Super Learner	2 [128], 10 [52], 13 [50]
Ensemble Methods	XGBoost	84 [78]
Linear Models	Lasso Regression	8 [24], 10 [52], 14 [35]
Linear Models	Logistic Regression	2 [128], 12 [30], 14 [35]
Linear Models	Ordinary Least Squares	8 [24]
Linear Models	Ridge Regression	8 [24]
Linear Models	SGD	9 [100]
Naive Bayes	Bernoulli Naive Bayes	9 [100]
Naive Bayes	Gaussian Naive Bayes	9 [100]
Neural Networks	Conv. Neural Network	4 [49]
Neural Networks	Multi-Layer Perceptron	9 [100], 22 [58]
Others	Apriori Algorithm	85 [9]
Others	Confident Learning	23 [42]
Others	Gaussian Process	9 [100]
Others	K-Nearest Neighbors	9 [100], 11 [115]
Others	Positive-Unlabelled	57 [41]
Others	Support Vector Machines	9 [100], 10 [52], 17 [118], 22 [58]
Tree-Based Models	Classification Tree	2 [128], 3 [127]
Tree-Based Models	Decision Tree	22 [58]

Table 7 Types of ML models used with NLP

Method group	Method	Papers (ID and reference)
Linear Models	Logistic Regression	30 [81]
Linear Models	Passive-Aggressive Model	27 [70]
Linear Models	Perceptron	27 [70]
Linear Models	Ridge Classifier	27 [70]
Linear Models	SGD	27 [70]
Naive Bayes	Bernoulli Naive Bayes	27 [70]
Naive Bayes	Complement Naive Bayes	27 [70]
Naive Bayes	Multinomial Naive Bayes	27 [70]
Naive Bayes	Naive Bayes	30 [81]
Nearest Neighbors	K-Nearest Neighbors	27 [70]
Nearest Neighbors	Nearest Centroid	27 [70]
Neural Networks	Bi-LSTM	27 [70], 31 [71]
Neural Networks	Bottleneck	27 [70], 31 [71]
Neural Networks	Conv. Neural Network	16 [119]
Neural Networks	Deep Neural Network	27 [70]
Neural Networks	Self-Organising Maps	6 [103]
Others	BERT	31 [71]
Others	Random Forest	27 [70]
Others	Support Vector Machines	27 [70], 30 [81]

Since its introduction in 2017 [123], the Transformer model has revolutionized NLP, achieving state-of-the-art performance across multiple tasks. Building on its success, several derivative models, such as GPT and BERT, have been developed [138]. However, only one study in our dataset utilized this class of models. Specifically, BERT was applied in

Table 8 Types of network science models used

Method group	Method	Papers (ID and reference)
Networks	Bipartite Network	48 [87], 51 [74], 60 [88], 61 [126]
Networks	Complex Network	33 [32], 43 [16], 86 [129]
Networks	GAAN	34 [15]
Networks	Multiplex Network	5 [124]
Networks	Social Network	21 [14], 28 [46], 62 [91], 81 [139], 90 [137]
Others	Network Science Methods	20 [96], 54 [34]
Others	Node Detection	33 [32]

[71] to extract red flags from public procurement data, demonstrating competitive results compared to traditional methods. This highlights a gap in the literature regarding the potential of Transformer-based models like GPT and BERT for improving fraud detection in public procurement.

Instead of Transformer models, these studies predominantly employed Deep Learning models, such as Deep Neural Networks (DNNs) and Long Short-Term Memory (LSTM) networks. Deep learning models consist of multiple layers that enable the hierarchical learning of complex data representations [68]. These algorithms are particularly effective at uncovering intricate patterns in massive amounts of unlabeled and unstructured data [85]. In the analyzed papers, NLP techniques were used to extract information from textual sources such as procurement notices and official documentation. With these unstructured inputs, DNN models were primarily trained to detect favoritism in public procurement processes.

4.2.2 Network science methods

Another identified field of study was NS, which focuses on the study and analysis of networks, their properties, and applications [60]. It can be broadly defined as the study of network representations of physical and social phenomena, leading to predictive models for understanding these complex systems [11]. NS was employed in 17.20% (16 papers) of the studies analyzed. Table 8 presents these results.

A considerable body of literature employs various types of NS methodologies to map relationships between entities with the aim of detecting criminal behavior. Social networks have been used in detecting social security fraud [122], identifying fraud rings [114], and uncovering criminal suspects [33]. Bipartite networks have been applied to identify leaders within criminal organizations [43], as well as to detect insurance fraud [93, 121] and telecommunication fraud [57]. Complex networks have been used to model interactions within criminal structures [116], detect high-crime areas within cities [113], and identify entities involved in financial fraud [48]. Our analysis of the primary studies indicated that network models are also frequently used to detect fraudulent behavior in public procurement. Social and bipartite network models were the most commonly adopted, primarily applied to characterize and reveal collusive relationships between bidders.

Social networks represent connections between individuals or organizations based on specific social relationships. In this context, nodes correspond to entities involved in public procurement, and edges represent their interactions. Social Network Analysis (SNA) can be employed to examine hidden relationships and expose associations between participants engaged in collusion [137].

Bipartite networks, on the other hand, consist of two distinct sets of nodes, with connections only between nodes of different types. Due to their heterogeneous nature, these networks provide a realistic model of real-world systems and have proven particularly useful in representing trade networks in economic systems [69].

Beyond network models, standard network science methods are also widely used in this field. Studies in this category often employ Pattern Mining algorithms for anomaly detection [96] and graph database technologies combined with market concentration indicators [34]. Unlike network-based solutions, which predominantly focus on detecting collusion, studies using standard network science methods are primarily aimed at identifying favoritism in public tenders.

4.2.3 Statistical methods

The final and most prominent field of study identified in our research is statistics. In this context, statistics encompasses a collection of methods for data collection, organization, and analysis, enabling researchers to derive conclusions and support decision-making [51, 54, 104]. This field accounts for 52.68% (49 papers) of the studies in our dataset, making it the dominant approach for analyzing fraud in public procurement. Table 9 presents these findings.

The first category within this field is linear regression, an explainable method used to construct functions that predict the value of a target variable based on the values of predictor variables [65]. The most prevalent model in this group is Ordinary Least Squares (OLS), a fundamental regression technique that fits a linear model by minimizing the residual sum of squares between predicted and observed values [83]. Most studies that applied OLS and other types of linear regression focused on detecting favoritism in public procurement.

Table 9 Types of statistics models used

Method group	Method	Papers (ID and reference)
Linear Regression	Basic Linear Regression	53 [112]
Linear Regression	Dynamic Panel Regression	18 [19]
Linear Regression	Fixed-Effects Model	36 [5], 83 [73]
Linear Regression	Linear Probability Model	39 [4]
Linear Regression	Multilevel Model	36 [5]
Linear Regression	Ordinary Least Squares	25 [22], 36 [5], 38 [59], 45 [72], 46 [31], 58 [125], 63 [21], 69 [38], 78 [27], 79 [6], 80 [23], 91 [135]
Linear Regression	Two-Stage Least Squares	64 [132]
Logistic Regression	Binary Regression	29 [36], 35 [20], 47 [29], 49 [1], 82 [18]
Logistic Regression	Bootstrap Regression	19 [12]
Logistic Regression	Linear Probability Model	56 [97]
Logistic Regression	Logit Estimate Model	56 [97]
Logistic Regression	Multinomial Regression	92 [131]
Others	Bootstrap Approach	93 [130]
Others	Clusters	24 [10], 76 [53]
Others	Gen. Method of Moments	91 [135]
Others	Heuristics	73 [92]
Others	IQR	40 [110]
Others	Item Response Theory	55 [37]
Others	Matching Estimators	59 [28]
Others	Modified Order Statistic	26 [105]
Others	Newcomb-Benford Law	1 [90], 74 [26], 88 [101]
Others	Probabilistic Method	42 [108], 66 [107], 67 [109]
Others	Reference Scenario	65 [106]
Others	Regression Discontinuity	37 [62], 41 [61]
Others	Statistical Analysis	15 [98], 44 [82], 52 [8], 70 [102], 71 [39], 87 [77], 89 [67]

Tenders with single bids are widely recognized as a red flag for favoritism in public procurement. This commonly used corruption proxy is readily available in procurement datasets, eliminating the need for time-consuming investigations. The accessibility of this feature allows OLS models to be used in formulating hypotheses regarding the impact of transparency measures [5, 6] and political connections [31, 38] on favoritism, among other factors.

The next group within this field is logistic regression, another predictive modeling technique. Unlike linear regression, logistic regression is designed for categorical target variables. The most frequently used approach within this category is binary logistic regression, where the target variable typically represents the presence of corruption (collusive or competitive tender) or favoritism (tender with a single bid or tender with multiple bids).

Finally, several studies employed a diverse range of methodologies that do not fit neatly into a single category. These works utilized Exploratory Data Analysis (EDA), visualizations (plots and tables), and pattern mining algorithms to uncover anomalies in procurement processes. Notable methods in this group include the Newcomb-Benford Law and the probabilistic method.

The first notable method is the Newcomb-Benford Law. This method states that in naturally occurring numerical datasets, smaller leading digits appear more frequently than larger ones [101]. A significant deviation from this expected distribution often signals data anomalies, making it a useful tool for fraud detection [90].

Another standout methodology employed in this group is the probabilistic method. This tool is used to prove that a structure with the desired properties exists. This is done by defining an appropriate probability space of structures and demonstrating that the properties hold in these structures with a positive probability [2]. The papers that used the probabilistic method applied this model intending to show the presence of collusive behavior in construction tenders.

4.3 RQ2 - types of fraud

The next research question we aimed to address in our study concerns the types of fraud currently being investigated. To categorize the studies, we extracted relevant information from the papers and classified them into four groups: collusion, favoritism, procurement anomalies, and other types of fraud. Figure 5 presents these results.

The first category, collusion, includes research focused on collusive behavior and irregular agreements between bidders, such as bid rigging and cartels. These fraudulent schemes involve bidders coordinating their actions to manipulate the procurement process, often leading to inflated prices and reduced competition. Bid rigging refers to agreements among bidders designed to artificially raise prices or lower the quality of goods and services. Such practices hinder market entry for new competitors and reduce incentives for innovation, ultimately leading to significant economic losses [22]. Cartels, in turn, are defined as arrangements among competing firms with the objective of eliminating competition and increasing profits. These agreements may involve price fixing, market division, bid manipulation, and other anti-competitive practices [64].

The second category, procurement anomalies, comprises studies that use generic or unspecified terms for fraud, such as anomaly or malfeasance. The term anomaly is typically used to describe deviations from the expected behavior in public procurement processes (e.g., unusual purchases or transactions) that may indicate irregularities. Malfeasance, on

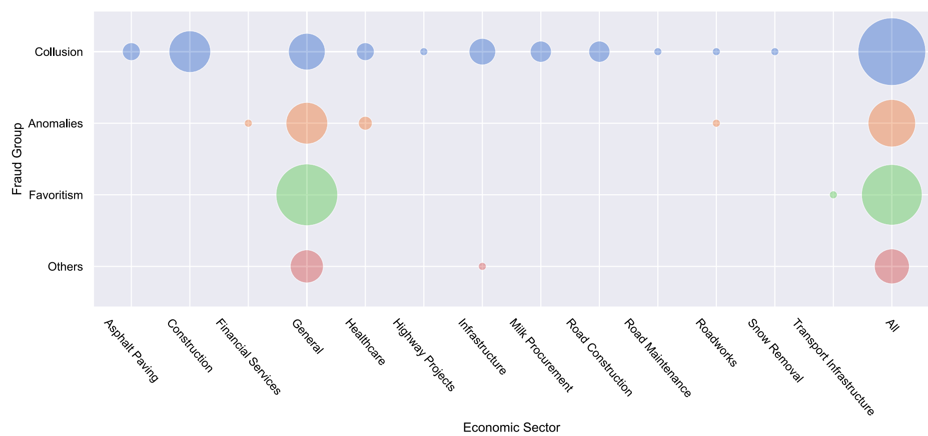

Figure 5 Groups of frauds studied by economic sector

the other hand, is a broad term encompassing various forms of misconduct, including cost overruns, favoritism, and collusion. This category also comprises studies that aim to detect multiple types of fraud simultaneously, rather than concentrating on a specific form of wrongdoing.

The third category, favoritism, refers to cases where the procurement process is manipulated to benefit a particular bidder. Such practices undermine the transparency and fairness of public tenders, often discouraging competitive participation and leading to inefficient outcomes, where the most advantageous offer may not win the contract [118, 119]. Studies in this category frequently use the number of bids as an indicator of favoritism, with single-bid tenders serving as a widely accepted proxy for this form of misconduct.

The final category, other types of fraud, includes studies that examine fraudulent activities not encompassed by the previous three categories. These works explore a diverse range of irregularities, each of which reflects a distinct aspect of corruption or manipulation in public procurement. The types of fraud in this category include:

- Corruption Risk Indicator (CRI): an objective metric designed to estimate the likelihood of fraud in a specific contract or tender. CRIs are typically calculated based on procurement characteristics that may be associated with corrupt practices;
- Contract Additions: refers to post-award increases in the monetary value of contracts, which may serve as a proxy for potential corruption or manipulation in the procurement process [97];
- Discrete Contract Value to Budget (DCVB) Ratio: this ratio measures the proportion of contract values assigned through discretionary methods relative to the total procurement budget of a public entity. A high DCVB ratio may indicate favoritism [112];
- Overpricing: characterized by a significant discrepancy between the estimated and final values of a procurement contract, which may suggest corruption, inefficiency, or mismanagement;
- Shell Companies: legally registered entities used to conceal the identity of their true owners [56]. In procurement, shell companies are often used to submit fictitious bids, creating the false appearance of competition;

- Tender Complaints: refers to protests, objections, or complaints filed by participants in the procurement process, which may serve as red flags indicating potential irregularities or fraudulent behavior.

The fraud category with the highest number of papers in our primary set is collusion, accounting for 38.70% of the studies (36 papers). This category also exhibits the greatest diversity in economic sectors, reflecting the varied sources of data used in these studies. Among the sectors analyzed, construction stands out, appearing in 18 of the 36 papers focused on collusion. As previously discussed, this prevalence may be due to the large financial sums involved in construction tenders, which makes them particularly susceptible to corrupt agreements between bidders.

The second-largest category is favoritism, comprising 31.18% of the studies (29 papers). Unlike collusion, this category is dominated by the general economic sector, as studies often use data from tenders spanning multiple industries. The objective of this research is to develop solutions that can detect favoritism across various sectors. A key factor in this approach is the identification of single bidding as a common proxy for favoritism, allowing for the creation of standardized fraud detection methods applicable to tenders across different economic domains.

The final two categories identified in our study are procurement anomalies and other types of fraud, representing 19.35% (18 papers) and 10.75% (10 papers) of the studies, respectively. Similar to the favoritism category, research in these groups primarily focuses on fraud in the general economic sector. Many of these studies leverage common corruption proxies to develop generalizable fraud detection models, ensuring applicability across tenders from different industries.

4.4 RQ3 - types of data

The next question in our study concerns the data used by the proposed models. We analyzed the types of data and categorized them accordingly. The identified data types include tenders, bids, companies, contracts, financial transactions, data from multiple stages of the procurement process, and other related categories. Figure 6 illustrates the types of data used for each identified fraud type.

Studies on collusion predominantly rely on bid data (18 papers) and tender data (14 papers). This aligns with expectations, as most collusion detection research focuses

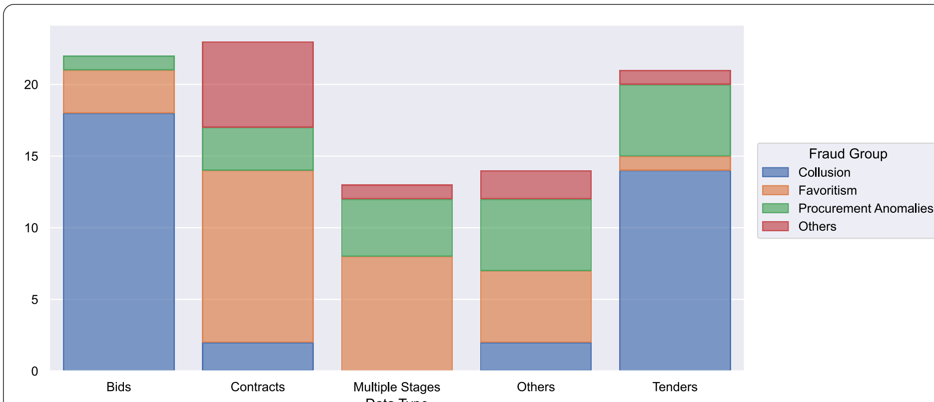

Figure 6 Types of data by fraud type

on identifying bidding patterns that suggest coordinated behavior among participants. A common approach to improving detection rates is the use of screening variables [50, 52, 58, 100, 111, 127, 128]. Screens are statistical measures derived from the distribution of bid values within a tender. They help uncover collusive patterns among bidders and enhance the effectiveness of machine learning models in detecting collusion [100].

The studies on favoritism primarily utilized data from multiple stages (8 papers) and contract data (12 papers). Many works in this category enriched their datasets with additional information about the companies involved in the procurement process. For instance, some studies incorporated firm-level data to identify partisan favoritism [19, 38]. Others employed political variables [28] and campaign donation records [27] to detect favoritism linked to political connections.

Several studies also leveraged data from multiple stages and contracts to construct graph-based datasets, enabling the identification of favoritism patterns [34, 96]. Network science methods were frequently applied to uncover suspicious relationships between buyers and suppliers [31, 46, 126]. Additionally, numerous papers explored the use of red flags to detect favoritism. Many relied on single-bid occurrences as a proxy for corruption, examining how corruption risk indicators—such as non-published procedures [29, 31], restricted procedures [25, 37], and non-competitive procedures [18, 27]—can serve as signals of favoritism.

Studies focused on detecting anomalies in the procurement process predominantly used data from multiple stages (4 papers) and tender data (5 papers). Many of these works applied graph and network analysis techniques to uncover irregularities [14, 15, 91]. Others relied on corruption indicators, such as the duration of the auction process and the number of bids, to identify potential fraud cases [82, 98]. Several studies also enriched their datasets with firm-related information to detect irregularities involving bidders [8, 17, 92].

For detecting other types of fraud, contract data was the most commonly used source (6 papers). A frequent approach in this category involved analyzing contract information from different countries to calculate the Corruption Risk Indicator (CRI) for procurement processes [21, 72, 115]. Another notable characteristic of this research group is the comparatively larger average sample size, as illustrated in Table 10.

The average sample size used in studies on other types of fraud (2,000,000) is only behind those focused on procurement anomalies (2,400,000) and ahead of those in the favoritism (700,000) and collusion (440,000) groups. This finding aligns with the expectation that studies on procurement anomalies and other fraud types tend to rely on unlabeled datasets or widely available target features, unlike those examining favoritism and, especially, collusion.

In addition to analyzing the types of data used, we also collected information on data availability. Figure 7 presents the number of papers that published the datasets used in their research, unveiling that only 22.58% of papers published the data. This is an expected

Table 10 Average sample size by data type

	Collusion	Favoritism	Others	Anomalies
Bids	43,000	1,900,000	–	378,000
Contracts	800	800,000	394,000	51,000
Others	7,500,000	750,000	8,800,000	8,200,000
Multiple Stages	–	60,000	149,000	94,000
Tenders	14,000	147,000	196,000	387,000

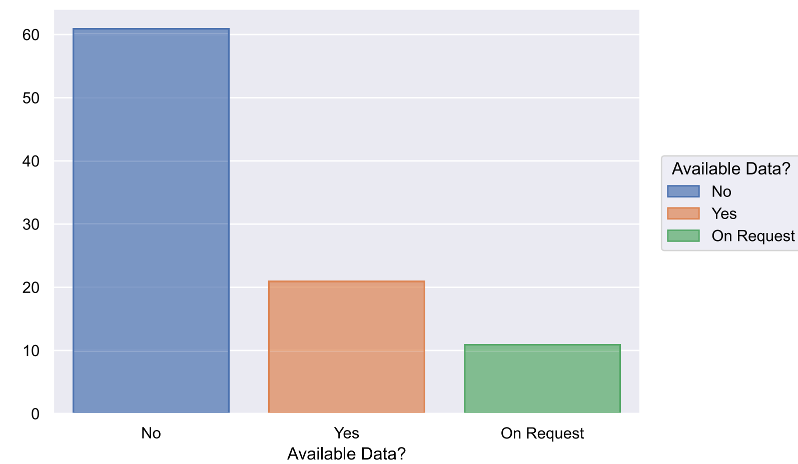

Figure 7 Number of papers that published data

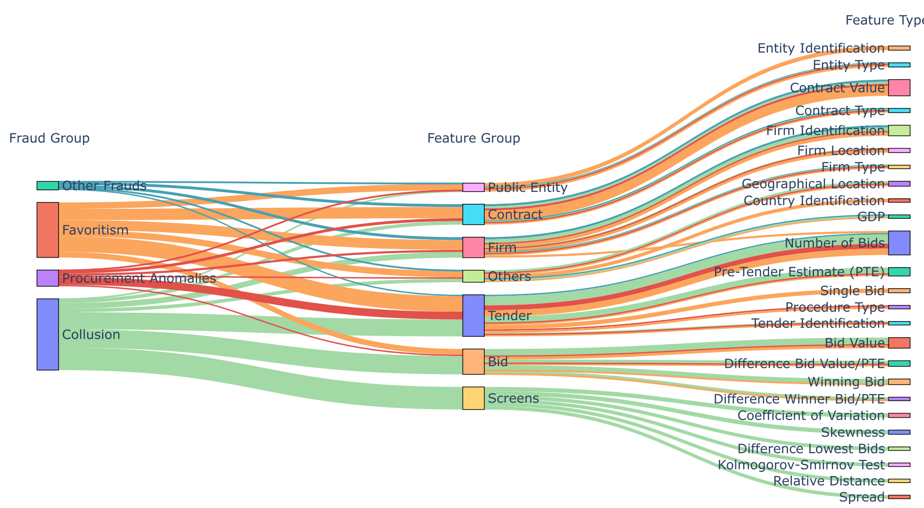

Figure 8 Simple features used in the primary works

outcome, given the confidential nature of the procurement process and the use of personal data of those involved in this process. Sharing this personal data is frequently restricted by data protection legislation in various countries.

4.5 RQ4 - features and target variables

The next question focuses on the features used to develop various models and methods. Overall, we identified 878 features and categorized them into two types: simple features and red flags. Given the high granularity of these features, we propose a taxonomy for their classification. This taxonomy consists of two levels: feature group and feature type. We also identified seven groups of simple features: public entity, contract, firm, tender, bid, screens, and others. Figure 8 presents the results.

The proposed taxonomy allows for a better visualization and understanding of the results. Figure 8 shows the features that appeared in at least 5 works. This filter was applied to show only features that appeared in multiple papers. These results also showed the use of multiple groups of features in most of the papers analyzed, with only 27 papers (29.03%) using just one group of features.

The most common simple features identified were the number of bids in a tender (27 papers), the value of the contract (21 papers), the pre-tender estimate (12 papers) and the value of each bid in a tender (11 papers). The results from Fig. 8 also highlight the lack of features from the public entity. As one of the main actors involved in the procurement process, it is important to understand how this type of information can help detect irregular activities.

Another possible gap is related to the features from the Screens group. These features are indices derived from the values of all bids in a given tender and can be used in the identification of collusion. Several works in the literature used these screens to improve the collusion detection rate of ML models [50, 52, 58, 100, 111, 127, 128]. However, the majority of works in this category employ screens to detect collusion in construction and infrastructure tenders. Further studies are needed to validate the use of these features in different economic sectors.

Red flags, on the other hand, can be divided into five groups: bidder relationships, firm, contract, procurement notice and tender. Red flags are used in the literature to measure the risk of corruption. These features are correlated with corruption instead of perfectly matching it. This preventive approach is used to detect potential weaknesses and alert for potential vulnerabilities [25]. Figure 9 shows the results.

The most common red flags are tenders with a single bid (11 papers), tenders with a restricted procedure (7 papers), an unpublished procurement notice or call for tender (6 papers) and the use of risky evaluation criteria to define the winner of a tender (5 papers). These results confirm the preventive nature of red flags, since the majority of red flags are raised during the first stages of the procurement process. Notable exceptions are the single bid red flag, which is a common proxy for favoritism and is widely used in the literature to detect this type of fraud, and contract (5 papers). We believe that is important to define features capable of identifying possible fraudulent behavior that might happen only after the winner of the procurement process is chosen.

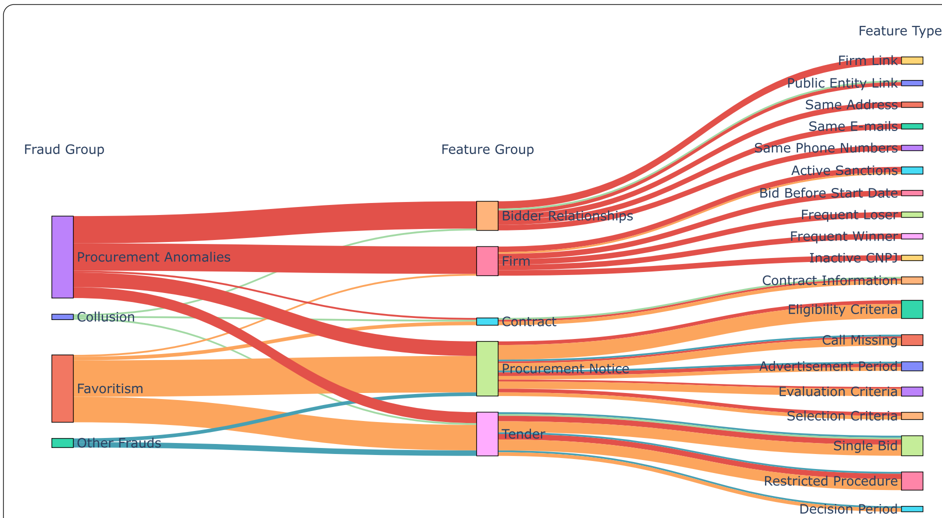

Figure 9 Red flags used in the primary works

Another gap found by analyzing the results shown in Fig. 9 is related to red flags in collusion. Even though almost 40% of the works in our mapping address this type of fraud, only one paper uses red flags to alert for possible collusion during the procurement process. The importance of red flags, or audit trails, comes from its capability of selecting relevant information from a large volume of bidding data that can be used by specialists in the detection of fraud [7].

Given the importance of collusion for the literature of public procurement fraud detection, new ways of alerting entities that a collusive practice is taking place must be defined, studied and disseminated. Further research is required to validate the potential use of features such as bid rotation and incumbency [62] and frequent winners and losers [7, 14, 91] for detecting collusion. Appendix C presents an overview of all groups of features and red flags used by the papers evaluated.

4.5.1 RQ4 - target variables

An analysis of the target variables used by the primary works was also conducted in the present study. We identified 41 different types of targets and divided them into three groups: binary targets, linear targets, and proxies. A proxy for corruption is a feature that, despite not confirming the presence of a certain type of fraud, can be used to indicate that this fraud might be occurring. Figure 10 shows the results. A complete overview of all the target variables analyzed can be found in Appendix D

In the case of studies focused on collusion, the most common target variable was a binary feature representing the presence of collusion. This target appeared in 16 papers, typically based on well-known datasets involving documented investigations, such as Operation Car Wash in Brazil [100, 105, 107], the Okinawa bid-rigging cartel in Japan [49, 52], the Ticino cartel in Switzerland [127, 128], and the Ohio school milk cartel in the United States [58, 62].

For studies investigating favoritism, the most frequently used target variable was the presence of single-bid tenders, which appeared in 10 papers. Single bidding also emerged as the most widely adopted proxy for corruption in the reviewed literature. The rationale

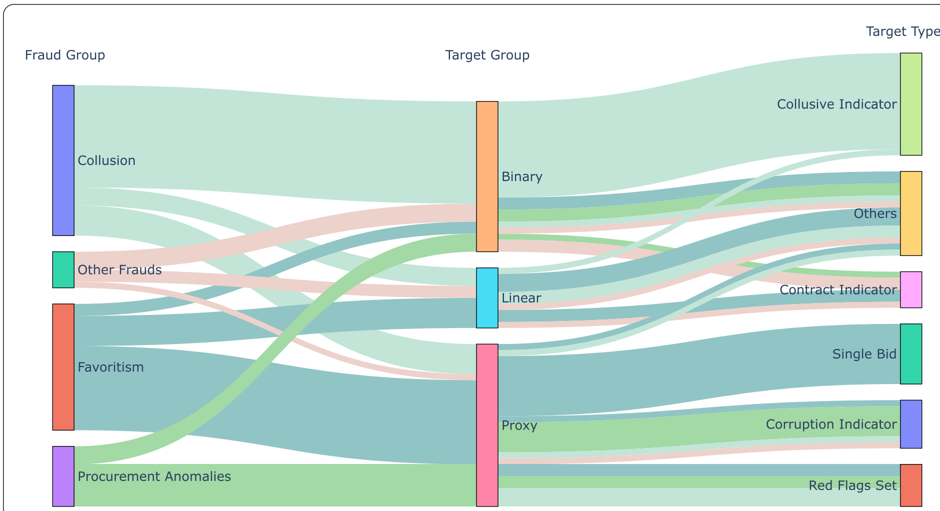

Figure 10 Target variables used in the primary works

behind this metric is that when only one firm submits a bid for a tender, it may suggest restricted competition or deliberate steering of the process toward a preferred company. A common approach in this group involves using statistical methods to identify features that increase the likelihood of single-bid outcomes.

In the case of procurement anomalies, proxy variables were the predominant type of target (6 papers). These studies often employed CRIs or sets of predefined red flags to quantify the risk or detect the presence of irregularities in the procurement process.

Studies categorized under other types of fraud primarily used contract-based indicators as target variables. These variables typically reflected contract characteristics that could signal fraudulent activity, such as unusual cost overruns [115] or post-award contract additions [97], which are often interpreted as proxies for overpricing or corruption.

Additionally, 31 papers did not use any explicit target variable, opting instead for unsupervised approaches. These studies relied on anomaly detection, clustering, and network science methodologies to uncover and characterize potentially fraudulent behavior without prior labeling. A complete overview of the works in this category is also provided in Appendix D.

4.6 RQ5 - results

The final question in our study examines the results reported by the evaluated papers. This section is divided into three groups, corresponding to the fields of study discussed in Sect. 4.2: ML models, NS methods, and statistical approaches. To answer the question “Q5: What are the results achieved by the primary works?” defined during the planning phase, we extracted the following information from the papers: the type of measure and the result achieved by all the models and methods employed during the research.

An important disclaimer in this section is that we found no evidence that any single model is demonstrably superior. Thus, the purpose of this section is to identify patterns and areas for improvement in model implementation rather than to suggest that one model outperforms others in detecting specific types of fraud. The following sections describe the methods employed in each of the fields previously described. Appendix E presents the results reported by the papers that used numerical metrics to evaluate the models employed.

4.6.1 Machine learning models

The first category of results evaluated consists of numerical metrics reported by ML models. For brevity, we focus on the three most commonly used metrics: accuracy, precision, and recall. These are standard performance measures in binary classification [99], where the target variable typically represents a fraudulent tender, bid, or contract (positive class) versus a non-fraudulent one (negative class). The first of these metrics is accuracy, shown in Equation (1), which measures the ratio of correct predictions over the total number of samples:

$$ACC = \frac{tp + tn}{tp + fp + tn + fn} \tag{1}$$

where tp represents the number of true positives, tn the number of true negatives, fp the number of false positives, and fn the number of false negatives. Accuracy is the most commonly used metric, appearing in 11 papers that reported ML model results. While it is

Table 11 Average accuracy results by method group

Method group	Anomalies	Collusion	Favoritism	Others
Clusters	–	–	90.0	–
Ensemble Methods	82.0	82.40	–	53.67
Linear Models	64.0	70.14	72.5	–
Naive Bayes	–	73.70	71.0	–
Neural Networks	–	78.60	82.5	–
Tree-Based Models	–	56.81	–	–
Others	–	65.37	79.0	46.33

Table 12 Average precision results by method group

Method group	Anomalies	Collusion	Favoritism	Others
Clusters	–	–	92.0	–
Ensemble Methods	98.55	88.6	–	75.33
Linear Models	24.63	89.92	57.50	–
Naive Bayes	–	88.07	42.50	–
Nearest Neighbors	–	83.15	–	–
Neural Networks	–	89.67	85.4	–
Others	–	89.35	73.45	75.33

straightforward to compute and interpret, it has significant limitations, particularly its tendency to favor the majority class [47]. This can be problematic when evaluating highly imbalanced datasets. Table 11 presents the average accuracy results reported by the ML models.

Ensemble methods achieved the highest accuracy in detecting procurement anomalies, collusion, and other types of fraud among all evaluated ML models. Five papers reported accuracy results for ensemble methods [50, 52, 100, 127, 128], while only one study provided accuracy results for ML models specifically applied to procurement anomaly detection [30]. This highlights the need for further research to determine whether alternative approaches could yield better results for this fraud category. Most studies in this area relied on tender and bid data to train their models. However, [118] achieved the highest accuracy using contract data. Future research should explore the integration of contract data to assess its potential in enhancing the effectiveness of machine learning algorithms for public procurement fraud detection.

The next metric evaluated is precision, shown in Equation (2). This metric is used to measure the positive patterns that are correctly predicted from the total predicted patterns in a positive class [47]:

$$PRE = \frac{tp}{tp + fp} \tag{2}$$

where tp denotes the number of true positives and fp the number of false positives. Seven papers reported results using this metric.

The precision metric measures the proportion of predicted positive samples that are actually positive. It is particularly useful when false positives are more costly than false negatives [80]. This is especially relevant in fraud detection, where a false positive could lead to an unnecessary and costly investigation to verify a fraud that was mistakenly highlighted by the model. Table 12 presents the average precision results, categorized by method group and fraud type.

Table 13 Average recall results by method group

Method group	Anomalies	Collusion	Favoritism	Others
Clusters	–	–	89.0	–
Ensemble Methods	85.95	95.30	–	55.67
Linear Models	26.60	95.98	28.0	–
Naive Bayes	–	96.70	7.0	–
Nearest Neighbors	–	92.95	–	–
Neural Networks	–	91.07	84.55	–
Others	–	91.0	60.70	43.67

Ensemble methods achieved the best results in detecting procurement anomalies, collusion, and other types of fraud. Meanwhile, the use of contract data and clustering techniques consistently yielded superior results in favoritism detection. Most studies reporting precision results relied on textual data sources, including contracts [118], procurement notices [70, 71], and tender documentation [81]. The highest overall precision results were reported by [24], which used ensemble methods and procurement data.

The final metric used was recall, also known as the True Positive Rate, as shown in Equation (3). This metric measures the number of positive class samples that were correctly classified by the model [80]:

$$REC = \frac{tp}{tp + fn} \tag{3}$$

where tp denotes the number of true positive samples and tn the number of true negative samples. This is a useful metric for fraud detection since a model with a higher recall is capable of correctly detecting more fraudulent cases. Six papers reported results using this metric. Table 13 shows the average results by method group and fraud type.

Ensemble methods reported the best results when compared to other methods in the detection of procurement anomalies [24]. Clusters achieved better results in the detection of favoritism [118] and linear models reported superior results in the prediction of collusive cases [70]. The majority of works in this group also used data in textual form to train the models, including procurement notices [70, 71], tender documentation [81] and contracts [118]. Other works also used procurement data from different steps of the procurement process [100, 115].

Some studies employing ML models did not use numerical evaluation metrics, such as accuracy or precision, to report their results. Instead, these works presented their findings through graphical plots or tables, followed by a discussion and interpretation by the authors. A common strategy among these studies is the use of clustering techniques to identify patterns that may indicate favoritism in the procurement process.

A variation of the K-Means algorithm designed for time series clustering, named Soft-DTW k-means, was used by [79] to group electoral campaign donors based on their revenue. The authors identified two clusters of companies that experienced a sudden revenue increase after making donations, suggesting possible favoritism. Similarly, [25] applied a Latent Markov model to analyze data from contracting authorities with a higher risk of corruption based on a set of red flags. This model enables dynamic clustering into latent states and estimates transitions between states over time [25]. Their findings revealed a group of authorities exhibiting a heightened corruption risk based on the identified red flags.

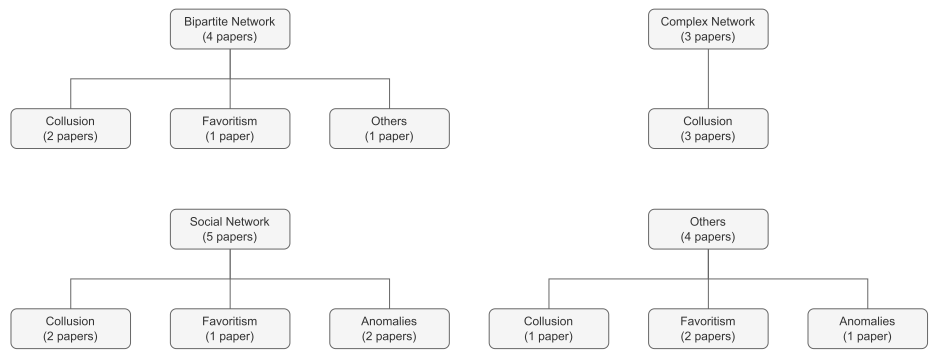

Figure 11 Overview of network science methods employed in the detection of different types of fraud

Another common strategy involves the use of the Isolation Forest algorithm to detect anomalies in the procurement process. It is an anomaly detection method based on binary trees, which iteratively splits the data, isolating outliers in the process. Unlike distance- or density-based methods, it offers reduced execution time and memory requirements while maintaining strong accuracy [13]. This model was employed by [17] to validate the results of the use of a knowledge graph built from public procurement data. This strategy was able to identify 27 possible cases of favoritism within the integrated data. [89] used the Isolation Forest method to flag public procurement processes that were protested by the entities involved or with complaints from external entities. The model was able to identify more than 90% of the processes with one of these anomalies during the tender and contracting stages.

4.6.2 Network science methods

The next field of study evaluated was NS. To enhance understanding of the results achieved with NS methods and enable a more robust comparison of similar studies, we categorized the papers based on the models used and the types of fraud analyzed. Figure 11 provides an overview of this classification.

The obtained results highlight several key use cases: the application of bipartite and complex networks for collusion detection, the use of social networks to identify collusion and public procurement anomalies, and the implementation of other network science methods to detect favoritism. The results for each of these use cases are detailed next.

Two studies employed bipartite networks to detect collusion. In [74], the authors used bid data to construct a co-bidding network, identifying potential collusion among firms. Their approach successfully detected communities of firms with similar bidding patterns. One particular community exhibited several concerning characteristics, including an unusually low number of single bids and high activity specialization. While these features can help narrow the scope of audits, they do not constitute definitive evidence of fraudulent behavior. Similarly, in [88], the authors built a bipartite network based on relationships between firms and their interactions with public entities to uncover potential corrupt schemes. Their findings revealed that companies involved in corruption cases often share personnel, and the authors introduced a corruption indicator to compare patterns across cases.

Three studies applied complex networks to detect collusion. In [16], the authors combined complex networks with centrality measures to identify companies engaged in collusion. Centrality measures help categorize network nodes based on their importance [16]. Using this strategy, the authors achieved 68% accuracy in detecting collusion within Brazilian federal tenders over a two-year period. Similarly, in [32], the authors also used complex networks to detect collusion using data from tenders. In this work, however, a different strategy called Nodes Detection using Network Science (NDNS) was employed. The NDNS is a model that uses complex networks to find the most relevant nodes in a multi-network scenario and is more efficient than traditional centrality measures when executing this task. The authors showed that this approach could achieve a 93% precision in detecting fraudulent values, compared to 38% precision with centrality measures.

Another widely used network science method for detecting fraud in public procurement is social network analysis. This approach was applied in 5 studies, with a focus on collusion detection (2 papers) [137, 139] and procurement anomaly detection (2 papers) [14, 91].

In [139], the authors built a bidder network to analyze characteristics that indicate collusion suspicion using data from construction project bidders. Based on these characteristics, they developed a classification mechanism with three levels of collusion suspicion, providing a tool to screen potential fraudulent behaviors. Similarly, in [137], the authors used social networks to detect collusion in the construction sector. They analyzed collusion case judgments to quantify the structural characteristics of collusion networks, revealing that these networks tend to be sparse, small, highly concentrated, and random. Their findings demonstrated that social networks can serve as a strong foundation for collusion detection.

Both works that targeted procurement anomaly detection leveraged red flags to identify suspicious bidders based on their characteristics and relationships. In [14], the authors constructed a social network using red flags such as shared addresses and partners among firms, effectively flagging anomalies in procurement. A case study demonstrated that this strategy successfully identified signs of potential fraud. Similarly, in [91], the authors ranked suspicious tenders based on red flags and bidder relationships, showing that this method can filter high-risk tenders, reducing the number of cases requiring manual review by experts.

The remaining studies applied alternative network science methods to detect favoritism in the procurement process. In [96], the authors developed a pattern-based framework for detecting anomalies in graphs, specifically identifying induced subgraphs, a pattern type often overlooked in prior research. Using a public procurement contract dataset, their framework successfully detected subgraph patterns indicative of fraudulent behavior. In [34], the authors also used a contract dataset in a graph-based solution to detect public procurement anomalies. Using a graph database, the authors implemented the proposed graph data model using data from more than 20,000 contracts. A market concentration indicator was used in the data, showing entities with a high score for this metric.

4.6.3 Statistical methods

Finally, we also evaluated the results achieved with statistical tools and models. Two primary groups stood out: linear regressions and logistic regressions. Figure 12 shows the results.

The first major group of statistical methods employed in the studies is linear regression (17 papers). Regression models are traditional statistical tools that identify relationships between a response variable and one or more predictors [133]. Linear regression

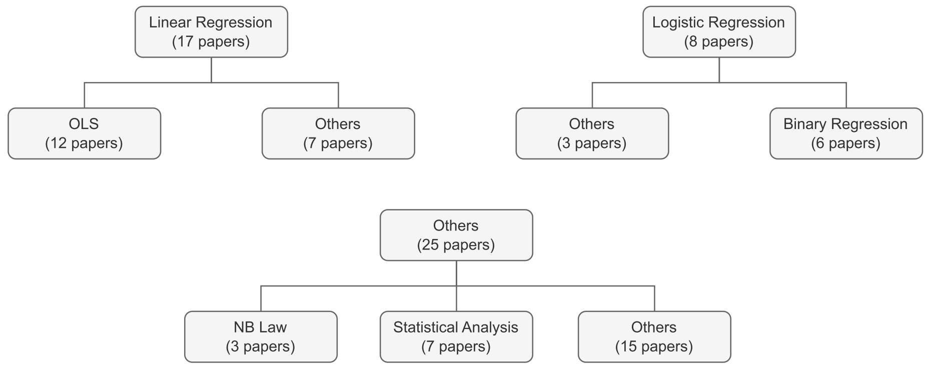

Figure 12 Overview of the statistics methods employed in the papers evaluated

specifically predicts a quantitative response based on one or multiple explanatory variables [55, 120]. Despite its simplicity, this model often provides a robust description of how predictors influence the response variable and can outperform more complex models, particularly when dealing with smaller datasets [45]. The Ordinary Least Squares (OLS) model was the most commonly used (12 papers) within the linear regression group, primarily for identifying favoritism in procurement processes. Several studies in this group explored the impact of transparency measures on fraud occurrence. In [5], the authors analyzed how transparency indexes, such as the publication of a call for tender, affect the proportion of contracts awarded with a single bid. Their results indicated that increased transparency reduces single bidding by 0.4–1.2%. Similarly, in [6], the authors examined the statistical relationship between the absence of a publicly available contract and undesirable procurement outcomes. Their findings revealed that projects without published contracts tend to have lower price savings and more frequent single bidding.

Other studies in this group used OLS to investigate how electoral campaign donations influence favoritism in public procurement, particularly after elections. In [38], the authors analyzed Croatian public procurement data and found that post-election political contributions increased donating firms' revenues by 27%, while firms affiliated with losing parties experienced a 12% revenue drop compared to pre-election levels. Similarly, in [27], the authors examined the influence of political donations on favoritism in federal procurement, revealing that firms donating to the president's party faced a higher favoritism risk. Their results also indicated that less independent public agencies were more susceptible to electoral donation influence.

The next major statistical method analyzed is logistic regression (8 papers), with 6 specifically employing binary regression. In [12], the authors examined company indicators associated with cartel behavior, finding that cartels exploit subcontracts and price similarities, with their tactics varying based on the number of participants. These findings highlight the need to analyze firm interactions when investigating collusive behavior. Similarly, in [36], the authors explored zero initial contract price decreases as a marker of collusion, concluding that higher contract values and increased participant numbers reduce firms' ability to engage in collusive practices. In [20], the authors used contract data to evaluate corruption control mechanisms in development aid, finding that increased donor oversight and broader tender access reduced single bidding by 3.6–4.3%. In [29], the authors developed a proxy measure for high-level corruption in European public procurement, based on single bidding in competitive markets. Their results showed that unpublished calls for tenders and non-quantitative evaluation criteria carried a higher corruption risk. In [18], the authors analyzed federal contract data to assess whether politicized agencies favor firms linked to political parties. Their findings suggested that political pressure leads to non-competitive procurement, with politicized departments more likely to engage in favoritism.

The final category of studies includes other statistical methods, encompassing 25 papers. The majority (7 papers) employed basic statistical tools, such as graphical plots and tables, to detect procurement fraud. Within this group, a significant emphasis was placed on identifying procurement anomalies. In [98], the authors used big data analytics to detect corruption, leveraging data visualization techniques to highlight procurement risks based on corruption indicators, such as the number of participants and auction duration. Their study identified geographic regions where over 60% of projects exhibited corruption indicators based on participation levels. Similarly, in [82], the authors applied statistical tools to analyze procurement data and flag suspicious transactions. Their results indicated that authorities should scrutinize tenders that involve frequent requests for documentation clarification and those where winning bids offer reductions exceeding 25%. In [8], the authors used statistical analysis of public bids to identify red flags, such as small companies exceeding annual revenue limits or linked to larger firms. In [39], the authors developed a data mining approach to detect anomalies in Brazilian municipalities' expenditures. Their methodology employed ranking and clustering techniques based on population size and micro-region, helping experts filter suspicious expenditures. Their results highlighted cities with spending patterns inconsistent with their peer groups, indicating potential fraud.

4.7 Real-world implementations

Despite numerous studies reporting promising results using various models for fraud detection, there is a clear gap in the real-world implementation of these methodologies for uncovering new cases of fraud and corruption. Among the 93 papers included in our systematic study, only 2 described the deployment of their proposed methodologies in systems designed for investigative agencies to detect fraud in public procurement. In [134], the authors collaborated with a procurement audit agency in Singapore to develop an explainable fraud detection system for public procurement. This system can analyze different fraud types, generate reports for auditors, and flag suspicious transactions for further investigation. The developed toolkit successfully identified 67.1% of transactions marked as suspicious. Similarly, in [7], the authors proposed a semi-automated pipeline for fraud detection in public bidding. Their approach included document classification and data quality evaluation modules. As a proof of concept, the proposed system achieved 75% accuracy in document classification. The framework was then applied to two real-world scenarios: (i) audit trails for fraud detection and (ii) a price database for overpricing detection. These applications demonstrated the system's potential in reducing specialists' workload by streamlining the search for irregularities.

5 Conclusions

This study aimed to systematically analyze primary research on the detection and characterization of corruption in government procurement. Using a systematic mapping review methodology, we established a structured protocol for identifying relevant studies. This protocol included defining research questions, setting inclusion and exclusion criteria, and selecting databases to conduct a comprehensive search using a query designed to retrieve as many relevant studies as possible.

The execution of this protocol initially identified 1529 studies. After screening and applying inclusion and exclusion criteria, 123 papers were deemed eligible. Following a further refinement step, 68 studies were included in our review. Additionally, we conducted five iterations of a snowballing procedure to capture studies that might have been missed in the database search, leading to the inclusion of 25 additional papers. In total, 93 studies were analyzed as part of our systematic review.

We analyzed and extracted information from all the papers in our set of primary works to answer the research questions defined during the planning phase of our systematic study. The initial evaluation highlighted the use of statistical methods to detect favoritism, ML models and network science to detect collusion. Works in these groups used, for the most part, data from bids and tenders to build the proposed solutions. We also found that papers that employed NLP techniques used larger datasets when compared to papers that did not employ NLP techniques.

We also proposed a two-level taxonomy to classify the features and red flags used in these studies. Given the large number of features found in the literature, this taxonomy improved the visualization and understanding of key results. Through this classification, we identified gaps in how public entity-related data is utilized in fraud detection, with few studies exploring its potential. We also noted a lack of research on red flags for collusion detection, which are significantly underutilized compared to red flags for other fraud types. We hope these findings will guide future research toward addressing these gaps and encourage the development of robust, data-driven anti-fraud systems.

Our analysis of reported results showed that ensemble methods outperformed other ML models in nearly all fraud categories. Network science methods also demonstrated strong results, despite being used in fewer studies compared to ML approaches. Additionally, statistical methods such as linear and logistic regression successfully identified key fraud indicators, with a particular emphasis on favoritism detection.

Despite the promising results reported in these studies, significant gaps remain in the practical application of fraud detection methodologies. One major limitation is the lack of real-world implementations, since very few studies reported the deployment of their methods in operational systems used by investigative agencies. Validating these models with real-world fraud investigation data, rather than in controlled research environments, is essential to improving the reliability and adoption of data-driven fraud detection tools. Addressing this gap will be crucial in advancing the development and deployment of effective public procurement fraud detection systems.

Appendix A: List of primary works

We extracted information related to the year, language and publication type for all works in our list of primary works. For the papers published in journals, we also extracted the impact factor of the publication vehicles. The Journal Citation Reports (JCR) measures how often articles in a journal are cited within a year.[6] Most of the primary works in our study were published in a journal with an impact factor between 0 and 3.

Table 14 List of primary works (1)

ID	Year	Language	Publication type	Impact factor	Published data?
1 [90]	2024	English	Journal	–	No
2 [128]	2023	English	Journal	6.3	No
3 [127]	2023	English	Journal	1.9	No
4 [49]	2023	English	Journal	1.7	On Request
5 [124]	2024	English	Journal	3.8	Yes
6 [103]	2024	English	Journal	1.2	No
7 [111]	2023	English	Journal	1.7	Yes
8 [24]	2022	English	Journal	3.0	No
9 [100]	2022	English	Journal	9.6	Yes
10 [52]	2021	English	Journal	0.9	No
11 [115]	2023	English	Conference	–	No
12 [30]	2022	English	Journal	2.6	Yes
13 [50]	2022	English	Journal	1.5	No
14 [35]	2021	English	Journal	6.9	No
15 [98]	2020	English	Conference	–	No
16 [119]	2023	English	Journal	2.0	Yes
17 [118]	2021	English	Journal	2.6	Yes
18 [19]	2020	English	Journal	1.6	No
19 [12]	2024	English	Journal	1.0	On Request
20 [96]	2023	English	Conference	–	Yes
21 [14]	2022	Portuguese	Conference	–	No
22 [58]	2023	Portuguese	Conference	–	No
23 [42]	2022	English	Conference	–	No
24 [10]	2021	English	Journal	3.0	No
25 [22]	2021	English	Journal	2.2	No
26 [105]	2022	English	Journal	4.6	Yes
27 [70]	2020	English	Conference	–	No
28 [46]	2023	English	Journal	1.0	Yes
29 [36]	2023	English	Journal	–	No
30 [81]	2020	English	Conference	–	No
31 [71]	2023	English	Conference	–	No

[6] https://clarivate.com/webofsciencegroup/essays/impact-factor/.

Table 15 List of primary works (2)

ID	Year	Language	Publication type	Impact factor	Published data?
32 [134]	2021	English	Journal	3.1	No
33 [32]	2020	Portuguese	Conference	–	Yes
34 [15]	2023	English	Conference	–	No
35 [20]	2020	English	Journal	1.8	Yes
36 [5]	2020	English	Journal	2.6	No
37 [62]	2023	English	Journal	5.9	Yes
38 [59]	2023	English	Journal	2.9	No
39 [4]	2020	English	Journal	1.7	No
40 [110]	2023	Portuguese	Conference	–	No
41 [61]	2022	English	Journal	6.9	No
42 [108]	2020	English	Journal	2.0	On Request
43 [16]	2022	Portuguese	Conference	–	No
44 [82]	2023	English	Journal	1.2	No
45 [72]	2022	English	Journal	2.3	On Request
46 [31]	2020	English	Journal	2.5	No
47 [29]	2020	English	Journal	4.6	No
48 [87]	2023	English	Journal	2.0	Yes
49 [1]	2023	English	Journal	–	No
50 [17]	2023	English	Journal	0.7	No
51 [74]	2021	English	Journal	1.3	Yes
52 [8]	2023	Portuguese	Conference	–	No
53 [112]	2024	English	Journal	1.6	Yes
54 [34]	2023	English	Conference	–	No
55 [37]	2023	English	Journal	2.8	No
56 [97]	2020	English	Journal	3.4	Yes
57 [41]	2021	English	Conference	–	No
58 [125]	2020	English	Journal	2.8	On Request
59 [28]	2023	English	Journal	1.8	No
60 [88]	2021	English	Journal	1.9	Yes
61 [126]	2021	English	Journal	3.4	No
62 [91]	2023	Portuguese	Conference	–	No

Table 16 List of primary works (3)

ID	Year	Language	Publication type	Impact factor	Published data?
63 [21]	2024	English	Journal	1.2	No
64 [132]	2023	English	Journal	1.2	No
65 [106]	2023	English	Journal	3.0	On Request
66 [107]	2020	English	Journal	4.1	Yes
67 [109]	2020	Portuguese	Journal	–	Yes
68 [25]	2022	English	Conference	–	No
69 [38]	2022	English	Journal	2.3	On Request
70 [102]	2024	English	Journal	3.3	No
71 [39]	2023	Portuguese	Conference	–	No
72 [79]	2023	Portuguese	Conference	–	No
73 [92]	2022	Portuguese	Conference	–	No
74 [26]	2021	English	Journal	5.3	On Request
75 [7]	2024	English	Journal	–	No
76 [53]	2024	English	Journal	3.1	No
77 [89]	2020	English	Conference	–	No
78 [27]	2023	English	Journal	5.2	Yes
79 [6]	2024	English	Journal	1.0	No
80 [23]	2023	English	Journal	2.1	On Request
81 [139]	2020	English	Journal	1.5	On Request
82 [18]	2023	English	Journal	5.0	Yes
83 [73]	2023	English	Journal	–	No
84 [78]	2024	English	Conference	–	No
85 [9]	2020	Portuguese	Journal	–	No
86 [129]	2023	English	Journal	1.9	No
87 [77]	2020	English	Journal	–	No
88 [101]	2022	Portuguese	Journal	–	No
89 [67]	2022	English	Conference	–	No
90 [137]	2021	English	Journal	3.1	On Request
91 [135]	2023	English	Journal	1.6	No
92 [131]	2021	English	Journal	5.3	On Request
93 [130]	2022	English	Journal	3.2	No

Appendix B: Overview of papers

Overview of all papers in the list of primary works by economic sector. Table 17 also includes the fraud group studied in the paper and the type of models employed. Three types of models were employed in the papers: Machine Learning (ML), Network Science (NS) and Statistics.

Table 17 Overview of papers by economic sector

Economic sector	Fraud group	Model type	Papers (ID and ref.)
Construction	Collusion	ML	2 [128], 3 [127], 9 [100], 13 [50], 22 [58]
Construction	Collusion	NS	81 [139], 86 [129]
Construction	Collusion	Statistics	19 [12], 26 [105], 41 [61], 42 [108], 66 [107], 67 [109], 92 [131], 93 [130]
Construction	Other Frauds	Statistics	63 [21]
General	Collusion	ML	4 [49], 10 [52], 27 [70], 85 [9]
General	Collusion	NS	5 [124], 43 [16], 51 [74], 60 [88], 90 [137]
General	Collusion	Statistics	87 [77]
General	Favoritism	ML	6 [103], 16 [119], 17 [118], 23 [42], 30 [81], 31 [71], 57 [41], 68 [25], 72 [79]
General	Favoritism	NS	20 [96], 28 [46], 54 [34], 61 [126]
General	Favoritism	Statistics	18 [19], 35 [20], 36 [5], 38 [59], 46 [31], 47 [29], 55 [37], 59 [28], 64 [132], 69 [38], 70 [102], 78 [27], 82 [18], 83 [73], 91 [135]
General	Other Frauds	ML	11 [115], 77 [89], 84 [78]
General	Other Frauds	NS	48 [87]
General	Other Frauds	Statistics	40 [110], 45 [72], 53 [112], 56 [97], 76 [53]
General	Anomalies	ML	12 [30], 14 [35], 50 [17]
General	Anomalies	NS	21 [14], 62 [91]
General	Anomalies	Statistics	15 [98], 49 [1], 52 [8], 58 [125], 71 [39], 73 [92], 88 [101]
Others	Collusion	NS	33 [32]
Others	Collusion	Statistics	24 [10], 25 [22], 29 [36], 37 [62], 80 [23]
Others	Anomalies	NS	34 [15]
Others	Anomalies	Statistics	1 [90], 44 [82]
Transport	Collusion	ML	7 [111]
Transport	Collusion	Statistics	39 [4], 65 [106], 74 [26], 89 [67]
Transport	Favoritism	Statistics	79 [6]
Transport	Anomalies	ML	8 [24]

Appendix C: Features

Groups of features and red flags used by the papers in the set of primary works.

Table 18 Red flags used by papers that studied favoritism

Fraud group	Feature group	Feature type	Papers (ID and ref.)
Favoritism	Contract	Contract Amendments	78 [27]
Favoritism	Contract	Contract Information	55 [37]
Favoritism	Contract	Missing Contract Information	36 [5]
Favoritism	Firm	Active Sanctions	78 [27]
Favoritism	Firm	Firm Location	64 [132]
Favoritism	Firm	Tax Haven Register	78 [27]
Favoritism	Others	Acquisition Period	70 [102]
Favoritism	Others	Bundled goods	70 [102]
Favoritism	Others	Conviction Rate	64 [132]
Favoritism	Others	Grant Period	70 [102]
Favoritism	Others	Predefined Labels	70 [102]
Favoritism	Others	Public Corruption Convictions	64 [132]
Favoritism	Others	State Integrity Investigation	64 [132]
Favoritism	Others	Unusual Transactions	70 [102]
Favoritism	Procurement Notice	Advertisement Period	46 [31], 55 [37]
Favoritism	Procurement Notice	Call Missing	36 [5], 46 [31], 47 [29], 78 [27]
Favoritism	Procurement Notice	Call Modified	46 [31]
Favoritism	Procurement Notice	Eligibility Criteria	31 [71], 46 [31]
Favoritism	Procurement Notice	Evaluation Criteria	46 [31], 47 [29], 55 [37], 68 [25]
Favoritism	Notice	Imprecise CPV Codes	36 [5]
Favoritism	Notice	Missing Contract Duration	36 [5]
Favoritism	Notice	Missing Language Information	36 [5]
Favoritism	Notice	Selection Criteria	6 [103], 36 [5]
Favoritism	Public Entity	Frequency of Same Winner	55 [37]
Favoritism	Tender	Decision Period	46 [31], 47 [29]
Favoritism	Tender	Evaluation Period	55 [37]
Favoritism	Tender	Exclusions	55 [37]
Favoritism	Tender	Modified Procedure	55 [37]
Favoritism	Tender	Restricted Procedure	46 [31], 47 [29], 55 [37], 68 [25], 78 [27]
Favoritism	Tender	Single Bid	35 [20], 46 [31], 47 [29], 55 [37], 68 [25], 78 [27]
Favoritism	Tender	Submission Period	47 [29], 68 [25]

Table 19 Red flags used by papers that studied procurement anomalies

Fraud group	Feature group	Feature type	Papers (ID and ref.)
Anomalies	Bidder Relationships	Firm Link	21 [14], 62 [91], 75 [7]
Anomalies	Bidder Relationships	Political Link	62 [91]
Anomalies	Bidder Relationships	Public Entity Link	62 [91], 75 [7]
Anomalies	Bidder Relationships	Same Address	21 [14], 62 [91], 75 [7]
Anomalies	Bidder Relationships	Same E-mails	21 [14], 62 [91], 75 [7]
Anomalies	Bidder Relationships	Same Phone Numbers	21 [14], 62 [91], 75 [7]
Anomalies	Contract	Contract Information	8 [24]
Anomalies	Firm	Active Sanctions	21 [14], 62 [91], 75 [7]
Anomalies	Firm	Bid Before Start Date	21 [14], 62 [91], 75 [7]
Anomalies	Firm	Contradictory CNAE	62 [91]
Anomalies	Firm	Frequent Loser	21 [14], 62 [91], 75 [7]
Anomalies	Firm	Frequent Single Bidder	58 [125]
Anomalies	Firm	Frequent Winner	21 [14], 62 [91], 75 [7]
Anomalies	Firm	Inactive CNPJ	21 [14], 62 [91], 75 [7]
Anomalies	Firm	Missing Electrical Register	62 [91]
Anomalies	Firm	No Employees	58 [125]
Anomalies	Firm	Small Company Revenue	62 [91]
Anomalies	Firm	Suspicious Procurement Share	58 [125]
Anomalies	Firm	Suspicious Public Subsidy	58 [125]
Anomalies	Procurement Notice	Advertisement Period	8 [24]
Anomalies	Procurement Notice	Call Missing	8 [24]
Anomalies	Procurement Notice	Call With Urgency Clauses	8 [24]
Anomalies	Procurement Notice	Eligibility Criteria	8 [24]
Anomalies	Procurement Notice	Evaluation Criteria	8 [24]
Anomalies	Procurement Notice	Selection Criteria	8 [24]
Anomalies	Procurement Notice	Subcontracting Rules	8 [24]
Anomalies	Procurement Notice	Tender Page Count	8 [24]
Anomalies	Procurement Notice	WV-Hours Share	8 [24]
Anomalies	Procurement Notice	Word Count	8 [24]
Anomalies	Procurement Notice	Worksite Verification	8 [24]
Anomalies	Public Entity	Firm List Preference	8 [24]
Anomalies	Public Entity	Firm Other Preference	8 [24]
Anomalies	Public Entity	Outsider Contact Point	8 [24]
Anomalies	Tender	Documents Verification	8 [24]
Anomalies	Tender	Lowest Bid Lost	62 [91]
Anomalies	Tender	Restricted Procedure	8 [24]
Anomalies	Tender	Single Bid	21 [14], 62 [91], 75 [7]
Anomalies	Tender	Single Offer Forbidden	8 [24]

Table 20 Red flags used by papers that studied collusion and other frauds

Fraud group	Feature group	Feature type	Papers (ID and ref.)
Collusion	Bidder Relationships	Public Entity Link	87 [77]
Collusion	Contract	Contract Amendments	87 [77]
Collusion	Contract	Contract Information	87 [77]
Collusion	Contract	Missing Contract Information	87 [77]
Collusion	Firm	Market Share	87 [77]
Collusion	Firm	Number of Victories	87 [77]
Collusion	Firm	Overpricing	87 [77]
Collusion	Firm	Position Average	87 [77]
Collusion	Firm	Ratio Dropout	87 [77]
Collusion	Firm	Waiver	87 [77]
Collusion	Tender	Dropout	87 [77]
Collusion	Tender	Exclusions	87 [77]
Collusion	Tender	Single Bid	87 [77]
Collusion	Tender	Weak Competition	87 [77]
Other Frauds	Firm	New Firm	11 [115]
Other Frauds	Firm	Tax Haven Register	11 [115]
Other Frauds	Procurement Notice	Advertisement Period	11 [115]
Other Frauds	Procurement Notice	Call Missing	11 [115]
Other Frauds	Public Entity	DCVB	53 [112]
Other Frauds	Tender	Decision Period	11 [115]
Other Frauds	Tender	Restricted Procedure	11 [115]
Other Frauds	Tender	Single Bid	11 [115]

Table 21 Features used by papers that studied anomalies and other frauds

Fraud group	Feature group	Papers (ID and ref.)
Other Frauds	Bid	45 [72]
Other Frauds	Contract	11 [115], 45 [72], 48 [87], 56 [97]
Other Frauds	Firm	11 [115], 48 [87]
Other Frauds	Others	45 [72], 56 [97], 63 [21], 76 [53], 77 [89], 84 [78]
Other Frauds	Notice	11 [115]
Other Frauds	Public Entity	45 [72], 48 [87], 53 [112]
Other Frauds	Tender	11 [115], 40 [110], 45 [72], 56 [97]
Anomalies	Bid	1 [90], 8 [24], 12 [30], 15 [98], 52 [8]
Anomalies	Relationships	21 [14], 62 [91], 75 [7]
Anomalies	Contract	8 [24], 12 [30], 14 [35], 44 [82], 88 [101]
Anomalies	Firm	12 [30], 14 [35], 21 [14], 32 [134], 34 [15], 49 [1], 52 [8], 58 [125], 62 [91], 73 [92], 75 [7]
Anomalies	Others	12 [30], 14 [35], 32 [134], 34 [15], 44 [82], 49 [1], 50 [17], 52 [8], 58 [125], 71]
Anomalies	Notice	8 [24]
Anomalies	Public Entity	8 [24], 12 [30], 14 [35]
Anomalies	Tender	8 [24], 12 [30], 14 [35], 15 [98], 21 [14], 32 [134], 34 [15], 44 [82], 52 [8], 58 [125], 62 [91], 73 [92], 75 [7]

Table 22 Features used by papers that studied collusion and favoritism

Fraud group	Feature group	Papers (ID and ref.)
Collusion	Bid	9 [100], 22 [58], 24 [10], 25 [22], 26 [105], 37 [62], 41 [61], 42 [108], 51 [74], 65 [106], 66 [107], 67 [109], 80 [23], 87 [77], 93 [130]
Collusion	Relationships	87 [77]
Collusion	Contract	9 [100], 13 [50], 29 [36], 80 [23], 87 [77], 89 [67]
Collusion	Firm	9 [100], 25 [22], 37 [62], 39 [4], 41 [61], 42 [108], 51 [74], 66 [107], 67 [109], 81 [139], 85 [9], 87 [77], 89 [67], 92 [131], 93 [130]
Collusion	Others	5 [124], 7 [111], 9 [100], 27 [70], 29 [36], 33 [32], 39 [4], 43 [16], 51 [74], 60 [88], 66 [107], 86 [129], 87 [77], 90 [137], 92 [131], 93 [130]
Collusion	Public Entity	29 [36], 87 [77]
Collusion	Screens	2 [128], 3 [127], 7 [111], 9 [100], 10 [52], 13 [50], 22 [58]
Collusion	Tender	2 [128], 3 [127], 4 [49], 7 [111], 9 [100], 13 [50], 19 [12], 22 [58], 25 [22], 26 [105], 29 [36], 37 [62], 39 [4], 41 [61], 42 [108], 51 [74], 65 [106], 66 [107], 67 [109], 74 [26], 81 [139], 85 [9], 86 [129], 87 [77], 92 [131], 93 [130]
Favoritism	Bid	17 [118], 18 [19], 23 [42], 38 [59], 54 [34], 57 [41], 72 [79]
Favoritism	Contract	17 [118], 18 [19], 20 [96], 28 [46], 35 [20], 36 [5], 38 [59], 46 [31], 47 [29], 54 [34], 55 [37], 59 [28], 64 [132], 69 [38], 78 [27], 79 [6], 82 [18]
Favoritism	Firm	17 [118], 18 [19], 23 [42], 28 [46], 38 [59], 46 [31], 47 [29], 54 [34], 57 [41], 64 [132], 69 [38], 72 [79], 78 [27], 82 [18], 91 [135]
Favoritism	Others	16 [119], 17 [118], 18 [19], 20 [96], 30 [81], 35 [20], 36 [5], 46 [31], 47 [29], 57 [41], 59 [28], 61 [126], 64 [132], 69 [38], 70 [102], 72 [79], 78 [27], 79 [6], 82 [18], 83 [73], 91 [135]
Favoritism	Notice	6 [103], 20 [96], 31 [71], 36 [5], 46 [31], 47 [29], 55 [37], 68 [25], 78 [27]
Favoritism	Public Entity	17 [118], 23 [42], 36 [5], 46 [31], 54 [34], 55 [37], 57 [41], 59 [28], 69 [38], 78 [27], 79 [6], 83 [73]
Favoritism	Tender	6 [103], 17 [118], 18 [19], 20 [96], 23 [42], 28 [46], 31 [71], 35 [20], 36 [5], 38 [59], 46 [31], 47 [29], 54 [34], 55 [37], 57 [41], 59 [28], 64 [132], 68 [25], 69 [38], 72 [79], 78 [27], 79 [6], 82 [18], 83 [73], 91 [135]

Appendix D: Targets

Groups of features that were used as targets for the methods analyzed and papers that used unsupervised methodologies.

Table 23 Targets used by the primary works evaluated

Fraud group	Type	Variable	Papers (ID and ref.)
Collusion	Binary	Collusive Indicator	2 [128], 3 [127], 4 [49], 9 [100], 10 [52], 13 [50], 19 [12], 22 [58], 26 [105], 29 [36], 33 [32], 37 [62], 66 [107], 67 [109], 90 [137], 92 [131]
Collusion	Binary	Others	39 [4]
Collusion	Linear	Collusive Indicator	93 [130]
Collusion	Linear	Others	25 [22], 80 [23]
Collusion	Proxy	Corruption Indicator	27 [70]
Collusion	Proxy	Others	43 [16]
Collusion	Proxy	Red Flags Set	5 [124], 87 [77], 89 [67]
Favoritism	Binary	Others	16 [119], 82 [18]
Favoritism	Linear	Contract Indicator	18 [19], 64 [132]
Favoritism	Linear	Others	38 [59], 46 [31], 91 [135]
Favoritism	Proxy	Corruption Indicator	78 [27]
Favoritism	Proxy	Others	6 [103]
Favoritism	Proxy	Red Flags Set	31 [71], 68 [25]
Favoritism	Proxy	Single Bid	23 [42], 30 [81], 35 [20], 36 [5], 47 [29], 57 [41], 59 [28], 61 [126], 79 [6], 83 [73]
Other Frauds	Binary	Contract Indicator	11 [115], 56 [97]
Other Frauds	Binary	Others	84 [78]
Other Frauds	Linear	Contract Indicator	45 [72]
Other Frauds	Linear	Others	53 [112]
Other Frauds	Proxy	Corruption Indicator	63 [21]
Anomalies	Binary	Contract Indicator	14 [35]
Anomalies	Binary	Others	49 [1], 58 [125]
Anomalies	Proxy	Corruption Indicator	8 [24], 12 [30], 14 [35], 44 [82]
Anomalies	Proxy	Red Flags Set	21 [14], 62 [91]

Table 24 Papers that employed unsupervised methodologies

Model type	Method group	Fraud group	Papers (ID and ref.)
Machine Learning	Clusters	Collusion	7 [111]
Machine Learning	Clusters	Favoritism	17 [118], 72 [79]
Machine Learning	Ensemble Methods	Other Frauds	77 [89]
Machine Learning	Ensemble Methods	Anomalies	50 [17]
Machine Learning	Others	Collusion	85 [9]
Machine Learning	Others	Favoritism	17 [118]
Network Science	Networks	Collusion	51 [74], 60 [88], 81, 86
Network Science	Networks	Favoritism	28 [46]
Network Science	Networks	Other Frauds	48 [87]
Network Science	Networks	Anomalies	34 [15]
Network Science	Others	Favoritism	20 [96], 54 [34]
Statistics	Linear Regression	Favoritism	69 [38]
Statistics	Others	Collusion	24 [10], 41 [61], 42 [108],65 [106], 74 [26]
Statistics	Others	Favoritism	55 [37], 70 [102]
Statistics	Others	Other Frauds	40 [110], 76 [53]
Statistics	Others	Anomalies	1 [90], 15 [98], 52 [8],71 [39], 73 [92], 88 [101]

Appendix E: Results

Average results reported by the ML models that reported the results using numeric evaluation metrics.

Table 25 Average results reported by the ensemble methods (1)

Method	Fraud group	Measure type	Result	Papers (ID and ref.)
Ada Boost	Collusion	Accuracy	84.05	9 [100]
Ada Boost	Collusion	Balanced Accuracy	71.915	9 [100]
Ada Boost	Collusion	False Negatives (%)	8.08	9 [100]
Ada Boost	Collusion	False Positives (%)	7.86	9 [100]
Extra Trees	Collusion	Accuracy	86.13	9 [100]
Extra Trees	Collusion	Balanced Accuracy	73.55	9 [100]
Extra Trees	Collusion	False Negatives (%)	7.15	9 [100]
Extra Trees	Collusion	False Positives (%)	15.50	9 [100]
Gradient Boosting	Collusion	Accuracy	80.35	9 [100]
Gradient Boosting	Collusion	Balanced Accuracy	70.91	9 [100]
Gradient Boosting	Collusion	False Negatives (%)	9.74	9 [100]
Gradient Boosting	Collusion	False Positives (%)	9.83	9 [100]
Random Forest	Collusion	Accuracy	79.77	2 [128], 3 [127], 9 [100], 10 [52], 13 [50]
Random Forest	Collusion	Area Under Curve	91.20	22 [58]
Random Forest	Collusion	Balanced Accuracy	73.20	9 [100]
Random Forest	Collusion	F1 Score	86.05	2 [128], 27 [70]
Random Forest	Collusion	False Negatives (%)	8.02	9 [100]
Random Forest	Collusion	False Positives (%)	7.40	9 [100]
Random Forest	Collusion	Precision	88.60	27 [70]
Random Forest	Collusion	Recall	95.30	27 [70]
Super Learner	Collusion	Accuracy	88.05	2 [128], 10 [52], 13 [50]
Super Learner	Collusion	F1 Score	79.60	2 [128]

Table 26 Average results reported by the ensemble methods (2)

Method	Fraud group	Measure type	Result	Papers (ID and ref.)
Gradient Boosting	Anomalies	Accuracy	85.0	12 [30]
Gradient Boosting	Anomalies	Area Under Curve	85.31	14 [35]
Gradient Boosting	Anomalies	Area Under Curve-PR	6.03	14 [35]
Gradient Boosting	Anomalies	Brier Score	5.22	14 [35]
Gradient Boosting	Anomalies	False Negative Rate	24.0	12 [30]
Gradient Boosting	Anomalies	False Positive Rate	9.0	12 [30]
Gradient Boosting	Anomalies	Mean Average Precision (100)	59.44	14 [35]
Gradient Boosting	Anomalies	Mean Average Precision (1000)	28.17	14 [35]
Gradient Boosting	Anomalies	Normalized Discounted Cumulative Gain (100)	81.33	14 [35]
Gradient Boosting	Anomalies	Normalized Discounted Cumulative Gain (1000)	65.89	14 [35]
Random Forest	Anomalies	Accuracy	79.0	12 [30]
Random Forest	Anomalies	Area Under Curve	84.58	14 [35]
Random Forest	Anomalies	F measure	91.75	8 [24]
Random Forest	Anomalies	False Negative Rate	10.0	12 [30]
Random Forest	Anomalies	False Positive Rate	28.0	12 [30]
Random Forest	Anomalies	Precision	98.55	8 [24]
Random Forest	Anomalies	Recall	85.95	8 [24]
Gradient Boosting	Others	Area Under Curve-PR	80.70	84 [78]
Random Forest	Others	Accuracy	53.67	11 [115]
Random Forest	Others	F1 Score	57.67	11 [115]
Random Forest	Others	Precision	75.33	11 [115]
Random Forest	Others	Recall	55.67	11 [115]
StackedEnsemble	Others	Area Under Curve-PR	82.70	84 [78]
XGBoost	Others	Area Under Curve-PR	81.10	84 [78]

Table 27 Average results reported by the linear models

Method	Fraud group	Measure type	Result	Papers (ID/ref.)
Lasso Regression	Collusion	Accuracy	88.0	10 [52]
Logistic Regression	Collusion	Accuracy	77.7	2 [128]
Logistic Regression	Collusion	F1 Score	78.9	2 [128]
Passive-Aggressive Model	Collusion	F1 Score	93.4	27 [70]
Passive-Aggressive Model	Collusion	Precision	90.7	27 [70]
Passive-Aggressive Model	Collusion	Recall	96.3	27 [70]
Perceptron	Collusion	F1 Score	92.2	27 [70]
Perceptron	Collusion	Precision	89.2	27 [70]
Perceptron	Collusion	Recall	95.4	27 [70]
Ridge Classifier	Collusion	F1 Score	92.9	27 [70]
Ridge Classifier	Collusion	Precision	89.7	27 [70]
Ridge Classifier	Collusion	Recall	96.3	27 [70]
SGD	Collusion	Accuracy	67.79	9 [100]
SGD	Collusion	Balanced Accuracy	53.71	9 [100]
SGD	Collusion	F1 Score	92.87	27 [70]
SGD	Collusion	False Negatives (%)	14.94	9 [100]
SGD	Collusion	False Positives (%)	17.28	9 [100]
SGD	Collusion	Precision	89.97	27 [70]
SGD	Collusion	Recall	95.97	27 [70]
Logistic Regression	Favoritism	Accuracy	72.5	30 [81]
Logistic Regression	Favoritism	Area Under Curve	59.5	30 [81]
Logistic Regression	Favoritism	Precision	57.5	30 [81]
Logistic Regression	Favoritism	Recall	28.0	30 [81]
Lasso Regression	Anomalies	Brier Score	6.17	14 [35]
Lasso Regression	Anomalies	F measure	25.5	8 [24]
Lasso Regression	Anomalies	Mean Average Precision (100)	35.833	14 [35]
Lasso Regression	Anomalies	Mean Average Precision (1000)	19.28	14 [35]
Lasso Regression	Anomalies	Normalized Discounted Cumulative Gain (100)	69.44	14 [35]
Lasso Regression	Anomalies	Normalized Discounted Cumulative Gain (1000)	60.67	14 [35]
Lasso Regression	Anomalies	Precision	24.55	8 [24]
Lasso Regression	Anomalies	Recall	26.55	8 [24]
Logistic Regression	Anomalies	Accuracy	64.0	12 [30]
Logistic Regression	Anomalies	Area Under Curve	65.07	14 [35]
Logistic Regression	Anomalies	Area Under Curve-PR	8.97	14 [35]
Logistic Regression	Anomalies	False Negative Rate	59.0	12 [30]
Logistic Regression	Anomalies	False Positive Rate	29.0	12 [30]
Ordinary Least Squares	Anomalies	F measure	25.90	8 [24]
Ordinary Least Squares	Anomalies	Precision	25.60	8 [24]
Ordinary Least Squares	Anomalies	Recall	26.20	8 [24]
Ridge Regression	Anomalies	F measure	25.25	8 [24]
Ridge Regression	Anomalies	Precision	23.75	8 [24]
Ridge Regression	Anomalies	Recall	27.05	8 [24]

Table 28 Average results reported by the Nearest Neighbors ML group

Method	Fraud group	Measure type	Result	Papers (ID and ref.)
K-Nearest Neighbors	Collusion	F1 Score	87.60	27 [70]
K-Nearest Neighbors	Collusion	Precision	80.90	27 [70]
K-Nearest Neighbors	Collusion	Recall	95.50	27 [70]
Nearest Centroid	Collusion	F1 Score	87.80	27 [70]
Nearest Centroid	Collusion	Precision	85.40	27 [70]
Nearest Centroid	Collusion	Recall	90.40	27 [70]

Table 29 Average results reported by the Neural Networks ML group

Method	Fraud group	Measure type	Result	Papers (ID and ref.)
Bi-LSTM	Collusion	F1 Score	91.65	27 [70]
Bi-LSTM	Collusion	Precision	91.20	27 [70]
Bi-LSTM	Collusion	Recall	92.15	27 [70]
Bottleneck	Collusion	F1 Score	93.0	27 [70]
Bottleneck	Collusion	Precision	92.80	27 [70]
Bottleneck	Collusion	Recall	93.20	27 [70]
Conv. Neural Network	Collusion	Accuracy	91.60	4 [49]
Conv. Neural Network	Collusion	Precision	87.73	4 [49]
Deep Neural Network	Collusion	F1 Score	89.4	27 [70]
Deep Neural Network	Collusion	Precision	89.53	27 [70]
Deep Neural Network	Collusion	Recall	89.63	27 [70]
Multi-Layer Perceptron	Collusion	Accuracy	77.08	9 [100]
Multi-Layer Perceptron	Collusion	Area Under Curve	95.43	22 [58]
Multi-Layer Perceptron	Collusion	Balanced Accuracy	53.56	9 [100]
Multi-Layer Perceptron	Collusion	False Negatives (%)	14.65	9 [100]
Multi-Layer Perceptron	Collusion	False Positives (%)	8.62	9 [100]
Bi-LSTM	Favoritism	F1 Score	85.20	31 [71]
Bi-LSTM	Favoritism	Precision	86.30	31 [71]
Bi-LSTM	Favoritism	Recall	84.0	31 [71]
Bottleneck	Favoritism	F1 Score	84.80	31 [71]
Bottleneck	Favoritism	Precision	84.50	31 [71]
Bottleneck	Favoritism	Recall	85.10	31 [71]
Conv. Neural Network	Favoritism	Accuracy	82.50	16 [119]
Conv. Neural Network	Favoritism	Loss	57.67	16 [119]

Table 30 Average results reported by the Tree-Based ML group

Method	Fraud group	Measure type	Result	Papers (ID and ref.)
Classification Tree	Collusion	Accuracy	56.81	2 [128], 3 [127]
Classification Tree	Collusion	F1 Score	80.60	2 [128]
Decision Tree	Collusion	Area Under Curve	97.02	22 [58]

Table 31 Average results reported by the Clusters ML group

Method	Fraud group	Measure type	Result	Papers (ID and ref.)
PCA	Favoritism	Accuracy	90.0	17 [118]
PCA	Favoritism	Area Under Curve	91.0	17 [118]
PCA	Favoritism	F Score	93.0	17 [118]
PCA	Favoritism	Precision	92.0	17 [118]
PCA	Favoritism	Recall	89.0	17 [118]

Table 32 Average results reported by the ML models in the method group 'Others'

Method	Fraud group	Measure type	Result	Papers (ID and ref.)
Gaussian Process	Collusion	Accuracy	55.90	9 [100]
Gaussian Process	Collusion	Balanced Accuracy	50.22	9 [100]
Gaussian Process	Collusion	False Negatives (%)	40.53	9 [100]
Gaussian Process	Collusion	False Positives (%)	5.81	9 [100]
K-Nearest Neighbors	Collusion	Accuracy	76.31	9 [100]
K-Nearest Neighbors	Collusion	Balanced Accuracy	60.90	9 [100]
K-Nearest Neighbors	Collusion	False Negatives (%)	12.83	9 [100]
K-Nearest Neighbors	Collusion	False Positives (%)	10.86	9 [100]
Support Vector Machines	Collusion	Accuracy	64.05	9 [100], 10 [52]
Support Vector Machines	Collusion	Area Under Curve	96.43	22 [58]
Support Vector Machines	Collusion	Balanced Accuracy	63.72	9 [100]
Support Vector Machines	Collusion	F1 Score	90.05	27 [70]
Support Vector Machines	Collusion	False Negatives (%)	17.31	9 [100]
Support Vector Machines	Collusion	False Positives (%)	21.80	9 [100]
Support Vector Machines	Collusion	Precision	89.35	27 [70]
Support Vector Machines	Collusion	Recall	91.0	27 [70]
BERT	Favoritism	F1 Score	86.2	[71]
BERT	Favoritism	Precision	83.8	[71]
BERT	Favoritism	Recall	88.8	[71]
Confident Learning	Favoritism	False Negatives (%)	14.0	23 [42]
Confident Learning	Favoritism	False Positives (%)	29.0	23 [42]
Confident Learning	Favoritism	True Negatives (%)	71.0	23 [42]
Confident Learning	Favoritism	True Positives (%)	86.0	23 [42]
Support Vector Machines	Favoritism	Accuracy	79.0	17 [118], 30 [81]
Support Vector Machines	Favoritism	Area Under Curve	70.0	17 [118], 30 [81]
Support Vector Machines	Favoritism	F Score	96.0	17 [118]
Support Vector Machines	Favoritism	Precision	70.0	17 [118], 30 [81]
Support Vector Machines	Favoritism	Recall	51.33	17 [118], 30 [81]
K-Nearest Neighbors	Others	Accuracy	46.33	11 [115]
K-Nearest Neighbors	Others	F1 Score	47.0	11 [115]
K-Nearest Neighbors	Others	Precision	75.33	11 [115]
K-Nearest Neighbors	Others	Recall	43.67	11 [115]

Abbreviations
ML, Machine Learning; SLR, Systematic Literature Review; SLM, Systematic Literature Mapping; NS, Network Science; AI, Artificial Intelligence; PRISMA, Preferred Reporting Items for Systematic Reviews and Meta-Analyses; BS, Backward Snowballing; FS, Forward Snowballing; RQ, Research question; NLP, Natural Language Processing; DNNs, Deep Neural Networks; LSTM, Long Short-Term Memory networks; SNA, Social Network Analysis; OLS, Ordinary Least Squares; EDA, Exploratory Data Analysis; DCVB, Discrete Contract Value to Budget; CRI, Corruption Risk Indicator; NDNS, Nodes Detection using Network Science; JCR, Journal Citation Reports.

Acknowledgements
Not applicable.

Author contributions
All authors contributed equally to this work.

Funding information
This research was funded by the Coordination for the Improvement of Higher Education Personnel Foundation (CAPES) and the Public Prosecutor's Office of Santa Catarina (MPSC).

Data availability
No datasets were generated or analysed during the current study.

Declarations

Ethics approval and consent to participate
Not applicable.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Author details
[1]Graduate Program in Computer Science, Federal University of Santa Catarina, Florianópolis, Santa Catarina, Brazil. [2]Graduate Program in Computer Science and Computational Mathematics, University of São Paulo, São Carlos, São Paulo, Brazil.

Received: 25 February 2025 Accepted: 6 July 2025 Published online: 22 July 2025

References

1. Alford LC, Hansen SC, Hermis JM (2023) Predicting federal contractor misconduct. J Forensic Account Res: 1–29
2. Alon N, Spencer JH (2015) The probabilistic method. Wiley, New York
3. Amthauer J, Fleiß J, Guggi F, et al (2023) Ready or not? A systematic review of case studies using data-driven approaches to detect real-world antitrust violations. Comput Law Secur Rev 49:105807
4. Barrus D, Scott F (2020) Single bidders and tacit collusion in highway procurement auctions. J Ind Econ 68(3):483–522
5. Bauhr M, Czibik Á, de Fine Licht J, et al (2020) Lights on the shadows of public procurement: transparency as an antidote to corruption. Governance 33(3):495–523
6. Bolcha P, Vozárová P, Fanta M (2024) Transparency meets efficiency: the case of public procurement in the Czech Republic. Public Works Manag Policy 29(3):296–331
7. Brandão MA, Reis AP, Mendes BM, et al (2024) Plus: a semi-automated pipeline for fraud detection in public bids. Digit Gov Res Pract 5(1):1–16
8. Braz CS, Mendes BM, Oliveira GP, et al (2023) Análise de irregularidades em licitações públicas com foco em empresas de pequeno porte. In: Anais do XI Workshop de Computação Aplicada em Governo Eletrônico, SBC, pp 94–105
9. Brito Ramalhoa HM, Cavalcanti de Almeidaa AT, Alves Fragaa A (2020) Detecção de casos suspeitos de conluio em licitações públicas: uma aplicação do algoritmo a priori de aprendizado de máquina para o estado da paraíba. Teor Prát Adm 10(2)
10. Busu M, Busu C (2021) Detecting bid-rigging in public procurement. A cluster analysis approach. Adm Sci 11(1):13
11. Buyya R, Calheiros RN, Dastjerdi AV (2016) Big data: principles and paradigms, 1st edn. Morgan Kaufmann Publishers Inc., San Francisco
12. Carbone C, Calderoni F, Jofre M (2024) Bid-rigging in public procurement: cartel strategies and bidding patterns. Crime Law Soc Change: 1–33
13. Chabchoub Y, Togbe MU, Boly A, et al (2022) An in-depth study and improvement of isolation forest. IEEE Access 10:10219–10237
14. Costa LL, Reis AP, Bacha CA, et al (2022) Alertas de fraude em licitações: Uma abordagem baseada em redes sociais. In: Brazilian workshop on Social Network Analysis and Mining (BraSNAM), SBC, pp 37–48
15. Cruz E, Ramirez N, Parra E, et al (2023) Detecting anomalous suppliers in electronic procurement systems using gaan. In: 2023 IEEE international conference on big data (BigData). IEEE, pp 4147–4155
16. da Silva Pereira AK, Vita YM, da Silva Felix G, et al (2022) Usando redes complexas na identificação de empresas fraudulentas em licitações públicas. In: Anais do X Workshop de Computação Aplicada em Governo Eletrônico, SBC, pp 13–24
17. Dahbi KY, Chiadmi D, Lamharhar H (2023) Knowledge graph based representation to extract value from open government data. Int J Adv Comput Sci Appl 14(3)
18. Dahlström C, Fazekas M, Lewis DE (2021) Partisan procurement: contracting with the United States federal government, 2003–2015. Am J Polit Sci 65(3):652–669
19. Dávid-Barrett E, Fazekas M (2020) Grand corruption and government change: an analysis of partisan favoritism in public procurement. Eur J Crim Policy Res 26(4):411–430
20. Dávid-Barrett E, Fazekas M, Hellmann O, et al (2020) Controlling corruption in development aid: new evidence from contract-level data. Stud Comp Int Dev 55:481–515
21. Davis DB, Mendoza RU, Yap JK (2024) Corruption risk and political dynasties: exploring the links using public procurement data in the Philippines. Econ Gov 25(1):81–109
22. de Andrade Lima RC, Resende GM (2021) Using the Moran's I to detect bid rigging in Brazilian procurement auctions. Ann Reg Sci 66(2):237–254
23. De Leverano A (2023) Clustered bids in first-price auctions: collusion or competition? Econ Lett 233:111393
24. Decarolis F, Giorgiantonio C (2022) Corruption red flags in public procurement: new evidence from Italian calls for tenders. EPJ Data Sci 11(1):16
25. Del Sarto S, Coppola P, Troìa M (2021) A latent Markov approach for clustering contracting authorities over time using public procurement red flags. In: Convegno della Società Italiana di Statistica. Springer, Berlin, pp 57–74
26. Erfani A, Zhang K, Cui Q (2021) Tab bid irregularity: data-driven model and its application. J Manag Eng 37(5):04021055
27. Fazekas M, Ferrali R, Wachs J (2023) Agency independence, campaign contributions, and favoritism in us federal government contracting. J Public Adm Res Theory 33(2):262–278
28. Fazekas M, Hellmann O (2023) Elections and corruption: incentives to steal or incentives to invest? Stud Comp Int Dev: 1–24
29. Fazekas M, Kocsis G (2020) Uncovering high-level corruption: cross-national objective corruption risk indicators using public procurement data. Br J Polit Sci 50(1):155–164
30. Fazekas M, Sberna S, Vannucci A (2022) The extra-legal governance of corruption: tracing the organization of corruption in public procurement. Governance 35(4):1139–1161
31. Fazekas M, Wachs J (2020) Corruption and the network structure of public contracting markets across government change. Polit Gov 8(2):153–166

32. Figueiredo B, Nakamura F, Felix G, et al (2020) Usando análises sociais na identificação de nós relevantes em um cenário multi-redes: Operação licitante fantasma, um estudo de caso. In: Anais do VIII Workshop de Computação Aplicada em Governo Eletrônico, SBC, pp 108–119
33. Florentino ÉS, Goldschmidt RR, Cavalcanti MC (2022) Identificando suspeitos de crimes por meio de interações implícitas no youtube. iSys-Braz J Inf Syst 15(1):3–14
34. Fortuny M, Sandoval R, Riofrío D, et al (2023) Building a graph database for electronic inverse auction: Ecuador's case study. In: 2023 IEEE seventh Ecuador Technical Chapters Meeting (ECTM). IEEE, pp 1–7
35. Gallego J, Rivero G, Martínez J (2021) Preventing rather than punishing: an early warning model of malfeasance in public procurement. Int J Forecast 37(1):360–377
36. Gilenko EV, Ivanov AE, Golubeva AA, et al (2023) Suppliers' collusion in public procurement: ex-ante determinants in zero initial contract price decrease auctions (the case of Russian healthcare). Int J Procure Manag 17(2):180–203
37. Gnaldi M, Del Sarto S (2023) Validating corruption risk measures: a key step to monitoring sdg progress. Soc Indic Res: 1–27
38. Gnip AG (2022) All you need is political love? Assessing the effects of partisan favouritism in Croatia's public procurement. Eur J Polit Econ 75:102170
39. Gomide LD, dos Santos GB, Costa LL, et al (2023) Mineração de dados sobre despesas públicas de municípios mineiros para gerar alertas de fraudes. In: Anais do XXXVIII Simpósio Brasileiro de Bancos de Dados, SBC, pp 378–383
40. Goodfellow I, Bengio Y, Courville A (2016) Deep learning. MIT Press, Cambridge. http://www.deeplearningbook.org
41. Goryunova N, Baklanov A, Ianovski E (2021) Detecting corruption in single-bidder auctions via positive-unlabelled learning. In: International conference on mathematical optimization theory and operations research. Springer, Berlin, pp 316–326
42. Goryunova N, Baklanov A, Ianovski E (2021) A noisy-labels approach to detecting uncompetitive auctions. In: International conference on machine learning, optimization, and data science. Springer, Berlin, pp 185–200
43. Grassi R, Calderoni F, Bianchi M, et al (2019) Betweenness to assess leaders in criminal networks: new evidence using the dual projection approach. Soc Netw 56:23–32
44. Hamisu M, Mansour A (2021) Detecting advance fee fraud using nlp bag of word model. In: 2020 IEEE 2nd international conference on cyberspac (CYBER Nigeria), pp 94–97
45. Hastie T, Tibshirani R, Friedman J (2017) The elements of statistical learning: data mining, inference, and prediction
46. Herrera Murillo DJ, Valerio-Ureña G, Silva Atencio G, et al (2023) Information technologies for countering corruption: analysis of the Costa Rican public procurement system. Rev Bras Adm Pública 57(1)
47. Hossin M, Sulaiman MN (2015) A review on evaluation metrics for data classification evaluations. Int J Data Min Knowl Manag Process 5:1–11. https://doi.org/10.5121/ijdkp.2015.5201
48. Huang D, Mu D, Yang L, et al (2018) Codetect: financial fraud detection with anomaly feature detection. IEEE Access 6:19161–19174
49. Huber M, Imhof D (2023) Flagging cartel participants with deep learning based on convolutional neural networks. Int J Ind Organ 89:102946
50. Huber M, Imhof D, Ishii R (2022) Transnational machine learning with screens for flagging bid-rigging cartels. J R Stat Soc, Ser A, Stat Soc 185(3):1074–1114
51. Ibe O (2014) Fundamentals of applied probability and random processes. Academic Press, San Diego
52. Imhof D, Wallimann H (2021) Detecting bid-rigging coalitions in different countries and auction formats. Int Rev Law Econ 68:106016
53. Interian R, Carpanese I, Mello B, et al (2024) Red flag algorithms for Brazilian electronic invoices: outlier detection and price risk classification. Int Trans Oper Res 31(3):1533–1553
54. Jambu M (1991) Exploratory and multivariate data analysis. Elsevier, Amsterdam
55. James G, Witten D, Hastie T, et al (2023) Linear regression. In: An introduction to statistical learning: with applications in python. Springer, Berlin, pp 69–134
56. Jancsics D (2023) Shell companies and government corruption. In: Global encyclopedia of public administration, public policy, and governance. Springer, Berlin, pp 11951–11956
57. Jiang Y, Liu G, Wu J, et al (2022) Telecom fraud detection via Hawkes-enhanced sequence model. IEEE Trans Knowl Data Eng 35(5):5311–5324
58. Júnior DPG, de Sousa Filho GF, Lucídio dos Anjos FC (2023) Classificação de fraudes em licitações públicas através do agrupamento de empresas em conluios. In: Anais do XI Workshop de Computação Aplicada em Governo Eletrônico, SBC, pp 13–24
59. Kang S, Kim D, Kim G (2023) Corporate entertainment expenses and corruption in public procurement. J Asian Econ 84:101554
60. Karyotis V, Khouzani MHR (2016) Malware diffusion models for modern complex networks: theory and applications, 1st edn. Morgan Kaufmann Publishers Inc., San Francisco
61. Kawai K, Nakabayashi J (2022) Detecting large-scale collusion in procurement auctions. J Polit Econ 130(5):1364–1411
62. Kawai K, Nakabayashi J, Ortner J, et al (2023) Using bid rotation and incumbency to detect collusion: a regression discontinuity approach. Rev Econ Stud 90(1):376–403
63. Keele S, et al (2007) Guidelines for performing systematic literature reviews in software engineering
64. Klimašauskienė D, Giedraitis V (2011) Fighting cartels: an application of economic theory and the effects of leniency policies. Ekonomika 90(1):39–52
65. Kotu V, Deshpande B (2018) Data science: concepts and practice. Morgan Kaufmann, San Mateo
66. Kratcoski P, Edelbacher M (2018) Fraud and corruption major types, prevention, and control: major types, prevention, and control. https://doi.org/10.1007/978-3-319-92333-8
67. Le C, Nguyen T, Le T (2022) Detecting red-flag bidding patterns in low-bid procurement for highway projects with pattern mining. In: Proc., 9th int. conf. on construction engineering and project management. Korea Institute of Construction Engineering and Management, Seoul, pp 11–17
68. LeCun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521(7553):436–444
69. Liew CY, Labadin J, Kok WC, et al (2023) A methodology framework for bipartite network modeling. Appl Netw Sci 8(1):6

70. Lima M, Silva R, de Souza Mendes FL, et al (2020) Inferring about fraudulent collusion risk on Brazilian public works contracts in official texts using a bi-lstm approach. In: Findings of the Association for Computational Linguistics: EMNLP 2020, pp 1580–1588
71. Lima W, Lira R, Paiva A, et al (2023) Methodology for automatic extraction of red flags in public procurement. In: 2023 International Joint Conference on Neural Networks (IJCNN). IEEE, pp 01–07
72. Lisciandra M, Milani R, Millemaci E (2022) A corruption risk indicator for public procurement. Eur J Polit Econ 73:102141
73. Lofaro RJ, Boykin EA, McCue CP, et al (2023) Year-end spending spikes and single bid procedures: an analysis of public procurement in the European economic area. Public Financ Manag 21(2):135–166
74. Lyra MS, Curado A, Damásio B, et al (2021) Characterization of the firm–firm public procurement co-bidding network from the state of Ceará (Brazil) municipalities. Appl Netw Sci 6:1–10
75. Lyra MS, Damásio B, Pinheiro FL, et al (2022) Fraud, corruption, and collusion in public procurement activities, a systematic literature review on data-driven methods. Appl Netw Sci 7(1):83
76. Lyra MS, Pinheiro FL, Bacao F (2022) Public procurement fraud detection: a review using network analysis. In: Complex networks & their applications X: volume 1. Proceedings of the tenth international conference on complex networks and their applications COMPLEX NETWORKS 2021 10. Springer, Berlin, pp 116–129
77. Maia P, Meira W Jr, Cerqueira B, et al (2020) Auditing government purchases with a multicriteria anomaly detection strategy. J Data Inf Manag 11(1)
78. Martín MLS, Benitez DRN, Coronel JMP, et al (2024) Quantifying the risk of complaints in public procurement tenders in Paraguay using machine learning. In: Proceedings of the 15th International Multi-conference on Complexity, Informatics and Cybernetics: IMCIC 2024
79. Mendes BM, Braz CS, Costa LL, et al (2023) Impacto de doaçoes eleitorais no faturamento de empresas: Um estudo nas eleiçoes municipais em minas gerais. In: Anais do XXXVIII Simpósio Brasileiro de Bancos de Dados, SBC, pp 420–425
80. Miller C, Portlock T, Nyaga DM, et al (2024) A review of model evaluation metrics for machine learning in genetics and genomics. Front Bioinform 4:1457619
81. Modrusan N, Rabuzin K, Mrsic L (2020) Improving public sector efficiency using advanced text mining in the procurement process. In: Data, pp 200–206
82. Mokeeva T, Yurko K (2023) Using digital tools in government procurement analysis: detecting suspicious purchases with control indicators. Int J Technol 14(8)
83. Montgomery DC, Peck EA, Vining GG (2021) Introduction to linear regression analysis. Wiley, New York
84. Nai R, Sulis E, Meo R, et al (2022) Public procurement fraud detection and artificial intelligence techniques: a literature review. In: Companion proceedings of the 23rd international conference on knowledge engineering and knowledge management, CEUR-WS, pp 1–13
85. Najafabadi MM, Villanustre F, Khoshgoftaar TM, et al (2015) Deep learning applications and challenges in big data analytics. J Big Data 2:1–21
86. Nakagawa EY, Scannavino KRF, Fabbri SCPF, et al (2017) Revisão sistemática da literatura em engenharia de software: teoria e prática. Elsevier, Brasil
87. Nicolás-Carlock J, Luna-Pla I (2023) Organized crime behavior of shell-company networks in procurement: prevention insights for policy and reform. Trends Organ Crime: 1–17
88. Nicolás-Carlock JR, Luna-Pla I (2021) Conspiracy of corporate networks in corruption scandals. Front Phys 9:667471
89. Niessen MEK, Paciello JM, Fernandez JIP (2020) Anomaly detection in public procurements using the open contracting data standard. In: 2020 seventh International Conference on eDemocracy & eGovernment (ICEDEG). IEEE, pp 127–134
90. Oliveira ABd, Silva PVJdG (2024) Governance in public procurement: accounting analysis of public acquisitions in Brazil during the covid-19 pandemic. Int J Procure Manag 19(3):315–335
91. Oliveira GP, Mendes BM, Braz CS, et al (2023) Ranqueamento de licitaçoes publicas a partir de alertas de fraude. In: Anais do XII Brazilian workshop on social network analysis and mining, SBC, pp 1–12
92. Oliveira GP, Reis AP, Freitas FA, et al (2022) Detecting inconsistencies in public bids: an automated and data-based approach. In: Proceedings of the Brazilian symposium on multimedia and the web, pp 182–190
93. Óskarsdóttir M, Ahmed W, Antonio K, et al (2022) Social network analytics for supervised fraud detection in insurance. Risk Anal 42(8):1872–1890
94. Page MJ, McKenzie JE, Bossuyt PM, et al (2021) The prisma 2020 statement: an updated guideline for reporting systematic reviews. BMJ 372
95. Petersen K, Vakkalanka S, Kuzniarz L (2015) Guidelines for conducting systematic mapping studies in software engineering: an update. Inf Softw Technol 64:1–18
96. Potin L, Figueiredo R, Labatut V, et al (2023) Pattern mining for anomaly detection in graphs: application to fraud in public procurement. In: Joint European conference on machine learning and knowledge discovery in databases. Springer, Berlin, pp 69–87
97. Pulido NR, Poveda AC, Carvajal JEM (2020) Corruption and institutions: an analysis for the Colombian case. Heliyon 6(9)
98. Purwanto A, Emanuel AWR (2020) Data analysis for corruption indications on procurement of goods and services. In: 2020 3rd International Conference on Information and Communications Technology (ICOIACT). IEEE, pp 56–60
99. Rainio O, Teuho J, Klén R (2024) Evaluation metrics and statistical tests for machine learning. Sci Rep 14(1):6086
100. Rodríguez MJG, Rodríguez-Montequín V, Ballesteros-Pérez P, et al (2022) Collusion detection in public procurement auctions with machine learning algorithms. Autom Constr 133:104047
101. Sampaio A, Figueiredo P, Loiola E (2022) Compras públicas no brasil: Indícios de fraudes usando a lei de newcomb-benford. Cad Gest Pública e Cid 27(86):1–20
102. Sanda MR, Siminică MI, Avram CD, et al (2024) Ghosts in the machine: how big data analytics can be used to strengthen online public procurement accountability. Sustainability 16(9):3698
103. Sanz IP, Iturriaga FJL, Blanco-Alcántara D (2024) A neural network approach for predicting corruption in public procurement. Eur J Int Manag 22(2):175–197

104. Siegel AF (2016) Practical business statistics. Academic Press, San Diego
105. Signor R, Ballesteros-Pérez P, Love PE (2021) Collusion detection in infrastructure procurement: a modified order statistic method for uncapped auctions. IEEE Trans Eng Manag 70(2):464–477
106. Signor R, Love PE, Ballesteros-Pérez P (2024) Detecting bid rigging in public auctions for procuring infrastructure projects: formulating the reference scenario for decision-making. Constr Manag Econ 42(6):545–563
107. Signor R, Love PE, Belarmino AT, et al (2020) Detection of collusive tenders in infrastructure projects: learning from operation car wash. J Constr Eng Manage 146(1):05019015
108. Signor R, Love PE, Oliveira Jr A, et al (2020) Public infrastructure procurement: detecting collusion in capped first-priced auctions. J Infrastruct Syst 26(2):05020002
109. Signor R, Oliveira A Jr, Lopes AO, et al (2020) Detecção e materialização de conluio em licitações de obras públicas. Rev Bras Ciênc Polic 11(3):113–180
110. Silva MO, Costa LL, Bezerra G, et al (2023) Análise de sobrepreço em itens de licitações públicas. In: Workshop de Computação Aplicada em Governo Eletrônico (WCGE), SBC, pp 118–129
111. Silveira D, de Moraes LB, Fiuza EP, et al (2023) Who are you? Cartel detection using unlabeled data. Int J Ind Organ 88:102931
112. Silverio-Murillo A, Prudencio D, Balmori-de-la Miyar JR (2024) The effect of the covid-19 pandemic on risk of corruption. Public Organ Rev: 1–21
113. Spadon G, Scabora LC, Araujo MV, et al (2018) Complex-network tools to understand the behavior of criminality in urban areas. In: Information technology-new generations: 14th international conference on information technology. Springer, Berlin, pp 493–500
114. Srivastava S, Singh AK (2018) Graph based analysis of Panama papers. In: 2018 fifth international conference on Parallel, Distributed and Grid Computing (PDGC). IEEE, pp 822–827
115. Tas BKO (2022) Weak supervision can help detecting corruption in public procurement. In: Proceedings of SAI intelligent systems conference. Springer, Berlin, pp 548–555
116. Toledo AS, Carpi LC, Atman AP, et al (2025) Outlier mining in criminal networks: the role of machine learning and outlier detection models. J Comput Soc Sci 8(2):36
117. Torres Berru Y, López Batista VF, Torres-Carrión P, et al (2020) Artificial intelligence techniques to detect and prevent corruption in procurement: a systematic literature review. In: Proceedings, part II 1, Applied Technologies: First International Conference, ICAT 2019, Quito, Ecuador, December 3–5, 2019. Springer, Berlin, pp 254–268
118. Torres-Berru Y, López Batista VF (2021) Data mining to identify anomalies in public procurement rating parameters. Electronics 10(22):2873
119. Torres-Berru Y, Lopez-Batista VF, Zhingre LC (2023) A data mining approach to detecting bias and favoritism in public procurement. Intell Autom Soft Comput 36(3)
120. Tranmer M, Elliot M (2008) Multiple linear regression. Cathie Marsh Cent Census Surv Res (CCSR) 5(5):1–5
121. Tumminello M, Consiglio A, Vassallo P, et al (2023) Insurance fraud detection: a statistically validated network approach. J Risk Insur 90(2):381–419
122. Van Vlasselaer V, Eliassi-Rad T, Akoglu L, et al (2015) Afraid: fraud detection via active inference in time-evolving social networks. In: 2015 IEEE/ACM international conference on Advances in Social Networks Analysis and Mining (ASONAM), pp 659–666. https://doi.org/10.1145/2808797.2810058
123. Vaswani A (2017) Attention is all you need. Adv Neural Inf Process Syst
124. Villamil I, Kertész J, Fazekas M (2024) Collusion risk in corporate networks. Sci Rep 14(1):3161
125. Vuković V (2020) Corruption and re-election: how much can politicians steal before getting punished? J Comp Econ 48(1):124–143
126. Wachs J, Fazekas M, Kertész J (2021) Corruption risk in contracting markets: a network science perspective. Int J Data Sci Anal 12:45–60
127. Wallimann H, Imhof D, Huber M (2023) A machine learning approach for flagging incomplete bid-rigging cartels. Comput Econ 62(4):1669–1720
128. Wallimann H, Sticher S (2023) On suspicious tracks: machine-learning based approaches to detect cartels in railway-infrastructure procurement. Transp Policy 143:121–131
129. Wang B, Zhu J, Xie J, et al (2023) Collusion suspicion among bidders of construction supervision projects in China: based on bidding big data and complex networks. KSCE J Civ Eng 27(1):13–26
130. Wang X, Long W, Sang M, et al (2022) Towards sustainable urbanization: exploring the influence paths of the urban environment on bidders' collusive willingness. Land 11(2):280
131. Wang X, Ye K, Chen M, et al (2021) A conceptual framework for the inclusion of exogenous factors into collusive bidding price decisions. J Manag Eng 37(6):04021071
132. Wei Z, Zhu Y (2023) Public corruption and the allocation of government contracts. Rev Financ Econ 41(1):3–22
133. Weisberg S (2005) Applied linear regression
134. Westerski A, Kanagasabai R, Shaham E, et al (2021) Explainable anomaly detection for procurement fraud identification—lessons from practical deployments. Int Trans Oper Res 28(6):3276–3302
135. Wittberg E, Fazekas M (2023) Firm performance, imperfect competition, and corruption risks in procurement: evidence from Swedish municipalities. Public Choice 197(1):227–251
136. Wohlin C (2014) Guidelines for snowballing in systematic literature studies and a replication in software engineering. In: Proceedings of the 18th international conference on evaluation and assessment in software engineering, pp 1–10
137. Xiao L, Ye K, Zhou J, et al (2021) A social network-based examination on bid riggers' relationships in the construction industry: a case study of China. Buildings 11(8):363
138. Xu P, Zhu X, Clifton DA (2023) Multimodal learning with transformers: a survey. IEEE Trans Pattern Anal Mach Intell 45(10):12113–12132
139. Zhu J, Wang B, Li L, et al (2020) Bidder network community division and collusion suspicion analysis in Chinese construction projects. Adv Civ Eng 2020(1):6612848
140. Zigo D, Vincent F (2021) "Beneficial owners" policy: comparison of its efficacy in the West with prospects for curbing corruption in China. Danube 12(4):273–292

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.