

# Public Procurement Fraud Detection and Artificial Intelligence Techniques: a Literature Review

Roberto Nai<sup>1,\*</sup>, Emilio Sulis<sup>1</sup> and Rosa Meo<sup>1</sup>

<sup>1</sup>Computer Science Department, University of Turin, Italy

## Abstract

Every year, a significant part of public and private organisations' revenues is lost to fraud. Recently, increasing digitisation has also brought more attention on organisational processes, procurement, and fraud data. Several automated methods have already been proposed to extract information from these kind of sources, including public procurement, also to develop predictive models for fraud detection. In addition, artificial intelligence techniques including machine learning, neural networks, natural language processing, and network analysis methodologies have been adopted to address the issue. This study offers a review of the most recent emerging studies on fraud detection for public organisations. Finally, it summarises the main existing research by proposing a review of current challenges in the field.

## Keywords

Corruption detection, artificial intelligence, literature review, public procurement,

## 1. Introduction

Organizations are increasingly focused on mitigating the chances of experiencing fraud, which represent a significant loss of revenue. For instance, an accredited biennial 2020 study carried out by the Association of Certified Fraud Examiners claims that on average 5% of a company's revenue is lost because of unchecked fraud every year. Among the reasons for these large losses is that it takes about 14 months for a fraud to be discovered and that audits capture only 3 percent of actual fraud. This necessitates the use of better tools and processes to quickly and inexpensively identify potential criminals [1].

Researches in political science, economics and sociology investigated the field, trying to highlight possible flaws in the systems that lead to such risks, with a view to prevention. Recently, the new possibilities offered by information technology allow for new studies in the area of fraud detection as well. Recent changes include the availability of large data sets at low cost, the use of increasingly powerful computing devices and the development of applications that enable the training of machine learning (ML) models [2, 3].

Public organizations are also subject to fraud risks, starting with public procurement. A major challenge is to be able to detect potential fraud automatically, through appropriate artificial intelligence (AI) techniques. Machine learning methods, in particular, have proven very effective

---

*The Knowledge Management for Law Workshop (KM4LAW), September 26, 2022, Bozen-Bolzano, Italy*


\*Corresponding author.

✉ roberto.nai@unito.it (R. Nai); emilio.sulis@unito.it (E. Sulis); rosa.meo@unito.it (R. Meo)

🆔 0000-0003-4031-5376 (R. Nai); 0000-0003-1746-3733 (E. Sulis); 0000-0002-0434-4850 (R. Meo)



© 2022 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

 CEUR Workshop Proceedings (CEUR-WS.org)

in a wide range of practical applications [4, 5, 6]. In addition, the most recent methodologies have also developed Natural Language Processing (NLP) [7] methodologies, as well as neural networks [8].

In order to systematise the existing research on the topic of AI techniques applied to fraud detection in public procurement, we propose a systematic review of the research literature. To summarize our goals, we explored the following three research questions:

- RQ1: Which disciplinary areas are more interested in investigating frauds in public procurement?
- RQ2: What AI techniques are being applied to investigate fraud in public procurement contracts?
- RQ3: Which research studies are most influential in the field?

The remainder of the paper is organised as follows: section 2 introduces some related works. In section 3 we describe the proposed methodology, while section 4 provides insights about the results of our review. Finally, section 5 provides a summary of main research, and section 6 concludes the paper.

## 2. Related works

There are previous works focused on the application of techniques to public procurement. In [9], 102 articles published between 2015 and 2019 have been selected from Scopus and WoS databases, focusing on the primary data mining techniques used to prevent corruption. It is observed that the main techniques of AI are those based on the theorem of Bayes, neural networks, Support Vector Machines (SVM), decision trees, Random Forest, logistic and linear regression.

In another recent survey, 147 articles published between 2015 and 2019 have been selected from Scopus and WoS [10]. These works focused the following types of corruption: fraud (77.49%), overpricing (7.05%), bribery (5.05%) and favoritisms (4.66%) generate greater citations in the articles. A large part of research analyses business intelligence literature for bank fraud using text mining. For the geographical distribution of the authors, the first author of the publication was considered; the leading countries with articles on data mining and corruption are United States (16,3%), China (10,9%) and United Kingdom (8,9%).

Another work focused on the most used methods to detect different types of “corruption” [11], exploring 23 articles published between 2016 and 2021. Data mining and ML methods are used in this segment over a large amount of data collected from different data sets, such as contract registers, blacklist economic operators, business registers and so on. The methods include classification techniques, with the aim of detecting connections between economic operators and contracting authorities, but also for finding companies that participated in collusion, as well as associations rules, and graph databases algorithms.

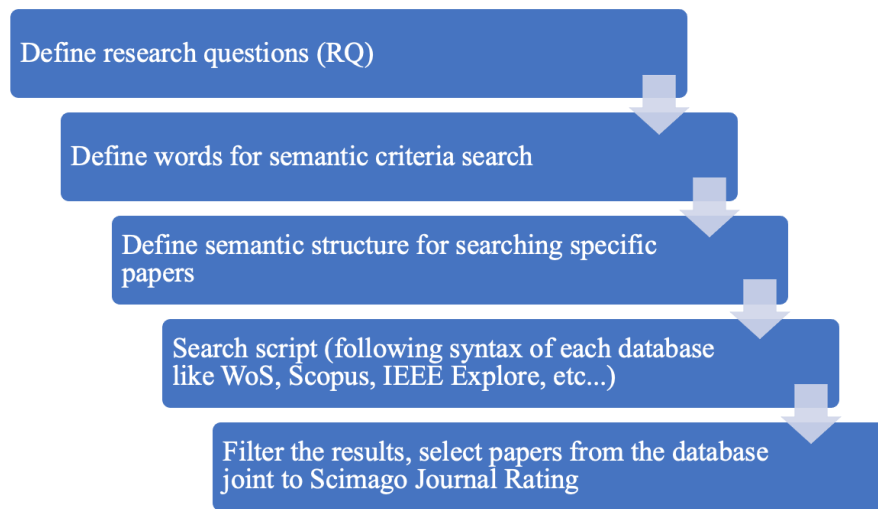
Another recent review is focused on Social Network Analysis (SNA) to capture the contributions of the scientific community to the topic of corruption in public procurement [12]. Authors identified the most recurrent authors, their interactions, number of citations, identification of keywords, and their repetitions. Authors analyzed 18 articles from 2011 to 2021. To perform

network analysis on the collected dataset and represent the interactions between the actors or nodes of the graph, the open-source engine VOSviewer<sup>1</sup> was used; the tool allowed the authors to identify the publications, authors, journals, institutions, keywords, and countries with the most significant impact on the research in repositories of scientific articles (from parameters such as centrality degree and edge weight).

Our work builds on previous researches by focusing on recent work (2016-2021) related to fraud detection with ML methods/techniques.

### 3. Methodology

For our survey, the methodology of [13] is followed, dividing the process into three phases: planning, conducting the review, and reporting the review. The workflow in figure 1 resumes the activity of this research.



**Figure 1:** Workflow for the planning and conducting the review.

Table 1 describes the objective of the survey through the research questions and the expected output variables.

#### 3.1. Semantic structure of the search

In the first phase of the workflow, the following keywords have been defined: public tenders, public competitions, public procurement, e-procurement, state laws, fraud detection, corruption, crime, criminal, prediction, predictive, modeling, detection, artificial intelligence, machine learning, deep learning, neural networks. In a second phase, various combinations of the keywords have been prepared as query for the scientific databases (Scopus, WoS and IEEE Explore). In a third phase, specific scripts have been prepared following the syntax of each database. For instance, one of the query in Scopus (very similar in WoS) has been:

<sup>1</sup><https://www.vosviewer.com>

Question	Type of answer sought
RQ1: Which disciplinary areas are more interested in investigating frauds in public procurement?	List of disciplinary areas
RQ2: What AI techniques are being applied to investigate fraud in public procurement contracts?	List of AI techniques
RQ3: Which research studies are most influential in the field?	Weighted graph of citations between articles

**Table 1**  
Research questions.

```
TITLE-ABS-KEY ((( "PUBLIC TENDER" OR "PUBLIC PROCUREMENT"
OR "E-PROCUREMENT" OR "public competitions"
OR "public regulations" OR "state laws" )
AND ( "DETECTION" OR fraud OR corruption OR crime OR criminal )
AND ( prediction OR predictive OR "machine learning"
OR "deep learning" OR "neural networks" OR "modeling"
OR "artificial intelligence" )))
```

### 3.2. Inclusion and exclusion criteria

The results have been filtered by categories: Computer Science, Engineering, Business, Politics and Business Management, joint with Scimago Journal Rating (Quartile and H-Index). The search has been automated in Python combining the Scopus APIs, WoS and IEEE Explore to get automatically the list of results joint with Scimago rating indexes<sup>2</sup>. Finally, an amount of 15 studies has been selected for this survey.

## 4. Results

We summarize the main results of our survey by exploring for each paper some features of interest: the disciplines involved, dataset used in input (Input), the AI techniques, methodologies and technologies adopted (Methods).

Table 2 and table 3 resume the results about RQ1. Disciplines involved are mostly Computer Science, but also Business, Management and Accounting or Engineering, from different venues, e.g. Conferences or Journal in Quartile 1 or 2 (Q). Regarding the geographical distribution of the authors, all the authors are considered. Most researchers come from Europe (Spain, Portugal, Italy) and America (Brazil, U.S.A., Paraguay).

Finally, the researches summarized in table 4 have been selected with respect to the RQ2 research question. Most frequent methods concern typical ML supervised and unsupervised algorithms. About technologies, python emerges as the most used programming language (libraries as Scikit-learn are becoming the state-of-the-art in the subject), but also Java or R.

<sup>2</sup>Python script using Scopus APIs and Scimago Journal Rating is available here: <https://github.com/roberto-nai-unito/scopus-api>

Among the tools used, we notice Neo4J or KNIME. Finally, some works explore the adoption of neural network, as well as social network analysis methods.

Paper	# Cit	Venue	Disciplines	Q
[14]	-	Electronics (Switzerland)	Computer Science Engineering	Q2
[15]	3	International Journal of Forecasting	Business and International Management	Q1
[16]	7	International Transactions in Operational Research	Business Management and Accounting Computer Science Decision Sciences	Q1
[17]	-	Conference paper	-	-
[18]	3	Conference paper	-	-
[19]	3	Governance	Public administration	Q1
[20]	-	Automation in Construction	Engineering	Q1
[8]	2	Conference paper	-	-
[21]	7	Proceedings	Computer Science	-
[22]	1	Proceedings	Computer Science	-
[23]	1	Proceedings	Computer Science	-
[24]	8	Complexity	Computer Science	Q1
[25]	4	Business and Politics	Business, Management and Accounting	Q1
[26]	5	Conference paper	-	-
[27]	10	SSRN Electronic Journal	-	-

**Table 2**

Disciplinary area of the selected papers with citation count (Cit), the Venue, and the Journal's Quartile (Q1)

Country	Amount
Spain	12
Brazil	11
United States	8
Portugal	5
Italy	4
Paraguay	3
Croatia	2
Australia, Austria, Colombia, Slovenia, United Kingdom	1

**Table 3**

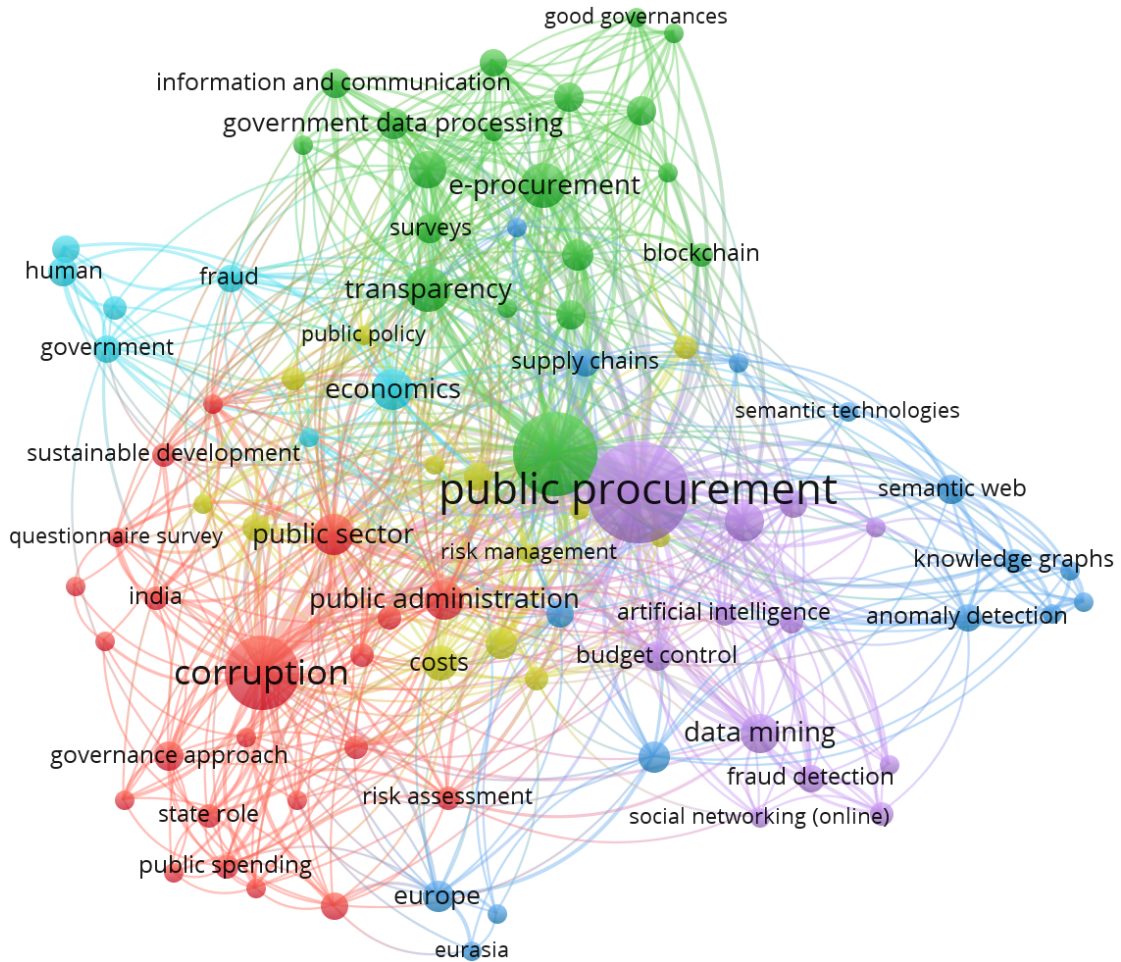
Geographical distribution of the authors.

A co-occurrence network has been generated in figure 2, starting from the main keyword “public procurement”. By applying a community detection algorithm (Louvain) [28], we detect most dense groups of terms. In particular, a first group includes terms related to corruption, risk, governance with respect to public administration, public spending, public sector. A second group includes terms about ICT technologies (blockchain, e-procurement, information and communication, data processing). Interestingly, a third group includes Artificial Intelligence,

<b>Paper</b>	<b>Input data</b>	<b>Methods &amp; Technologies</b>
[14]	Public Procurement System (SERCOP) of Ecuador	Clustering (K-Means), Self-Organizing map (SOM), Support Vector Machine (SVM) and Principal Component Analysis (PCA). Technologies: Python Scikit-learn library, Mini-Som, Self Organizing Map (SOM), AZURE Machine Learning.
[15]	Sistema Electronico de Contratación Pública (SECOP) of Columbia	Lasso classification model, gradient boosting classification model (GBM). Technologies: n.a.
[16]	Various from the states of Brazil.	Graph theory, network analysis, clusterization, regression analysis. Technologies: n.a.
[17]	Diario Oficial da Uniao (DOU) of Brazil	Bottleneck deep neural network and Bi-LSTM. Technologies: n.a.
[18]	Public procurement open data from Spain	Machine learning, pattern detection. Technologies: Python, R, Neo4j, others
[19]	Italian dataset managed by the ANAC	Binary logistic regression, random Forest, and Gradient Boosting Machines (GBM). Technologies: R.
[20]	Public procurement open data from Brazil, Italy, Japan, Switzerland and USA	SGD (Stochastic Gradient Descent), Extra Trees (Extremely Randomized Trees), Random Forest, Ada Boost, Gradient Boosting, SVC (C-Support Vector Classification), K Neighbors, MLP (Multi-Layer Perceptron), Bernoulli Naive Bayes and Gaussian Naive Bayes, Gaussian Process. Technologies: Python and Scikit-learn library.
[8], [21]	Electronic Public Procurement of Croatia	NLP, naïve Bayes (NB), logistic regression (LR), support vector machines (SVM). Technologies: Python.
[22]	Public procurement open data from Paraguay	Unsupervised learning model for anomaly detection based on the Isolation Forest algorithm. Technologies: KNIME framework.
[23]	Portuguese Public Procurement	Supervised machine learning, graph-oriented database. Technologies: Python Scikit-learn library, Neo4j.
[24]	Public procurement open data from Spain and EU	Random forest regression method. Technologies: Random Forest Regressor from Scikit-learn (Python).
[25]	European Economic Area members and associate countries	Random forest. Technologies: n.a.
[26]	Various (private and public)	Text Analytics, Social Network Analysis, Unsupervised learning, Online probabilistic learning. Technologies: Python, Java, DB/2.
[27]	Italian dataset managed by the ANAC	Lasso, ridge regression, and random forest. Technologies: n.a.

**Table 4**  
Papers selected for literature review.

data mining, social network (and fraud detection). Fourth, another group of terms includes network methods: semantic web, knowledge graphs, semantic technologies (and anomaly detection). These results confirm the effectiveness of our approach, and we will use these categories in the presentation of the research in section 5.



**Figure 2:** Visualization of a term co-occurrence network.

## 5. Summary of main research

We can summarise the main results of the papers by grouping them into three classes. Most of the papers adopt typical ML methods, while two smaller groups deal mainly with neural networks and network analysis.

*Typical machine learning methods.* In [14], a multi-phase model was used (the identification of anomalies and generation of the detection model), which uses different algorithms, such as clustering (K-Means), Self-Organizing map (SOM), Support Vector Machine (SVM) and Principal

Component Analysis (PCA). Following this methodology, a semi-supervised learning model is built for the detection of anomalies, which obtains an accuracy of 95%, allowing the detection of procedures where the aim is to benefit a particular supplier by means of the qualification assignment parameters.

Two machine learning models have been used in [15], to predict whether a contract will result in malfeasance, breach of contract, or inefficiency: a lasso classification model [29] and a gradient boosting classification model [30]. The methods used allow to describe which variables—and in which way these variables—contribute to the likelihood that a contract will be problematic, which is very useful from the perspective of policymakers; for instance, variables associated with projects such as their size or duration were important predictors of malfeasance. Also, the time lag between adjudicating the contract and the nearest election showed high predictive value.

Alternative predictive models were estimated in [19], and the best method was selected based on prediction accuracy as well as fit with existing organized crime proxies. The predictive models included both traditional regression analysis and machine learning: binary logistic regression, random forest, and Gradient Boosting Machines (GBM). Testing prediction accuracy on unseen data, our best model, Gradient Boosting Machine, achieves 85%. Looking at external validity, the model's predicted EGO score also significantly and moderately strongly correlates with established indicators of organized criminality both within Italy and across Europe.

The accuracy of eleven ML algorithms for detecting collusion using collusive datasets obtained from Brazil, Italy, Japan, Switzerland and the United States is tested in [20]. While the use of ML in public procurement remains largely unexplored, its potential use to identify collusion is promising. The three top-performing ML algorithms have been the Extra Trees, Random Forest and Ada Boost (ensemble methods). In the scenario where all auction information was available, these algorithm's accuracy (detection rates) ranged between 81% and 95%, with a balanced accuracy generally above 73% (excluding the US dataset).

In [18], a prototype called SALER is proposed. Inside SALER, several internal and external data sources are analysed and assessed to explore possible irregularities in budget and cash management, public service accounts, salaries, disbursement, grants, subsidies, etc. SALER combines descriptive and predictive machine learning models and the results can be accessed with a web interface. Finally, the authors mention two frameworks similar to SALER: zIndex<sup>3</sup>, a public procurement benchmarking tool for rating contracting authorities which is being developed in the Czech Republic by researchers from the Charles University of Prague and Arachne<sup>4</sup>, considered by the European Commission as a good tool amongst anti-fraud measures; this risk-scoring tool generates more than 100 risk indicators sorted into specific risk categories to help managing authorities and intermediate bodies to prevent and detect errors and irregularities among projects, beneficiaries, contracts and contractors.

The relation between the award price and the bidding price is investigated by [24]. It is proposed an award price estimator that uses the random forest [31] regression method over the Spanish open data from 2012 to 2018. Finally, a similar analysis, employing a dataset from

---

<sup>3</sup><https://www.zindex.cz>

<sup>4</sup><https://ec.europa.eu/social/main.jsp?catId=325&intPageId=3587&langId=en>



European countries (TED<sup>5</sup>), is presented to compare and generalise the results. The article illustrates how a machine learning algorithm can be useful. Particularly, random forest predicts the award prices with less uncertainty, adapting to the real market.

Machine learning tools are used to analyze by [25], a large dataset of public contracts from across Europe, in order to identify the conditions under which close connections, defined both in terms of repeated interaction, as well as geographical dispersion, appear. In this case, random forest models have been used.

In [27], three main results through detailed data on the content of calls for tenders involving roadwork contracts in Italy are presented. The prediction capability of the various corruption indicators using standard ML algorithms have been tested: lasso, ridge regression, and random forest. The article shows that, among ML methods, the random forests algorithm provides the most accurate prediction. At a more general level, the article suggests that a higher standardization of call for tenders documents can contribute to reduce corruption risks. For this purpose, sector authorities or specialized public bodies can play a crucial role.

In [22] the initial results of an anomaly detection experiment by applying Isolation Forrest algorithm to a publicly available dataset, i.e. the public procurement of Paraguay are discussed. An in-depth study of the diversity of ties between buyers and sellers in public contracts adopted a statistical analysis with Random Forest models starting with 3.3 million European Union contracts between 2009 and 2015. The effectiveness of the model is validated with local known anomalous procurement processes, which are: a) processes protested by entities involved in the contracting process, which were determined in favor of the protestant, and b) complaints about the contracting process from external entities with the possibility of anonymity. The results show an accuracy of over 90% in detecting these known anomalies as early as in the tender stage and during the contracting stage.

*Network analysis and text mining.* A Decision Support System (DSS) is proposed in [16] to allow law enforcement agencies to establish priorities concerning the companies to be investigated. This DSS incorporates data mining algorithms for quantifying dozens of corruption risk patterns for all public contractors inside a specific jurisdiction, leading to improvements in the quality of public spending and to the identification of more cases of fraud. These algorithms combine operations research tools such as graph theory, clusterization, and regression analysis with advanced data science methods to allow the identification of the main risk patterns. Starting from various dataset and social network analysis (graph model based), an unsupervised learning model has been developed for clustering fraudulent employees by [26].

In [8] and [21], the use of advanced text mining to improve the procurement process is explored. Based on Public Procurement of Croatia<sup>6</sup>. The authors introduce the use of NLP to improve the research of frauds, comparing common classification algorithms: Naïve Bayes (NB), Logistic regression (LR) and Support Vector Machines algorithm (SVM). The models have been trained and tested on all data, and by groups of procurement lots (food, medical equipment, construction, IT services, etc.) defined in the unique Public Procurement Dictionary (CPV). Groups such as IT services, repair and maintenance services, and health and social work services have good prediction results; conversely, groups such as architecture, construction, engineering

---

<sup>5</sup><https://data.europa.eu/data/datasets/ted-csv?locale=en>

<sup>6</sup><https://eojn.nn.hr>

and inspection services provide bad metrics, precisely because of the lack of information on technical and professional abilities.

*Neural networks.* The types of fraud investigated by [17] are mainly collusion (bid-rigging), over pricing, and delivery fraud (quality and quantity of services and materials). To evaluate the reference dataset, bottleneck Deep Neural Networks and Bidirectional Neural Networks [32] was chosen. Deep neural network models were built using the Tensorflow [33]. Both bottleneck deep neural network and Bi-LSTM proved to be competitive with traditional classifiers and achieved better precision, which is more desirable (over recall) in a criminal fraud investigation. In [23], starting from the Portuguese Public Procurement portal, a graph-oriented user interface is proposed to support decision-making, using Cypher queries. Beside this, supervised machine learning methods are used to find suspicious procurement.

After summarize the main research with respect to the methods and technologies used, it's important to note how this kind of researches can have limitations, i.e. it is widely known that ML algorithms are akin to a black box from which it is difficult to explain to not-experts. Lawyers and stakeholders can be interested in explanation about the results, while the inherent complexity of the problem being analyzed do not facilitate the task (at least not in a straightforward manner) [34]. Another issue is the need of a substantial amount of reliable historical data, some of which (especially the collusion-related) may not always be made available by competition commissions or law enforcement agencies [20].

## 6. Conclusions and future work

We provided a review of the most recent studies on fraud detection for public organisations. We detect typical methods based on ML algorithms or network analysis, with some emerging interest on neural network. As a future work, we are interested in detect relevant authors in the field, according to [35]. In particular, we aim to perform a bibliometric network analysis (graph-based and timeline-based) to find centrality and density of authors connected about the topic of machine learning and fraud detection in public procurement.

## References

- [1] ACFE, 2020 report to the nations—the acfe's 11th study on the costs and effects of occupational fraud (2020). URL: <https://acfe-public.s3-us-west-2.amazonaws.com/2020-Report-to-the-Nations.pdf>.
- [2] H. R. Varian, Big data: New tricks for econometrics, *Journal of Economic Perspectives* 28 (2014) 3–28. URL: <https://www.aeaweb.org/articles?id=10.1257/jep.28.2.3>. doi:10.1257/jep.28.2.3.
- [3] S. Mullainathan, J. Spiess, Machine learning: An applied econometric approach, *Journal of Economic Perspectives* 31 (2017) 87–106. URL: <https://www.aeaweb.org/articles?id=10.1257/jep.31.2.87>. doi:10.1257/jep.31.2.87.
- [4] Z. Zhou, *Machine Learning*, Springer, 2021. URL: <https://doi.org/10.1007/978-981-15-1967-3>. doi:10.1007/978-981-15-1967-3.

- [5] E. Sulis, L. Humphreys, F. Vernerero, I. A. Amantea, D. Audrito, L. D. Caro, Exploiting co-occurrence networks for classification of implicit inter-relationships in legal texts, *Inf. Syst.* 106 (2022) 101821. doi:10.1016/j.is.2021.101821.
- [6] E. Sulis, L. B. Humphreys, D. Audrito, L. Di Caro, Exploiting textual similarity techniques in harmonization of laws, in: S. Bandini, F. Gasparini, V. Mascardi, M. Palmonari, G. Vizzari (Eds.), *AIxIA 2021 – Advances in Artificial Intelligence*, Springer International Publishing, Cham, 2022, pp. 185–197.
- [7] S. Bird, E. Klein, E. Loper, *Natural Language Processing with Python*, 1st ed., O’Reilly Media, Inc., 2009.
- [8] N. Modrušan, K. Rabuzin, L. Mrcic, Improving public sector efficiency using advanced text mining in the procurement process, 2020, pp. 200–206. doi:10.5220/0009823102000206.
- [9] Y. Torres Berru, V. F. López Batista, P. Torres-Carrión, M. G. Jimenez, Artificial intelligence techniques to detect and prevent corruption in procurement: A systematic literature review, in: M. Botto-Tobar, M. Zambrano Vizueté, P. Torres-Carrión, S. Montes León, G. Pizarro Vásquez, B. Durakovic (Eds.), *Applied Technologies*, Springer International Publishing, Cham, 2020, pp. 254–268.
- [10] Y. Torres Berrú, V. Batista, P. Torres-Carrion, Data mining to detect and prevent corruption in contracts: Systematic mapping review, *RISTI - Revista Iberica de Sistemas e Tecnologias de Informacao* (2020) 13–25.
- [11] N. Modrusan, K. Rabuzin, L. Mrcic, Review of public procurement fraud detection techniques powered by emerging technologies, *International Journal of Advanced Computer Science and Applications* 12 (2021). doi:10.14569/IJACSA.2021.0120272.
- [12] Public procurement fraud detection: A review using network analysis”, in: R. Benito, C. Cherifi, H. Cherifi, E. Moro, L. Rocha, M. Sales-Pardo (Eds.), *Complex Networks & Their Applications X*, volume I of *Studies in Computational Intelligence*, Springer, Cham, 2022”, pp. 116–129. URL: <https://complexnetworks.org>. doi:10.1007/978-3-030-93409-5\_11".
- [13] P. V. Torres-Carrión, C. S. González-González, S. Aciar, G. Rodríguez-Morales, Methodology for systematic literature review applied to engineering and education, in: 2018 IEEE Global Engineering Education Conference (EDUCON), 2018, pp. 1364–1373. doi:10.1109/EDUCON.2018.8363388.
- [14] Y. Torres-Berru, V. F. López Batista, Data mining to identify anomalies in public procurement rating parameters, *Electronics* 10 (2021). URL: <https://www.mdpi.com/2079-9292/10/22/2873>. doi:10.3390/electronics10222873.
- [15] J. Gallego, G. Rivero, J. Martínez, Preventing rather than punishing: An early warning model of malfeasance in public procurement, *International Journal of Forecasting* 37 (2021) 360–377. URL: <https://www.sciencedirect.com/science/article/pii/S0169207020300935>. doi:https://doi.org/10.1016/j.ijforecast.2020.06.006.
- [16] R. B. Velasco, I. Carpanese, R. Interian, O. C. G. P. Neto, C. C. Ribeiro, A decision support system for fraud detection in public procurement, *Int. Trans. Oper. Res.* 28 (2021) 27–47. URL: <https://doi.org/10.1111/itor.12811>. doi:10.1111/itor.12811.
- [17] M. Lima, R. Silva, F. Lopes de Souza Mendes, L. R. de Carvalho, A. Araujo, F. de Barros Vidal, 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*, Association for Computational Linguistics, Online, 2020, pp. 1580–1588. URL:

<https://aclanthology.org/2020.findings-emnlp.143>. doi:10.18653/v1/2020.findings-emnlp.143.

- [18] F. Plumed, J. Casamayor, C. Ferri, J. Gómez, E. Vendrell Vidal, SALER: A Data Science Solution to Detect and Prevent Corruption in Public Administration, 2019, pp. 103–117. doi:10.1007/978-3-030-13453-2\\_9.
- [19] M. Fazekas, S. Sberna, A. Vannucci, The extra-legal governance of corruption: Tracing the organization of corruption in public procurement, *Governance* (2021).
- [20] M. J. García Rodríguez, V. Rodríguez-Montequín, P. Ballesteros-Pérez, P. E. Love, R. Signor, Collusion detection in public procurement auctions with machine learning algorithms, *Automation in Construction* 133 (2022) 104047. URL: <https://www.sciencedirect.com/science/article/pii/S0926580521004982>. doi:<https://doi.org/10.1016/j.autcon.2021.104047>.
- [21] K. Rabuzin., N. Modrušan., Prediction of public procurement corruption indices using machine learning methods, in: *Proceedings of the 11th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management - KMIS, INSTICC, SciTePress, 2019*, pp. 333–340. doi:10.5220/0008353603330340.
- [22] M. Niessen, J. Paciello, J. Fernandez, Anomaly detection in public procurements using the open contracting data standard, 2020, pp. 127–134. doi:10.1109/ICEDEG48599.2020.9096674.
- [23] D. Carneiro, P. Veloso, A. Ventura, G. Palumbo, J. Costa, Network Analysis for Fraud Detection in Portuguese Public Procurement, 2020, pp. 390–401. doi:10.1007/978-3-030-62365-4\\_37.
- [24] M. J. García Rodríguez, V. Montequín, F. Ortega-Fernández, J. Balsera, Public procurement announcements in spain: Regulations, data analysis, and award price estimator using machine learning, *Complexity* 2019 (2019) 1–20. doi:10.1155/2019/2360610.
- [25] M. Popa, Uncovering the structure of public procurement transactions, *Business and Politics* 21 (2019) 351–384. doi:10.1017/bap.2019.1.
- [26] A. Dhurandhar, B. Graves, R. Ravi, G. Maniachari, M. Ettl, Big data system for analyzing risky procurement entities, in: *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '15, Association for Computing Machinery, New York, NY, USA, 2015*, p. 1741–1750. URL: <https://doi.org/10.1145/2783258.2788563>. doi:10.1145/2783258.2788563.
- [27] F. Decarolis, C. Giorgiantonio, Corruption red flags in public procurement: New evidence from italian calls for tenders, *SSRN Electronic Journal* (2020). doi:10.2139/ssrn.3612661.
- [28] X. Que, F. Checconi, F. Petrini, J. A. Gunnels, Scalable community detection with the Louvain algorithm, in: *2015 IEEE International Parallel and Distributed Processing Symposium, IEEE, 2015*, pp. 28–37.
- [29] R. Tibshirani, Regression shrinkage and selection via the lasso, *Journal of the royal statistical society series b-methodological* 58 (1996) 267–288.
- [30] J. H. Friedman, Greedy function approximation: a gradient boosting machine, *Annals of statistics* (2001) 1189–1232.
- [31] L. Breiman, Machine learning, volume 45, number 1 - springerlink, *Machine Learning* 45 (2001) 5–32. doi:10.1023/A:1010933404324.
- [32] M. Schuster, K. Paliwal, Bidirectional recurrent neural networks, *IEEE Transactions on Signal Processing* 45 (1997) 2673–2681. doi:10.1109/78.650093.
- [33] M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. Corrado, A. Davis,

- J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Y. Yu, X. Zheng, Tensorflow: Large-scale machine learning on heterogeneous distributed systems, 2015. URL: <http://download.tensorflow.org/paper/whitepaper2015.pdf>.
- [34] R. Guidotti, A. Monreale, S. Ruggieri, F. Turini, F. Giannotti, D. Pedreschi, A survey of methods for explaining black box models, *ACM Comput. Surv.* 51 (2018). URL: <https://doi.org/10.1145/3236009>. doi:10.1145/3236009.
- [35] A. Perianes-Rodriguez, L. Waltman, N. J. van Eck, Constructing bibliometric networks: A comparison between full and fractional counting, *Journal of Informetrics* 10 (2016) 1178–1195. URL: <https://www.sciencedirect.com/science/article/pii/S1751157716302036>. doi:<https://doi.org/10.1016/j.joi.2016.10.006>.