COMPUTER LAW & SECURITY REVIEW 51 (2023) 105887

Available online at www.sciencedirect.com

ScienceDirect

journal homepage: www.elsevier.com/locate/CLSR

Computer Law & Security Review

Public tenders, complaints, machine learning and recommender systems: a case study in public administration

*Roberto Nai**, *Rosa Meo*, *Gabriele Morina*, *Paolo Pasteris*

Computer Science Department - University of Turin, Corso Svizzera, 185, Turin (TO), 10149, Italy

ARTICLE INFO

Keywords:
Public procurement
Legal prediction
Complaint detection
Knowledge discovery
Natural language processing
Machine learning
Recommender system

ABSTRACT

With the proliferation of e-procurement systems in the public sector, valuable and open information sources can be jointly accessed. Our research aims to explore different legal Open Data; in particular, we explored the data set of the National Anti-Corruption Authority in Italy on public procurement and the judges' sentences related to public procurement, published on the website of the Italian Administrative Justice from 2007 to 2022. Our first goal was to train machine learning models capable of automatically recognizing which procurement has led to disputes and consequently complaints to the Administrative Justice, identifying the relevant features of procurement that correspond to certain anomalies. Our second goal was to develop a recommender system on procurement to return similar procurement to a given one and find companies for bidders, depending on the procurement requirements.

1. Introduction

There are currently major digital management and transformation initiatives in all countries. Information systems are crucial for an organization, as it is becoming increasingly easy to store multiple sources of information in different formats and accumulate big data volumes. Data management systems can become decisive in providing meaningful decision-making information, improving the analysis of daily activities and decision-making processes (Dumas et al., 2018). Information systems research is increasingly focused on exploiting large databases in disparate domains. The availability of digital data stored in enterprise information systems (EISs) also favors the growth of automatic tools for the analysis and exploitation of data through increasingly sophisticated techniques, such as Artificial Intelligence (AI) applied to organizations (Collins et al., 2021; K¨appel et al., 2021). Automated methods of extracting knowledge from data collected by EISs are increasingly used in several sectors, such as the legal domain. The laws are typically structured texts and follow rather defined procedural processes. Therefore, the applicative tasks can easily include compliance analysis and anomalies with Artificial Intelligence (AI) methods. Some of the main AI techniques include Machine Learning (ML) and Deep Learning (DL) to automatically extract knowledge by exploiting data from experience, or the Natural Language Processing (NLP) subfield, that can achieve results in many natural language tasks, e.g., text modeling and parsing. Several practical applications demonstrated their success: classifiers trained on textual data can distinguish between classes of messages, the users' intent in automatic question-answering systems, or their opinions in social media platforms; classifiers that predict

* Corresponding author.
E-mail address: roberto.nai@unito.it (R. Nai).

https://doi.org/10.1016/j.clsr.2023.105887

the behavior of software systems on past execution traces; fraud detection methods, automatic translation systems, etc. (Muhammad and Yan, 2015). Researchers in political science, economics, and sociology investigated the field of public procurement, trying to highlight possible flaws in the systems that lead to the risk of an inefficient, anomalous contract awarding practice and costly performance for public administrations, to the risk of loosing some tender opportunities or proposing offers with an incongruent amount for economic operators. From the viewpoint of the policies, the big data accumulated in databases on past tenders can be exploited and analyzed with a view of risk prevention (Varian, 2014). 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 train ML models in order to be applied in real-time to the tender cases (Mullainathan and Spiess, 2017).

These premises allow us to pose and search for the answer to the following research questions:

- *RQ1: How can we join different legal data sets to obtain a single labelled one?*
- *RQ2: Do these juridical online data sets contain useful information to generate recommender systems that find similar cases regarding tenders, economic operators, and public administrations distilling similar practices?*
- *RQ3: Is it possible to set up an experiment to predict the possible complaints to administrative courts through the features of public procurement?*

Given the exploratory and application-driven nature of our research, we have chosen a real-world case study to carry out our investigation. As a case study, we applied AI techniques on two real legal data sets about procurement process in Italy. First, a link was sought between the two data sets in order to label procurements with or without complaints. Next, we applied machine learning algorithms to verify the performance of classifiers and the features of procurements that have the most impact on classification. We finally developed a recommendation system by applying machine learning algorithms and deep neural networks to return similar procurement to a given one and find companies for bidders, depending on the procurement requirements.

The remainder of the paper is organized as follows: Section 2 introduces the related works. In Section 3 we describe the case study; in Section 4 we describe the proposed methodology, while Section 5 provides insights about the results of the research. Section 6 briefly discusses some explanations of the predictive outcomes. Finally, Section 7 concludes the paper.

2. Related work

2.1. Study retrieval

We conducted a *Systematic Literature Review* (SLR) according to the approach described in Kitchenham (2004) to retrieve and select the previous studies related to our research (Section 1). Next, guided by these goals, we developed relevant search strings for querying a database of academic papers. We considered the following relevant *keywords*: “public procurement” or “public tender” or “e-procurement” (a relevant study that takes as input a public procurement data set), “machine learning” or “deep learning” (a relevant study that concerns the use of machine- or deep learning techniques), “recommender systems” or “information retrieval” (a relevant study that concerns the use of techniques for searching similar cases described in large corpora of documents), “prediction” (a relevant study focused on the prediction of what can happen in the future).

We combined the above strings to the Google Scholar academic database and retrieved all studies that contained at least one of the phrases in the title, keywords, abstract, or the full text of the paper. We used Google Scholar, a well-known electronic literature database search engine. It encompasses all relevant databases such as *ACM Digital Library* and *IEEE Xplore*, and also allows searching within the full text of a paper. The following *inclusion* criteria were applied to the retrieved studies:

- the study is concerned with predictions in the context of public procurement (this criterion was assessed by reading title and abstract);
- the study is cited at least five times.

The search was conducted in January 2023 and returned 298 distinct documents using initially only the search strings (“public procurement” etc., mentioned above). From this first total, we applied the *inclusion* criteria, obtaining 32 unique papers. The initial search with strings and inclusion filters was automated using a script in Python[1] that exploits Google Scholar’s API to automatically return a data set of publications. We are aware that the five-citation inclusion filter might exclude new, potentially interesting but not yet cited articles. Since the API filter on citations is automatic, we plan to make it *weighted* according to year of publication and number of citations. The studies that passed the inclusion criteria were further assessed according to some *exclusion* criteria. Determining if the exclusion criteria are satisfied required a deeper analysis of the study, e.g., examining the paper’s approach and/or results sections. The applied exclusion criteria are:

- the study does not propose a *predictive method*.
- the study does not concern *outcome*-oriented prediction;
- the study does not take a *public procurement data set* as input.

Applying the exclusion criteria to the 32 studies selected in the previous step resulted in 12 relevant ones.[2] The research was carried out by two of the authors, who shared the results without disagreement.

[1] A sample script can be found at https://github.com/roberto-nai/CLSR2023.

[2] All the retrieved papers that satisfy the inclusion criteria and exclusion can be found at https://github.com/roberto-nai/CLSR2023.

2.1.1. Study retrieval summary

Two ML models have been used in Gallego et al. (2021) to predict whether a contract will result in malfeasance, breach of contract, or inefficiency: a Lasso classification model (Tibshirani, 1996) and an eXtreme Gradient Boosting (XGB) classification model (Friedman, 2001). The study in Rabuzin and Modrusan (2019) explores using advanced text mining to improve the procurement process. It is based on the Public Procurement of Croatia. The authors introduce the use of NLP to improve the research of frauds, comparing common classification algorithms: Naive Bayes (NB), Logistic Regression (LR), and Support Vector Machines (SVM).

The authors of Decarolis and Giorgiantonio (2022) present three main results through detailed data on procurement content involving roadwork contracts in Italy. The prediction capability of the various corruption indicators using standard ML algorithms has been tested: Lasso, Ridge Regression, and Random Forest (RF).

The relation between the award and bidding prices is investigated by García Rodríguez et al. (2019). An award price estimator is proposed using the RF regression method (Breiman, 2001) over the Spanish open data from 2012 to 2018.

The accuracy of eleven ML algorithms for detecting collusion using collusive data sets obtained from Brazil, Italy, Japan, Switzerland, and the United States is tested in Rodríguez et al. (2022).

The authors of Popa (2019) used ML tools to analyze a large data set of public contracts from across Europe to identify the conditions under which close connections among public administrations and economic operators appear, defined in terms of repeated interaction and geographical dispersion. In this case, RF models were used.

The authors of Wang (2016) used SVM and LR to find out the relationship between fraud risk, competition, and performance monitoring using the American SAM (System for Award Management) and FCMD (Federal Contractor Misconduct Database).

Alternative predictive models have been estimated in Fazekas et al. (2021); the article traces the organization of corruption in public procurement by theoretically and empirically assessing the contribution of Extra-legal Governance Organizations (EGO). They used traditional regression and supervised machine- learning methods for identifying and validating proxy indicators for EGO presence in public procurement, such as single bidding or municipal spending concentration.

In Ovsyannikova and Domashova (2020), the use of machine learning methods have been explored, specifically neural networks, to predict public procurement contract outcomes. A Python application was developed to classify public contracts in the pipe industry, identifying high-risk contracts with high non-performance risk.

In Pamučar et al. (2022), a Neuro-Fuzzy neural network is presented for evaluating and predicting the success of a construction company in public procurement.

Finally, there are some literature reviews on this topic; in Torres-Berru et al. (2020) articles from 2015 to 2019 are analysed, in Lyra et al. (2022) articles from 2011 to 2021 are analysed, and in Nai et al. (2022) articles from the last five years (2017 - 2022) are analysed.

As regards recommender systems, their services to retrieve quickly relevant documents to specific cases are often asked for in the legal domain. In Dhanani et al. (2021) the authors propose a method based on graph clustering that forms clusters of referentially similar judgments and within those clusters, it finds semantically relevant judgments. It exploits a highly scalable *Louvain* approach to cluster the judgment citation network, and *Doc2Vec* to capture the semantic relevance among judgments within a cluster. The efficacy and efficiency of the proposed method are evaluated using the large real-life judgments of the Supreme Court of India. Doc2Vec is a form of sentence embeddings that we introduce in our work too but with a different method (SBERT).

In Thomas et al. (2020) the authors propose *Quick Check* a system that extracts the legal arguments from a user's brief and recommends highly relevant case law opinions. It uses a combination of full-text search, citation network analysis, click-stream analysis, and a hierarchy of ranking models. It returns cases that are similar in legal issues and facts.

In Zheng et al. (2022) the authors propose a law recommendation framework, called *LawRec*, based on Bidirectional Encoder Representation from Transformers (BERT) and Skip-Recurrent Neural Network (Skip-RNN) models. It integrates the knowledge of legal provisions with the case description and uses the BERT model to learn the case description text and legal knowledge, respectively. At last, laws and regulations for cases are recommended. Experiment results on the Fayan Cup data set show that it can achieve better performance than state-of-the-art methods. We also used BERT models in our system but used a different version (SBERT) able to rank similar textual content. We used it to retrieve similar procurement and contract descriptions in the ANAC database starting from a given case. We also adopted it to retrieve similar descriptions in the procurement database of contracts starting from the description of the object of complaint in the Administrative Justice in order to find the relevant contract with a non-exact match starting from a complaint.

After summarizing the main research concerning the used methods and technologies, it is important to note how this research can have limitations. It is widely known that ML algorithms are akin to a black box whose working and outcomes are difficult to explain and justify to non-experts. Lawyers and stakeholders can be interested in explaining the results. At the same time, the inherent complexity of the problem being analyzed does not facilitate the task (at least not in a straightforward manner), as explained in Bibal et al. (2021).

2.2. The motivations of our work

Taking the advantage of previous research, which has shown how the use of machine learning and algorithms such as RF, XGB, etc. is also applicable on local government public data, our research attempts to fill a gap in the literature by proposing the evaluation of ML models, not limiting ourselves to detect corruption or estimate a suitable award price. Our goal is also to propose a smart engine to identify cases of similar procurement. If the smart engine recognizes that a public administration received a complaint because of a tender, the following stipulated contracts could be at risk of being stopped by the Administrative Justice action, a fact that could cause an

Table 1 – Main features of the ANAC data set.

Table	Feature	Description
Procurement	Procurement ID: CIG	Alphanumeric value
	Procurement object	Textual summary of the procurement
	Framework agreement between PA and EO	1 if yes, else 0
	Number of lots	Integer value {1..n}
	Procurement type	Supplies Works Services
	Procurement area	Ordinary Special
	Procurement amount	Float value
	Date of publication	Date in format yyyy-mm-dd
	EO selection criterion	Integer value {1..122}
	Realization method	Integer value {1..19}
	Region (NUTS - nomenclature of territorial units for statistics 2022)	Italian region names + Central Government
	CPV division code (CPV codes and nomenclatures 2022)	Integer ID (XX000000-Y)
Award	EO consortium (group of EOs)	1 if it's a group of EOs, else 0 (individual)
	Award date	Date in format yyyy-mm-dd
	Awarded amount (bid amount)	Float value
	Awarded amount drop (bid drop)	Float value
	Number of bids admitted	Integer value {1..n}
	Subcontracting admitted	1 if yes, else 0
Contract Authority	PA denomination	Textual string
Participants	EO denomination	Textual string
Economic Operator	EO denomination	Textual string

increase in costs and execution times for the Public Administration (PA) and the Economic Operator (EO) with an evident loss of economic power; to do this, our research proposes to merge two legal data sets to extract the knowledge needed to achieve one of the goals: to recognize ahead of time that a public administration launching tender or awarding a contract is at risk of receiving a complaint in front of the Administrative Justice.

3. Case study

Our work is based on two legal data sets involving the public procurement process in Italy. The first data set was obtained from the National Anti- Corruption Authority, abbreviated to ANAC, an independent Italian administrative authority whose task is to prevent corruption in the Italian public administration, implement transparency and supervise public contracts. ANAC collects data on calls for procurement from the public contract authority and provides a catalog of Open Data describing public procurement, public authorities (public administrations which create the procurement), and economic operators (contractors who win the procurement). Currently, the ANAC website[3] provides data on approximately 7.5 million of public procurement collected from 2007 to 2022 whose amount is above 40 thousand euros.

The second data set was obtained from the Italian Administrative Justice (IAJ) that contains the judges' sentences related to the public procurement complaints; currently, the IAJ website[4] provides about 67, 850 sentences collected from 2007 to 2022.

3.1. *Data overview*

The ANAC data set[5] contains a table Procurement that store the procurement published in the various Italian regions from 2007 to 2022, The table Contract Authority stores the PAs who created the procurement; the table Participants, stores all EO bidders of each procurement; the table Economic Operator stores the successful EO bidders (winners) and finally the table Award contains data on the procurement award (date of the award, amount of the award, etc.). The complete list of features of the ANAC data set is listed in Table 1; below, the extended description of some feature. Each procurement is identified by an alphanumerical value called CIG (the key ID value), used to connect the above-mentioned tables. A procurement can be defined inside a *framework agreement*, meaning that the PA and EO have a previous agreement to provide services for further procurements for a defined duration of time (e.g. 1 year). Often, a procurement task is split into *lots*, with a lower amount. Procurements have a *type* (Works, Supplies, Services) and a *sector* (Ordinary or Extraordinary) based on whether they are planned or due to extraordinary events (e.g. floods). Each procurement has a well-defined *selection criterion* to choose the EO that will win, and *implementation criterion* which the winning EO will have to comply with. EOs can form a *consortium*, i.e. participate in procurement as an association of EOs.

The IAJ is a textual data set containing the judges' sentences saved in HTML format, DOC/DOCX, and PDF files. In addition to the texts, the sentence files contain some useful metadata: the ECLI code (ECLI 2022) of the sentence, the court

[3] https://dati.anticorruzione.it/opendata.
[4] https://www.giustizia-amministrativa.it/web/guest/dcsnprr.

[5] Open Data collected from the various sections of the ANAC website are publicly available at https://github.com/roberto-nai/CLSR2023.

Table 2 – Quantitative description of the ANAC data set.

Topic	Description	Values
General values	Total procurement (2007–2022)	7, 551, 113
	Total number of PAs	43, 394
	Total number of winning companies (EOs)	192, 009
	Mean number of offers received per procurement	4.5
	Total number of awards	1, 628, 414
Number of procurement by received offers (bidders)	Only 1 offer (bidder)	68.3%
	From 2 to 5 offers (bidders)	19.6%
	More than 5 offers (bidders)	12.1%
Number of procurement by CPV (top 5)	33: Medical and pharmaceutical equipments	24.1%
	45: Construction work	13%
	30: Office and computing machinery	5.7%
	50: Repair and maintenance services	5.6%
	79: Business services	3.7%
Number of procurement by type	Supplies	50%
	Services	35.8%
	Works	14.2%
Number of procurement by region (top 5)	1: Central (Government)	29.9%
	2: Lombardy	12.3%
	3: Piedmont	6.6%
	4: Veneto	5.6%
	5: Lazio	5.3%
Participants	Number of distinct participants	530, 645
	Number of procurement with participants denomination	506, 473

region (that corresponds to the region of the public authority that created the tender), the year and the progressive number of the judge's sentence. Thanks to the ECLI code, it is possible to trace the metadata of complaints related to the sentences: there is the complaint object, the year, and the progressive number (from which the litigation started).

Following the indications of the domain experts, these features were then used as input for the recommender system and the classifiers (described in Section 4.3 and 4.4 respectively).

3.1.1. *Italian case vs. European case*

The "Tenders Electronic Daily" (TED) website[6] publishes 676 thousand procurement notices a year, including 258 thousand calls for tenders which worth approximately 670 billion euro. In this context, the main features of a procurement (e.g., ID, object, type, lots, regions/NUTS,[7] amounts, bidders, awards, etc.) of the ANAC dataset find a generalization in the Open Data of the calls for tenders of other European countries. The ANAC dataset is a specialization of EU data to which national laws apply. Regarding the complaints from IAJ, in all countries there is the possibility of calling a review for a public procurement award and obtaining a judgment[8] with a certain date; in some states such as France and Spain, the judicial review system is similar to the Italian one, because it is devolved to the jurisdiction of the administrative judge (OECD 2000).

In this perspective, our research can be extended to other European countries, as the features of the ANAC data set are the same for all EU countries.

[6] https://ted.europa.eu/TED/main/HomePage.do.
[7] https://ec.europa.eu/eurostat/web/nuts/background.
[8] Directive 2007/66/EC

3.2. *Descriptive analysis of the data set*

In the following Section 3.1, the most relevant information about the ANAC data set is explained. Table 2 shows a quantitative description of the tables contained in the data set, joined together for this research work using the shared key value CIG. The main distinction between procurement is their type: Supplies (50%), Services (35.8%), and Works (14.2%). The main observed issue is that there are data quality problems in table Award because it does not contain the winners for all the procurement (only 21.5% of the procurement also contain information on the award); similarly, table Participants contains only 6.7% of the names of the EOs participating in the procurement. Moreover, the features of the EOs are very limited (there is only the denomination and the VAT ID). Taking the incomplete data into account, however, it can be observed that there is a high number of tenders with only one bidder (68%), also reported by the European public reports (Fazekas, 2019) that in 2019 estimated it in 66.7%; this ratio is higher than in other countries, like, for example, Poland (37.5%), Romania (34%), or Czech Republic (26.6%).

We recall here some broad statistics on procurement in Italy. All the categories of procurement matters are Supplies, Services, and Works and are described by the Common Procurement Vocabulary (CPV). These categories are organized into an ontology (in a hierarchical organization) whose elements are identified by codes. If we use the most significant digits of the codes (that correspond to the upper part of the ontology and the coarsest grain categories) they provide five CPV *divisions* that account for 52% of the total number of procurement. The Central Government accounts for about 30% of the procurement and, together with the top 4 regions, about

60% of the total number of procurement (we recall that, overall, Italy has 20 regions).

4. Methodology

4.1. Problem definition

Following the RQ1, the join between ANAC and IAJ data set is not a naive one since the IAJ database is made by textual documents that could refer to ANAC procurement in different ways: the procurement identifier (CIG), the denomination of the PAs, the EOs (participant or winner), the region and the year of the sentence. Following the RQ3, the problem statement is given: find a ML model such that, given the characteristics (features) of a procurement and the respective label (positive/negative case), it predicts with the highest expected accuracy the presence of a possible complaint related to that procurement to the IAJ courts.

4.2. Merge of data sets

According to Section 4.1, the join between ANAC and IAJ data set was carried out following Information Retrieval (IR) (Belkin and Croft, 1987) and NLP (Nadkarni et al., 2011) techniques. First, the extracted texts from sentences files were indexed with specialized IR tools; *Lucene* (McCandless et al., 2010) and its expansions, *Solr* (Grainger and Potter, 2014) and *ElasticSearch* (ES) (Dixit and Essentials, 2016), represent the major open source IR toolkits used in Industry (Azzopardi et al., 2017). Pursuing the work proposed by Nai et al. (2022) and according to the DB-Engines Ranking of Search Engines (DB-engines ranking of search engines 2023) of February 2023, Elasticsearch (Gormley and Tong, 2015) is the most popular search engine software. Following that indication, the texts and the metadata of complaints and sentences were extracted from the documents and inserted into Newline Delimited JSON (NDJSON (NDJSON - Newline Delimited JSON 2022)) files,[9] one of the input formats accepted by ES for data indexing; the NDJSON schema, where the complaint and the sentence texts have been serialized, is publicly available.[10]

We then explored deep learning techniques to improve the connection between the procurement from the ANAC data set and the sentences from IAJ. We adopted NLP techniques by using *LaBSE BERT* model (Feng et al., 2020) to create sentence embeddings of *procurement object* from ANAC Procurement table (Table 1). Sentence embedding is a collection of techniques based on artificial neural networks that transform textual sentences into vectors of real numbers on which we can compute distances. These real numbers represent the probabilities that the words appear in the sentence. The probabilistic nature of sentence embeddings explains their power when we measure the distances among two vectors representing sentences. These distances depend meaningfully on their similarity in meaning. Lower the vector distance, the more similar the sentences. The same technique was adopted for the *complaint object* of sentences published by IAJ. Cosine similarity (Salton and Buckley, 1988) and TF-IDF fuzzy matching (Tata and Patel, 2007) were then applied on sentence embeddings to collect the corresponding similar subject of the *procurement object* and the *complaint object* collectively. Although ES reduces the match results via the BM25 (Robertson and Zaragoza, 2009) score algorithm, exact results in ES can be obtained by using the *match phrase* query which will only return documents that precisely match the phrase that a user is searching for (this is even more strict than a match query using the AND operator); through this type of query, fuzzy cases were avoided. For the NLP phase, pairs of subjects {procurement object, complaint object} whose similarity and TF-IDF fuzzy matching exceeded 0.85 were considered; the Python source code is publicly available.[11]

When the match between the entries of the two data sets was successful (via IR or NLP), we used the presence of complaint on procurement as a *positive case* on that procurement entry; otherwise, as a *negative case*.

Fig. 1 summarises the workflow described above.

4.3. The recommender system on procurement

Following the research question *RQ2*, we relied on an abstract and general representation of the brief, textual description of procurement provided by the responsible person of the awarding procedure from the PA to find similar procurement in the database. The goal and benefit of such a system could be to have the advantage of comparing the conditions of the respective contracts (awarded amount, discount, contract conclusion time) and acquire more knowledge regarding the public administrations that perform similar calls and the economic operators that answer. To build an abstract and general representation of the semantic content of the contract description, we trained the numerical vectors called *sentence embeddings* using BERT (Devlin et al., 2018). We used as input sentences the brief descriptions in natural texts of procurement in the ANAC database. We obtained vectors with 768 dimensions. Successively, given a case of an individual procurement, we searched for the most similar and relevant ones in the rest of the database using SBERT (Reimers and Gurevych, 2019) and LaBSE (Feng et al., 2020): they are a multilingual version of BERT, using siamese networks able to work on multilingual and Italian corpora.[12] They are often used as tools to rank a set of sentences from the most similar to a given sentence, called *query*, to the less relevant ones.

Fig. 2 shows the principle guiding SBERT. The pooling operation to the output of BERT is used to derive a fixed-sized sentence embedding that can later be compared on similarity by the well-known cosine similarity measure. The outcome is used to perform the ranking of the sentences. We used the linear search approach to find the k best exemplars (neighbors) of a given procurement with a cosine score as a similarity mea-

[9] NDJSON is a convenient format for storing or streaming structured data that may be processed one record at a time; it is a format suitable for the data exchange in software client/server applications.

[10] https://github.com/roberto-nai/CLSR2023.

[11] https://github.com/roberto-nai/CLSR2023.

[12] SBERT and LaBSE can work even on other languages for which we do not have a training set because the internal model on embeddings is independent on the language.

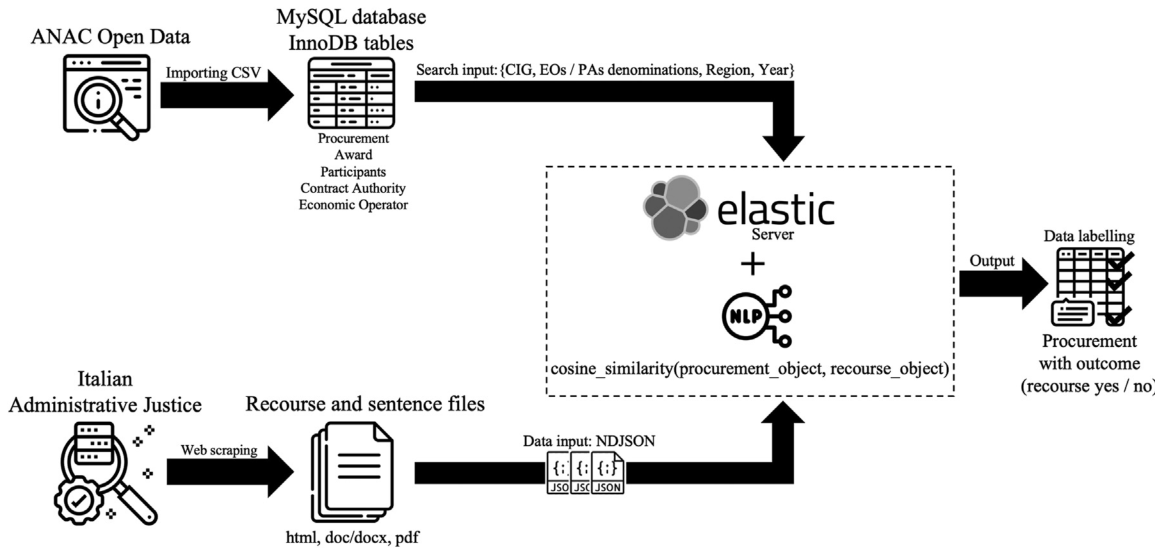

Fig. 1 – Methodology workflow from data collection to merging and labeling.

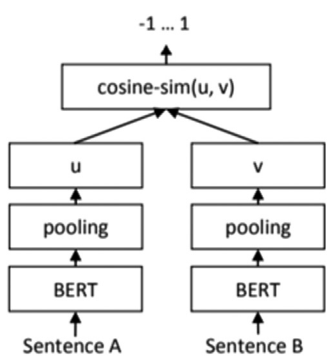

Fig. 2 – The method used by SBERT to evaluate the similarity between two sentence embeddings.

sure. To overcome the scalability issue related to the computation of the similarity score for a large number of cases, we calculated the cosine similarities in batches: saving only the k so far best results found for any procurement example, we can forget all the other results, free the RAM, and proceed to the next batch.

4.4. *ML prediction models training*

The No Free Lunch Theorem (NFLT) (Adam et al., 2019) states that no ML models work best in all situations and data sets. Following it, the best approach to find out the ML model whose prediction is the most accurate is to test multiple

ML models, and tune and compare them for the specific scenario. Since the problem is predicting whether a procurement will have a complaint, a binary classification algorithm will be used. Identified the solution as a supervised learning classification task (Cord and Cunningham, 2008), the following classifiers were explored: K- Nearest Neighbours (KNN) (Peterson and neighbor, 2009), LR (Wright, 1995), NB (Rish et al., 2001), SVM (Wang, 2005), Decision Tree (DT) (Kotsiantis, 2013), RF (Breiman, 2001), and XGB (Chen et al., 2015). The input features of the ML models are described in Table 1, to which the dependent variable "complaint" needs to be added, which takes the value 1 if the procurement has experienced complaint, 0 otherwise (Section 4.2). The ML models were implemented in Python (version 3.8) and the *Scikit-learn* library (version 1.2); the script is publicly available.[13]

5. Results

5.1. *Merge of data sets*

As described in Section 4.2, we performed three types of searches, aiming to increasingly improve the matches between IAJ sentences and the ANAC procurement to be labeled: by {CIG}, by {EO participant denomination, EO winner denomination, PA denomination, Region/Court, Year}, by the similarity between {procurement object, complaint object}. The results in Table 3 show how the methods, used incrementally, improve the reference between the ANAC and IAJ data sets based on the available sentences (67, 850).

5.2. *ML data set input*

The finally obtained labeled data set consists of 15, 117 rows; the distribution of *positive cases* (procurement with complaint) was analyzed in terms of type of procurement, region, CPV division, and year. Following, the complaint distribution; the year with the highest number of complaints was 2018 (13.1%), followed by 2020 (11.3%) and 2017 (11.2%); type of procurement with the highest number of appeals is Services (49.5%), followed by 'Works (33.1%) and Good/Supplies (12.29%); the region with the highest number of appeals is the special Central region (19.2%), followed by Lombardy (10.4%) and Campania

[13] https://github.com/roberto-nai/CLSR2023.

Table 3 – Reference found between ANAC procurement and IAJ sentences.

Reference found by feature	Total	Overall percentage
Procurement identifier: CIG	8, 418	12.4%
Denomination: EO participant, EO winner, PA, region/court, year	4, 178	18.5%
Similarity: procurement object, complaint object	2, 491	22.3%

(8.61%); the CPV division with the highest number of appeals is number 45 (represents the construction works), followed by number 71 (architectural, construction, engineering, and inspection services) and 90 (sewage-, refuse-, cleaning-, and environmental services).

The labeled data set was then divided into three smaller data sets containing procurement grouped by type: a data set for Works of 10, 150 rows (5, 075 positive/ negative cases), one for Services of 15, 028 rows (7, 514 positive / negative cases) and one for Supplies of 5, 232 rows (2, 616 positive / negative cases). The smaller labeled data sets were used as input for the ML algorithms, balancing the same number of *positive cases* and *negative cases* keeping negative cases distributed like the positive cases (the negative cases, which are more frequent than positive ones, were randomly sampled keeping the balance with the positive cases per year, region, and CPV). This balancing effort is necessary because we do not want that ML models are biased toward the negative cases that represent the most frequent class. The bias toward the majority class is a well-known side effect of the statistical evaluation measures used to evaluate the ML model training that can be eliminated by a random selection of the cases from the majority class and making balanced the frequency of the cases from the classes in the training set. A sample of the three extracted data sets is publicly available.[14]

5.3. *ML models performance measures*

For validating the classification model, three different ratios between the training and test subsets (*train: test* in percentage) were randomly chosen with values 90: 10, 80: 20, and 70: 30. As XGB and RF were the most promising algorithms, the models they learned from data were fitted to obtain the best hyper-parameters (the parameters driving the algorithms): the Python library *Hyperopt*[15] proved to be practical and effective for this goal (Bergstra et al., 2013). Fig. 3 sh; Bergstra et al., 2015ows ROC/AUC[16] (Géron, 2022) curves for each of the models using data in the hold-out set. The figures represent the *true positive rate* (the proportion of detected complaint cases that were correctly classified) and the *false positive rate* (the proportion of incorrectly guessed complaint cases) achieved by the classifiers. As a reminder, the classifier performs better as it approaches the top-left corner of the plot where all cases are correctly classified. Instead, it is indistinguishable from random guessing as it comes closer to the diagonal (45-degree line). As can be seen from the pictures, for the Services and Works data sets the best results were obtained by XGB with a maximum peak of 0.928 with a *train: set* value of 80: 20; for the Supplies data set, the best results were obtained by RF and XGB with a maximum peak of 0.864 with a *train: test* set value of 80: 20. Table 4 also shows the results in terms of Accuracy (Géron, 2022) and F1-Score[17] (Géron, 2022) of the models; consistent with ROC/AUC values, XGB and RF have the best performance. As our task is a classification with unbalanced class distributions (the number of cases with complaint is smaller than those without), we also adopted the *Stratified K-Fold* technique to validate the models (Raschka and Mirjalili, 2019); Table 5 shows the results of the three best models with a 10-fold cross-validation (a good standard value for k is 10, as suggested by Kohavi et al., 1995); as can be seen, the final values decrease slightly compared to those in Table 4, which was used as a baseline.

5.4. *Recommender system performance evaluation*

To evaluate the performance of our recommender system, we decided to evaluate its Precision at 10.[18] To compute it, a panel of 3 persons worked independently on a test set composed of 100 random procurement examples of each type of Works, Services, and Supplies. Since each example is a possible complex sentence that refers to multiple elements (the kind of procedure for the award of the contract, the location, the subject, and various features depending on the procurement matter), the panel agreed in advance, case by case, on the elements of judgment and their weight (in percent) depending if they believe that particular key element should be found in a recommendation. The number of these elements of judgment span from 2 to 5. Successively, each member of the panel, independent from the others, gave a relevance score of actual similarity between the query tender and its recommendations on each of the key elements spotted in advance. An example of this procedure for procurement of the type Works can be seen in Table 6. We notice that the school's name is given a low weight compared with the other elements (type of intervention and type of building).

14 https://github.com/roberto-nai/CLSR2023.

15 https://hyperopt.github.io/hyperopt-sklearn/.

16 Receiver Operating Characteristic (ROC) is the graph that represents the fraction of the correct positive predictions (True Positive Rate) out of the positive cases and the fraction of erroneous positive predictions (False Positive Rate) out of the negative cases; the Area Under the ROC Curve (AUC) corresponds to the accuracy of the prediction model.

17 F1-Score is the harmonic mean between recall and precision and it is often used for combining precision and recall in a unique measure of the prediction performance.

18 Precision at 10 for the recommendation of one item is the fraction of the ten returned results, which is relevant. If the test set comprises n items, it is the average over them.

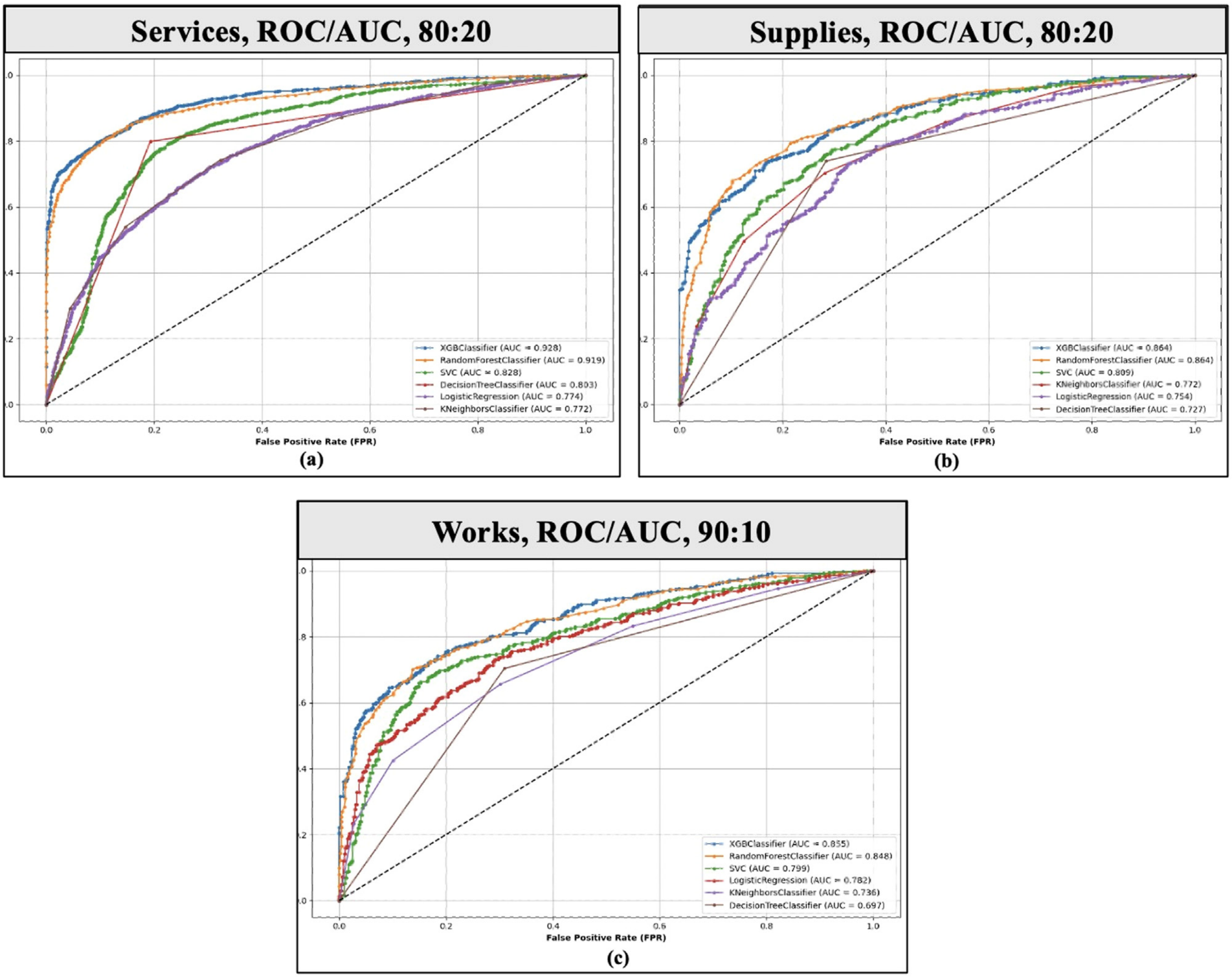

Fig. 3 – ROC/AUC in the best case for the ML models applied to Services (a), Supplies (b), and Works (c) data sets. Full size image available at https://github.com/ roberto-nai/CLSR2023.

Table 4 – ML models performance measures on *train: set* best case.

Data set	Classifier	Accuracy	F1-Score	ROC/AUC
Services (15,028 rows, 80: 20)	XGB	0.847	0.841	0.928
	RF	0.847	0.841	0.919
	SVM	0.802	0.801	0.828
Works (10,150 rows, 90: 10)	XGB	0.773	0.761	0.855
	RF	0.768	0.756	0.848
	SVM	0.756	0.745	0.799
Supplies (5232 rows, 80: 20)	RF	0.780	0.774	0.864
	XGB	0.778	0.770	0.864
	SVM	0.751	0.748	0.809

Gathered together the scores, we can define a final relevance score by computing the mean of the scores given by the people of the panel.

The results of the Precision at 10 will then depend on which threshold θ for the score just created we choose; the lower the threshold, the higher the precision will be. We can think of this threshold as a measure of how strictly similar we want the recommended procurement and the query. A score of 0.9 means that the description of the procurement must be almost identical, with at most minor differences (an example is given by Table 6). A summary of the precision values at 10 with different thresholds can be seen in the chart at Fig. 4.

We see, as expected, that the precision tends to decrease as we increase the threshold. Moreover, we can note that the

Table 5 – ML models performance measures on Stratified K-Fold (average values).

Data set	Classifier	Accuracy	F1-Score	ROC/AUC
Services (15,028 rows, $k = 10$)	XGB	0.844	0.835	0.919
	RF	0.836	0.830	0.909
	SVM	0.728	0.709	0.795
Works (10,150 rows, $k = 10$)	XGB	0.743	0.726	0.815
	RF	0.74082	0.720	0.812
	SVM	0.707	0.679	0.769
Supplies (5232 rows, $k = 10$)	RF	0.759	0.747	0.828
	XGB	0.758	0.735	0.823
	SVM	0.700	0.684	0.773

Table 6 – Example of the scoring procedure for one recommended item for a tender query.

Type	Element 1	Element 2	Element 3	Element 4
Tender query	upgrading	and safety measures for the	Crespellani	Elementary school
Weights	30	30	10	30
recommendation	upgrading	and safety measures for the		Elementary school

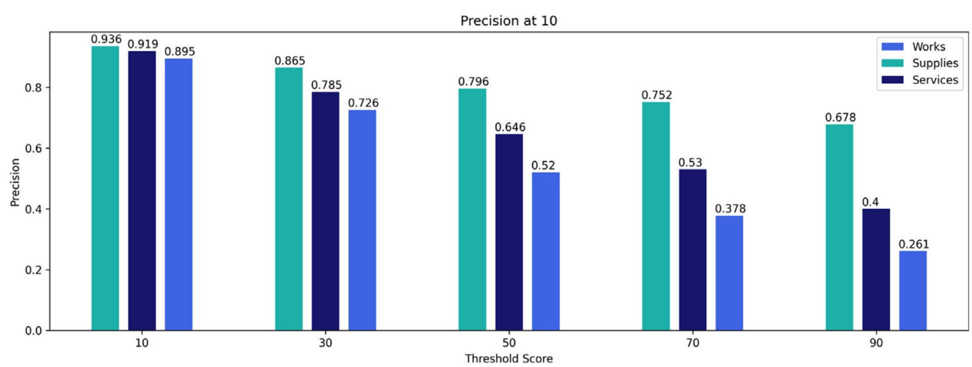

Fig. 4 – Results of precision at 10 for Works, Supplies, Services with different thresholds.

top 10 recommended items almost always have at least something in common with the query; it has to be noted, however, that for some tenders, there could not exist any other tender which is relevant to the recommender systems. It can do nothing but fail in its task. Moreover, we can observe how the recommendation system seems to work better for tenders of the type of Supplies. This makes sense looking at their description: they tend to be shorter concerning Works and Services, with just a couple of words that describe the object of the supply (some found examples are: “Laboratory materials”, “Vertical signage supply”, “Toner”).

Additionally to the precision at 10 for services, the trend of the precision at K, for $K = 1, \ldots, 10$ with different thresholds, can be seen in the chart at Fig. 5. Similar trends are observed for the other procurement types. We observe that the obtained precision values for the first 10 examples are satisfactory and in line with the possibility of providing a set of a conspicuous number of examples that share many of the description elements in procurement and for all the procurement types.

6. Explainable models

Quite often ML models are black boxes, as is the case of ensemble methods[19] like RF and XGB, because they make predictions that are difficult to explain or justify because the outcomes are due to a multitude of features. In particular, in the domain of justice, it is very important to justify the outcomes of the predictions since the end-users need to be informed about the reasons why the procurement risk a complaint.

According to surveys on Explainable AI models, such as Mohseni et al. (2021), an explainable model is global if it explains the behavior of a ML model in its entirety, or it is local if it explains the predictions of individual instances. The locality is particularly useful in the legal domain because it allows a counterfactual explanation (Dandl and Molnar, 2020): in the legal domain it explains which are the smallest changes in the

[19] ensemble models assemble a large number of simpler models.

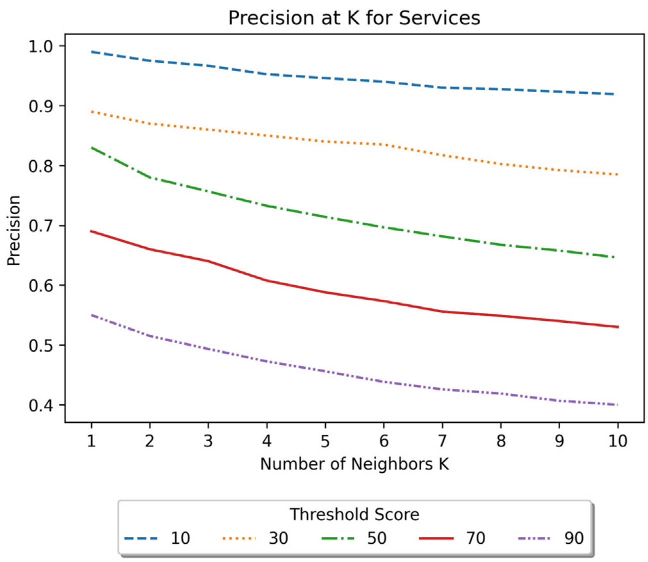

Fig. 5 – Results of precision at 10 for Services with different number K of neighbours.

feature values of procurement that change the prediction to a predefined output, e.g., might turn the presence of a complaint into an absence.

An ML model is transparent if its predictions are immediately explainable; it is post-hoc if the explanation is obtained a-posteriori of the predictive model, with the adoption of further procedures. In this work, we adopt a post-hoc method, called SHapley Additive exPlanations or SHAP (Lundberg and Lee, 2017) that can be used both for the explanation of prediction on single examples and for providing the features that are most decisive for the predictions of a ML model. SHAP is an explanatory method based on a solid mathematical foundation, which illustrates individual predictions based on the Shapley values of game theory. S. Lundberg and S. Lee in Lundberg and Lee (2017) reframe the problem of computing how each member of a team or coalition contributes to a coalition value, into the problem of computing how much a feature value in a given instance contributes to the model output. The idea is to explain the prediction of the original ML model (denoted here by f) on an instance x through a surrogate and simpler model, the explanation model (denoted by g), by a score computed as the sum of the contributions of a subset of the original features, each with a unit weight multiplied by a coefficient that is the Shapley value. Prediction is given by:

$$f(x) \approx g(x') = \Phi_0 + \sum_{i}^{M} \Phi_i \qquad (1)$$

where x' represents the instance x projected on a subset composed by M original features and Φ_i are the Shapley coefficients.

In this way, the outcome score of the ML model is obtained by an additive formula that can be used both for explaining the prediction of single instances (in which the features that have the highest Shapley coefficients weigh more) and for the global prediction.

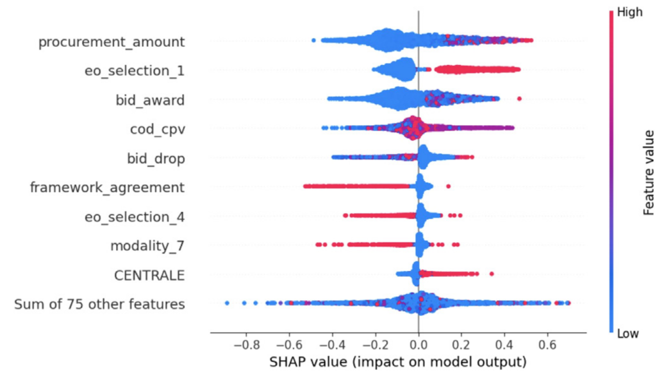

Fig. 6 – SHAP model for explaining a ML model outcome.

We computed the SHAP values for explaining the outcomes of the RF using the Python SHAP library[20]; the script is publicly available.[21] As a reminder, SHAP shows the contribution or the importance of each feature on the prediction of the model, it does not evaluate the quality of the prediction itself.

Fig. 6 shows the SHAP values for the RF model applied to the Services data set; in the bee-swarm plot the features are ordered by their effect on prediction, but we can also see how higher and lower values of the feature will affect the result. Each dot in the plot represents a single observation; the horizontal axis represents the SHAP value, while the color of the point shows if that observation has a higher or a lower value when compared to other observations. In Fig. 6, higher procurement amount and the maximum frequency of EO selection criterion (of value 1) have an impact on the prediction of *positive cases*, while lower values have an impact on the prediction of *negative cases*; the third most important feature affecting the model is the bid award, whose logic is similar to the procurement amount. As also expected from the domain experts, the procurement amounts (amount and bio-award) and the tender selection criteria (eo selection), are the features that most influence the litigation of the tender; this consideration of the domain experts, reinforces the need for an increased explainability of the results of the IA models.

7. Conclusion and future work

This work demonstrates the possibility to manage a huge juridical data set from the Italian National Public Authority to automatically extract meaningful knowledge to address ML experiments (*RQ1*); aside from *RQ1*, this research work showed that the current Italian data set of public procurement has data quality problems (i.e. missing data on awards and participants, not very descriptive data features on EOs). In addition, for *RQ2*, we explored the results of a recommender system that we trained with the successful technology of deep neural networks with sentence embeddings and show that their results are actually reliable and potentially useful. We trained and

[20] https://shap.readthedocs.io/en/latest/.
[21] https://github.com/roberto-nai/CLSR2023.

tested a predictive experiment to estimate the prediction of a complaint presence in front of the administrative courts on the basis of the features of public procurement (RQ3). Finally, we tried to provide an explanation of the prediction model using SHAP (a method from the game theory computing the effect of procurement features in deciding the prediction).

This research work shows that increasingly established methods and technologies (Open Data, IR systems, ML, NLP, etc.) can improve the public administration sector systems, proceedings, and services.

In future work, we plan to explore contemporary approaches like few-shot learning, in-context learning, and large language models (e.g., GPT-3, T5) to test whether they improve the performance of the models used so far, also for the recommendation system to address scalability and robustness. We plan to investigate furthermore the explainable AI techniques. As stated by Carvalho et al. (2019), ML systems are becoming increasingly ubiquitous. These systems' adoption has been expanding, accelerating the shift towards a society informed by decisions driven by algorithms. However, most of these accurate decision support systems remain complex black boxes, meaning their internal logic is hidden and even experts might not fully understand the rationale of their predictions. Moreover, new regulations have made mandatory the audit and the explainability of the decisions, increasing the demand to question, understand, and trust ML systems (Meo et al., 2022). Another direction of study concerns the adoption of Process Mining techniques (Van Der Aalst and van der Aalst, 2016) for conformance checking and predictive process monitoring applied to a log of temporal events obtained by ANAC and IAJ data sets.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

data/code linked in the article

REFERENCES

Adam SP, Alexandropoulos S-AN, Pardalos PM, Vrahatis MN. No free lunch theorem: a review, approximation and optimization (2019) 57–82.

Azzopardi L, Moshfeghi Y, Halvey M, Alkhawaldeh RS, Balog K, Di Buccio E, et al. Lucene4ir: developing information retrieval evaluation resources using lucene. ACM SIGIR Forum; 2017. p. 58–75 ACM New York, NY, USA.

Belkin NJ, Croft WB. Retrieval techniques; 1987.

Bergstra J, Komer B, Eliasmith C, Yamins D, Cox DD. Hyperopt: a python library for model selection and hyperparameter optimization. Comput Sci Discovery 2015;8(1).

Bergstra J, Yamins D, Cox DD, et al. Hyperopt: a python library for optimizing the hyperparameters of machine learning algorithms. Proceedings of the 12th Python in Science Conference. Citeseer; 2013. p. 20.

Bibal A, Lognoul M, De Streel A, Frénay B. Legal requirements on explainability in machine learning. Artif Intell Law 2021;29:149–69.

Breiman L. Random forests. Mach Learn 2001;45:5–32.

Carvalho DV, Pereira EM, Cardoso JS. Machine learning interpretability: a survey on methods and metrics. Electronics 2019;8(8):832.

Chen T, He T, Benesty M, Khotilovich V, Tang Y, Cho H, et al. Xgboost: extreme gradient boosting, R package version 0.4-2 1 (4) (2015) 1–4.

Collins C, Dennehy D, Conboy K, Mikalef P. Artificial intelligence in information systems research: a systematic literature review and research agenda. Int J Inf Manage 2021;60.

Cord M, Cunningham P. Machine learning techniques for multimedia: case studies on organization and retrieval. Springer Science & Business Media; 2008.

CPV codes and nomenclatures, https://simap.ted.europa.eu/web/simap/cpv, visited: 2022-12-01 (2022).

Dandl S, Molnar C. Counterfactual explanations (2020).

DB-engines ranking of search engines, https://db-engines.com/en/ranking/search+engine, visited: 2023-02-01 (2023).

Decarolis F, Giorgiantonio C. Corruption red flags in public procurement: new evidence from Italian calls for tenders. EPJ Data Sci 2022;11(1):16.

Devlin J, Chang M-W, Lee K, Toutanova K. Bert: pre-training of deep bidirectional transformers for language understanding, arXiv preprint arXiv:1810.04805 (2018).

Dhanani J, Mehta R, Rana D. Legal document recommendation system: a cluster based pairwise similarity computation. J Intell Fuzzy Syst 2021;41(5):5497–509.

Dixit B. Elasticsearch essentials. Packt Publishing Ltd; 2016.

Dumas M, La Rosa M, Mendling J, Reijers HA, et al. Fundamentals of business process management, 2. Springer; 2018.

ECLI, https://e-justice.europa.eu/content_european_case_law_identifier_ecli-175-en.do, visited: 2022–12-01 (2022).

Fazekas M. Single bidding and non-competitive tendering procedures in Eu co-funded projects. Publications Office of the European Union; 2019.

Fazekas M, Sberna S, Vannucci A. The extra-legal governance of corruption: tracing the organization of corruption in public procurement. Governance 2021.

Feng F, Yang Y, Cer D, Arivazhagan N, Wang W. Language-agnostic Bert sentence embedding, arXiv preprint arXiv:2007.01852 (2020).

Friedman JH. Greedy function approximation: a gradient boosting machine. Ann Stat 2001:1189–232.

Géron A. Hands-on machine learning with Scikit-learn, Keras, and TensorFlow. O'Reilly Media, Inc.; 2022.

Gallego J, Rivero G, Mart´ınez J. Preventing rather than punishing: an early warning model of malfeasance in public procurement. Int J Forecast 2021;37(1):360–77.

Garc´ıa Rodr´ıguez MJ, Rodr´ıguez Montequ´ın V, Ortega Fernández F, Villanueva Balsera JM. Public procurement announcements in spain: regulations, data analysis, and award price estimator using machine learning. Complex 2019;2019:1–20.

Gormley C, Tong Z. Elasticsearch: the definitive guide: a distributed real-time search and analytics engine. O'Reilly Media, Inc.; 2015.

Grainger T, Potter T. Solr in action. Manning Publications Co.; 2014.

K¨appel M, Jablonski S, Sch¨onig S. Cost-sensitive predictive business process monitoring. European conference on advances in databases and information systems. Springer; 2021. p. 14–26.

Kitchenham B. Procedures for performing systematic reviews, Keele, UK, Keele University 33 (2004) (2004) 1–26.
Kohavi R, et al. A study of cross-validation and bootstrap for accuracy estimation and model selection. Ijcai; 1995. p. 1137–45 Montreal, Canada.
Kotsiantis SB. Decision trees: a recent overview. Artif Intell Rev 2013;39:261–83.
Lundberg SM, Lee S-I. A unified approach to interpreting model predictions. Adv Neural Inf Process Syst 2017;30.
Lyra MS, Pinheiro FL, Bacao F. Public procurement fraud detection: a review using network analysis. Complex networks & their applications X: volume 1, proceedings of the tenth international conference on complex networks and their applications complex networks 2021 10. Springer; 2022. p. 116–29.
McCandless M, Hatcher E, Gospodnetić O, Gospodnetić O. Lucene in action, Vol. 2. Manning Greenwich; 2010.
Meo R, Nai R, Sulis E. Explainable, interpretable, trustworthy, responsible, ethical, fair, verifiable AI... What's next?. European conference on advances in databases and information systems. Springer; 2022. p. 25–34.
Mohseni S, Zarei N, Ragan ED. A multidisciplinary survey and framework for design and evaluation of explainable ai systems. ACM Trans Interact Intell Syst (TiiS) 2021;11(3–4):1–45.
Muhammad I, Yan Z. Supervised machine learning approaches: a survey. ICTACT J Soft Comput 2015;5(3).
Mullainathan S, Spiess J. Machine learning: an applied econometric approach. J Econ Perspect 2017;31(2):87–106.
Nadkarni PM, Ohno-Machado L, Chapman WW. Natural language processing: an introduction. J Am Med Inform Assoc 2011;18(5):544–51.
Nai R, Sulis E, Meo R, et al. Public procurement fraud detection and artificial intelligence techniques: a literature review. Companion proceedings of the 23rd international conference on knowledge engineering and knowledge management. CEUR-WS; 2022. p. 1–13.
Nai R, Sulis E, Pasteris P, Giunta M, Meo R. Exploitation and merge of information sources for public procurement improvement. Joint European conference on machine learning and knowledge discovery in databases. Springer; 2022. p. 89–102.
NDJSON - Newline Delimited JSON, https://dataprotocols.org/ndjson, visited: 2022-12-01 (2022).
NUTS - nomenclature of territorial units for statistics, https://ec.europa.eu/eurostat/web/nuts/background/, visited: 2022-12-01 (2022).
OECD. Public procurement review procedures (30); 2000.
Ovsyannikova A, Domashova J. Identification of public procurement contracts with a high risk of non-performance based on neural networks. Procedia Comput Sci 2020;169:795–9.
Pamu˘car D, Bozanic D, Pu˘ska A, Marinković D. Application of neuro- fuzzy system for predicting the success of a company in public procurement, decision making. Appl Manage Eng 2022;5(1):135–53.
Peterson LE. K-nearest neighbor. Scholarpedia 2009;4(2):1883.
Popa M. Uncovering the structure of public procurement transactions. Bus Politics 2019;21(3):351–84.
Rabuzin K, Modrusan N. Prediction of public procurement corruption indices using machine learning methods. KMIS; 2019. p. 333–40.
Raschka S, Mirjalili V. Python machine learning: machine learning and deep learning with python, Scikit-learn, and Tensorflow 2. Packt Publishing Ltd; 2019.
Reimers N, Gurevych I. Sentence-bert: sentence embeddings using Siamese Bert-networks, arXiv preprint arXiv:1908.10084 (2019).
Rish I, et al. An empirical study of the naive Bayes classifier. IJCAI 2001 workshop on empirical methods in artificial intelligence; 2001. p. 41–6.
Robertson S, Zaragoza H, et al. The probabilistic relevance framework: Bm25 and beyond. Found Trends® Inf Retrieval 2009;3(4):333–89.
Rodr´ıguez MJG, Rodr´ıguez-Montequ´ın V, Ballesteros-Pérez P, Love PE, Signor R. Collusion detection in public procurement auctions with machine learning algorithms. Autom Constr 2022;133.
Salton G, Buckley C. Term-weighting approaches in automatic text retrieval. Inf Process Manag 1988;24(5):513–23.
Tata S, Patel JM. Estimating the selectivity of tf-IDF based cosine similarity predicates. ACM Sigmod Record 2007;36(2):7–12.
Thomas M, Vacek T, Shuai X, Liao W, Sanchez G, Sethia P, et al. Quick check: a legal research recommendation system. NLLP@KDD; 2020.
Tibshirani R. Regression shrinkage and selection via the lasso. J R Stat Soc Ser B-Methodological 1996;58:267–88.
Torres-Berru Y, Batista VFL, Torres-Carrión P. Data mining to detect and prevent corruption in contracts: systematic mapping review, RISTI-Revista Iberica de Sistemas e Tecnologias de Informacao 2020 (E29) (2020) 13–26.
Van Der Aalst W, van der Aalst W. Data science in action. Springer; 2016.
Varian HR. Big data: new tricks for econometrics. J Econ Perspect 2014;28(2):3–28.
Wang L. Support vector machines: theory and applications, 177. Springer Science & Business Media; 2005.
Wang Y. Detecting fraud in public procurement. State University of New York at Stony Brook; 2016 Ph.D. thesis.
Wright RE. Logistic regression. (1995).
Zheng M, Liu B, Sun L. Lawrec: automatic recommendation of legal provisions based on legal text analysis. Comput Intell Neurosci 2022;2022.