AI Applied to the Analysis of the Contracts of the Italian Public Administrations

Roberto Nai[1,*], Ishrat Fatima[1], Gabriele Morina[1], Emilio Sulis[1], Laura Genga[2], Rosa Meo[1] and Paolo Pasteris[1]

[1]Computer Science Department, University of Turin, Corso Svizzera 185, Torino (TO), 10149, Italy
[2]Eindhoven University of Technology, De Zaale, Eindhoven, Netherlands

Abstract

The proliferation of e-procurement systems in the public sector allows for joint access to useful and open information sources. Our research explores ways to improve the quality and correctness of the public procurement process and the efficiency of administrations, the reduction of the time spent by economic operators, and the costs of public administrations. In particular, we explored the dataset of the National Anti-Corruption Authority in Italy on public procurement and the judges' sentences related to public procurement. Our first goal was to identify which procurement led to disputes and recourse to Administrative Justice by identifying relevant procurement features. Our second goal was to develop a recommender system on procurement by applying machine learning algorithms and deep neural models to return similar procurement to a given one and find companies as potential bidders, depending on the procurement requirements. Our third goal is to automate the analysis of a dataset of public procurement, contract awards, and appeal procedures. Process discovery techniques were applied to the dataset, considering control-flow, organizational (resource), and time perspectives. The results demonstrate the importance of applying these techniques in the legal field.

Keywords

Public Procurement, Legal Prediction, Natural Language Processing, Machine Learning, Process Mining, Variant Analysis

1. Introduction

Legal informatics is expanding due to the digitization of law, allowing for the exploitation of computational technologies and algorithms.

Therefore, the applicative tasks can easily include compliance analysis and anomaly detection with Artificial Intelligence (AI) methods. Some of the main AI techniques successfully applied in the legal fields include *Machine Learning* (ML) and specifically *Deep Learning* (DL) models that can automatically extract some knowledge about the semantics in texts. Before the advent of DL models, the application of *Natural Language Processing* (NLP) already achieved good results in many tasks involving natural languages, such as text modeling, parsing, machine translation, and automatic query answering.

In the same perspective, a relatively new approach for providing knowledge about data registered in information systems is *Process Mining* (PM), aimed at discovering, monitoring and analyzing organization processes. It exploits the data generated during the process execution, stored in the form of *event logs* [1]. As a bridge between data mining and business process management (BPM), the discipline provides a meaningful process-oriented perspective. Thus, temporal series of legal events can be investigated to *automatically discover* and visualize the sequential execution (or *control-flow*) of activities belonging to a legal process, such as the procurement award process, or the issuance of a sentence.

These premises allow us to pose and search for the answer to the following research questions: RQ1) *How can we automatically extract information from different national juridical online data sets?* 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 event of recourse to administrative courts through the features of public procurement?* RQ4) *Can we obtain meaningful insights from applying process discovery techniques on legal data from information systems?*

This paper focuses on two research approaches with complementary methodologies on the same datasets. In the following, Section 2 provides the background of our work by introducing some related work in Section 2.1 and the case study in Section 2.2. The first project considers automatic textual information extraction with Information Retrieval and ML techniques, described in Sec-

Ital-IA 2023: 3rd National Conference on Artificial Intelligence, organized by CINI, May 29–31, 2023, Pisa, Italy

*Corresponding author.
roberto.nai@unito.it (R. Nai); ishrat.fatima@unito.it (I. Fatima); gabriele.morina@edu.unito.it (G. Morina); emilio.sulis@unito.it (E. Sulis); L.Genga@tue.nl (L. Genga); rosa.meo@unito.it (R. Meo); paolo.pasteris@unito.it (P. Pasteris)
0000-0003-4031-5376 (R. Nai); 0000-0002-3807-4107 (I. Fatima); 0009-0009-6033-3441 (G. Morina); 0000-0003-1746-3733 (E. Sulis); 0000-0001-8746-8826 (L. Genga); 0000-0002-0434-4850 (R. Meo); 0000-0001-9638-9273 (P. Pasteris)

tion 3. A second contribution addresses an organizational perspective with the automatic discovery of activity sequences using process mining algorithms in Section 4. Finally, Section 5 presents conclusions and future work.

2. Background

2.1. Related work

We conducted a *Systematic Literature Review* (SLR) according to the approach described in [2, 3] to retrieve and select the previous studies related to our research, starting by specifying the research questions.[1].

In [4], the authors describe how to find connections between the procurement data and the appeals and how to exploit the resulting data for the measurement of litigation and clustering into communities, the nodes representing entities having similar interests. How network analysis can improve prediction on legal data has been described in [5].

Alternative predictive models have been estimated in [6]; Extra-legal Governance Organizations (EGOs) have been identified as major contributors to Italian corruption in public procurement.

As regards recommender systems, in [7] 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.

Our goal is to propose a smart engine to identify cases of similar procurement. If the smart engine recognizes that a public administration received a recourse because of a tender, the following stipulated contracts could be at risk of being stopped by the Administrative Justice action.

To the best of our knowledge, very few works investigated a process-oriented approach to legal cases.

At the intersection between PM and law, some works explore real-world cases of process discovery involving public procurement: a case study focuses on a heuristic algorithm revealing a concept drift in the publication of contracts in the Philippines [8].

An application of process discovery in the legal field is discussed in [9], where the authors applied knowledge discovery techniques for the extraction of lawsuit processes from the information system of the Court of Justice of the State of Sao Paulo, Brazil.

We build on these works and, as proof of concept for PM, we analyzed the results considering each of the following PM *perspectives*: control-flow, organizational (resource), and time [1].

[1]For completeness, all the retrieved papers that satisfy the inclusion and exclusion criteria can be found at: https://tinyurl.com/ksxaz7uv

2.2. Case study

Our work is based on two legal data sets involving the public procurement process in Italy.

The first dataset was obtained from the National Anti-Corruption Authority (ANAC), which collects data on calls for procurement from the public contract authority and provides a catalog of Open Data describing public procurement, Public Administrations (PAs) which are responsible for the procurement, and economic operators (EOs) participating in the tender awarding process or being awarded the procurement. Currently, the ANAC website[2] provides data on approximately 7.5 million of public procurement collected from 2007 to 2022.

The second dataset comes from the Italian Administrative Justice (IAJ) and contains judges' sentences related to public procurement appeals. Currently, the IAJ website[3] provides about $67,850$ sentences collected from 2007 to 2022.

2.2.1. Data sets overview

In the ANAC dataset, each procurement is identified by an alphanumeric key value called CIG and it has the following relevant features: the *procurement object*, a textual summary of the procurement; the *sector* to which it belongs, of three different types: Goods/Supplies (50%), Services (35.8%), and Public Works (14.2%); the administrative *region*[4] that issued the procurement; the *amount* of the procurement, from 40k euro upwards; the number of *lots* in the procurement, the *CPV* code[5] describing the main object of the contract obtained from a public ontology aligned in multiple languages.

The IAJ is a textual dataset containing the administrative judges' sentences saved in HTML format (91.5%), DOC/DOCX (8.4%), and PDF files (0.1%). In addition to the texts, the sentence files contain some useful metadata: the ECLI code[6] of the sentence, the court region (that corresponds to the region of the PA 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 appeals related to the sentences: the *recourse object*, the year, and the progressive number (from which the litigation started).

[2]https://dati.anticorruzione.it/opendata
[3]https://www.giustizia-amministrativa.it/web/guest/dcsnprr
[4]NUTS: https://ec.europa.eu/eurostat/web/nuts/background
[5]https://simap.ted.europa.eu/web/simap/cpv
[6]https://e-justice.europa.eu/content_european_case_law_identifier_ecli-175-it.do

3. AI Applied to the Analysis of the Contracts of the Italian Public Administrations

3.1. Methodology

3.1.1. Merge of data sets

Following the RQ1, the join between ANAC and IAJ datasets was carried out using Information Retrieval (IR) [10] and Natural Language Processing (NLP) [11] techniques.

First, the extracted texts from the sentences files were indexed with specialized IR tools, with Elasticsearch [12] being the most popular[7]. The texts and metadata of appeals and sentences were serialized into Newline Delimited JSON (NDJSON[8]) and indexed by the internal engine of the tool. We also employed Named Entity Recognition (NER) methods in Elastic Search to recognize the involvement of economic operators and PAs in recourses.

NLP techniques were then used to create sentence embeddings of *procurement objects* from ANAC and *recourse objects* from IAJ, to improve the connection between the two datasets; for this purpose, *LaBSE BERT* model [13] has been used. Cosine similarity [14] was then applied on sentence embeddings to collect the corresponding similar subjects of a *procurement object* and a *recourse object*. When the match between the entries of the two data sets was successful (via IR or NLP), we used the presence of an appeal on procurement as an indication of a *positive case* on that procurement entry; otherwise, it was treated as a *negative case*.

Figure 1 summarises the workflow described above.

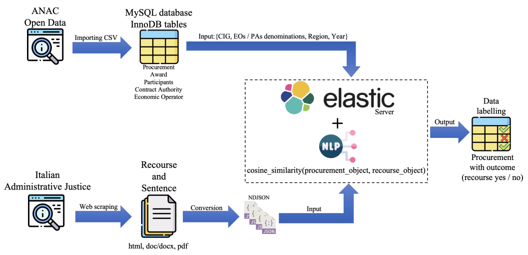

Figure 1: Data collection, merging, and labeling workflow methodology

3.1.2. The recommender system on procurement

Following the RQ2, we relied on *procurement object* (a summary textual description) to find similar procurement in the database. To build an abstract and general representation of the semantic content of the contract description by training the numerical vectors called *sentence embeddings* using BERT [15]. 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 [16] and LaBSE: they are a multilingual version of BERT and uses siamese networks to work on multilingual and Italian corpora. They are often used as tools to rank a set of sentences for their similarity to a given sentence, denoted as a *query*.

3.1.3. ML prediction models training

Following RQ3, a binary classification model will be trained to predict whether a procurement will have a recourse. Identified the solution as a supervised learning classification task [17], the following classifiers [18] were explored: K-Nearest Neighbours (KNN), Logistic Regression (LR), Naive Bayes (NB), Support Vector Machines (SVM), Decision Tree (DT), Random Forest (RF), and eXtreme Gradient Boosting (XGB).

3.2. Results

3.2.1. Merge of data sets

To better correlate IAJ sentences with the ANAC procurement (Section 3.1), we conducted three different types of searches: 1) by {CIG} (the procurement identifier); 2) by {EO participant, EO winner, PA, Region/Court, Year}; 3) by the similarity between {procurement object, recourse object}.

The results in Table 1 show how the methods, used incrementally, improve the ability to recognize a reference between the ANAC and the IAJ data sets based on the available sentences (67, 850).

Table 1
Reference found between ANAC procurement and IAJ sentences

Reference found by {feature}	Total	Overall perc.
Procurement identifier: {CIG}	8, 418	12.4%
Denominations: {EO participant, EO winner, PA, Region/Court, Year}	4, 178	18.5%
Similarity: {procurement object, recourse object}	2, 491	22.3%

[7]https://db-engines.com/en/ranking/search+engine
[8]http://ndjson.org

3.2.2. Recommender system performance evaluation

To evaluate the performance of our recommender system (Section 3.1.2), we decided to evaluate its Precision at 10. Precision at 10 was calculated by a panel of three individuals working separately on a test set of recommendations for 100 random procurement instances for Public Works, Services, and Goods/Supplies. Since each example refers to multiple elements (e.g., the awarding procedure, the location, the subject), the panel agreed in advance, case by case, on the elements of judgment (span from 2 to 5). Each panel member gave a relevance score of similarity between the query tender and its recommendations on each key element. The final relevance score is the mean of the scores given by the panel.

The results of Precision at 10 depend on the threshold θ for the relevance score; the lower the threshold, the higher the precision. We can think of this threshold as a measure of how strictly similar we want the recommended procurement and the query. A summary of the precision values at 10 is in Figure 2. We observe how the recommendation system works better for tenders of Goods/Supplies (orange bars). This makes sense because their descriptions are shorter than Public Works (green bar) or Services (yellow bar).

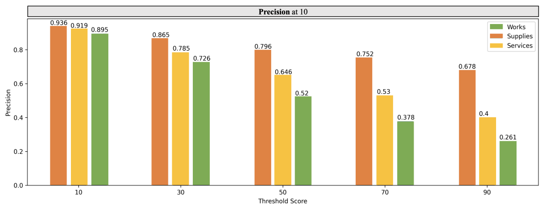

Figure 2: Results of precision at 10 for Public Works, Goods/-Supplies, Services with different thresholds

3.2.3. ML models performance measures

The labeled dataset obtained at the previous stage (Section 3.2.1) was divided into three smaller data sets containing procurement grouped by type: a dataset for Public Works of 10, 150 rows,one for Services of 15, 028 rows, and one for Goods/Supplies of 5, 232 rows. The smaller labeled data sets were used as input for the ML algorithms, balancing the number of *positive cases* and *negative cases*, keeping the negative cases distributed like the positive ones. 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.

Table 2 shows the results in terms of Accuracy [19] of the models; consistent with ROC/AUC[9] values, XGB and RF have the best performance.

Table 2
ML models performance measures on (*train:set*) best case

Dataset	Classifier	Accuracy	AUC
Services (80:20)	XGB	0.847	0.928
	RF	0.847	0.919
	SVM	0.802	0.828
Public Works (90:10)	XGB	0.773	0.855
	RF	0.768	0.848
	SVM	0.756	0.799
Goods/Supplies (80:20)	RF	0.780	0.864
	XGB	0.778	0.864
	SVM	0.751	0.809

4. Process Mining and Law

The typical main basic step in a PM search is the construction of the *log file* that includes the time sequence of events. Each event in an event log includes at least three basic features: the identifier of the process it belongs to, the name of the activity which generated the event, and the corresponding execution timestamp [20].

4.1. Methodology

4.1.1. Pre-processing and event log creation

CIG denotes the procurement identifier. The different activities involved in the process are 12, starting from procurement creation, and until the end of the contract. Each event includes the date at the level of granularity of the day on which the event occurred. In order to have a consistent event log, we removed cases too short, or with few activities, which are not meaningful according to domain experts. We then selected cases with at least 5 events: creation, publication, win, contract start, and contract end. The ANAC and IAJ merged data sets have been converted into an event log fulfilling the basic requirements for applying PM techniques, as represented in Figure 3.

In terms of technology, we imported the initial CSV log files for further analysis in the free and open source tool ProM[10] and DISCO from Fluxicon[11].

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

[10] https://promtools.org/

[11] https://fluxicon.com/disco/

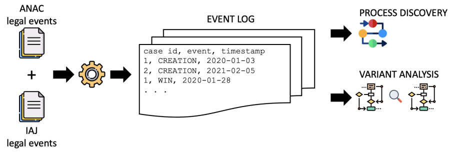

Figure 3: Methodological steps in our approach

4.1.2. Process Mining techniques

To answer RQ4, discovery algorithms can be applied to automatically derive process models. In the wide range of discovery methods proposed in the literature, we focus on the Fuzzy Miner implementation [21]. As a proof of concept for PM, the results were analyzed considering each of the following PM *perspectives*: control-flow, organizational (resource), and time [1].

4.2. Results

4.2.1. Process discovery

Control-flow perspective. The discovered process model from the event log provides a complete overview of the actual legal process flow, as shown in Figure 4; in the process map, the activity with the highest frequency, PUBLICATION, is indicated as the starting point. A group of events has a higher frequency, where darker rectangles in the diagrams correspond to "standard" events existing in all procurement. A second group of particular events occurs with a lower frequency (lighter color in the map), i.e. subcontracting or suspensions.

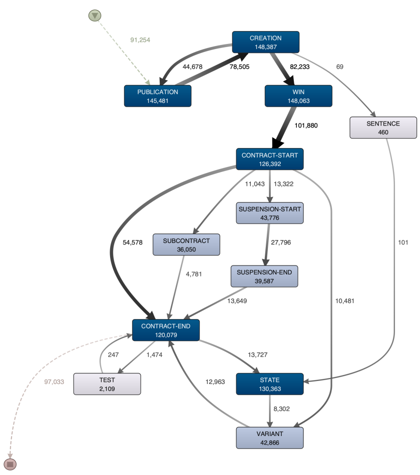

Figure 4: Process diagram based on absolute frequency metrics for the entire dataset

Organizational perspective. Since procurement is organized at the regional level, the "resource" taken into consideration is the region that issued the call. The results indicate the importance first of all of the "Central" region, which includes the administrative and governmental bodies of the Italian state (27,503 cases, i.e. 17%), as well as Lombardy (19,650 cases, i.e. 14.7%) and Emilia-Romagna (82,540 cases, i.e. 8.39%); the mean case duration for this three regions is between 12 and 14 months.

Time perspective. The diagram showing the average duration of transactions between activities makes it possible to identify bottlenecks. As highlighted by thicker arcs in Figure 5, the main critical transitions are represented by: procurement CREATION to WIN (69.9 days on average); procurement CREATION to SENTENCE (35.8 weeks on average); procurement CONTRACT-START to VARIANT (30.7 weeks on average).

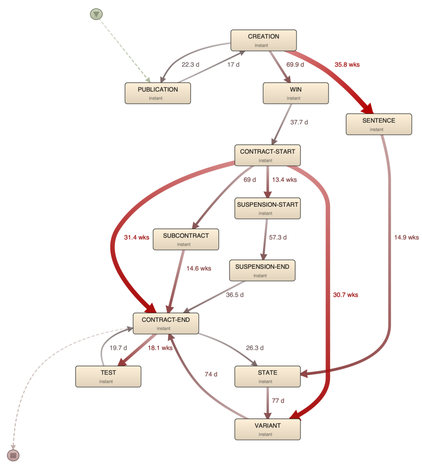

Figure 5: Time perspective highlighting slow transitions and bottlenecks in some activities of the legal process

5. Conclusions

This work demonstrates the possibility to manage a huge juridical dataset from the Italian National Public Authority to automatically extract meaningful knowledge to address Machine Learning experiments (RQ1).

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 tested a predictive experiment to estimate the prediction of the presence of recourse in front of the administrative courts on the basis of the features of public procurement (RQ3).

Responding to RQ4, discovery techniques allowed us to gain relevant insights into the main process behaviors.

In future work, we plan to investigate furthermore the *explainable AI* techniques. ML systems are becoming increasingly ubiquitous [22], increasing the demand to question, understand, and trust ML systems [23].

From the PM perspective, future work concerns the prediction of features of interest from an organizational perspective. First, we consider investigating the remaining time after the activity of interest (i.e., the awarding), as well as the successful or unsuccessful outcome of a tender.

References

[1] W. M. P. van der Aalst, Process Mining - Data Science in Action, Second Edition, Springer, 2016. URL: https://doi.org/10.1007/978-3-662-49851-4. doi:10.1007/978-3-662-49851-4.

[2] B. Kitchenham, Procedures for performing systematic reviews, Keele, UK, Keele University 33 (2004) 1–26.

[3] R. Nai, E. Sulis, R. Meo, Public procurement fraud detection and artificial intelligence techniques: a literature review (2022).

[4] R. Nai, E. Sulis, P. Pasteris, M. Giunta, R. Meo, Exploitation and merge of information sources for public procurement improvement, in: Machine Learning and Principles and Practice of Knowledge Discovery in Databases: International Workshops of ECML PKDD 2022, Grenoble, France, September 19–23, 2022, Proceedings, Part I, Springer, 2023, pp. 89–102.

[5] E. Sulis, L. Humphreys, F. Vernero, I. A. Amantea, D. Audrito, L. D. Caro, Exploiting cooccurrence networks for classification of implicit inter-relationships in legal texts, Inf. Syst. 106 (2022) 101821. URL: https://doi.org/10.1016/j.is.2021.101821. doi:10.1016/j.is.2021.101821.

[6] M. Fazekas, S. Sberna, A. Vannucci, The extra-legal governance of corruption: Tracing the organization of corruption in public procurement, Governance (2021).

[7] J. Dhanani, R. Mehta, D. Rana, Legal document recommendation system: A cluster based pairwise similarity computation, Journal of Intelligent & Fuzzy Systems 41 (2021) 5497–5509.

[8] M. J. Sangil, Heuristics-based process mining on extracted philippine public procurement event logs, in: 2020 7th International Conference on Behavioural and Social Computing (BESC), 2020, pp. 1–4. doi:10.1109/BESC51023.2020.9348306.

[9] A. J. Unger, J. F. d. S. Neto, M. Fantinato, S. M. Peres, J. Trecenti, R. Hirota, Process mining-enabled jurimetrics: analysis of a brazilian court's judicial performance in the business law processing, in: Proceedings of the Eighteenth International Conference on Artificial Intelligence and Law, 2021, pp. 240–244.

[10] N. J. Belkin, W. B. Croft, Retrieval techniques. (1987).

[11] P. M. Nadkarni, L. Ohno-Machado, W. W. Chapman, Natural language processing: an introduction, Journal of the American Medical Informatics Association 18 (2011) 544–551.

[12] C. Gormley, Z. Tong, Elasticsearch: the definitive guide: a distributed real-time search and analytics engine, OŔeilly Media, Inc., 2015.

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

[14] G. Salton, C. Buckley, Term-weighting approaches in automatic text retrieval, Information processing & management 24 (1988) 513–523.

[15] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, BERT: Pre-training of deep bidirectional transformers for language understanding, arXiv preprint arXiv:1810.04805 (2018).

[16] N. Reimers, I. Gurevych, Sentence-bert: Sentence embeddings using siamese bert-networks, arXiv preprint arXiv:1908.10084 (2019).

[17] P. Cunningham, M. Cord, S. J. Delany, Supervised learning, in: Machine learning techniques for multimedia, Springer, 2008, pp. 21–49.

[18] R. Saravanan, P. Sujatha, A state of art techniques on machine learning algorithms: a perspective of supervised learning approaches in data classification, in: 2018 Second international conference on intelligent computing and control systems (ICICCS), IEEE, 2018, pp. 945–949.

[19] A. Géron, Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow, " O'Reilly Media, Inc.", 2022.

[20] W. M. van der Aalst, J. Carmona, Process mining handbook, Springer Nature, 2022.

[21] C. W. Günther, W. M. Van Der Aalst, Fuzzy mining–adaptive process simplification based on multi-perspective metrics, in: Business Process Management: 5th International Conference, BPM 2007, Brisbane, Australia, September 24-28, 2007. Proceedings 5, Springer, 2007, pp. 328–343.

[22] D. V. Carvalho, E. M. Pereira, J. S. Cardoso, Machine learning interpretability: A survey on methods and metrics, Electronics 8 (2019) 832.

[23] R. Meo, R. Nai, E. Sulis, Explainable, interpretable, trustworthy, responsible, ethical, fair, verifiable ai... what's next?, in: Advances in Databases and Information Systems: 26th European Conference, ADBIS 2022, Turin, Italy, September 5–8, 2022, Proceedings, Springer, 2022, pp. 25–34.