

Chapter

Machine Learning in Procurement with a View to Equity

Ishrat Fatima, Roberto Nai and Rosa Meo

Abstract

The application of machine learning to big data from tenders published in Italy provides significant benefits to public administrations and economic operators, including improved procurement processes. Quantitative results from our study show a 96.5% accuracy using XGBoost models for predicting the presence of tender variations during contract execution. These models estimate the likelihood of variations in upcoming tenders: their correct prediction is a valuable tool because variations avoidance reduces completion time and costs of public contracts. Additionally, explainable AI tools help the description graphically and intuitively of the analyzed data. They also allow the analyst to highlight potential biases in tender participation and their awards, contributing to fairer public procurement. The results of their application to public tenders show that strong differences in the Italian country exist with a consequent lack of equity. Finally, the application of recommendation systems on the tender notices shows they are an effective cognitive tool to search for similar tenders and retrieve the actors involved, such as public administrations or economic operators. The precision score of the answers is above the value of 90% for the 74.15% of the queries. The chapter describes the tasks that permit the achievement of the above objectives.

Keywords: tenders, big data, machine learning, explainable artificial intelligence, descriptive models, ethical fairness, predictive models, deep neural network embedding, recommendation systems, prescriptive models

1. Introduction

Thanks to the use of digital information systems, organizations have stored data about their business processes. The large amounts of data collected can be used for many purposes, from monitoring progress and assessing risks to setting targets and making comparisons with other contexts (especially when these data are cross-checked with open data from other realities).

This chapter discusses the Italian case of tender notices published by public administrations.

Public procurement in Italy concerns contracts stipulated by the Public Administration with economic operators for the supply of goods, services, or works.

The award process takes place according to precise rules, mainly regulated by the Procurement Code.¹

The process of awarding a public contract follows these main phases:

Tender notice: The Public Administration publishes a tender notice that defines the project and the requirements for participation.

Presentation of offers (or bids): Interested economic operators present their offer, which includes technical and economic proposals.

Evaluation of offers: A special commission evaluates the offers according to specific criteria (lowest price or most economically advantageous offer).

Awarding: The contract is awarded to the economic operator who submitted the best offer, compliant with the established criteria.

Execution of the contract: The winning operator executes the contract, respecting the agreed conditions.

The entire process is subject to principles of transparency, competition, and non-discrimination to guarantee fairness and correctness.

Tender notices are published in the Official Gazette and a few data on them is sent by the public administrations to the regional public expenditure observatories or directly to the National Anti-Corruption Agency (ANAC). ANAC was set up by the Italian government in 2010 with the aim of monitoring and reducing public expenditure. The ANAC database on tenders is public and can be queried online [1].

The accumulated big data, consisting of more than seven million notices, on the expenditure of tenders of all the Italian public administrations, whether they are works, supplies, or services, is an extremely valuable resource. It provides data and evidence to help the data analysis expert formulate and test the validity of the hypothesis, from the awarding process to the litigation in front of the Regional Administrative Courts, to the execution of the contract, testing, and so on.

The tenders in the ANAC database are highly diverse, covering works, supplies, and services, and each tender has its specific type and sector. A tender may fall under public works, services like consultancy, or supply contracts. These tenders are significant as they determine the allocation of public funds and play a key role in economic development. Understanding the tendering process is critical to ensuring transparency, competitiveness, and fairness. For instance, consultancy tenders, which are the focus of this chapter, can be prone to variability and require close monitoring to prevent cost overruns. The data collected on these tenders, such as the *CIG* identifier, tender type, and award amounts, serve as a foundation for our study. There are two types of motivations for the focus of the chapter on the analysis of consulting tenders. The first one is Simpson's paradox.² According to this principle, a trend or correlation among variables appears in heterogeneous groups of data but it could disappear when the groups are merged. Therefore, we would like to reduce the risk of obtaining erroneous results by restricting the data analysis to a single group composed of more homogeneous types of tenders. The second is the awareness of the fact that the phenomenon of money laundering could be facilitated by the existence of a consultancy cost that is not easy to quantify or limit directly. Finally, we leverage a recent trend in research on Artificial Intelligence (AI) known as "Trustworthy AI" [2] in which ethical values and human rights are taken into account when AI tools are developed and

¹ The text of Procurement Code in Italy, which was updated several times, is available at <https://www.codiceappalti.it/>

² A description of the phenomenon is at: https://en.wikipedia.org/wiki/Simpson's_paradox.

used in society. Among the guaranteed rights, there is the right to obtain explanations for the outcome of a (technological) process, accountability such as the right to find responsibilities for every mistake of the AI tools, equity interpreted as the guarantee that everyone receives equal treatment regardless of ethnic group, the provenance, or other sensible features. In other terms, fairness and non-discrimination have a central role and should be used to evaluate the good performance of a human or technological process in which AI is involved [3].

The situation is particularly delicate in the case of tenders by public administrations (PA) involving the transfer of public funds. On the one hand, there is the right of every citizen of the territory to receive public money as a reward for the commitment and professionalism involved in a project for public administration. On the other hand, there are different contexts in which public administrations and economic operators operate in their territories. Not all situations are comparable, especially if we take into account the different economic backgrounds that exist in the Italian territory. The south, the so-called “Mezzogiorno”, is more disadvantaged, but also suffers more from the presence of criminal organizations originating from these regions. Criminals try to exploit and extort money from the legal and public administrations. As a result, southern Italy may suffer more from coexistence with criminal organizations and be suspected of involvement with them. This chapter studies in an objective way the available data on the tendering process on the Italian territory to measure with the lens of data analysts and tools for measuring AI fairness [3] if inequalities and asymmetries exist. Knowledge about the existence of the phenomenon is the first step to correct it.

The secondary objective of this chapter is to demonstrate that the ANAC database is a valuable resource. When analyzed, the available data can be used to improve the tendering process. Two directions are followed. The first uses predictive AI models trained on the ANAC data, to predict whether a contract awarded by a tender will be involved in a variation. This prediction could help public administrations avoid the variation, which often leads to delays and increased costs. In the second direction, a tender recommendation system is being developed on the ANAC database. This is a valuable resource for both the PA and economic operators, who can use it to search for similar cases in the past, similar competitors who could be possible partners for similar projects, and obtain useful information on costs, time, constraints, award procedures, etc. Moreover, recommendation systems help to increase the competitiveness and diversity of the bidding pool by analyzing the companies that typically bid on specific tenders by helping to suggest new companies that have not participated before but might be interested in bidding.

The remainder of this chapter is organized as follows: Section 2 introduces the related works; Section 3 describes the case study; Section 4 describes the study on equity of awarding tenders on the territory; Section 5 provides insights about predicting the presence of a variation in the procurement; Section 6 outlines the recommendation system and the obtained results. Finally, Section 7 concludes the chapter.

2. Related work

In the context of Machine Learning (ML) in procurement, there is a growing focus on leveraging AI technologies both at the National [4] and European level [5]. One study emphasizes the importance of integrating ethical and sustainable sourcing practices into supply chain management, highlighting how ML and big data analytics

can improve decision-making processes [6]. These technologies enable more socially responsible procurement by ensuring transparency, reducing biases, and fostering fair supplier selection and contract management [7].

Another perspective is offered through research on supply chain resilience, which examines how AI and ML can enhance information exchange among supply chain partners. This approach reduces risks and supports socially equitable outcomes by promoting transparency and fairness in procurement processes. AI technologies are seen as key to driving social equity by improving supplier diversity and advocating for fair labor practices across supply chains [8, 9].

A systematic literature review of social procurement in the construction and infrastructure sectors explores the evolution of procurement practices to include social value creation. This shift focuses on using government initiatives to mandate social outcomes in procurement contracts. The review identifies barriers and enablers to social procurement and provides strategies for overcoming challenges, highlighting the potential of procurement practices to deliver social benefits and promote equity [10, 11].

Additionally, the role of AI in procurement is explored in terms of its potential to strengthen diversity and inclusion. Data analytics and ML are highlighted as tools to identify opportunities to engage diverse suppliers and ensure equitable procurement processes [12]. This approach aligns with global trends in utilizing technology to support socially equitable procurement practices and underscores the importance of integrating equity considerations into ML models used in public procurement [13, 14].

Recommendation systems are increasingly utilized in the legal domain to swiftly retrieve relevant documents for specific cases. In the study of Dhanani et al. [15], a graph clustering method is proposed to group referentially similar judgments and identify semantically relevant ones within these clusters. In the study of Nai et al. [16], numerical vectors known as *sentence embeddings* [17] were trained using BERT (Bidirectional Encoder Representations from Transformer) [18] to build an abstract and general representation of the semantic content of contract descriptions. Input sentences were taken from the brief descriptions in natural texts of procurement in the ANAC database, resulting in 768-dimensional vectors. Subsequently, for an individual procurement case, the most similar and relevant ones in the rest of the database were searched using SBERT [19] and LaBSE [20]; similarly, in the study of Nai et al. [21], the performance of a recommendation system based on the embeddings provided by two different commercial models was compared.

Bert LaBSE model is a language-agnostic Bidirectional Transformer Sentence Embedding [20]. A transformer has one or more encoders, used to represent in a latent model the context of the sentence, while the decoder parts are usually applied to generate sentences from the abstract representation and to change the language in automatic translation. BERT connects the encoder and decoder through a token-based attention mechanism. In BERT, consecutive sentences from the input are separated into tokens, transformed into numerical vectors from which the system performs different tasks such as learning the words' context through random masked words and learning if sentences are consecutive (or related in a question-answering task).

As regards equity in public procurement, van Dijk and Wilke [22] examined the effect of different interests on public-good provision by adopting the equity theory. The latter aims at determining whether the distribution of resources is fair taking into account the contributions (or costs) and the benefits (or rewards) for each individual in an organization. Decarolis and Giorgiantonio [23] employ ML in studying the effectiveness of indicators used in police investigations on corruption (called red flags) in public tenders for roadworks. The survey [24] provides an overview of the

many control measures for fairness that researchers in AI proposed to monitor the fairness of a process. The monitored process might be assisted by AI and ML tools and its fairness needs to be verified with respect to the membership to a group (that expresses a sensitive feature) for some of the involved actors. The measures for fairness are guided by different concepts: (a) parity-based metrics that compare the predictive positive rates across groups. An instance is statistical *independence* between the predicted score and the group; (b) Confusion matrix-based metrics that compare groups by taking into account the potential differences across groups. An instance is *separation* that evaluates the equal opportunity of comparable individuals from the different groups; (c) Calibration-based metrics that compare groups by the probability that the predicted scores are emitted. An instance of this metric is *sufficiency* with tests for fairness. In Section 4.1, we apply the above three concepts to assess the fairness of the participation of economic operators in public tenders and the subsequent award of contracts, taking into account their different origins: north of Italy, center of Italy, or “Mezzogiorno” (south or islands).

3. Case study: Tenders of Italian anti-corruption authority: ANAC

The primary dataset pertains to ANAC, the Italian government authority collecting public tenders. A dedicated section on the ANAC website [1] allows users to access data in a standard format, enabling category selection and downloading in compressed files. In the ANAC Open Data catalog, the central dataset concerns the creation of a *Tender Notice*, organized by year and month. Besides the tender dataset, four additional significant datasets are available: the list of *Contracting Authorities* (CA) that have issued a tender, the list of tenders that have received an *Award*, the *Economic Operators* (EO) who have won a tender, and the *activities* associated with a tender post-award (e.g., *contract-start*, *contract-end*, *subcontract*, etc.). Each of these four datasets is briefly described below. CAs³ are public bodies or entities acting on behalf of public institutions, responsible for procuring goods, services, or works through tenders. They oversee and manage the entire public tender process. CAs are categorized into three types: Central (e.g., ministries), Regional and Local (e.g., municipalities), and other entities (e.g., hospitals). Awards include a list of awarded tenders, with details such as the final awarded amount, the award date, and the EOs receiving the tender. EOs can be individual enterprises, artisans, partnerships or corporations, cooperatives, etc. A 10-character alphanumeric variable called *CIG* identifies each tender and its related activities. CAs and EOs are recognized by their *tax code* (an alphanumeric string), and CAs also have a unique *ISTAT code*. Finally, the *Variants* dataset contains information on any variation authorized during execution from the original contract, as a result of unforeseen circumstances. **Figure 1** presents a portion of the *Tender Notice* and *Award* where some tenders have not been awarded to an EO (with *DATE_AWARD* and *TENDER_AWARD* fields left empty).

3.1 Main data overview

This section presents a short description of the most relevant features from three tables of ITH, whereas their features are listed in **Table 1**. In the *TENDER_NOTICE*

³ Contracting authorities are often referred to more generally as public administrations.

CIG	OBJECT	CONTRACTING AUTHORITY	DATE PUBLICATION	DATE AWARD	TENDER AMOUNT	TENDER AWARD	TENDER TYPE
50854040	ATTREZZATURE SERVIZIO ANESTESIA E RIANIMAZIONE	AZIENDA UNITA' SANITARIA LOCALE N. 6 SANLUIGI	13/03/2014	27/03/2014	€48.980,00	€48.980,00	FORNITURE
50768731A	P23-216-TLP-LATRONICO-PZ-RIPISTINO STATICO FUNZIONALE DEL SOLAIO DI COPERTURA	3 REPARTO GENCO A.M.BARI	26/07/2023	N/A	€87.367,16	N/A	LAVORI
AA40877889	POLIZZA ASSICURATIVA ALL RISKS BENI IMMOBILI E MOBILI 24 MESI + 12 MESI RINNOVO	COMUNE DI LONGARE	19/12/2023	N/A	€90.367,72	N/A	SERVIZI
AA0FEA7C2E	CONVENZIONE PER LA FORNITURA MATERIALI DI PULIZIA MONOUSO ED IGIENE PERSONALE PER I COMANDANTI DELLA GIURISDIZIONE DEL COMANDO INTERREGIONALE MARITTIMO SUD E DELLE ULU/NN. DIPENDENTI DALL'AREA OPERATIVA DI CINQUEV	DIREZIONE DI COMMISSARIATO MARINA MILITARE - TARANTO	19/12/2023	N/A	€215.000,00	N/A	FORNITURE
0075908809	S3.S5. NN 33-330-041 E N.S.A. N.26	ANAS - SOCIETA' PER AZIONI	17/09/2007	N/A	€50.000,00	N/A	LAVORI
8912403989	MISURA IN SICUREZZA DELL'AUDITORIUM SCUOLA N.100 MONO DI VIA XXIV MAGGIO - OPERE EDILI	COMUNE DI CIRIAGO	23/06/2023	18/07/2023	€129.089,87	€119.890,12	LAVORI
0861410806	S.C. 2001689 CAPO 2001689 MMF ANEAPPIHDEX ISOLATO IN XLP/E (D) HP/TE 1300 X/L TPO ASPINALE VISIBILE. CONDUTTORI E SCHERMO IN ALLUMINIO. ISOLAMENTO ESTRUSO A SPESORE RIDOTTO, GUAINA TERMOPLASTICA RESISTENTE ALL'URTO GALLABRASIONE. ALTRE CARATTERISTICHE SECONDO S.T. ACIA DISTRIBUZIONE/DMX7. ED. LUGLIO 2016 (IMPORTO COMPRENSIVO DELL'OPZIONE DI ESTENSIONE DEL 20%).	ARETI S.P.A.	13/12/2016	23/12/2016	€285.465,80	€229.488,00	FORNITURE

Figure 1. ANAC Open Data-CSV file data preview, where it is possible to see the main features of a tender. Full-size image available here: <https://bit.ly/3M75Pch>.

Table	Feature	Description
T_N	CIG	Alphanumeric value (key value)
	Tender object	Textual summary of the tender
	Framework agreement between PA and EO	1 if yes, else 0
	Number of lots	Integer value {1..n}
	Tender type	Supplies (U) Works (W) Services (S)
	Tender area	Ordinary (O) Special (S)
	Tender amount	Float value
	Date of publication	Date in format yyyy-mm-dd
	EO selection mode	Integer value {1..122}
	Execution mode	Integer value {1..19}
	Region	Italian region names + Central Government
	CPV	String ID (XX000000-Y)
	CPV division code (first two digits of CPV)	String ID (XX)
	PNNR flag	1 if yes, else 0
AW	CIG + AWARD_ID	Alphanumeric value (key value)
	EO consortium (group of EOs)	1 if it is 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
C_A	Number of bids admitted	Integer value {1..n}
	Subcontracting admitted	1 if yes, else 0
	Tax Code	Alphanumeric value (key value)
	ISTAT Code	Alphanumeric value
	CA denomination	Textual string

Table 1. Main features of the tables TENDER_NOTICE (T_N), AWARD (AW), and CONTRACTING_AUTHORITIES (C_A).

table, each tender is identified by an alphanumeric value called the *CIG* (the key ID value), which is used to link most of the remaining tables. The main distinction between tenders is their *type* and *sector*: types can be “Services” (S), “Supplies” (U), or “Works” (W), while sectors can be “Ordinary” (O) or “Extraordinary” (E) based on whether they are planned or due to extraordinary events (e.g., floods, earthquakes, etc.). All types of tenders are described by the CPV code, i.e., the *Common Procurement Vocabulary*.⁴ These categories are organized into an ontology (a hierarchical organization) whose elements are identified by codes; using the first two digits of the codes (which correspond to the upper part of the ontology and to the coarsest-grain categories), they provide the CPV *divisions* useful for distinguishing the product categories purchased by CAs (e.g., “90” represents cleaning services, while “9040” represents sewer cleaning). A tender can be defined within a *framework agreement*, meaning that the CA and EO have a prior agreement to provide services for further tenders for a defined time (e.g., 1–5 years). Often, a tender is split into *lots*, with a lower amount that can be awarded separately to different companies because they could be pertinent to different economic activities. Finally, each tender has a well-defined *selection criterion* that will be applied by the CA to choose the EO who will be awarded; an *implementation criterion* will also be considered and the winning EO must comply with it.

The table AWARDS contains the list of relevant features related to the tender award, including the awarding entity, date, amount, etc. As expected, non-awarded tenders are not reported in this table (so they are only available in the TENDER_NOTICE table).

The table CONTRACTING_AUTHORITIES contains information about the name of the authority and the main keys that allow to link the information of the tender with other tables. In this respect, the *tax code* determines the type of CA by joining this table with the ISTAT table necessary to study the relationship between the tender process and the regional population and geographical dimensions.

3.2 Italian National Institute of Statistics—ISTAT

An important aspect involves determining the scope of each tender administrative aggregation. The National Institute of Statistics (ISTAT)⁵ offers a comprehensive range of statistical information about Italy. Among these are the *Nomenclature of Territorial Units for Statistics* (NUTS)⁶ and the distribution of the *population per municipality*. The NUTS system is structured into three hierarchical levels: NUTS 1 comprises socio-economic regions, such as large economic areas (that in this chapter we call macro-areas); NUTS 2 pertains to smaller regions for implementing regional policies, such as provinces or large metropolitan areas; NUTS 3 covers the smallest areas, including regions, provinces, and municipalities.⁷ The distribution of inhabitants can be interesting for understanding, for example, the quote of investment per population in a certain area.⁸ Including NUTS and population in ANAC Open Data

⁴ <https://ted.europa.eu/en/simap/cpv>.

⁵ <https://www.istat.it/en>.

⁶ <https://www.istat.it/classificazione/codici-dei-comuni-delle-province-e-delle-regioni>.

⁷ <https://www.istat.it/it/archivio/6789>.

⁸ <https://www.istat.it/it/archivio/156224>.

facilitates, for instance, the comparison and analysis of territorial investments at the different geographic granularity levels, possibly normalizing costs by the population.

4. Descriptive tasks and fairness study

When analyzing a phenomenon from data, it is suggested to create descriptive models (such as geographical heat maps, histograms, etc.) that allow a first look at the phenomenon with a summary of an immediate interpretation. **Figure 2** summarizes the total amount spent on counseling tenders in the studied period (2016–2019). The outlier over the other regions, whose expenses are instead more similar, is the region Lazio, in the central macro-area. The capital of Italy, Rome, is located in this region and this could be a justification for the higher number of counseling contracts for the government of the country.

In **Figure 3**, we show the result of a predictive model (linear regression) that we trained to predict, as a target, the fraction of tenders for counseling in the same period on the whole Italian territory. The regression model was trained from the predictive features that are macroeconomic variables of the economic richness of the regions. These variables are shown in **Table 2** with the coefficients in the regression model.

These economic variables were chosen by the financial and economic institutions in Italy, like Banca d'Italia and Sistema Conti Pubblici Territoriali [25]. Data on these variables were downloaded from ISTAT.⁹ To train the regression, we expressed the



Figure 2.

The heat map of the Italian regions on the total amount of expenditure on contracts awarded to economic operators in the regions. Darker colors indicate higher values and lighter colors have lower values. Full-size image available here: <https://bit.ly/3M75Pch>.

⁹ <https://esploradati.istat.it/databrowser/#/it/dw/categories>.



Figure 3. Predicted fraction of the procurement awarded by economic operators of each region. Darker colors indicate higher values and lighter colors have lower values. Full-size image available here: <https://bit.ly/3M75Pch>.

Regression variables	Coefficient
Gross Domestic Product of the region (in Italian - PIL)	1.592985
internal consumption	356.007253
expenses for consumption on the territory by families	-273.787177
expenses of private non-profit social institutions serving families	0.275970
expenses for consumption of the public administrations	-44.654956
gross fixed capital formation	-0.661344
number of economic operators resident in the region	0.127807
land area	-0.874896
resident population	0.926562
constant term	0.0145896

Table 2. Input variables for the regression of the fraction of counseling tenders in the regions with their coefficient.

value of each variable for each region as a fraction of the global value for the Italian territory, so that the ranges of the variables are comparable and none of them can dominate the others in the regression model.

The results of the regression in terms of root mean squared errors (RMSE) are good because RMSE accounts for 0.3%. The intuitive meaning of using linear regression to predict the presence of contracts awarded by economic operators in the regions, based on socio-economic data of the regions, is that the presence of awarded contracts is proportional to the economic wealth of the people living

Region	Macro-area	Amount (Euro)	Predicted fraction
Abruzzo	Mezzogiorno	734,973,355.25	-0.006508
Aosta Valley	North	1,273,734,015.66	0.010198
Apulia	Mezzogiorno	5,881,228,868.15	0.026437
Basilicata	Mezzogiorno	930,324,140.31	-0.006734
Calabria	Mezzogiorno	751,751,372.06	-0.009588
Campania	Mezzogiorno	3,380,554,743.60	0.034156
Emilia-Romagna	North	16,192,602,344.01	0.061392
Friuli Venezia Giulia	North	2,538,965,895.64	0.008922
Lazio	Center	83,322,638,530.51	0.436136
Liguria	North	5,416,448,787.22	0.035922
Lombardy	North	18,370,982,885.57	0.244621
Marche	Center	2,986,739,413.73	0.019560
Molise	Mezzogiorno	147,358,035.15	0.003210
Piedmont	North	3,987,636,944.23	0.014454
Sardinia	Mezzogiorno	10,822,040,833.39	-0.015567
Sicily	Mezzogiorno	3,483,991,756.14	0.014344
Trentino-South Tyrol	North	13,090,022,541.79	0.118560
Tuscany	Center	13,255,987,874.61	0.039263
Umbria	Center	1,819,818,815.68	-0.000578
Veneto	North	7,081,496,000.08	0.049062

Table 3.

[left & center columns] Region, macro-area, and amounts awarded to economic operators in the region shown in Figure 2; [right column] the fraction of tenders in Italy for each region predicted by regression shown in Figure 3.

in the region and to the number of economic operators. The coefficients of the regression agree with this interpretation (see right column of Table 2). Among the variables, the most important ones are those with a higher and positive coefficient, that are *internal consumption*, *GDP*, and *number of economic operators* while those that are negatively correlated with the target have a large negative coefficient: *expenses for consumption on the territory by families* and *expenses for consumption of the public administrations*.

The total amounts for counseling tenders for each region are shown in Table 3 together with the proportion of tenders predicted by the regression model. If the predicted value of the fraction of tenders awarded is negative, it means that for those regions, the estimated value is higher than the actual value. This means that in these regions, we would expect proportionally more tenders to be awarded: this is the case for many regions in the Mezzogiorno. In Section 4.1, we go deeper into the analysis by considering the fairness of the distribution of tenders in the three macro-areas mentioned.

4.1 Analysis of the fairness of participation in tenders and the award of contracts to economic operators of different origins

In this analysis, we consider three variables:

1. S represents the sensitive variable, also called group, which, in the case study of this chapter, denotes the macro-region of origin (North, Center, or Mezzogiorno) of the economic operators that won the tenders.
2. Y represents the ground truth of the analyzed process, which, in the case study, denotes the actual proportion of tenders awarded to economic operators in the macro-areas.
3. R denotes the score, i.e., the result of the AI/ML model that predicts the target given the observed characteristics or given an assumption (often called the null hypothesis in statistics). In the case of this chapter, R is the score obtained by the linear regression that predicts the proportion of tenders awarded to economic operators in the macro-area according to the socio-economic characteristics of the area in **Tables 2** and **3**.

The three concepts introduced in the study of Caton and Haas [24] are applied. The formulae that must be satisfied for all groups to prove the fairness of the process are reported.

1. *Independence*: the score of the predictive model must be independent of the sensitive variable: $R \perp S$
2. *Separation*: the score of the predictive model must be independent of the sensitive variable, given the ground truth variable: $R \perp S|Y$
3. *Sufficiency*: the ground truth variable must be independent of the sensitive variable given the predicted score: $Y \perp S|R$

There is some difficulty in testing the statistical independence of the variables because of the limited number of observations (20, the number of administrative regions in Italy). Some statistical tests such as χ^2 [26] give reliable results if data obeys some assumptions that are not always valid in our data (for instance, no value in the contingency table should be less or equal to five but this occurs since data are spread into the different groups according to the macro-area). As a solution [26], we grouped the continuous values of variables R and Y in two bins (low, high) separated by the median value of their frequency distribution. After this transformation, the problem of testing independence according to the above formulae for testing fairness transforms into a new test described as follows.

The test aims to calculate the probability that a Bernoulli process resulted in the given observations shown in **Table 4**. The Bernoulli random events represent the frequency of tender awards to economic operators in a region occurring above or below the median, assuming they are independent and are not different in the country regions.

Since the independence test for separation (sufficiency) is conditional on the Y (R) variable, whose values are divided into two bins, the independence test in the two bins is repeated separately. Each observation on a region is treated as a Bernoulli trial in which we made the null hypothesis that counseling tender awards have the same probability of occurring in the regions of the three macro-areas. The probability that the Bernoulli process results as observed in **Table 4** is given by the well-known formula:

			Y low bin	Y high bin
R low bin	Center		2	0
	Mezzogiorno		6	0
	North		2	1
R high bin	Center		1	1
	Mezzogiorno		0	2
	North		0	7
			R low bin	R high bin
Y low bin	Center		2	0
	Mezzogiorno		7	0
	North		1	1
Y high bin	Center		1	1
	Mezzogiorno		0	1
	North		0	8

Table 4. Distribution of the regions of the three macro-areas to the lower or higher bin for the independence tests according to sufficiency (upper contingency table) and separation (lower contingency table).

$$\Pr = \frac{p^k (1 - p)^{(n-k)} n!}{k!(n - k)!} \tag{1}$$

where n is the number of trials (regions, 20) and k is the observed number of occurrences in the lower bin (or the higher) for the regions of the three groups (macro-areas). $p = 0.5$ is the probability that an observation falls in the lower bin according to the null hypothesis. Bins were created on the values of the variables Y (and R) according to the threshold of the median value that, by definition, is the value that leaves half of the observations below it and the other half above. In our case, it is the probability that tender awards occur in a region with values of Y (and R) lower than the median.

Since our sample has a limited size, for the independence tests, Fisher’s exact test¹⁰ must be applied using the formula 1 for the three cases in which we observe one or fewer regions in one of the two bins and all the remaining regions in the other. In both the cases of separation and sufficiency, the probability of obtaining the observed results from the Bernoulli process under the null hypothesis is extremely low and equal to $3.433 \cdot 10^{-5}$. For the simpler, independence test $S \perp R$, we grouped the two macro-areas of Center and Mezzogiorno in which the regions tend to have a lower proportion of awarded tenders and kept the regions of the North in a separate category. The above reasoning was repeated and concluded that the probability of observing a situation so unbalanced or more extreme is again very low ($6 \cdot 10^{-5}$).

Conclusions are that the estimation of the counseling tender awards by the linear regression model is not fair according to the definitions of fairness in the field of

¹⁰ For the formulation of Fisher’s exact test, see for instance: https://en.wikipedia.org/wiki/Fisher's_exact_test.

Explainable AI. The unfairness does not denote limits of the model itself: it can foresee precisely the ground truth of the actual tender awarding process in the three macro-areas. On the contrary, the unfairness of an accurate model in predicting the ground truth denotes the unfairness of the distribution of ground truth variable across groups.

Several other analyses were repeated (which cannot be described here for reasons of space), also taking into account the amount of each tender (with the amount divided into four levels) and the size of the economic operators (divided into four levels) according to the share capital of the company: similar results are always confirmed. Whatever the macro-area of the public administration, and especially when the size of the company is large and the amount of the tender is high, the public administrations tend to award the tenders with a higher probability to the regions of the North (or Center), but not to the Mezzogiorno. This result is particularly unbalanced and unfavorable for Mezzogiorno, given that the total number of enterprises in the Mezzogiorno and in the Center is not proportionally lower than that in the North, but approximately equal (even if the enterprises in the North tend to be larger).

The socio-ethical consequences of the inequalities observed are a lack of equal opportunities for economic operators in the Mezzogiorno. The observations confirm the historical gap that exists in Italy between the Mezzogiorno and the rest of Italy (the so-called southern question). On the one hand, it highlights the increasing difficulty and persistent disadvantage of Mezzogiorno operators in competing at the national level. On the other hand, it could highlight the existence of a possible reduced confidence on the part of contracting authorities in the ability of economic operators in the Mezzogiorno to carry out a public contract successfully, at a reduced cost and in a shorter period of time, given their persistent difficulties of operation and the difficult socio-economic context. The recommendations for the stakeholders are to carefully monitor the fairness indicators of the tender awarding process. For policymakers, it is a matter of proposing regulatory correctives, but also of carefully monitoring the efficiency of the use of resources by economic operators once they have been awarded public contracts. Here too, the large amount of data available in the ANAC database is useful, as it provides many examples of contract performance that can be used for comparison and as examples of best practice.

5. Predictive tasks

In the predictive tasks, several ML models were employed to evaluate their performance and accuracy in handling complex datasets. The models included Logistic Regression (LR), Decision Tree (DT), Random Forest (RF), XGBoost (XGB), and Naive Bayes (NB), each offering distinct advantages and trade-offs explored in this research.

LR, a linear model, is often used for binary classification tasks. It is straightforward and interpretable, making it a suitable baseline model. However, its simplicity limits its effectiveness in capturing complex relationships in high-dimensional or non-linear data [27].

DT, a non-linear model, splits data based on feature importance, making it more flexible than LR. DT is easy to interpret and can handle both numerical and categorical data. However, it tends to overfit the training data if not properly tuned, leading to reduced generalization on unseen data [28].

RF builds on the decision tree model by creating an ensemble of trees, improving robustness, and reducing overfitting. It performs well on complex datasets due to its ability to capture non-linear patterns and interactions between features. RF also provides feature importance scores, making it useful for interpretability in large datasets [29].

XGB, an advanced gradient boosting algorithm, often outperforms other models in terms of accuracy, especially with complex datasets. It builds trees sequentially, optimizing for errors made by previous trees, which allows for a more refined learning process. XGBoost is highly efficient in handling large-scale data but requires careful tuning of hyperparameters to achieve optimal performance [30].

NB is a probabilistic model based on Bayes' theorem, and while it performs well in certain text classification tasks or cases where feature independence can be assumed, it is generally less effective with complex, non-linear datasets. Its main advantage is its computational efficiency, which makes it useful for large datasets with simple relationships [31].

In particular, the described experiment aims to identify procurement that has undergone variations. Variations could be used by economic operators as a “kind of strategy” to gain back part of the lot amount that they originally lowered to have more chances of being awarded the contract. Thus, variations impact the final contract price and time and alter the budget allocations. Specifically, funds allocated to a variant are no longer available for other activities, thus unbalancing the overall expenditure. Therefore, procurement for services (S), namely *consultancy services*, was selected for this experiment. To keep the dataset balanced in the consultancy-related variants, procurement with the same CPV divisions (73 and 85) was chosen 4 years before COVID-19 (2016–2019); the resulting dataset thus consists of 25,732 tenders.

The dataset contains procurement marked by whether they include a variant (*label* = 1) or not (*label* = 0). The two classes are inherently imbalanced, with one possibly more prevalent than the other. The dataset was balanced by sampling equal numbers from each class to address this. This ensures that the ML models do not favor the more frequent class, thus providing a fair evaluation. The balanced data was randomly divided into training, validation, and test sets, with an 80–10–10 split, where the test set was reserved for unseen data to evaluate model performance on new, unobserved examples. Features were standardized and encoded to improve the performance of certain models, ensuring they operate effectively across varying scales of input data. In detail, based on the dataset described in **Table 1**, several preprocessing steps were applied to prepare the data for ML models. *Numerical fields*, such as the Tender amount and Awarded amount, were normalized to bring all values within a standard range, ensuring that no single variable dominated the learning process due to differences in scale. *Categorical fields* like Tender type were encoded using one-hot encoding, which converts the categories into binary vectors for compatibility with ML algorithms. For *boolean fields*, such as the Framework Agreement and Subcontracting admitted, a boolean encoding was used, where 1 indicates true and 0 indicates false. These preprocessing steps are standard in ML to ensure that the data is in a format suitable for training, reducing bias, and improving the accuracy of the models [32]. As described at the beginning of this section, various models were trained: LR, DT, RF, XGB, and NB. These models were chosen for their diverse capabilities, from simple interpretations to handling complex, non-linear relationships. Hyperparameter tuning was conducted on the models using the Hyperopt library to optimize their performance and identify the best combination of parameters for accurately predicting procurement affected by variants. The models were evaluated using *cross-validation* [33] to test their generalization capabilities. This method ensures that models perform well across different subsets of data, reducing the risk of overfitting the training set. Model performance

was assessed using Accuracy [34], Precision [34], Recall [34], and F1-Score.¹¹ These metrics provide insights into how well each model can identify procurement with variations, which is crucial for understanding their budgetary impacts. The data and scripts for this part of the experiments are publicly available.¹²

5.1 Predictive tasks: Results

The cross-validation metrics provide a robust evaluation of model performance by accounting for variability in the data, as highlighted in **Figure 4** and **Table 5**. Among the models analyzed, XGB demonstrates superior performance, with the highest cross-validated accuracy of 0.965 and F1-Score of 0.965. This indicates high precision in its predictions and a strong balance between precision and recall, with precision and recall values of 0.983 and 0.947, respectively, making it an ideal choice for datasets where both false positives and false negatives are costly. The RF model also performs well, with an accuracy of 0.947 and an F1-Score of 0.948. Although slightly lower than XGB, it still offers a commendable trade-off between recall and precision, reflected in its precision of 0.957. This suggests that RF is reliable in applications where precision is prioritized, albeit with slightly less balance than XGB. The DT model showcases a moderate performance, with an accuracy of 0.953 and an F1-Score of 0.954. Despite having lower overall precision than the RF and XGB models, it achieves a high precision of 0.965, indicating its effectiveness in scenarios

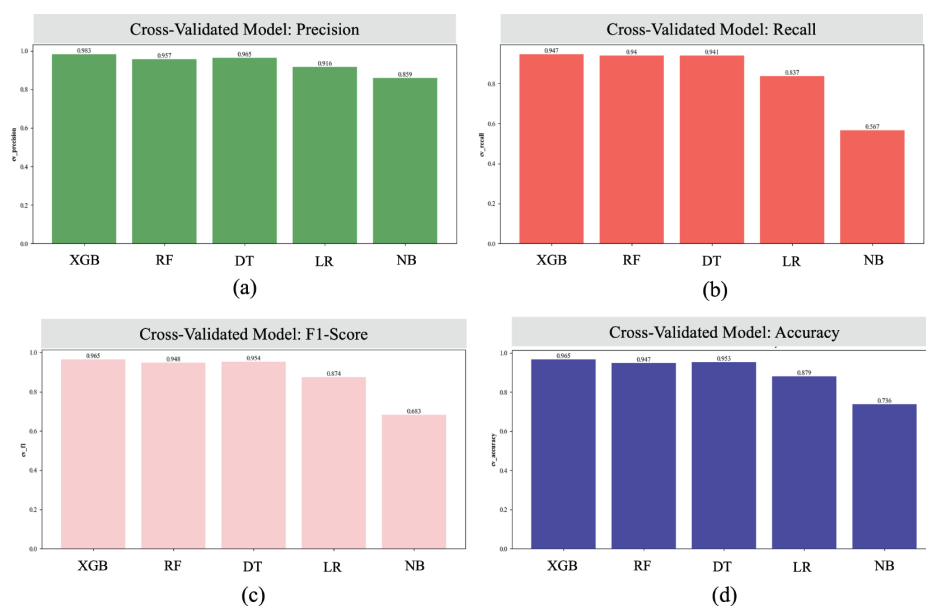


Figure 4. Predictive model results by metric: (a) Precision, (b) Recall, (c) F1-Score, (d) Accuracy. The x-axis contains the model name while the y-axis contains the obtained value of the metric. Full-size image available here: <https://bit.ly/3M75Pch>.

¹¹ 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.

¹² <https://github.com/roberto-nai/ANAC-OD-ETHICAL>.

Model	Accuracy	F1-Score	Precision	Recall
XGBoost	0.965	0.965	0.983	0.947
Random Forest	0.947	0.948	0.957	0.940
Decision Tree	0.953	0.954	0.965	0.941
Logistic Regression	0.879	0.874	0.916	0.837
Naive Bayes	0.736	0.683	0.859	0.567

Table 5. Predictive models results: (a) Precision, (b) Recall, (c) F1-Score, (d) Accuracy. In bold the best model, underlined the second.

where correct positive predictions are more critical than capturing all potential positive instances. In contrast, LR and NB present substantially lower cross-validation metrics. LR achieves an accuracy of 0.879 and an $F1$ -Score of 0.874, reflecting its limited ability to capture the complexities in the data compared to ensemble methods. Similarly, NB records the lowest performance across the board, with an accuracy of 0.736 and an $F1$ -Score of 0.683, indicating that its assumptions may not align well with the underlying data distribution. Overall, the cross-validation metrics emphasize the robustness of ensemble methods, particularly XGB and RF, in delivering consistent and reliable predictions across varying datasets.

The recommendations for stakeholders are to carefully monitor the incidence of variations in total costs and contract completion times. For policymakers, it is a matter of proposing regulatory corrections, but also of carefully monitoring the overall impact of variations on public expenditure. Again, the big data available in the ANAC database can help, as it can be used to model the general behavior of economic operators in proposing and executing tenders, and compare them specifically for the frequency of variation requests. Data analysts and machine learning tools applied to the big data of public contracts could detect economic operators that represent anomalies in proposing several variations in a recurring manner and, on the contrary, propose best practices and examples of virtuous behavior.

6. Recommendation systems

By filtering the database for a description of CPV codes related to “consultancy” tenders, we obtained 72 CPV codes. In particular, the CPV codes used for filtering the ANAC database are mainly from these divisions based on the first two digits: **Table 6** shows the list of CPV codes retrieved by using the recommendation system on the description of codes. Results show that there are 25 categories of CPV codes starting with the 71XXX000-Y division which are related to counseling.

6.1 Item retrieval and precision score

The recommendation system trained on the original ANAC database was used to obtain the top 10 tenders most similar to each of the “consultancy” tenders used as a query example.

For item retrieval, we used a Bert LaBSE model (Language-agnostic Bidirectional Transformer Sentence Embedding) [20]. Embedding [18] is a transformative

Divisions	Description of CPV code	Categories
71XXX000-Y	Architectural consulting services, Construction consulting services, Geological and geophysical consulting services, Construction consulting services	25
72XXX000-Y	Software consulting services, Software integration consulting services, Hardware integration consulting services	11
73XXX000-Y	Research and development-related consulting services, Consulting in the field of research and development	4
66XXX000-Y	Financial advisory services, Financial transaction management and clearing services	4
79XXX000-Y	Software copyright consulting services	21
85XXX000-Y	Counseling services provided by nursing staff, Guidance and counseling services	3
66XXX000-Y	Consulting services in the field of insurance	4
30XXX000-Y	Computer equipment	1
22XXX000-Y	Computer manuals	1
98XXX000-Y	Equal opportunity counseling services	1

Table 6. CPV division codes related to counseling and their category, used to filter the ANAC database for tenders to train the recommendation system.

approach to natural language that has become popular with deep artificial neural networks trained on large volumes of documents. It is used to represent the semantic content of documents written in natural language. As it is the case of language-agnostic models, documents might be written even in many languages. Embeddings were employed to treat the short description of procurement provided with the tender subject field in the ANAC database. Embeddings of the tender subjects represent, in a concise yet quite general and powerful way, the topic referred to in the tender subject. Embeddings are numerical vector representations that capture the semantic essence of the data. Cosine similarity was later used to find similar items.

To retrieve tenders related to the users' interest from ANAC-approved tenders, a query in natural language can be submitted by the user and processed to retrieve the top k tenders more similar to the description provided by the query. For example, **Table 7** shows the top 10 most similar tenders related to the query "legal counseling".¹³ Cosine similarity [35], a mathematical metric used to measure the similarity between two vectors in a multi-dimensional space, is then used to compare these embeddings. It involves calculating the cosine of the angle between the embeddings, providing a measure of how similar they are, based on orientation rather than magnitude. **Table 7** shows the results obtained by the user when a query is made. The system, when processed the user's query and transformed it into an embedding using the deep neural network of Bert LaBSE, uses cosine similarity to measure the relevance of tenders in its database to the query. This ensures that the most similar and relevant tenders are returned, enhancing the efficiency and effectiveness of the

¹³ In the original database, the tender object was in Italian: "SERVIZI LEGALI" and "SERVIZI DI ASSISTENZA LEGALE".

Query	Cig	Top k Tenders subject	Code CPV	Accuracy
1	69665502ED	LEGAL SERVICES	79,111,000-5	TP
2	8,018,638,568	LEGAL SERVICES	79,111,000-5	TP
3	76317249DE	LEGAL SERVICES	79,111,000-5	TP
4	8,139,176,483	LEGAL ASSISTANCE SERVICES	79,110,000-8	TP
5	78756296A5	LEGAL ASSISTANCE SERVICES	79,110,000-8	TP
6	8139089CB5	LEGAL ASSISTANCE SERVICES	79,110,000-8	TP
7	7993481D2F	LEGAL ASSISTANCE SERVICES	79,110,000-8	TP
8	7,912,174,489	LEGAL ASSISTANCE SERVICES	79,110,000-8	TP
9	79122199AA	LEGAL ASSISTANCE SERVICES	79,110,000-8	TP
10	7856558CC0	LEGAL ASSISTANCE SERVICES	79,110,000-8	TP

Table 7.

Top k tenders retrieved by the recommendation system for the query “LEGAL SERVICES” with $k = 10$.

search. However, these results can be explained by looking at the texts of the notices taken into consideration. Clearly, for the example provided by the query “legal counseling” for the recommendation system, it is easy to recognize that the sets of tenders shown in **Table 7** whose descriptions are “LEGAL SERVICES” and “LEGAL ASSISTANCE SERVICES” are similar to the query and to each other. These are assigned a cosine similarity of 0.99 and the two calls appear to be, each other, among the top 10 of their neighbors. If we consider as true positives (TP) the neighbors who are also of the “consultancy” type, we can calculate the $precision@10$ using the following formula:

$$Precision@10 = \frac{\text{Relevant tenders in top 10 recommendations}}{10} = \frac{TP}{10} \quad (2)$$

We tested the recommendation system for 100 queries randomly selected from the database with CPV in **Table 6**. We also filled the database with irrelevant examples randomly selected from the rest of the ANAC database to give the system also examples of irrelevant cases for the query. Their number is equal to the number of cases with CPV in **Table 6**. The obtained results of the recommendation system are that $precision@10$ values are, on average, equal to 0.85 considering the list of the top 10 most similar results to the queries. Results indicate that *more than half* of the recommended tenders out of the top 10 are of the “consultancy” type, and are similar to the tender under investigation (query tender). **Table 8** and the side picture show the results of precision for the tender recommendation system for random queries with CPV in **Table 6**. It results that 74.15% of queries have an extremely high precision score (equal to 0.9) while there is almost 8% of queries whose recommended tenders were less similar (with the lowest precision scores, in the two ranges 0.20–0.40 and 0.40–0.60).

The results allow to make claims for stakeholders: they should spread the possibility of using machine learning tools and recommendation systems, making the search for cases more effective and speeding up the bidding and awarding processes. If contracting authorities had a complete collection of cases similar to their own at

Precision score range	Percentage
> 0.90	74.15
0.80–0.90	4.49
0.60–0.80	8.98
0.40–0.60	3.37
0.20–0.40	4.49
<0.20	4.49

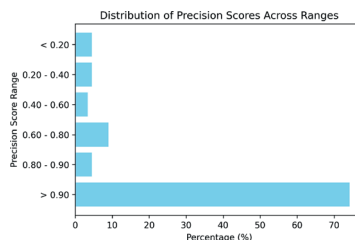


Table 8. Distribution of precision scores for the tender recommendation system for random queries with CPV in Table 6.

the moment they were drawing up their tender, they could prepare their tender with the most appropriate constraints. As a result, we would expect to see a lower number of no bids. Similarly, if economic operators had a collection of similar cases to their own, they would propose more appropriate bids and see a higher success rate in their bidding, with a consequent reduction in effort and resources.

7. Conclusions

The case of tenders issued by public administrations in Italy in the field of consultancy over a period of 4 years is considered. A full analysis of the data in the ANAC database was carried out. A descriptive analysis of the data was applied in the form of a geographical heat map, which highlighted that Lazio was the Italian region with the highest number of tenders in consultancy.

Ensuring ethical and fair practices in public procurement is essential, particularly in light of regional economic disparities. The proposed study leverages ML models to assess fairness in the awarding of contracts. By applying fairness tests (independence, separation, and sufficiency), the discussed study shows that contracts tend to be awarded more frequently in northern regions compared to southern ones, despite similar economic conditions. This raises concerns about equity and fairness in the allocation of public funds, making it crucial to implement AI systems that promote fairness and reduce biases. The obtained results confirm the persistence of the historical southern question in Italy, the socio-economic consequences of which are inequality and unequal opportunities for the population throughout the Italian territory. Future procurement processes should consider these findings to ensure fair opportunities for all regions.

As regards the predictive task, the experiments successfully identified procurement cases with variants, highlighting their effects on budget distribution. When a variant causes an increase in the final price, it alters expenditure allocations, potentially depriving other activities of necessary funds. The predictive tasks highlight that XGBoost (accuracy 96.5%) outperformed all other models, making it ideal for predicting tender contract variations where both false positives and negatives are costly. RF (accuracy 94.7%) also showed strong performance, while Logistic Regression and Naive Bayes were less effective, with lower metrics indicating their unsuitability for complex datasets. For public authorities, implementing XGBoost or Random Forest models can help prevent delays and cost overruns by accurately predicting variations

in bids. Policymakers should focus on integrating these AI tools to ensure fair and efficient public procurement processes. Contractors can use the insights from these models to better structure their bids.

Another discussed task is to train recommendation systems on the description of the tender topics. The recent technology of transformers (SBERT) was used, which is very successful in grouping texts with similar content. This type of recommendation service could be extremely useful to speed up the search for interesting cases in a large volume of examples, as well as to find similar situations or economic operators working in the same domain. The recommendation system boosts bidders' experience to access and design with more competitive techniques by providing insights into previous bidders and winners for similar tenders. The results obtained by the recommendation system trained on the same consultancy data show that the majority of the queries (74.15%) submitted to the system have a high precision score (higher than 0.90), which serves as an evaluation of the relevance of the answers within the top 10 results. If the remaining precision score ranges are considered (below 0.90) the percentage of queries drops to values similar to 4.49, this indicates that a very small number of queries resulted in less relevant outputs and shows that the performance of the system follows a skewed distribution, skewed toward the high precision scores. However, analyzing the characteristics of these lower-performing queries can reveal specific patterns or contexts where the system struggles.

The recommendation to stakeholders and policymakers is to enlarge and enforce the use of AI/ML tools to assist actors in the public contract bidding process and execution in order to effectively use the available big data and make the process more efficient in the use of resources. A recommendation system based on a Large Language Model (LLM) specifically trained on tenders and their related data (such as bid amount, successful bidders, and contracting authorities) can improve the tendering process for the various stakeholders in decision-making and improve the efficiency of the operations. It indicates the most relevant tenders by analyzing the supplier's or company's history and the expertise related to the area of interest. The LLM-based recommendation system can offer insights into risk assessment by keeping a track record of specific sectors of winning bidders and allowing the financial authorities to access the risk of investment and finances for specific companies. The future work is to employ the presented tools to discover anomalies, propose best practices, and recommend suitable actors in order to enlarge, especially in the Mezzogiorno, the economic operators' participation in the public tenders.

As Artificial Intelligence continues to transform delicate sectors of society, from healthcare to finance, it is crucial to ensure transparency, trust, and acceptance of AI technology [2]. Unfortunately, the complexity of many AI models often renders them "black boxes" with limited interpretability. Future works consider integrating Explainable AI (XAI) in predictive and recommendation models.

Acknowledgements

The authors thank the Department of Management for supporting this research with legal domain experts in public tenders, particularly Prof. Gabriella Margherita Racca and Dr. Francesco Gorgerino for contributing to understanding the procurement domain.

Conflict of interest

The authors declare no conflict of interest.

Appendices and addenda must be cited in the main text (example: See Appendix A). The section containing them must be titled accordingly (“Appendices”, “Appendix A”, “Addendum”, “Nomenclature”, etc). An example of appendix/addendum/nomenclature is given below:

Abbreviations

ANAC	National authority for anti-corruption
CPV	common procurement vocabulary
ISTAT	Italian statistical institute
PIL	gross domestic product


Author details

Ishrat Fatima^{*†}, Roberto Nai[†] and Rosa Meo[†]
Computer Science Department, University of Turin, Torino, IT, Italy

*Address all correspondence to: ishrat.fatima@unito.it

†These authors contributed equally.

IntechOpen

© 2025 The Author(s). Licensee IntechOpen. This chapter is distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. 

References

- [1] ANAC. Catalog of Open Data ANAC. 2024. Available from: <https://dati.anticorruzione.it/opendata> [Accessed: August 10, 2024]
- [2] Meo R, Nai R, Sulis E. Explainable, interpretable, trustworthy, responsible, ethical, fair, verifiable AI. What's next? In: *Advances in Databases and Information Systems - 26th European Conference, ADBIS 2022*; 5-8 September 2022; Turin, Italy. Lecture Notes in Computer Science. Vol. 13389. Heidelberg, DE: Springer; 2022. pp. 25-34. DOI: 10.1007/978-3-031-15740-0_3
- [3] Barocas S, Hardt M, Narayanan A. *Fairness and Machine Learning Limitations and Opportunities*, 2018. Cambridge, MA: The MIT Press ebook; 19 December 2023. pp. 55-60. Available from: <https://api.semanticscholar.org/CorpusID:113402716>
- [4] Nai R, Sulis E, Pasteris P, Giunta M, Meo R. Exploitation and merge of information sources for public procurement improvement. In: *International Workshops of ECML PKDD 2022 Grenoble, France*; 19-23 September 2022. Proceedings, Part I. Heidelberg, DE: Springer; 2023. pp. 89-102. DOI: 10.1007/978-3-031-23618-1_6
- [5] Nai R, Sulis E, Genga L. Automated analysis with event log enrichment of the European public procurement processes. In: Sales TP, Guizzardi G, Araújo J, Borbinha J, Guizzardi G, editors. *Advances in Conceptual Modeling: ER 2023 Workshops, CMLS, CMOMM4FAIR, EmpER, JUSMOD, OntoCom, QUAMES, and SmartFood*; 6-9 November 2023; Lisbon, Portugal. Lecture Notes in Computer Science (LNCS). Vol. 14319. Heidelberg, DE: Springer; 2023. pp. 178-188. DOI: 10.1007/978-3-031-47112-4_17
- [6] Goebel P, Reuter C, Pibernik R, Sichtmann C. The influence of ethical culture on supplier selection in the context of sustainable sourcing. *International Journal of Production Economics*. 2012;**140**(1):7-17. DOI: 10.1016/j.ijpe.2012.01.021
- [7] Kim S, Colicchia C, Menachof D. Ethical sourcing: An analysis of the literature and implications for future research. *Journal of Business Ethics*. 2018;**152**(4):1033-1052. DOI: 10.1007/s10551-016-3369-4
- [8] Hazen BT, Boone CA, Ezell JD, Jones-Farmer LA. Artificial intelligence and big data analytics for supply chain resilience: A systematic literature review. *Annals of Operations Research*. 2021;**299**(1):277-315. DOI: 10.1007/s10479-019-03472-8
- [9] Belhadi A, Kamble S, Wamba SF, Gunasekaran A, Ndubisi NO, Venkatesh M. Artificial intelligence and big data analytics for supply chain resilience: A systematic literature review. *Annals of Operations Research*. 2021;**302**(1):1-52. DOI: 10.1007/s10479-020-03626-1
- [10] Goodwin D, Bok B, Zhao F, Zhang P. A systematic literature review of research on social procurement in the construction and infrastructure sector: Barriers, enablers, and strategies. *Sustainability*. 2023;**15**(17):12964. DOI: 10.3390/su151712964
- [11] Nai R, Sulis E, Meo R. Public procurement fraud detection and artificial intelligence techniques: A literature review. In: Symeonidou D,

- Yu R, Ceolin D, Poveda-Villalón M, Audrito D, Caro LD, et al., editors. Companion Proceedings of the 23rd International Conference on Knowledge Engineering and Knowledge Management. Vol. 3256. Bozen-Bolzano, Italy: CEUR-WS.org, 2022. Available from: <https://ceur-ws.org/Vol-3256/km4law4.pdf>
- [12] Nai R, Fatima I, Morina G, Sulis E, Genga L, Meo R, et al. AI applied to the analysis of the contracts of the Italian public administrations. In: Falchi F, Giannotti F, Monreale A, Boldrini C, Rinzivillo S, Colantonio S, editors. Proceedings of the Italia Intelligenza Artificiale - Thematic Workshops co-Located with the 3rd CINI National lab AIIS Conference on Artificial Intelligence (Ital IA 2023), Pisa, Italy, May 29-30, 2023, CEUR Workshop Proceedings. CEUR-WS.org. Vol. 3486. 2023. pp. 255-260. Available from: <https://ceur-ws.org/Vol-3486/100.pdf>
- [13] Hoejmose SU, Adrien-Kirby PAJ. Socially and environmentally responsible procurement: A literature review and future research agenda of a managerial issue in the 21st century. *Journal of Purchasing and Supply Management*. 2012;**18**(4):232-242. DOI: 10.1016/j.pursup.2012.06.002
- [14] Lozano R. A holistic perspective on corporate sustainability drivers. *Corporate Social Responsibility and Environmental Management*. 2015;**22**(1):32-44. DOI: 10.1002/csr.1325
- [15] Dhanani J, Mehta R, Rana D. Legal document recommendation system: A cluster based pairwise similarity computation. *Journal of Intelligent and Fuzzy Systems*. 2021;**41**(5):5497-5509. DOI: 10.3233/JIFS-202576
- [16] Nai R, Meo R, Morina G, Pasteris P. Public tenders, complaints, machine learning and recommender systems: A case study in public administration. *Computer Law and Security Review*. 2023;**51**:105887. DOI: 10.1016/j.clsr.2023.105887
- [17] Reimers N, Gurevych I. Sentence-BERT: Sentence embeddings using Siamese BERT-networks. In: Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP); 3-7 November 2019; Hong Kong, 2019. pp. 3982-3992. DOI: 10.18653/v1/d19-1410
- [18] Devlin J, Chang M-W, Lee K, Toutanova K. BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding. 2019. eprint: 1810.04805. arXiv, cs.CL. Available from: <https://arxiv.org/abs/1810.04805>
- [19] SBERT.net. SentenceTransformers Documentation. GitHub; 2023. Available from: <https://sbert.net/> [Accessed: August 10, 2024]
- [20] Feng F, Yang Y, Cer D, Arivazhagan N, Wang W. Language-Agnostic BERT Sentence Embedding. 2022. eprint: 2007.01852. arXiv, cs.CL. Available from: <https://arxiv.org/abs/2007.01852>
- [21] Nai R, Sulis E, Fatima I, Meo R. Large language models and recommendation systems: A proof-of-concept study on public procurements. In: Rapp A, Di Caro L, Meziane F, Sugumaran V, editors. *Natural Language Processing and Information Systems*. Switzerland, Cham: Springer Nature; 2024. pp. 280-290. DOI: 10.1007/978-3-031-70242-6_27
- [22] van Dijk E, Wilke H. Differential interests, equity, and public good provision. *Journal of Experimental Social Psychology*. 1993. ISSN

0022-1031;**29**(1):1-16. DOI: 10.1006/jesp.1993.1001

[23] Decarolis F, Giorgiantonio C. Corruption red flags in public procurement: New evidence from Italian calls for tenders. *EPJ Data Science*. 2022;**11**(1):1-38. DOI: 10.1140/epjds/s13688-022-00325-x

[24] Caton S, Haas C. Fairness in machine learning: A survey. *ACM Computing Surveys*. 2024;**56**(7), Art. No.: 166:1-38. DOI: 10.1145/3616865

[25] Lombardini S. Modello macroeconomico previsionale per il PIL delle regioni italiane. Italy: CPT Ricerca and University of Genova; 2022. ISBN 9791280477170

[26] Khatun M, Siddiqui S. Testing pairs of continuous random variables for independence: A simple heuristic. *Journal of Computational Mathematics and Data Science*. 2021;**1**:100012. DOI: 10.1016/j.jcmds.2021.100012. ISSN 2772-4158

[27] Cox DR. The regression analysis of binary sequences. *Journal of the Royal Statistical Society: Series B (Methodological)*. 1958;**20**(2):215-232. Wiley Online Library. DOI: 10.1111/j.2517-6161.1958.tb00292.x

[28] Breiman L, Friedman J, Stone CJ, Olshen RA. *Classification and Regression Trees*. Belmont, CA: CRC Press; 1984. DOI: 10.1201/9781315139470

[29] Breiman L. Random forests. *Machine Learning*. 2001;**45**(1):5-32. DOI: 10.1023/A:1010933404324

[30] Chen T, Guestrin C. XGBoost: A scalable tree boosting system. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New

York, NY: ACM; 2016. pp. 785-794. DOI: 10.1145/2939672.2939785

[31] Domingos P, Pazzani M. On the optimality of the simple Bayesian classifier under zero-one loss. *Machine Learning*. 1997;**29**(2-3):103-130. DOI: 10.1023/A:1007413511361

[32] Géron A. *Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. 2nd ed. Sebastopol, CA: O'Reilly Media; 2019. ISBN 978-1492032649. Available from: <https://www.oreilly.com/library/view/hands-on-machine-learning/9781492032632/10.5555/3380750>

[33] Kohavi R. A study of cross-validation and bootstrap for accuracy estimation and model selection. In: *Proceedings of the 14th International Joint Conference on Artificial Intelligence (IJCAI)*. Vol. 2. 1995. pp. 1137-1145. DOI: 10.5555/1643031.1643047

[34] Powers DMW. Evaluation: From precision, recall and F-measure to ROC, informedness, markedness and correlation. *Journal of Machine Learning Technologies*. 2011;**2**(1):37-63

[35] Duarte GN. The Cosine Similarity and its Use in Recommendation Systems. *Medium*; 2023. Available from: <https://naomy-gomes.medium.com/the-cosine-similarity-and-its-use-in-recommendation-systems-cb2ebd811ce1> [Accessed: August 10, 2024]