Fraud Detection in Public Procurement Using Open Government Data and Red Flags

Abstract. Fraud in Public Procurement (PP) poses significant challenges, leading to financial losses and inefficiencies worldwide. The complexity of PP systems and the vast amounts of data involved make detecting irregularities difficult without advanced tools. This paper presents an approach to detect PP fraud using red flags derived from open data available on Brazilian government portals. By analyzing datasets such as company registrations, procurement contracts, sanctions, and political affiliations, the study identifies key risk indicators structured into categories, including financial, administrative, and political factors that may suggest fraudulent behavior. To ensure practical relevance, the proposed red flags were evaluated through a structured questionnaire with experienced auditors, allowing adjustments to address the specific characteristics of the Brazilian PP environment. This evaluation step is critical for strengthening the reliability of the framework and enhancing its applicability in real-world scenarios. The findings demonstrate that integrating open data with practical expertise significantly improves fraud detection strategies. This work contributes to digital government research by providing a systematic approach to leveraging open data for public oversight, promoting transparency, and increasing accountability in PPP activities. Ultimately, the proposed framework serves as a valuable tool for practitioners and researchers aiming to enhance fraud detection and risk management in government procurement.

Keywords. Open Government Data, Digital Government, Fraud Detection, Red Flags, Public Procurement

1. Introduction

Public Procurement (PP) is a fundamental process through which governments acquire goods and services necessary to fulfill their public functions. These processes are governed by rules and procedures designed to ensure transparency, equality among competitors, and efficient use of public resources (Mahmood, 2010). However, fraud in Public Procurement represents a significant global challenge, leading to substantial financial losses, inefficiencies, and erosion of trust in public administration (Ferguson, 2018). In Brazil and around the world, fraudulent practices such as overpricing, favoritism, and supplier collusion result in resource misallocation and undermine the principles of fair competition (Menezes, 2021; Santos & de Souza, 2016).

To address these challenges, the use of risk indicators known as red flags has gained prominence as an effective tool for detecting irregularities in PPs. Red flags act as warnings, highlighting patterns or behaviors that deviate from the norm, signaling potential fraudulent activities (Munteanu et al., 2024; Potin, Figueiredo, et al., 2023). These indicators are particularly valuable as they allow auditors and automated systems to focus on potentially suspicious activities before significant financial losses occur. These indicators can also be used to feed technologies based on AI to support data analysis (Modrušan et al., 2021). However, constructing red flags is not a trivial task, as it requires identifying relevant risk indicators, ensuring data quality, and adjusting them to the specific context of PP sempre tem (Kawashita et al., 2024)

In recent years, the global movement toward open government data has presented new opportunities to improve fraud detection mechanisms (Sampaio et al., 2024). In Brazil, the availability of open data from public portals has played a critical role in increasing transparency and accountability (Lourenço, 2015; Salm Junior et al., 2024). These data repositories, such as the Transparency Portal and Open Data Portals in federal, state

and local governments, offer extensive datasets on government expenditures, contracts, and bids. Despite the potential of these resources, the literature lacks comprehensive approaches that demonstrate how to systematically transform open data into actionable red flags for fraud detection.

This work aims to bridge this gap by presenting a systematic process for developing and evaluating red flags tailored to the Brazilian PP context. The proposed approach involves the collection and integration of datasets from official government portals, the construction of red flags through an ad hoc literature review and expert interviews, and their validation by experienced auditors. This evaluation ensures the relevance and applicability of the indicators to real-world oversight processes. By focusing on the procurement and contracting stages, particularly the post-purchase transactions, this study contributes to developing a more effective framework for fraud detection. The integration of advanced data analysis techniques with expert evaluation underscores the potential of open government data to enhance transparency, accountability, and the integrity of public procurement systems.

This work is organized as follows. Section 2 presents the Literature Review, discussing existing approaches to fraud detection in public procurement and the role of open data and red flags. Section 3 describes Our Proposed Process, outlining the methodology for collecting, integrating, and analyzing open data to identify fraudulent patterns. Section 4 details the Methodology for Data Collection and Fraud Detection in Public Procurement, covering the identification, preprocessing, and structuring of relevant datasets, as well as the development of risk indicators. Section 5 presents the Auditors Evaluation, which assesses the applicability and effectiveness of the proposed red flags through a structured questionnaire. Finally, Section 6 provides the Conclusions, summarizing the key contributions and suggesting directions for future research.

2. Literature Review

Traditionally, fraud detection in PPs has been done manually (Brandão et al., 2024), through audits based on samples and complaints (Velasco et al., 2021), which makes the process slow, ineffective, and prone to errors. This method cannot keep up with the large volume of data generated by large-scale PP. Given this, the need arises for a data-driven approach (Lyra et al., 2022), where large-scale data analysis and the use of technologies such as AI -based and automation tools can detect fraud patterns more quickly and accurately, allowing the proactive identification of fraud and other crimes before they cause significant damage.

Consequently, several studies have been conducted to transform this scenario, promoting fraud detection based on data analysis that can be implemented in an automated and effective manner. Niessen et al. (2020) propose a model for detecting anomalies in PP in Paraguay using the international Open Contracting Data Standard (OCDS). Applying the Isolation Forest unsupervised learning algorithm, the model identifies anomalies in PP based on open data. The study focuses on the planning, bidding, and contracting phases, where processes are analyzed to determine unusual patterns that may indicate fraud or irregularities. Potin, Labatut, et al. (2023) describe the creation of an open database that gathers information on public contracts in France. The data were collected from the Tenders Electronic Daily (TED) [1] to assist fraud detection of fraud and monitoring of public policies. The article discusses the challenges of treating raw data, such as identification and consistency problems. It presents the solutions adopted to standardize and improve the quality of the information. Sanda et al. (2024), the use of big data analytics to improve accountability and transparency in online PP. The study explores on the case of the Romanian PP portal and analyzes millions of direct procurement procedures carried out in 2023. The central objective is to identify anomalies and possible fraud, especially in exclusive commercial relationships between public sellers and buyers, using big data techniques to detect unusual patterns. All these authors use raw data, which means they are not focusing on the data enginnering process in order to enrich their data from diverse data source, as auditors usually work in the fraud detection in PPs scenarios.

So as data analysis and the use of automated tools advance, the concept of red flags has been increasingly integrated into these solutions (Decarolis & Giorgiantonio, 2022), functioning as an essential complement to fraud detection. While machine learning models and anomaly detection algorithms bring a proactive and automated approach (Nai et al., 2022), red flags offer an additional layer of contextualization and interpretation, guiding auditors and automated systems to focus on critical areas of PP. Combining these two approaches

[1] https://ted.europa.eu/en/

strengthens the monitoring of bids and contracts, allowing for more robust and accurate detection of possible irregularities.

Therefore, the use of red flags as risk indicators has been recognized as an effective strategy for identifying potential fraud in PP. These indicators, based on unusual or anomalous patterns, have been highlighted in several studies for their ability to proactively detect irregularities and guide audits more effectively. For instance, Modrušan et al. (2021) emphasize the importance of specific red flags, such as purchase splitting and reduced competition in bidding processes, coupled with machine learning algorithms to automate anomaly detection. This approach demonstrates how targeted indicators can enhance the identification of fraud patterns. Similarly, Velasco et al. (2021) propose a decision support system (DSS) incorporating data mining techniques to detect corruption risks in Brazil's PP context. The DSS relies on red flags like conflicts of interest and collusion among bidders, offering a structured method to prioritize investigations. Additionally, Potin, Figueiredo, et al. (2023) employ pattern mining techniques in procurement data graphs to detect collusive behaviors, using red flags such as shared addresses or concealed corporate ties between companies to highlight suspicious activities.

While last studies provide valuable insights, they exhibit certain limitations. Many approaches focus on isolated datasets or specific types of fraud, lacking a comprehensive integration of diverse data sources, specially to construct red flags. Moreover, the contextual relevance and practical validation of red flags are often overlooked, particularly in environments the possibility of as complex as Brazil's PP system.

Considering these foundations and limitations, our work addresses these gaps by integrating multiple public datasets into a centralized repository, enabling a more holistic analysis of procurement data. Scientifically, we advance the state of the art by contextualizing red flags to the Brazilian PP environment, ensuring their applicability through collaboration with local auditors. Technically, our approach combines advanced data analysis techniques with a systematic framework for red flag evaluation, delivering a scalable and robust solution for fraud detection.

3. Our Proposed Process

This section presents the proposed methodology for data collection, integration, and analysis focused on fraud detection in PP in Brazil. The approach follows three main steps: (i) the collection and centralization of open data obtained from open government data portals, (ii) the definition of risk indicators (red flags), developed through a literature review and a series of interviews conducted with auditors and public controllers, and (iii) the evaluation of both the red flags and the collected data, in collaboration with Brazilian auditors. This methodology ensures that the results are both accurate and relevant by combining advanced data analysis techniques, rigorous validation of red flags, and the practical expertise of Brazilian auditors. This combination addresses the challenges of data quality, contextual adaptation, and the operational nuances of fraud detection in Public Procurement.

This method aims to ensure accurate and relevant results regarding the data engineering process, combining data scientists' practical and theoretical expertise, the use of advanced technologies for this task, and the practical expertise of auditors. Our method is mainly defined by a structured process, as illustrated in Figure 1. The process involves the following tasks (T):

- **T1 -** Identification and collection of multimodal open data: Relevant public procurement data is identified and extracted from open government data (OGD) portals, including documents, structured data, and other accessible resources.
- **T2 -** Standardization and centralization of data: The collected data is standardized, ensuring consistency and reliability, and centralized into a unified repository to enable seamless integration.
- **T3 -** Transformation of multimodal data into structured formats: The process involves transforming text-based and tabular data into harmonized structured formats suitable for subsequent analysis.
- **T4 -** Preprocessing of structured data: This step involves cleaning and refining the structured data to ensure it is ready for constructing and validating red flags.
- **T5 -** Literature review for red flag identification: A targeted literature review is conducted to identify red flags commonly used in fraud detection within PP contexts.
- **T6 -** Interviews with auditors and controllers: Collaborators provide domain-specific insights to contextualize and adapt red flags to the characteristics of Brazilian PP.

- **T7 -** Construction of red flags: Using the insights gathered, red flags are developed and validated, leveraging the processed structured data.
- **T8 -** Evaluation and refinement: The generated red flags and data collections are evaluated by auditors. If necessary, refinements are applied iteratively to enhance the precision and applicability of the framework.

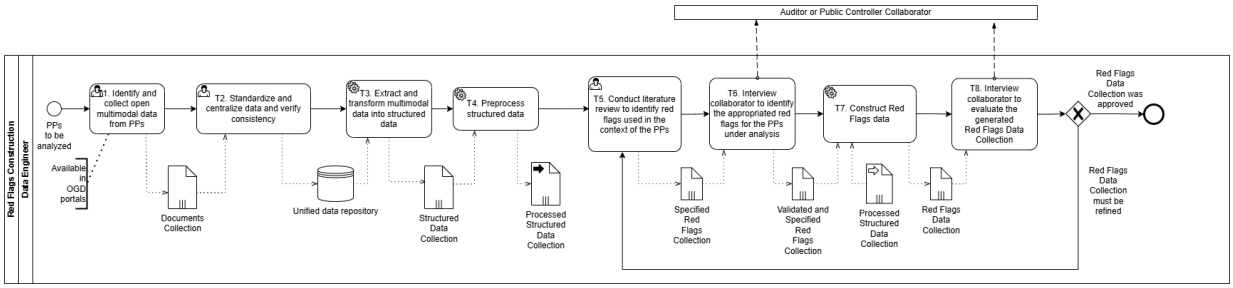


Fig. 1 – Public Procurement Data Process

4. Methodology for Data Collection and Fraud Detection in Public Procurement

This section presents the adopted methodology in detail, encompassing both the collection and preparation of data as well as the construction and validation of red flags in public procurement. The framework is divided into three subsections. The first subsection focuses on task **T1**, which involves identifying and collecting multimodal open data from public procurement processes, with an emphasis on selecting relevant and high-quality data sources. The second subsection addresses tasks **T2**, **T3**, and **T4**, which cover the standardization, transformation, and preprocessing of the collected data, ensuring it is structured, consistent, and ready for analysis. Finally, the third subsection outlines tasks **T5**, **T6**, and **T7**, describing the design, evaluation, and iterative refinement of red flags using insights from literature reviews, expert interviews. This comprehensive and structured approach ensures clarity at every stage, while integrating data engineering techniques and expert input to enhance the detection of fraud in public procurement processes.

4.1. Open Data Portals

For the development of this work, several Brazilian public portals were used, providing access to open data relevant to the analysis of fraud in PP. Below, we describe each of the data sources:

1. **Transparency Portal – Bolsa Família Payments** [2]**:** This portal provides detailed data on payments made by the Bolsa Família Program. This information includes beneficiaries, amounts paid, and payment dates, which are useful for identifying possible fraud or irregularities in public resources allocated to social programs.
2. **Superior Electoral Court (TSE)** [3]**:** The TSE portal provides data on candidates, parties, campaign donations, and election results. These data are crucial for verifying potential conflicts of interest in PP, such as the involvement of candidates or elected officials in government contracts.
3. **Government Procurement Portal – CATMAT/CATSER** [4]**:** This portal provides information on goods and services purchased by the federal government. The data are categorized (CATMAT for materials and CATSER for services). They are used to monitor the procurement of goods and services, facilitating the detection of supplier overpricing or collusion.
4. **Comprasnet Contracts API** [5]**:** The Comprasnet API allows access to detailed data on contracts established through federal public tenders. This database includes information on active contracts, contract values, and suppliers, making it essential for identifying anomalies and fraud patterns.
5. **National Registry of Legal Entities (CNPJ)** [6]**:** This dataset provides detailed information on companies registered in Brazil. Access to CNPJ data is essential for verifying the backgrounds of suppliers and identifying suspicious relationships between companies and public officials.

[2] https://portaldatransparencia.gov.br/download-de-dados/bolsa-familia-pagamentos
[3] https://dadosabertos.tse.jus.br/dataset/
[4] https://www.gov.br/compras/pt-br/acesso-a-informacao/consulta-detalhada/planilha-catmat-catser
[5] https://contratos.comprasnet.gov.br/api/docs
[6] https://dados.gov.br/dados/conjuntos-dados/cadastro-nacional-da-pessoa-juridica---cnpj

6. **Accounts System of the Federal Court of Accounts (TCU)** [7]**:** The TCU accounts system provides data related to the audit and oversight of public accounts. These data are used to validate information on government contracts and identify potential irregularities in the use of public resources.
7. **Registry of Suspended and Debarred Companies (CEIS)** [8]**:** The CEIS contains information on companies that have been declared ineligible or suspended from participating in public tenders. Analyzing these data helps prevent penalized companies from continuing to contract with the government.
8. **National Registry of Punished Companies (CNEP)** [9]**:** Similar to CEIS, the CNEP provides data on companies punished for fraud or irregularities, serving as a key resource for monitoring and preventing further fraud.
9. **Transparency Portal – Invoices** [10]**:** This portal provides data on invoices issued in connection with public contracts, enabling detailed analysis of goods and services acquired and the verification of irregularities in their execution.
10. **Safra Guarantee Program** [11]**:** This dataset presents information on the Safra Guarantee Program, which assists farmers in cases of significant losses. Analyzing this data helps ensure the correct allocation of resources and identify possible fraud within the program.

4.2. Open Data Collection

The data collection process for this study was conducted by extracting publicly available datasets from government portals. Each portal provided structured data differently, either through APIs, bulk downloads, or manually curated datasets in formats such as *csv*, *txt*, and *xls*. The data collected included information on public contracts, company registrations, invoice details, government spending, and social program payments, among others. The focus was on acquiring data that would identify patterns in PP. For example, contract data from the Comprasnet API provided detailed insights into procurement activities, while the CNPJ database offered comprehensive information on company ownership and structure, which is crucial for detecting conflicts of interest. All datasets were pre-processed to ensure consistency and relevance, involving data cleaning, normalization, and integration into a centralized repository. Furthermore, all codes used in this work are available in the same repository [12] as the dataset, ensuring transparency and facilitating replication or subsequent analysis by other researchers.

Due to the high volume of data in each dataset, distributed processing was employed using PySpark and executed on AWS EMR to efficiently handle the large-scale data transformation and integration tasks. This allowed for the parallel processing of extensive datasets, ensuring timely and accurate pre-processing. After this pre-processing phase, the final output consisted of 27 datasets, totaling 37.8 GB in *parquet* files.

Figure 2 presents the data architecture for collecting, processing, and storing information from the aforementioned Brazilian government portals. The structure is divided into three layers: the Data Layer encompasses data capture, initially stored in the Landing Zone on Amazon S3, which serves as a staging area before processing; the Computing Layer uses Amazon EMR with PySpark to perform ETL (Extraction, Transformation, and Loading) operations, generating refined datasets in Parquet format in the Refined Zone; and the Code Layer, represented by GitHub, which manages the Python and Spark scripts responsible for automating and versioning the pipeline.

The public portals described in this section provide a comprehensive and reliable basis for collecting and analyzing data relevant to PP in Brazil. By centralizing and integrating data from multiple sources—from social program payments to company registries and contract management—this work leverages a diverse set of open data to identify potential fraud and irregularitiesHowever, certain types of data, such as non-public information, financial transactions, tax records, and data protected under the Brazilian General Data Protection Law (LGPD), were excluded due to their restricted access and the focus on publicly available datasets. These omissions are justified by the focus on publicly accessible data, which ensures transparency and replicability of the methodology. Despite these limitations, using the selected portals facilitates the systematic analysis of procurement activities and increases transparency and accountability within the public sector. This data-driven

[7] https://contas.tcu.gov.br/ords/f?p=1660:2:::NO:::
[8] https://portaldatransparencia.gov.br/download-de-dados/ceis
[9] https://portaldatransparencia.gov.br/download-de-dados/cnep
[10] https://portaldatransparencia.gov.br/download-de-dados/notas-fiscais
[11] https://portaldatransparencia.gov.br/download-de-dados/garantia-safra
[12] https://github.com/s-fontes/projeto_capivara

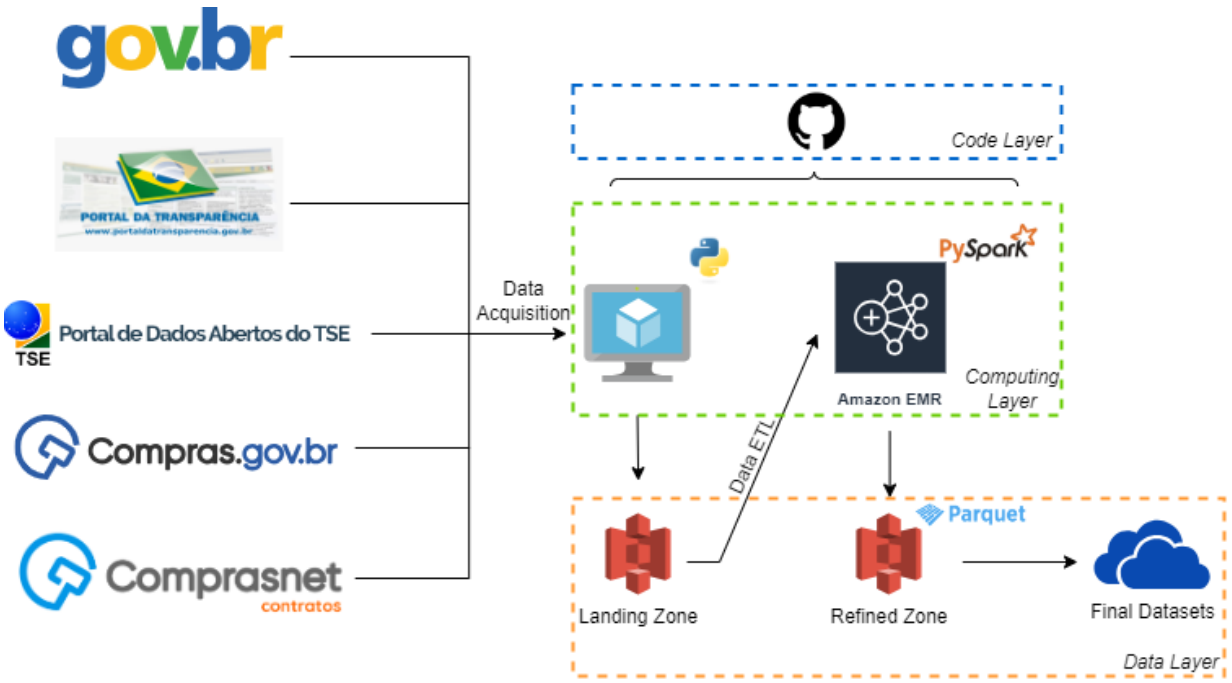


Fig. 2 – Data Collection Architecture

approach enables more efficient and robust monitoring of procurement processes, ultimately contributing to detecting fraud and misuse of public funds.

4.3. Red Flags: Key Risk Indicators for Fraud

Using the integrated and centralized data mentioned in the previous section, it is possible to identify patterns and behaviors that deviate from expectations, indicating potential risks of fraud and irregularities in PP. These patterns are captured through red flags — warning indicators that act as early signs that something may be wrong (Modrušan et al., 2021). Red flags allow managers and auditors to focus more precisely on analyzing contracts and suppliers, acting as an additional layer of verification in combating the inappropriate use of public resources. In this section, we explore the main red flags identified based on the data collected, detailing how each contributes to detecting suspicious practices and supporting more informed and adequate decision-making in monitoring Brazilian PP.

To identify the main red flags applicable to the context of PP in Brazil, an ad hoc literature review was conducted, focusing on studies and reports that address fraud practices and irregularities in bidding processes. This approach allowed us to identify key risk indicators used in different contexts, ensuring that the selected red flags are relevant and appropriate for identifying suspicious behavior and anomalous patterns within the available public data. Table 1 shows the leading base indicators for constructing red flags in PPs fraud detection.

Table 1 presents a set of base indicators used to construct red flags to detect fraud in PP. These indicators are grouped into categories that reflect different aspects of the Company, its contracts, and its partners, covering financial and administrative information and personal and political data. Each row in the table describes a specific attribute. The data sources for each base indicator are shown in numerical format, referencing the database collections used in Section 3.

Specific indicators, by themselves, may not present apparent risks and may be interpreted as normal when analyzed in isolation. However, when evaluated with other factors or within a specific context, these indicators may have a different meaning, pointing to possible irregularities. This change in interpretation occurs when the analysis considers the interrelationships between variables and their contextual nuances, allowing the identification of patterns that would be invisible in a simplified observation. Therefore, in identifying red flags, we must consider not only the data individually but also how this data connects within a broader scenario to

Tab. 1 – Base indicators for constructing red flags in PPs fraud detection

Tag	Name	Category	Source
partner_benefit	Partner as Social Program Beneficiary	Personal	1 and 10
company_date	Company Opening Date	Company	5
company_capital	Company Share Capital	Company	5
partner_politician	Political Partner	Personal	2 and 5
company_sanctions	Company that has some economic sanction	Company	8
partner_public_servant	Partner as a Public Servant	Personal	5
registration_data	Registration Data	Company	5
bidding_contracts	Contracts with and without Bidding	Company	4 and 5
contract_value	Contract Value	Contract	4
company_register	Company Registration Number	Company	5
company_tax_regime	Company Tax Regime	Company	5
company_donor	Electoral Campaign Donor Companies	Company	2 and 5
partner_donor	Electoral Campaign Donor Partner	Personal	2 and 5
partner_political_party	Partner as a Member of Political Party	Personal	2 and 5
company_electoral_campaign	Company Providing Services to the Electoral Campaign	Company	2 and 5
unreputable_company	Unreputable Company	Company	6 and 8
company_punished	Company Punished	Company	7
cnae	Code of Products and Services sold by the Company	Company	3, 9 and 5

provide a more complete and accurate view of the risks involved.

The **Partner as Social Program Beneficiary** red flag aims to identify whether any of the Company's partners have ties to a social program, which may indicate a possible mismatch between the declared financial condition and the position of partner in a company that participates in public tenders. This verification is relevant since a beneficiary, in theory, must be in a situation of economic vulnerability, and their association as a partner in a company involved in government contracts may suggest signs of the use of leads or fraud in the registration and participation process.

The **Company Opening Date** red flag seeks to assess the time interval between the date the Company was opened and the date the public contract was signed. Newly created companies that quickly obtain government contracts may indicate the existence of shell companies or companies explicitly created to defraud the bidding process. This check allows us to detect situations where the Company's short period of operation would not be consistent with its expected experience and capacity to participate in bids, suggesting a possible irregular scheme.

The **Company Share Capital** red flag aims to analyze the discrepancy between the Company's declared share capital and the value of the established public contract. Suppose the value of the contract is significantly higher than the share capital. In that case, this may indicate that the Company does not have the financial capacity to perform the service or supply the contracted goods, suggesting a possible risk of default, overpricing, or fraud. This verification is essential to assess the Company's financial solidity and suitability for the contract's size, ensuring greater security in the contracting processes.

The **Political Partner** red flag seeks to identify whether any of the Company's partners is, or has been, a candidate for political office, which may indicate a possible conflict of interest or use of political influence to favor the Company in bidding processes. The presence of a partner with political involvement may suggest practices of favoritism or collusion, making it necessary to analyze in more detail the conditions under which the contracts were signed.

The **Company that has some economic sanction** red flag aims to check whether the Company has any history of sanctions applied by government agencies, such as a ban on participating in public tenders or contracts due to previous irregularities. The presence of sanctions indicates a high risk of illicit practices, such as corruption, fraud, or breach of contract, and hiring companies with this history may compromise the integrity of the bidding process. This verification is essential to ensure that companies with a history of misconduct do not continue to benefit from public resources.

The **Partner as a Public Servant** red flag checks whether any of the Company's partners is, or has been, a public servant, which may indicate the existence of a conflict of interest, especially if the public servant is or has recently been in a position of influence within bodies related to the contracting process. The presence of public servants as partners may suggest the possibility of undue favoritism or privileged information in the bidding process. This monitoring is essential to ensure impartiality and transparency in public contracts.

The **Registration Data** red flag checks whether the company shares registration data, such as telephone, email, address, or even partners, with other companies. The existence of contact information or corporate structure in common may indicate hidden links between companies, such as the formation of groups that act in a coordinated manner to manipulate bidding processes or create the illusion of competition. This analysis helps to identify possible cases of collusion, the use of shell companies, or even the creation of multiple companies with the same partners to circumvent restrictions on participation in bidding processes and increase the chance of winning public tenders.

The **Contracts with and without Bidding** red flag analyzes the percentage of contracts that the Company won through bidding processes compared to contracts obtained without Bidding. A high proportion of contracts without Bidding may indicate favoritism or inappropriate use of bidding waivers, practices that may violate the principles of fair competition and transparency in PP. This indicator allows for monitoring suspicious contracting patterns and identifying companies that may be receiving undue benefits to the detriment of competitiveness and equality.

The **Contract Value** red flag examines the representativeness of the contracted value in different contexts, such as the contract value per 100,000 inhabitants and the contract value of the total budget of the contracting agency. By calculating the percentage that each contract represents per inhabitant, it is possible to identify whether the contracted amount is disproportionate to the size of the location served, suggesting possible exaggerations or misappropriation of resources. The analysis of the contract value compared to the agency's budget allows us to assess whether there is an excessive concentration of resources in a single supplier, which may indicate financial dependence, favoritism, or even a risk of overestimating the contract. These checks are essential to ensure that the values are appropriate to the context and that public resources are allocated balanced.

The **Company Registration Number** red flag checks the Company's registration status in the National Registry of Legal Entities (CNPJ) to identify whether the Company's registration is active or inactive. An inactive CNPJ suggests that the Company does not have legal authorization to operate, which makes its participation in public tenders and contracts irregular. An inactive CNPJ may indicate fraud, such as using shell companies or closed companies to formalize contracts, compromising the legality and security of public contracting processes.

The **Company Tax Regime** red flag assesses whether the value of contracts signed by the Company is compatible with the declared tax regime. Tax regimes have specific revenue limits and characteristics that determine the Company's operational and financial capacity. Contracts with values much higher than those permitted by the Company's regime may indicate attempts to hide revenue or simulate a smaller size to obtain tax advantages. This verification ensures participating companies' tax compliance and financial regularity.

The **Electoral Campaign Donor Companies** red flag seeks to identify whether the Company has made financial donations to political campaigns. This information is relevant, as it may indicate the existence of close ties between the Company and candidates or parties, suggesting possible favoritism or financial return in the form of public contracts. Verifying these donations is essential to ensure that contracts are being made impartially and not influenced by political interests, preserving the integrity of the bidding process and equality among competitors.

The **Electoral Campaign Donor Partner** red flag checks whether any of the Company's partners have made donations to political campaigns. Identifying partners who contribute financially to candidates or political parties may indicate the existence of personal ties that could influence the awarding of public contracts or the creation of exclusive benefits for the Company.

The **Partner as a Member of Political Party** red flag checks whether any of the Company's partners are affiliated with a political party. Party affiliation of partners indicates possible political connections that influence

the outcome of bidding processes, especially if the party in question is in government or a position of influence.

The **Company Providing Services to the Electoral Campaign** red flag seeks to identify whether the Company participated as a supplier of services or products for a political campaign. This type of involvement may indicate close relations between the Company and candidates or parties, raising suspicions of favoritism or financial return through public contracts.

The **Unreputable Company** red flag verifies the Company's suitability before control and inspection bodies, such as Audit Courts and Comptroller's Offices. Companies classified as unreliable or with a compromised reputation may have been penalized for involvement in fraud, irregularities in contracts, or corrupt practices.

The **Company Punished** red flag seeks to identify whether the Company has been punished by competent authorities, such as suspensions, fines, or a ban from participating in public tenders. These penalties may result from illegal practices or non-compliance with contractual obligations, indicating a history of misconduct.

The **Code of Products and Services sold by the Company** red flag verifies whether the economic activities registered in the National Registry of Economic Activities (CNAE) of the Company correspond to the purpose of the contract signed. The analysis of the CNAE allows us to confirm whether the Company has the legal qualification to provide the service or supply the specified product, avoiding hiring companies without technical capacity or operating outside their areas of expertise.

In summary, identifying the red flags described in this section allows for more effective monitoring of suspicious patterns and behaviors in PP. Each red flag plays a specific role in detecting potential irregularities, and combining these indicators allows for a more comprehensive view of the risks involved. By structuring these red flags based on different categories and data sources, it is possible to create an alert system that acts as an additional layer of control and oversight, contributing to greater transparency and integrity in bidding processes. Thus, using red flags is essential for building mechanisms to prevent and combat fraud in public contracts.

Despite the effectiveness of red flags mapped from public data, much of the information critical to detecting fraud is not available in open sources, which limits the scope of the analyses. There are red flags that depend on sensitive or restricted data, such as tax and banking information, which, for reasons of confidentiality and privacy protection, cannot be accessed directly. Indicators such as suspicious financial movements between companies, transfers of undeclared funds, or detailed cash flow analysis are essential to identify more complex fraud schemes. However, they can only be obtained by breaking banking and tax secrecy, which requires judicial authorization and the participation of regulatory and investigative bodies.

In addition, many fraudulent schemes involve interpersonal relationships and informal agreements that are difficult to capture by documentary records or financial transactions. Hidden conflicts of interest, the use of intermediaries (leads), and the exchange of favors between public and private agents are practices that cannot be evidenced solely based on data from public portals. To detect these irregularities, it would be necessary to resort to more comprehensive investigative techniques, such as interviews, internal audits, or cooperation between institutions to cross-check confidential information and ongoing monitoring by control and compliance agencies. Thus, although red flags constructed from public data are a good starting point, they represent only one part of a broader system to combat corruption and fraud.

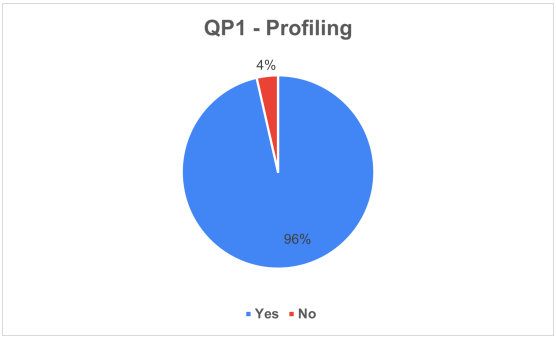
5. Auditors evaluation

In this section, we present a qualitative evaluation process with 28 experienced auditors, using a structured questionnaire to assess the relevance and applicability of the proposed red flags. This section corresponds to task **T8**, which focuses on evaluating the generated red flags and their associated data collections to ensure their adequacy and applicability. The goal is to ensure that the mapped indicators are technically sound and make sense in the practical context of PP oversight and control. Through this form, auditors were invited to give their opinions on the effectiveness of each red flag in identifying fraud and the adequacy of the databases used to build them, providing valuable insights into limitations, necessary adjustments, and possible additions. This feedback is essential to align the modeling of red flags with the operational reality faced by professionals, ensuring that the set of indicators is robust, relevant, and contextualized to the particularities of the PP environment in Brazil.

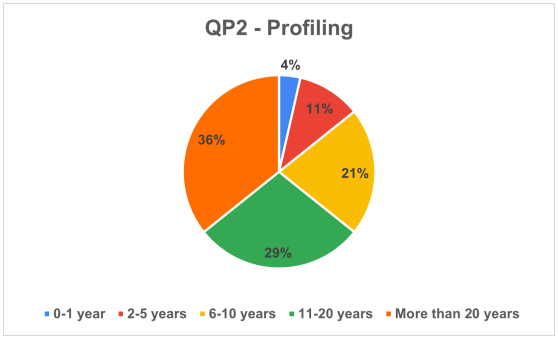
The questionnaire was divided into three sections, each with a specific objective. The first section aimed to understand the respondents' relationship with the topic. The second section aimed to assess whether the data sets presented were relevant for fraud detection. Finally, the third section sought to validate the relevance of the proposed red flags, verifying whether they made sense in the context of fraud detection in PP.

5.1. Profiling

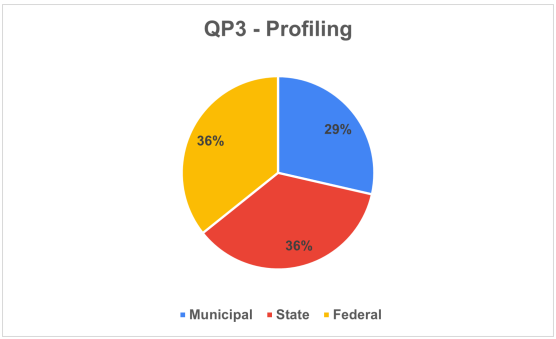
The questionnaire began with questions aimed at understanding the profile of the participants and their experience in the auditing area. The first question (Q1) investigated whether the respondents worked directly in auditing or controlling, allowing us to identify the practical relevance of their answers in the study context. Next, the auditors were asked (Q2) about their sphere of activity — municipal, state, federal, or other —allowing us to map the scope of the feedback collected. The third question (Q3) sought to understand the participants' experience by asking how many years they had been in the auditing area, categorizing them into levels of seniority. Finally (Q4), the questionnaire explored the degree of contact of the auditors with fraud in PP, ensuring that the analysis of red flags considered the perspective of professionals familiar with irregular practices. The results of these questions can be seen in Figure 3, and the translation of the questions and answers can be seen in Table 2. Considering that the questionnaire was addressed to Brazilians, all the answers collected were initially given in Portuguese, so each figure representing one or more answers will be accompanied by a translation table into English.



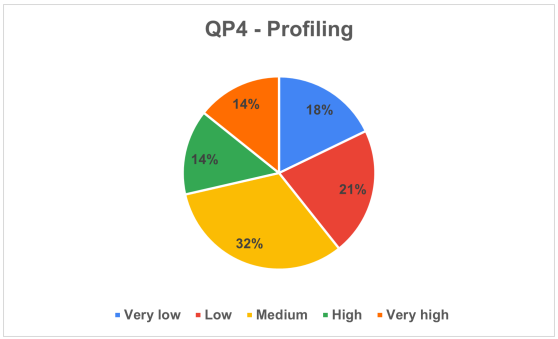
(a) Question Profiling 1



(b) Question Profiling 2



(c) Question Profiling 3



(d) Question Profiling 4

Fig. 3 – Answers to profiling questions

Tab. 2 – Profiling Questions

Question	Translated Question Text	Translated Alternatives
PQ1	Do you work in auditing or controlling?	Yes No
PQ2	In what sphere of activity do you perform your role?	Municipal State Federal
PQ3	How many years of experience do you have in auditing?	0-1 year 2-5 years 6-10 years 11-20 years Over 20 years
PQ4	What is your degree of contact with fraud in public procurement?	Very low Low Medium High Very high

5.2. Datasets

The second section of the questionnaire focused on assessing the relevance of various datasets in the context of fraud detection. Participants were presented with a list of public datasets. For each dataset, respondents were asked whether they considered it relevant for detecting fraud, with an optional follow-up question allowing them to justify their answers. At the end of this section, a final open-ended question invited participants to suggest any additional public datasets they deemed relevant for the study. This structure provided valuable insights into the practical applicability of existing datasets and potential gaps in the resources available for fraud detection. Each question about the relevance of a specific dataset included three response options: "Yes," "No," and "I do not know or prefer not to answer." This structure allowed participants to express their agreement or disagreement with the dataset's applicability while providing an option for uncertainty or reservation, ensuring inclusive and comprehensive feedback.

The results of the subsequent questions in this section are presented in Figure 4, which summarizes the responses for each dataset. The exact wording of the questions and their corresponding response options can be found in Table 3 for reference.

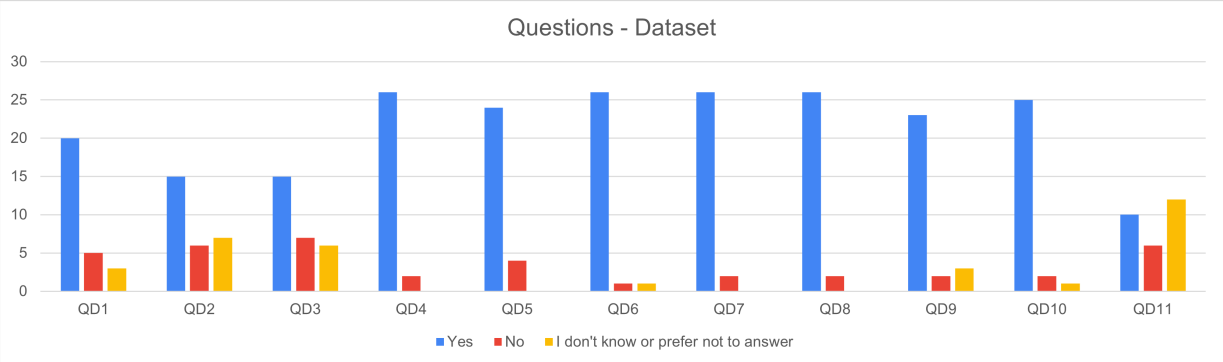


Fig. 4 – Answers to datasets questions

To avoid cluttering the main text while ensuring that all valuable insights are preserved for reference and given the large number of optional justifications provided for the answers in this section, these comments are available in our repository already mentioned in the subsection 4.2.

Tab. 3 – Datasets Questions

Question	Translated Question Text
DQ1	Bolsa Família - Payments
DQ2	TSE
DQ3	CATMAT/CATSER
DQ4	Contracts (GOV BR Contracts)
DQ5	Companies (CNPJ Registration)
DQ6	Unreliable Companies
DQ7	Unreliable and Suspended Companies
DQ8	Punished Companies
DQ9	Suppliers
DQ10	Invoices
DQ11	Safra Guarantee

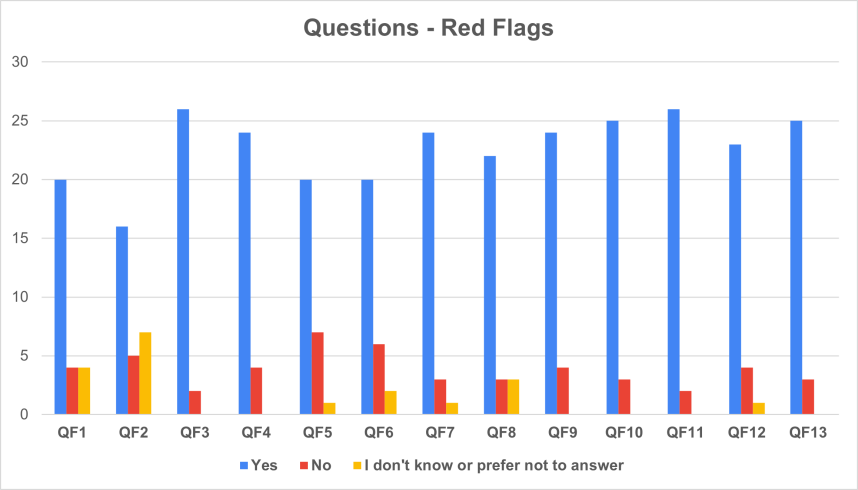
5.3. Red Flags

The third section of the questionnaire was designed to validate the red flags proposed in the study directly. Auditors were asked to assess each red flag's practical relevance in detecting fraud in PP. This step sought to capture insights into the applicability of the indicators in day-to-day audits and identify possible gaps or adjustments needed to improve the indicators presented. By adopting a practical approach, this section also explored auditors' perception of the ability of red flags to reflect patterns and behaviors typical of fraudulent practices, providing a solid basis for contextualizing and refining the metrics used in the study.

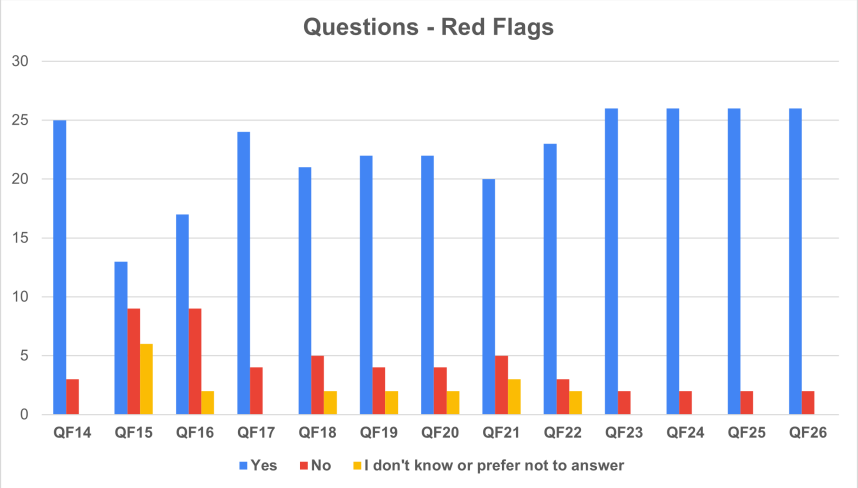
The results of the subsequent questions in this section are presented in Figure 5, which summarizes the responses for each dataset. The exact wording of the questions and their corresponding response options can be found in Table 4 for reference.

Question	Red Flag Indication	Description
FQ1	Bolsa Família Beneficiary Member	Check if any of the company's partner members and/or the person responsible for the supplier has any connection with the Bolsa Família Social Program
FQ2	Garantia Safra Beneficiary Member	Check if any of the company's partner members have any connection with the Garantia Safra Social Program
FQ3	Company Opening Date	Check the difference between the company's opening date and the contract signing date
FQ4	Company Share Capital	Check the difference between the company's share capital and the contract value
FQ5	Political Member/Candidate	Check if any of the partner members is or has been a candidate for any political office
FQ6	Company Tax Exemption	Check if the company has any tax exemption
FQ7	Company Sanctions Against	Check if the company has any sanctions by the Government
FQ8	Active Public Servant Member	Check if any of the partner members is a public servant
FQ9	Company Telephone	Check if the company has the same telephone number as another company(ies)
FQ10	Company Email	Check if the company has the same email as another company(ies)
FQ11	Addresses Related to the Company	Check if the company has its headquarters at the same address as another company(ies)
FQ12	Company Partners	Check if the partners are partners in other companies
FQ13	Contracts With Bidding	Check the percentage of contracts won through bidding
FQ14	Contracts Without Bidding	Check the percentage of contracts won without bidding
FQ15	Contract Value per 100k Inhabitants	Check the percentage that represents the value of the contract divided by 100k inhabitants
FQ16	Contract Value by Agency	Check the percentage that represents the value of the contract in relation to the agency's budget
FQ17	Inactive CNPJ	Check if the company registration number (CNPJ) is inactive.
FQ18	Tax Regime vs. Contract Value	Check if the value of the contacts falls within the tax regime described by the company
FQ19	Companies Donating to Political Campaigns	Check whether the company has donated any amount to political campaigns in recent years
FQ20	Partners Donating to Political Campaigns	Check whether any of the partners' members have donated any amount to political campaigns in recent years
FQ21	Members Participating in Political Parties	Check whether any of the partners are members of any political party
FQ22	Company that provided services to any electoral campaign	Check whether the company provided services to any political campaign
FQ23	Ineligible Company (CGU)	Check the company's suitability before the CGU
FQ24	Ineligible Company (TCU)	Check the company's suitability before the TCU
FQ25	Punished Company (CGU)	Check the company's punishment before the CGU
FQ26	CNAEs associated with the company	Check whether the CNAEs associated with the company match the sale reported in the contract

Tab. 4 – Red Flags Questions



(a) Red Flags Questions 1-13



(b) Red Flags Questions 14-26

Fig. 5 – Answers to red flags questions

5.4. Discussions

The results of the questionnaire applied to Brazilian auditors revealed a high level of agreement with the proposed red flags, consolidating the practical validity of these indicators in detecting fraud in PP. Most participants considered the red flags relevant and applicable, highlighting their usefulness in identifying patterns and behaviors associated with fraudulent practices. This convergence between the proposed indicators and the auditors' experience reinforces the methodological adequacy adopted in the study. It highlights the potential of red flags as practical tools for combating irregularities.

In addition, the auditors provided qualitative insights that complemented the quantitative responses, allowing for the refinement of the red flags and the identification of possible improvements. Specific comments pointed to the need for additional data cross-referencing in some cases, but without compromising the overall reliability of the indicators. This practical evaluation reinforces the relevance of the proposed red flags and demonstrates the importance of incorporating professional perspectives in creating more robust and contextualized fraud detection models.

6. Conclusions

This work presented an approach for detecting potential fraud in PP using red flags constructed from open data available on Brazilian government portals. By categorizing and defining various red flags based on fi-

nancial, administrative, and political factors, we highlighted the relevance of these indicators for identifying suspicious behaviors and anomalies. The structured analysis of public data, combined with risk indicators such as company sanctions, political ties, and inconsistencies in financial capacity, enables the detection of patterns that suggest potential irregularities, thus contributing to more transparent and accountable procurement processes.

An important aspect of this research was the evaluation of the proposed red flags through structured feedback from experienced auditors. This process provided valuable insights into each indicator's practical applicability and relevance in real-world audits. The auditors' expertise was instrumental in refining the red flags to reflect the nuances of PP in Brazil better, ensuring that the set of indicators aligns with the operational challenges faced in the field. This collaboration enhanced the credibility of the proposed framework and highlighted areas for improvement, contributing to a more robust and context-sensitive tool for fraud detection.

However, the work also emphasizes that the effectiveness of red flags is limited when constrained to open data sources. Many complex fraud schemes involve confidential information, such as banking transactions and tax records, which are not publicly accessible. Additionally, one of the key limitations of this study is its focus exclusively on the Brazilian context. While the findings and framework contribute significantly to addressing procurement fraud in Brazil, the applicability and performance of these red flags in other countries or procurement environments remain unexplored.

As a result, while red flags provide a valuable layer of oversight, a more comprehensive approach, including secure data-sharing mechanisms and collaboration with regulatory bodies, is essential to combat sophisticated fraud schemes. Future research should explore integrating advanced data sources and methodologies to enhance detection capabilities further and ensure the robustness of fraud prevention frameworks in PP. Furthermore, expanding this study to international contexts could provide valuable insights into how procurement fraud manifests across different legal and administrative systems, ultimately refining and broadening the applicability of the proposed approach.

References

Brandão, M. A., Reis, A. P., Mendes, B. M., De Almeida, C. A. B., Oliveira, G. P., Hott, H., Gomide, L. D., Costa, L. L., Silva, M. O., Lacerda, A., et al. (2024). Plus: A semi-automated pipeline for fraud detection in public bids. *Digital Government: Research and Practice*, *5*(1), 1–16.

Decarolis, F., & Giorgiantonio, C. (2022). Corruption red flags in public procurement: New evidence from italian calls for tenders. *EPJ Data Science*, *11*(1), 16.

Ferguson, G. (2018). *Global corruption: Law, theory & practice* [3rd ed. Available at https://canlii.ca/t/27td]. University of Victoria.

Kawashita, I., Baptista, A. A., Soares, D., & Andrade, M. (2024). Open government data use: The brazilian states and federal district cases. *Plos one*, *19*(3), e0298157.

Lourenço, R. P. (2015). An analysis of open government portals: A perspective of transparency for accountability. *Government information quarterly*, *32*(3), 323–332.

Lyra, M. S., Damásio, B., Pinheiro, F. L., & Bacao, F. (2022). Fraud, corruption, and collusion in public procurement activities, a systematic literature review on data-driven methods. *Applied Network Science*, *7*(1), 83.

Mahmood, S. A. I. (2010). Public procurement and corruption in bangladesh confronting the challenges and opportunities. *Journal of public administration and policy research*, *2*(6), 103.

Menezes, A. P. V. C. (2021). Inteligência artificial para identificação de indícios de fraude e corrupção em compras públicas no tcu.

Modrušan, N., Rabuzin, K., & Mršic, L. (2021). Review of public procurement fraud detection techniques powered by emerging technologies. *International Journal of Advanced Computer Science and Applications*, *12*(2), 1–9.

Munteanu, V., Zuca, M.-R., Horaicu, A., Florea, L.-A., Poenaru, C.-E., & Anghel, G. (2024). Auditing the risk of financial fraud using the red flags technique. *Applied Sciences*, *14*(2), 757.

Nai, R., Sulis, E., Meo, R., et al. (2022). Public procurement fraud detection and artificial intelligence techniques: A literature review. *Companion Proceedings of the 23rd International Conference on Knowledge Engineering and Knowledge Management*, 1–13.

Niessen, M. E. K., Paciello, J. M., & Fernandez, J. I. P. (2020). Anomaly detection in public procurements using the open contracting data standard. *2020 Seventh International Conference on eDemocracy & eGovernment (ICEDEG)*, 127–134.

Potin, L., Figueiredo, R., Labatut, V., & Largeron, C. (2023). Pattern mining for anomaly detection in graphs: Application to fraud in public procurement. *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*, 69–87.

Potin, L., Labatut, V., Morand, P.-H., & Largeron, C. (2023). Foppa: An open database of french public procurement award notices from 2010–2020. *Scientific Data*, *10*(1), 303.

Salm Junior, J. F., Araujo, V. M. d., Schommer, P. C., & Raupp, F. M. (2024). Padrão de dados abertos em compras e contratações públicas: Uma construção colaborativa. *Revista de Administração Contemporânea*, *28*, e230175.

Sampaio, I., Fontes, S., Andrade, E., Bernardini, F., & Viterbo, J. (2024). Enhancing transparency through explainable artificial intelligence: An exploratory analysis on collusion and corruption scenario in digital government. *Proceedings of the 25th Annual International Conference on Digital Government Research*, 52–62.

Sanda, M.-R., Siminică, M.-I., Avram, C.-D., & Popescu, L. (2024). Ghosts in the machine: How big data analytics can be used to strengthen online public procurement accountability. *Sustainability*, *16*(9), 3698.

Santos, F. B., & de Souza, K. R. (2016). *Como combater a corrupção em licitações*. Editora Forum.

Velasco, R. B., Carpanese, I., Interian, R., Paulo Neto, O. C., & Ribeiro, C. C. (2021). A decision support system for fraud detection in public procurement. *International Transactions in Operational Research*, *28*(1), 27–47.