SEAP Fraud Catcher

Bojescu Bianca-Daniela, Constantin Ana-Maria, Gălățianu Mihai,
Munteanu Denisa-Maria-Ionela, Țiplea Matei

January 16, 2025

Abstract

The SEAP Fraud Detection Application is designed to identify and address fraudulent activities within Romania's public procurement platform, SEAP (e-licitatie.ro). By employing artificial intelligence and clustering algorithms, the system detects irregularities such as inflated prices, collusive bidding practices, and repeated contract awards. It integrates a Django-based backend and a Vue.js-powered Chrome extension for seamless data collection, analysis, and user interaction. This hybrid approach ensures scalability, adaptability, and user accessibility while enhancing the transparency and efficiency of public procurement processes. Through customized methods and domain-specific adaptations, the SEAP Fraud Catcher aims to reduce fraud and establish a benchmark for accountable governance.

1 Introduction

1.1 Background

Public procurement is a cornerstone of government operations, enabling the acquisition of goods, services, and works necessary for public functions. In Romania, the Sistemul Electronic de Achiziții Publice (SEAP), through its platform *e-licitatie.ro*, is the designated online space for public procurement activities. The platform fosters transparency and competition by enabling public institutions to announce procurement needs and inviting private entities to submit bids.

However, despite these safeguards, SEAP is vulnerable to fraudulent practices. Public procurement fraud disrupts market fairness, wastes taxpayer resources, and undermines public trust in governmental operations. It takes various forms, including bid rigging, overpricing, and collusion between bidders and officials [1, 2]. These fraudulent activities inflate costs and divert resources from their intended purposes, negatively affecting the quality and delivery of public services [2].

The digitization of public procurement has increased the availability of data, creating opportunities for leveraging advanced techniques such as artificial intelligence (AI) and machine learning to detect fraud and anomalies [1, 2]. These technologies have demonstrated significant potential in identifying collusive bidding, conflicts of interest, and other suspicious patterns [1].

1.2 Problem Statement

Fraud within the SEAP platform persists as a significant obstacle to transparent and equitable procurement processes. Common issues include:

- Inflated contract prices far exceeding market value [1, 2].
- Bid rigging, characterized by predetermined winners or manipulation of competition [1, 2].
- Repeated awarding of contracts to the same companies without justification [2].

These challenges undermine the principles of fairness and accountability in public procurement. Furthermore, the manual oversight mechanisms currently in place are resource-intensive and often insufficient to identify complex fraud patterns in large datasets [1, 2].

1.3 Objectives

The SEAP Fraud Detection Application aims to address these challenges by developing an advanced, web-based fraud detection system. The primary objectives are:

1. **Detection of Fraudulent Activities**: Automate the identification of irregular procurement practices using AI algorithms. The system will flag anomalies, including overpricing, bid rigging, and repeated awards [2].

2. **Public Accessibility**: Ensure transparency by making fraud detection results publicly accessible without user account requirements. A user-friendly interface will allow stakeholders, including government officials and the public, to explore flagged cases effortlessly [1].

3. **Real-Time Data Analysis**: Provide near-instantaneous updates and insights by continuously monitoring procurement data from the SEAP platform [2].

4. **Enhancing Transparency and Efficiency**: Strengthen trust in public procurement by offering a robust tool to highlight and deter fraudulent behaviors, thus optimizing resource allocation and ensuring compliance with ethical standards [1, 2].

By achieving these objectives, the SEAP Fraud Detection Application not only aims to mitigate fraud but also establishes a benchmark for transparency and accountability in public procurement systems.

2 State-of-the-Art

2.1 Overview of Existing Solutions

Public procurement fraud detection has received significant attention because of its critical role in ensuring transparency and efficiency in government spending. Various technological solutions and methodologies have been developed to address fraud, each leveraging advancements in data analysis, artificial intelligence, and machine learning. Below are some prominent examples of existing solutions that have shaped the field:

ARACHNE Risk Scoring Tool:

ARACHNE is an integrated IT tool developed by the European Commission to support managing authorities in the European Structural and Investment Funds (ESIF) in detecting and preventing fraud.[3] Its primary objective is to enhance the efficiency and effectiveness of management verifications and controls by providing risk-based insights into procurement projects. Key features include:

- **Risk Scoring:** ARACHNE calculates risk scores for projects, beneficiaries, contracts, and contractors using predefined indicators. These scores enable managing authorities to focus their oversight efforts on high-risk areas, improving resource allocation.

- **Data Integration:** The tool combines internal ESIF data with external data sources, such as financial and business registries, to provide a holistic view of potential fraud risks.

- **User Interface:** A user-friendly interface enables users to interact with risk scores and explore underlying data, facilitating decision-making and reporting.

ARACHNE represents a proactive approach to fraud detection by focusing on early identification of anomalies, which can help prevent financial mismanagement and fraud in cohesion policy funds. However, its effectiveness depends on the availability and quality of integrated data sources.

Fraud Detection System for Public Procurement (FDPP):

The FDPP is an advanced system developed under a European initiative to address irregularities in public tenders. This system employs state-of-the-art machine learning algorithms to analyze procurement data and identify potential fraud patterns. Key features include:

- **Anomaly Detection:** The FDPP identifies deviations in bid pricing from expected market values, which may indicate inflated prices or collusive behavior.
- **Pattern Recognition:** The system employs machine learning models to uncover repetitive award patterns, irregular bidding behaviors, or other anomalies that suggest favoritism or collusion.
- **Graphical Dashboards:** To make results accessible to non-technical users, FDPP offers visual dashboards that present anomalies and trends in an intuitive format. These dashboards allow users to quickly identify suspicious activities and dig deeper into the underlying data when needed.

FDPP has been instrumental in raising awareness about procurement fraud risks across multiple European countries. Its strengths lie in its scalability and the use of advanced algorithms for fraud detection. However, its generalized design may reduce its effectiveness when applied to region-specific regulatory environments.

Fraud.net:

Fraud.net[4] is a cloud-based fraud detection platform designed for use across various industries, including public procurement. Its capabilities are built on a foundation of AI and real-time analytics, offering comprehensive fraud management solutions. Distinctive features include:

- **Entity Screening & Monitoring:** Fraud.net continuously evaluates businesses and individuals, identifying potential risks during onboarding and throughout their lifecycle. This functionality is especially useful for ensuring compliance with Anti-Money Laundering (AML) and Know Your Customer (KYC) regulations.
- **AI-Powered Rules Engine:** The platform allows organizations to customize fraud detection rules to align with their unique requirements, enabling tailored risk assessments.
- **Real-Time Transaction Monitoring:** Fraud.net provides near-instantaneous alerts for suspicious transactions, helping organizations respond promptly to potential threats.
- **Collective Intelligence Network:** By aggregating data from a global network of sources, the platform enhances its ability to detect known fraud patterns, offering additional protection against fraudulent activities.

Fraud.net stands out for its scalability and adaptability, making it suitable for organizations of varying sizes. However, its focus on broader fraud management across industries means it lacks procurement-specific functionality, such as detecting bid rigging or collusion.

2.2 Limitations and Challenges

Despite advancements in fraud detection technologies, existing solutions for public procurement fraud detection face several limitations and challenges:

- **Lack of Regional Specificity:** Many fraud detection systems, such as FDPP, are designed for multi-country use and prioritize generalizability over region-specific needs. This limits their effectiveness when applied to local regulatory environments like Romania's SEAP platform, which has unique procurement laws and practices. The absence of customization for local languages, formats, and market standards can hinder the detection of fraud patterns.
- **Data Quality and Accessibility:** The reliability of fraud detection systems depends heavily on the availability of high-quality, structured, and comprehensive datasets. However, public procurement data is often fragmented, unstandardized, and incomplete. Missing data points, such as unrecorded bids or vague procurement descriptions, can reduce the accuracy of predictive models and anomaly detection algorithms.

- **Dynamic Nature of Fraud Tactics:** Fraudsters continually adapt their methods to exploit weaknesses in detection systems. Current models often lag behind emerging fraud schemes, making it difficult to predict and prevent new types of fraudulent activities. For instance, systems may struggle to identify subtle collusion patterns or hidden relationships between bidders without constant updates and retraining of models.

- **Black-Box Nature of AI Models:** Many AI-powered fraud detection systems operate as ”black boxes,” producing outputs without clear explanations for flagged anomalies. This lack of transparency makes it difficult for users, particularly non-technical stakeholders, to interpret the results or trust the system’s decisions. Explainability remains a critical challenge for machine learning models in fraud detection.

- **Resource-Intensive Implementations:** Advanced systems like Fraud.net often require substantial financial and technical investments for deployment and maintenance. Smaller organizations or regions with limited budgets may find these solutions cost-prohibitive. Additionally, the integration of such tools with existing procurement platforms can be complex and time-consuming.

- **Evolving Legal and Compliance Requirements:** Public procurement operates within a framework of evolving regulations. Fraud detection systems must continuously adapt to new laws, policies, and standards to remain effective. This necessitates frequent updates to rule-based systems, data pipelines, and compliance modules, adding to the system’s operational complexity.

- **Ethical and Privacy Concerns:** The use of AI and data analytics in fraud detection raises ethical considerations, particularly regarding data privacy and security. Ensuring that sensitive procurement data is handled responsibly, with safeguards against misuse, is essential to maintaining trust in the system.

By addressing these limitations, the SEAP Fraud Catcher project aims to deliver a more adaptive and robust solution, capable of overcoming these challenges while catering specifically to the needs of Romania’s public procurement system.

3 Proposed Solution

3.1 Concept and Methodology

The proposed solution for detecting procurement fraud on the SEAP platform is designed to identify irregularities and anomalies in public procurement data. The concept leverages clustering algorithms and fraud detection techniques to group procurement items and evaluate their likelihood of being fraudulent based on specific patterns.

The proposed solution for detecting procurement fraud on the SEAP platform involves a structured methodology consisting of several key steps. First, data retrieval is performed, where public procurement data is collected from the SEAP platform (e-licitatie.ro). This data includes essential information such as tender names, contract values, company names, and product descriptions. Next, during the data preprocessing phase, the retrieved data is cleaned and standardized to ensure consistency and accuracy for subsequent analysis. This step enhances the reliability of the insights derived from the data.

The third step, clustering, involves grouping procurement items based on their attributes, such as price, quantity, and description similarity, using clustering algorithms. This process identifies patterns and helps detect potential outliers that may indicate irregularities. Following clustering, a fraud scoring mechanism is applied. For each item, a fraud score is calculated to quantify its deviation from expected norms. This score acts as a measure to flag transactions that display unusual or suspicious characteristics.

Finally, the results undergo interpretation, where flagged items are presented for further analysis. This allows users to focus their attention on transactions with the highest likelihood of fraud, streamlining the identification process and enhancing transparency in public procurement practices.

This approach combines statistical and analytical techniques to automate the detection of anomalies in public procurement data. By focusing on data-driven patterns, the system enhances transparency and efficiency in identifying potential fraudulent activities.

3.2 Implementation Details

The C4 Level 2 Diagram below illustrates the architecture of the SEAP Fraud Detection Application. This diagram highlights the interactions between the system's components, including the Django API, Chrome Extension, and the backend database, along with how data flows through the system to support the fraud detection process.

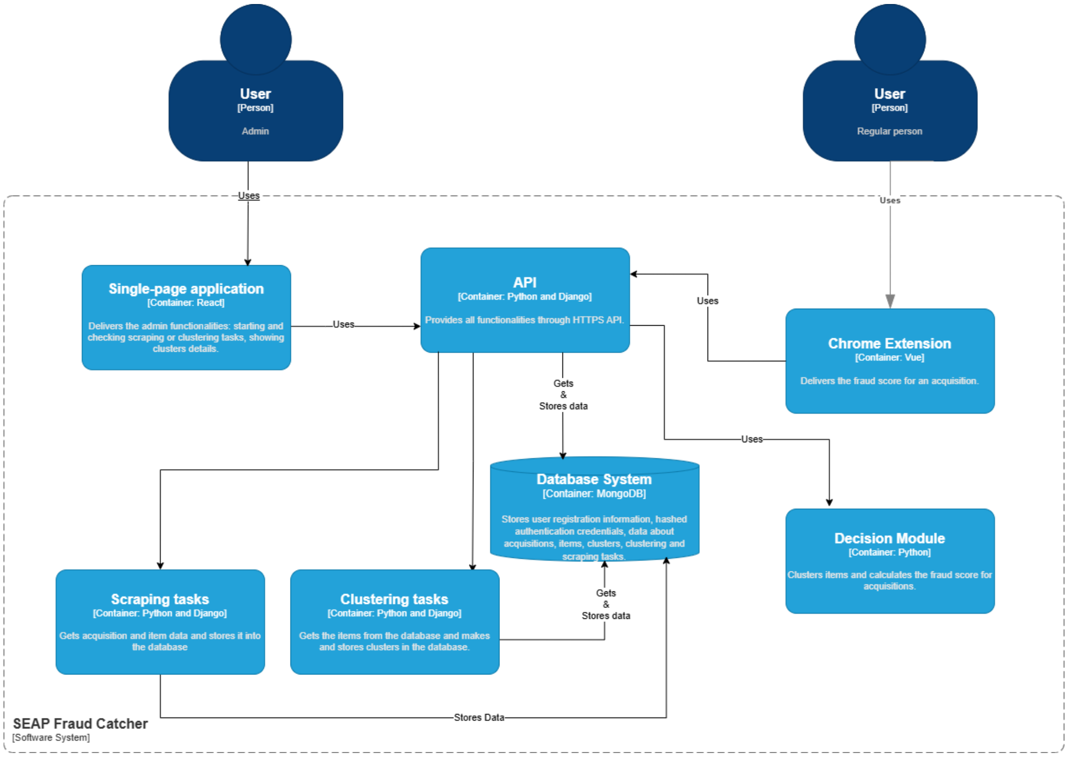

User
[Person]
Admin
User
[Person]
Regular person
Uses
Uses
Single-page application
[Container: React]
Delivers the admin functionalities: starting and checking scraping or clustering tasks, showing clusters details.
Uses
API
[Container: Python and Django]
Provides all functionalities through HTTPS API.
Uses
Chrome Extension
[Container: Vue]
Delivers the fraud score for an acquisition.
Gets & Stores data
Uses
Database System
[Container: MongoDB]
Stores user registration information, hashed authentication credentials, data about acquisitions, items, clusters, clustering and scraping tasks.
Decision Module
[Container: Python]
Clusters items and calculates the fraud score for acquisitions.
Scraping tasks
[Container: Python and Django]
Gets acquisition and item data and stores it into the database
Clustering tasks
[Container: Python and Django]
Gets the items from the database and makes and stores clusters in the database.
Gets & Stores data
Stores Data
SEAP Fraud Catcher
[Software System]

Figure 1: C4 Level 2 Diagram of the SEAP Fraud Detection Application architecture. This diagram shows the relationships and communication flows between the key components of the system.

The diagram emphasizes the modular design of the system, with the following key components:

- **Django API:** Acts as the central hub, managing data collection, clustering tasks, and fraud score computation. It connects the backend database and handles requests from the Chrome Extension.
- **Chrome Extension:** Serves as the user-facing interface, allowing users to interact with the fraud detection system in real time. It communicates with the Django API to retrieve fraud scores and display results.
- **Backend Database:** Stores procurement data, pre-computed clusters, and fraud scores to support efficient analysis and real-time responses.
- **Scraping Tasks:** Responsible for periodically collecting data from the SEAP platform and updating the backend database without interrupting user interactions.
- **Clustering Tasks:** Operated by admins, these tasks process procurement data into clusters for fraud detection, ensuring updated and accurate fraud scores.

This architecture supports the seamless operation of the SEAP Fraud Detection Application, enabling accurate and efficient fraud detection while maintaining a user-friendly interface.

3.2.1 Django API

The Django API serves as the backbone of the SEAP Fraud Detection Application, enabling the seamless integration of data collection, processing, and communication between the front-end and backend systems. Its modular design ensures that each component handles a specific task, contributing to the overall functionality of the application.

The **scraping tasks** module is responsible for retrieving procurement data from the SEAP platform (e-licitatie.ro) and preparing it for analysis. This data is then processed by the **clustering tasks** module, which applies advanced clustering algorithms to group similar procurement items and detect anomalies. The clustered data is passed to the **decision module**, where fraud scores are calculated based on predefined metrics, highlighting potentially suspicious activities for further review.

To ensure secure access to the system, the API incorporates the **custom authentication** module, which manages user authentication and authorization. Additionally, the **aspects** module handles auxiliary tasks such as error logging, performance monitoring, and debugging, ensuring the stability and reliability of the API. All functionalities are exposed through the **API module**, which defines RESTful endpoints that allow users to interact with the system. These endpoints facilitate operations such as retrieving procurement data, analyzing flagged items, and querying fraud detection results.

The API is built on Django's scalable and maintainable architecture, with data stored securely in a database. To ensure robustness, the **tests** module performs rigorous unit and integration testing on all components. Together, these modules create a comprehensive and efficient system for detecting procurement fraud, enhancing transparency and accountability in public acquisition processes.

Artificial intelligence tools such as ChatGPT and Claude played a significant role in the development of the Django API by providing valuable insights and assistance throughout the implementation process. These tools were instrumental in helping the development team familiarize themselves with Django and adopt best practices for building scalable and maintainable applications. AI also assisted in configuring complex settings and creating tasks for slower methods, such as optimizing scraping workflows and designing efficient mechanisms to make and add clusters to the database. Furthermore, AI tools contributed to the testing phase by offering guidance on writing comprehensive test cases, ensuring the reliability and robustness of the API. By leveraging AI, the development process was accelerated, and the overall quality of the module was enhanced.

Challenges

During the development of the Django API, several challenges were encountered, primarily related to data collection and database updates. One significant challenge was determining how to update the database without disrupting ongoing API interactions. To address this, a dedicated scraping task was implemented, which could be initiated by an admin. This task was designed to run separately from the API's primary functions, ensuring that user interactions with the API remained uninterrupted while the database was being updated.

Another notable challenge was the inconsistent performance of the SEAP public API during peak working hours (approximately 12:00 PM to 6:00 PM). During these times, the SEAP API experienced significant slowdowns, likely due to increased usage by other users or systems. For instance, a scraping task that took 49 seconds to complete in the evening could take up to 7 minutes when run in the middle of the day. To mitigate this issue, scraping tasks were scheduled to run in the evening when the SEAP public API was significantly faster. This scheduling optimization ensured efficient data retrieval while minimizing delays in database updates.

3.2.2 Clustering Method

The clustering method in the SEAP Fraud Detection Application employs a hybrid approach, integrating multiple clustering strategies to identify patterns and anomalies in procurement data. This approach ensures robust and flexible analysis, adapting to the varying characteristics of datasets and providing meaningful insights into potential fraudulent activities.

The methodology begins with the computation of a distance matrix, where each element represents the absolute difference in procurement attributes, such as prices or quantities. This matrix forms the foundation for various clustering strategies, each tailored to specific use cases. The primary strategies employed include Agglomerative Clustering, K-Means, K-Means++ initialization, and OPTICS. These algorithms are orchestrated through a hybrid framework, allowing the system to dynamically select the most suitable strategy based on the data characteristics.

Agglomerative Clustering groups items hierarchically, using linkage criteria to form clusters with similar characteristics. K-Means and K-Means++ provide efficient clustering, with the latter leveraging optimized initialization to improve convergence and accuracy. OPTICS (Ordering Points To Identify the Clustering Structure) is employed for its ability to identify clusters with varying densities and detect outliers effectively. These strategies are further complemented by the HybridClustering module, which combines clustering results and refines them by reclustering subgroups to enhance precision.

The system handles small datasets differently, bypassing traditional clustering and instead calculating a fraud score directly based on deviations from the mean price. For larger datasets, clusters are analyzed to identify the largest group, and the average price within this cluster serves as the benchmark. Each procurement item's fraud score is then computed as a percentage deviation from this benchmark, highlighting anomalies.

This hybrid clustering method is implemented within the `FraudDetectionClustering` and the `HybridClustering` classes, ensuring a modular and extensible design. By combining multiple algorithms, the methodology offers a comprehensive and adaptive framework for fraud detection. Suspicious procurements are flagged based on their fraud scores, enabling targeted investigation and enhancing transparency in public procurement processes.

Artificial intelligence tools were integral in optimizing the development of the clustering method. They provided valuable recommendations for selecting clustering algorithms and improving their performance, ensuring the most suitable techniques were applied to the specific requirements of the SEAP Fraud Detection Application. These tools also assisted in personalizing the algorithms to address the unique challenges of clustering procurement data, such as mobile phone tenders, where nuanced differences in specifications and pricing required tailored solutions. Furthermore, AI was used to incorporate best practices in designing and implementing the hybrid framework, facilitating a robust, modular, and extensible architecture for clustering tasks. By leveraging AI, the development process was enhanced, resulting in a more effective and adaptable clustering methodology.

Challenges

The development and implementation of the clustering method presented several challenges, particularly related to efficiency and customization. Initially, the fraud score calculation required creating clusters dynamically and then searching for the specific item within those clusters. However, this process was time-intensive, often taking more than 30 minutes, which made it impractical for real-time applications. To address this, we opted to create clusters beforehand and store them in the database. While this reduced the time needed for fraud score calculation, generating the clusters and storing them in the database took approximately 40 minutes. To ensure this process did not disrupt ongoing API requests, clustering tasks were designed to be operated by admins, running separately from the API's primary functionalities.

Another significant challenge was the diversity of procurement items, which made clustering difficult as each type of item required a tailored approach to determine relevant attributes. Given time constraints, we decided to focus on a specific category—mobile phones—and specialized the clustering methodology to cater to this category. By refining the clustering process for mobile phones, we were able to achieve more precise and reliable results, ensuring the system's effectiveness in detecting anomalies within this domain.

3.2.3 Chrome Extension

The Chrome Extension of the SEAP Fraud Detection Application acts as the primary interface for user interaction, enabling quick and efficient verification of whether a specific acquisition on the SEAP platform (e-licitatie.ro) is potentially fraudulent. Developed with Vue.js and Pinia, the extension provides an intuitive user experience, seamlessly connecting to the backend API for fraud detection. It detects whether the user is on a valid SEAP acquisition page through URL validation using regular expressions. If the user is not on an appropriate page, an informative message is displayed, guiding them to navigate to the correct page. The extension manages acquisition data, such as bidder name, contracting authority, description, and estimated value, by utilizing Chrome's local storage. This data is updated in real time whenever new information is received, ensuring the interface reflects the latest acquisition details.

The primary functionality of the extension is the fraud detection process, which is initiated by the user through a button in the interface. The extension sends a request to the Django API to retrieve the fraud score for a specific acquisition. The fraud score, along with additional insights, is then displayed visually through a dynamic FraudMeter component, allowing users to quickly assess the risk of fraud. The extension also provides real-time feedback through success or error messages, handled by the Vue Toastification library, enhancing the overall user experience. Application states, such as whether fraud detection is active or if the interface is loading, are managed with Pinia stores to ensure smooth transitions and responsiveness.

The Chrome Extension plays a crucial role in the SEAP Fraud Detection Application ecosystem by serving as a bridge between the user interface and the backend data analysis. It allows users to interact with the system in a fast and accessible manner, significantly reducing the time needed to identify potential fraud and enhancing the transparency and efficiency of public procurement processes.

Artificial intelligence tools were instrumental in the development of the Chrome Extension, particularly in improving its styling and ensuring seamless integration with the Django API. These tools provided recommendations for implementing modern design practices, enhancing the extension's visual appeal and user experience. Additionally, AI offered guidance on effectively structuring API requests and handling responses, which streamlined the integration process between the extension and the backend system. By leveraging AI, the team was able to expedite the development process, create a more intuitive and aesthetically pleasing interface, and ensure robust communication with the backend for accurate and timely fraud detection.

Challenges

The development of the Chrome Extension posed several challenges, particularly related to ensuring seamless functionality and efficient interaction with the backend API. One significant challenge was managing real-time communication between the extension and the Django API, especially when handling large volumes of requests or potential latency issues. This required careful optimization of API calls and response handling to ensure the extension remained responsive during fraud score calculations. Another challenge was implementing URL validation for SEAP pages, as the structure of URLs on the platform was not entirely uniform, necessitating a flexible yet reliable approach using regular expressions.

Additionally, the need to store and update procurement data in real-time using Chrome's local storage introduced complexities in data synchronization. Ensuring that the displayed information reflected the latest acquisition details required meticulous handling of state updates and transitions within the extension, especially when multiple users interacted with the same data concurrently. Styling the extension to create a user-friendly interface that remained consistent across different screen sizes and browser environments added further complexity, as did debugging and testing the extension to account for diverse user behaviors and edge cases.

Despite these challenges, the use of AI tools significantly expedited problem-solving during development, helping the team address issues like efficient API integration and modern design practices. These efforts ensured that the Chrome Extension delivered a smooth and intuitive user experience while effectively supporting the system's fraud detection goals.

4 Results and Evaluation

To evaluate the performance of the proposed fraud detection solution, several key aspects of the system's efficiency were measured, focusing on task durations and data handling across different stages of the pipeline. The performance tests were conducted using data extracted from the SEAP platform over varying time periods, including one month, six months, and a full year. These tests provide insights into how the system scales and performs as the dataset size increases.

A significant optimization to improve performance has been the addition of a clusters object in the database. This object stores pre-grouped procurement items, which makes it more efficient to assign new items to existing clusters during fraud scoring. By organizing items into clusters, the system can quickly evaluate the likelihood of fraud for a given procurement item by calculating a fraud score based on its relationship to the nearest cluster. This reduces the time needed to process individual items, especially those that are already categorized, and improves the overall responsiveness of the application.

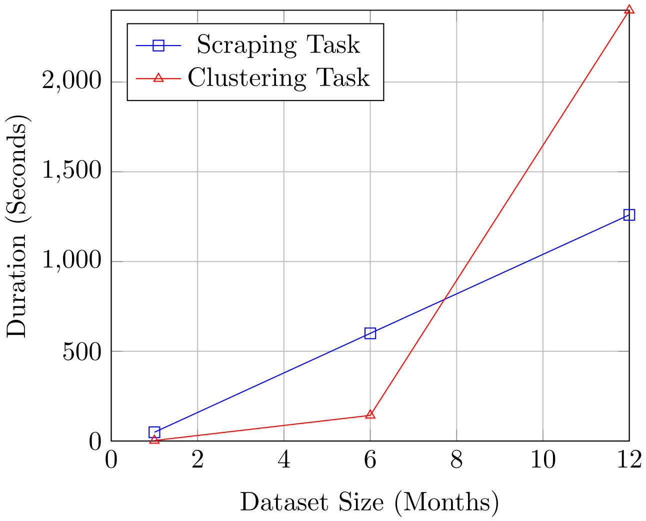
4.1 Results

To assess the system's scalability and efficiency, performance measurements were taken for the following key tasks:

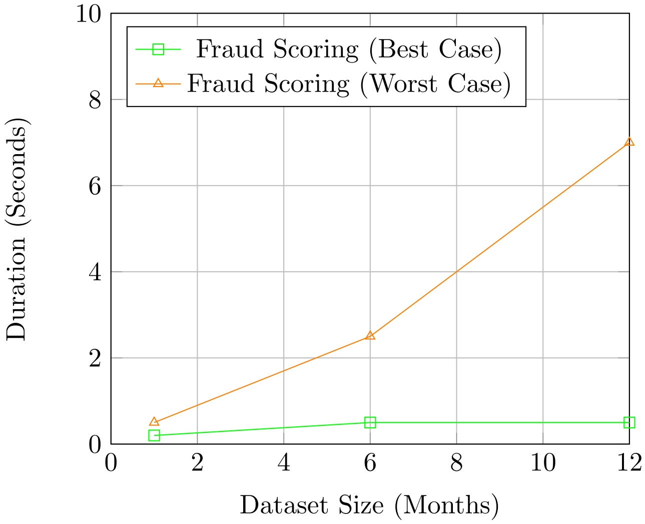
- **Scraping Task Duration:** The time it takes to scrape procurement data from the SEAP platform.
- **Clustering Task Duration:** The time required to cluster procurement items, depending on the number of relevant procurements.
- **Fraud Scoring Task Duration:** The time it takes to calculate fraud scores for items, specifically:
  - For items already grouped within clusters.
  - For items that do not belong to any existing cluster.

The next section will present the results of these performance tests, visualized through diagrams that illustrate the system's efficiency with different dataset sizes and task types.

Performance Comparison: Scraping vs. Clustering Tasks



Performance Analysis of Fraud Scoring: Best and Worst Case Scenarios



4.2 Evaluation

To evaluate the system's ability to detect anomalies, fraud scores were generated for various procurement items, and the results were analyzed to assess their accuracy in identifying irregularities.

- **Strengths of the System:**
  - **Accurate Detection of Anomalously Low Prices and High Prices:** The system effectively identifies items with prices that deviate significantly from the expected range, whether they are unusually low or excessively high. For items priced significantly lower than the norm, the system assigns appropriate fraud scores, flagging them as potential anomalies. Similarly, for items with unusually high prices, the fraud score reflects their deviation, helping to identify cases where overpricing may indicate fraudulent activity.
  - **Cluster-Based Fraud Scoring for Improved Reliability:** The clustering mechanism groups procurement items based on their prices, and fraud scoring is calculated using the cluster with the largest number of items. This approach reduces the impact of outliers and enhances the accuracy of the scoring mechanism by focusing on the most representative data points.
  - **Handling of Large Datasets:** The system maintains strong performance across datasets of varying sizes, demonstrating scalability and consistency in detecting irregularities.
- **Identified Limitations and Areas for Improvement:**
  - **Lack of Comparable Data:** In cases where fraudulent procurements (e.g., items with excessively high prices) do not have comparable data within the dataset, the system assigns a fraud score of 0. This is a limitation of the current comparison-based mechanism.
  - **Incomplete or Inconsistent Data:** Errors such as incomplete fields (e.g., missing quantities or incorrect unit descriptions) further impact the fraud detection algorithm. For instance, the absence of critical fields can skew the calculation of fraud scores, leading to false positives or negatives.

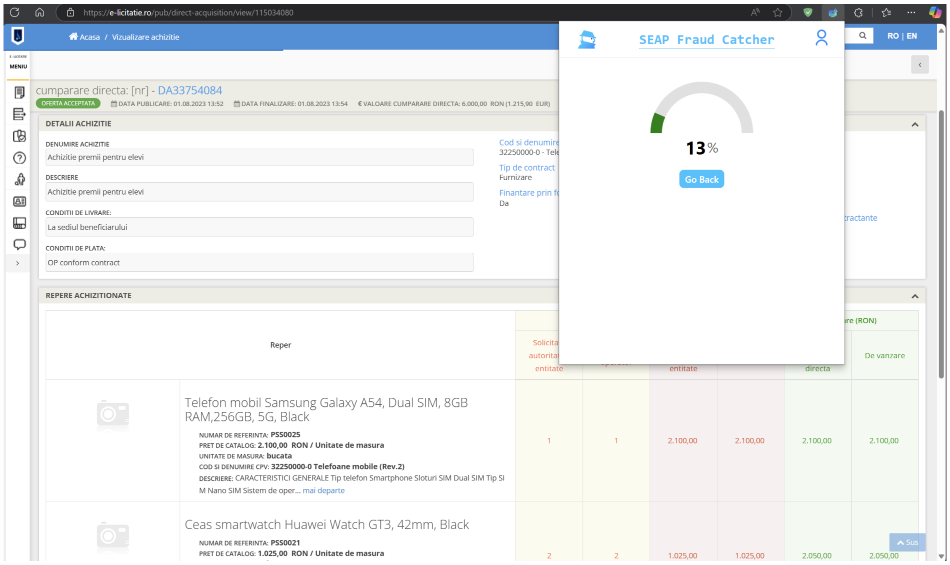
While the system performs well in detecting price-related anomalies and leveraging clustering to reduce the impact of outliers, challenges related to human errors in data entry must be addressed. By implementing data validation, quality metrics, and enhanced clustering techniques, the solution can further improve its robustness and reliability in detecting fraud in public procurement data.

Examples

To demonstrate the effectiveness of the system in detecting potential fraudulent procurements, we present three examples with varying fraud scores: low, medium, and high. Each example includes a screenshot of the procurement details along with the calculated fraud score.

Examples

To illustrate the system's ability to detect anomalies and its clustering methodology, we present examples of procurement items with varying fraud scores: low, medium, and high. Each example includes a screenshot of the procurement item followed by a screenshot of the cluster containing that item.



https://e-licitatie.ro/pub/direct-acquisition/view/115034080

cumparare directa: [nr] - DA33754084

OFERTA ACCEPTATA DATA PUBLICARE: 01.08.2023 13:52 DATA FINALIZARE: 01.08.2023 13:54 € VALOARE CUMPARARE DIRECTA: 6.000,00 RON (1.215,90 EUR)

DETALII ACHIZITIE

DENUMIRE ACHIZITIE: Achizitie premii pentru elevi

DESCRIERE: Achizitie premii pentru elevi

CONDITII DE LIVRARE: La sediul beneficiarului

CONDITII DE PLATA: OP conform contract

Cod si denumire: 32250000-0 - Tele

Tip de contract: Furnizare

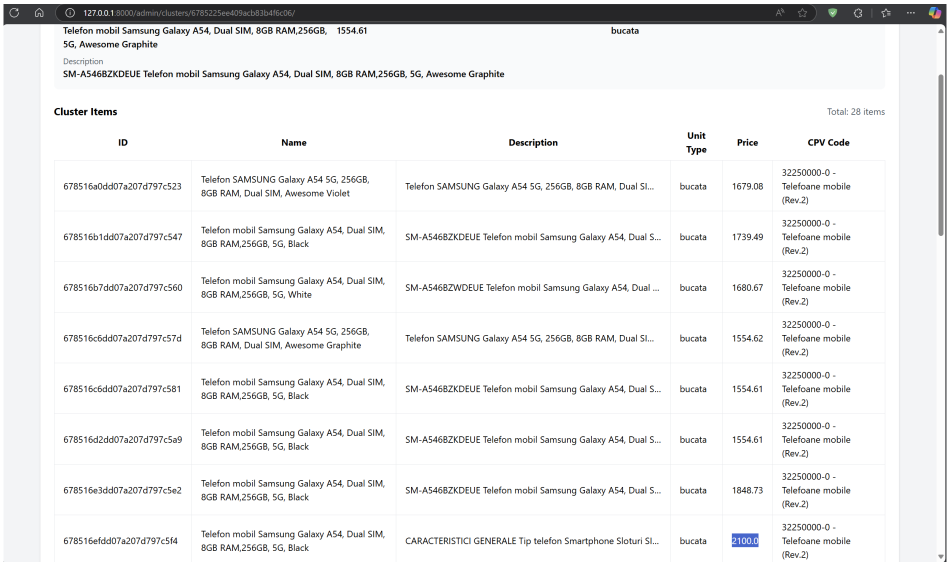
Finantare prin f: Da

tractante

REPERE ACHIZITIONATE

Reper	Solicita autorita entitate		entitate		directa	De vanzare
Telefon mobil Samsung Galaxy A54, Dual SIM, 8GB RAM,256GB, 5G, Black NUMAR DE REFERINTA: PSS0025 PRET DE CATALOG: 2.100,00 RON / Unitate de masura UNITATE DE MASURA: bucata COD SI DENUMIRE CPV: 32250000-0 Telefoane mobile (Rev.2) DESCRIERE: CARACTERISTICI GENERALE Tip telefon Smartphone Sloturi SIM Dual SIM Tip SIM Nano SIM Sistem de oper... mai departe	1	1	2.100,00	2.100,00	2.100,00	2.100,00
Ceas smartwatch Huawei Watch GT3, 42mm, Black NUMAR DE REFERINTA: PSS0021 PRET DE CATALOG: 1.025,00 RON / Unitate de masura	2	2	1.025,00	1.025,00	2.050,00	2.050,00

Figure 2: Procurement item with a low fraud score. The item is consistent with expected procurement norms, showing minimal deviation from the average price.



127.0.0.1:8000/admin/clusters/6785225ee409acb83b4f6c06/

Telefon mobil Samsung Galaxy A54, Dual SIM, 8GB RAM,256GB, 5G, Awesome Graphite 1554.61 bucata

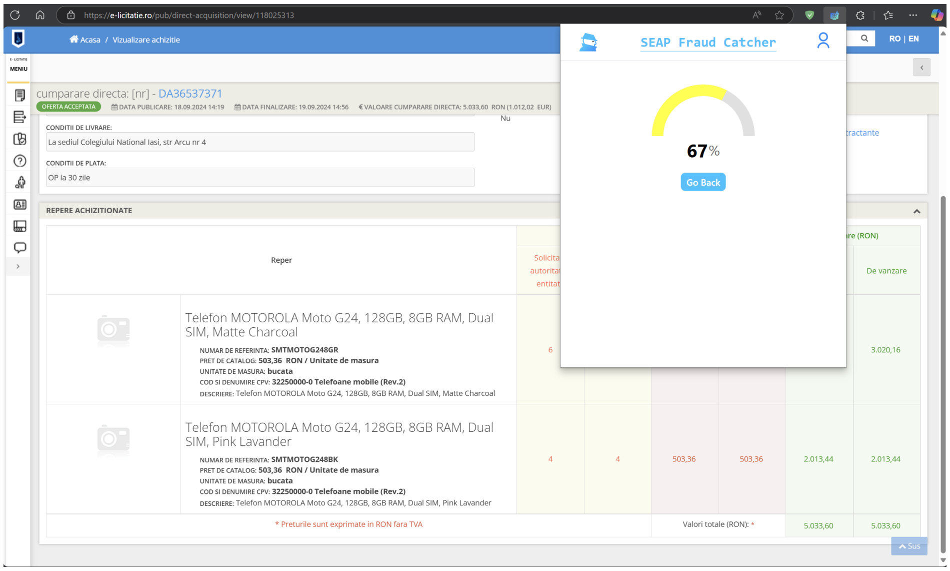
Description
SM-A546BZKDEUE Telefon mobil Samsung Galaxy A54, Dual SIM, 8GB RAM,256GB, 5G, Awesome Graphite

Cluster Items

Total: 28 items

ID	Name	Description	Unit Type	Price	CPV Code
678516a0dd07a207d797c523	Telefon SAMSUNG Galaxy A54 5G, 256GB, 8GB RAM, Dual SIM, Awesome Violet	Telefon SAMSUNG Galaxy A54 5G, 256GB, 8GB RAM, Dual SI...	bucata	1679.08	32250000-0 - Telefoane mobile (Rev.2)
678516b1dd07a207d797c547	Telefon mobil Samsung Galaxy A54, Dual SIM, 8GB RAM,256GB, 5G, Black	SM-A546BZKDEUE Telefon mobil Samsung Galaxy A54, Dual S...	bucata	1739.49	32250000-0 - Telefoane mobile (Rev.2)
678516b7dd07a207d797c560	Telefon mobil Samsung Galaxy A54, Dual SIM, 8GB RAM,256GB, 5G, White	SM-A546BZWDEUE Telefon mobil Samsung Galaxy A54, Dual ...	bucata	1680.67	32250000-0 - Telefoane mobile (Rev.2)
678516c6dd07a207d797c57d	Telefon SAMSUNG Galaxy A54 5G, 256GB, 8GB RAM, Dual SIM, Awesome Graphite	Telefon SAMSUNG Galaxy A54 5G, 256GB, 8GB RAM, Dual SI...	bucata	1554.62	32250000-0 - Telefoane mobile (Rev.2)
678516c6dd07a207d797c581	Telefon mobil Samsung Galaxy A54, Dual SIM, 8GB RAM,256GB, 5G, Black	SM-A546BZKDEUE Telefon mobil Samsung Galaxy A54, Dual S...	bucata	1554.61	32250000-0 - Telefoane mobile (Rev.2)
678516d2dd07a207d797c5a9	Telefon mobil Samsung Galaxy A54, Dual SIM, 8GB RAM,256GB, 5G, Black	SM-A546BZKDEUE Telefon mobil Samsung Galaxy A54, Dual S...	bucata	1554.61	32250000-0 - Telefoane mobile (Rev.2)
678516e3dd07a207d797c5e2	Telefon mobil Samsung Galaxy A54, Dual SIM, 8GB RAM,256GB, 5G, Black	SM-A546BZKDEUE Telefon mobil Samsung Galaxy A54, Dual S...	bucata	1848.73	32250000-0 - Telefoane mobile (Rev.2)
678516efdd07a207d797c5f4	Telefon mobil Samsung Galaxy A54, Dual SIM, 8GB RAM,256GB, 5G, Black	CARACTERISTICI GENERALE Tip telefon Smartphone Sloturi SI...	bucata	2100.0	32250000-0 - Telefoane mobile (Rev.2)

Figure 3: Cluster containing the item with a low fraud score. The cluster demonstrates overall consistency in prices, with no significant anomalies detected.


https://e-licitatie.ro/pub/direct-acquisition/view/118025313
Acasa / Vizualizare achizitie
E-LICITATIE
MENIU
RO | EN
SEAP Fraud Catcher
67%
Go Back
Sus

cumparare directa: [nr] - DA36537371

OFERTA ACCEPTATA DATA PUBLICARE: 18.09.2024 14:19 DATA FINALIZARE: 19.09.2024 14:56 € VALOARE CUMPARARE DIRECTA: 5.033,60 RON (1.012,02 EUR)

Nu

CONDITII DE LIVRARE:
La sediul Colegiului National Iasi, str Arcu nr 4

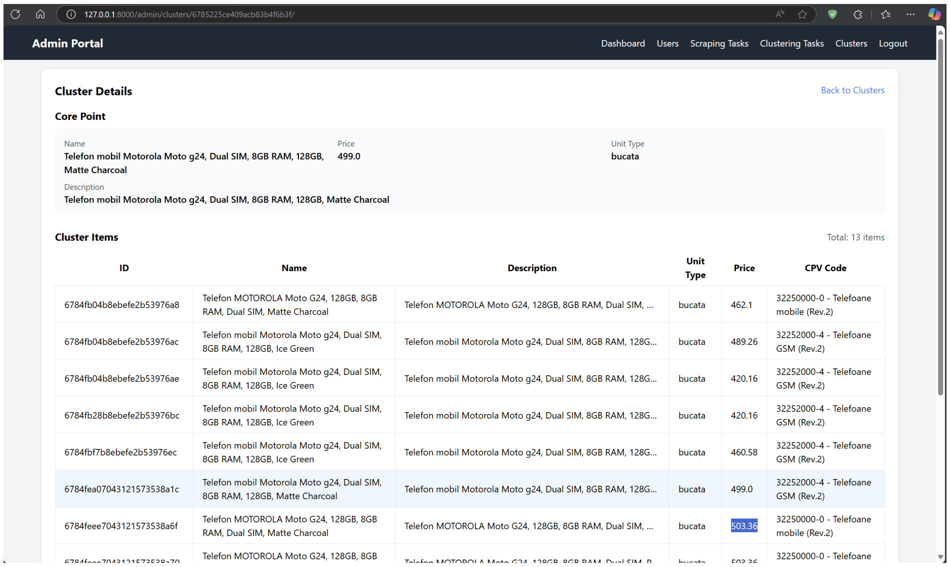
CONDITII DE PLATA:
OP la 30 zile

tractante

REPERE ACHIZITIONATE

Reper	Solicita autoritat entitat				re (RON)	
						De vanzare
Telefon MOTOROLA Moto G24, 128GB, 8GB RAM, Dual SIM, Matte Charcoal NUMAR DE REFERINTA: SMTMOTOG248GR PRET DE CATALOG: 503,36 RON / Unitate de masura UNITATE DE MASURA: bucata COD SI DENUMIRE CPV: 32250000-0 Telefoane mobile (Rev.2) DESCRIERE: Telefon MOTOROLA Moto G24, 128GB, 8GB RAM, Dual SIM, Matte Charcoal	6					3.020,16
Telefon MOTOROLA Moto G24, 128GB, 8GB RAM, Dual SIM, Pink Lavander NUMAR DE REFERINTA: SMTMOTOG248BK PRET DE CATALOG: 503,36 RON / Unitate de masura UNITATE DE MASURA: bucata COD SI DENUMIRE CPV: 32250000-0 Telefoane mobile (Rev.2) DESCRIERE: Telefon MOTOROLA Moto G24, 128GB, 8GB RAM, Dual SIM, Pink Lavander	4	4	503,36	503,36	2.013,44	2.013,44
* Preturile sunt exprimate in RON fara TVA			Valori totale (RON): *		5.033,60	5.033,60

Figure 4: Procurement item with a medium fraud score. The item shows moderate deviation from the expected price range, warranting further investigation.



Cluster Details

Core Point

Name
Telefon mobil Motorola Moto g24, Dual SIM, 8GB RAM, 128GB, Matte Charcoal

Price
499.0

Unit Type
bucata

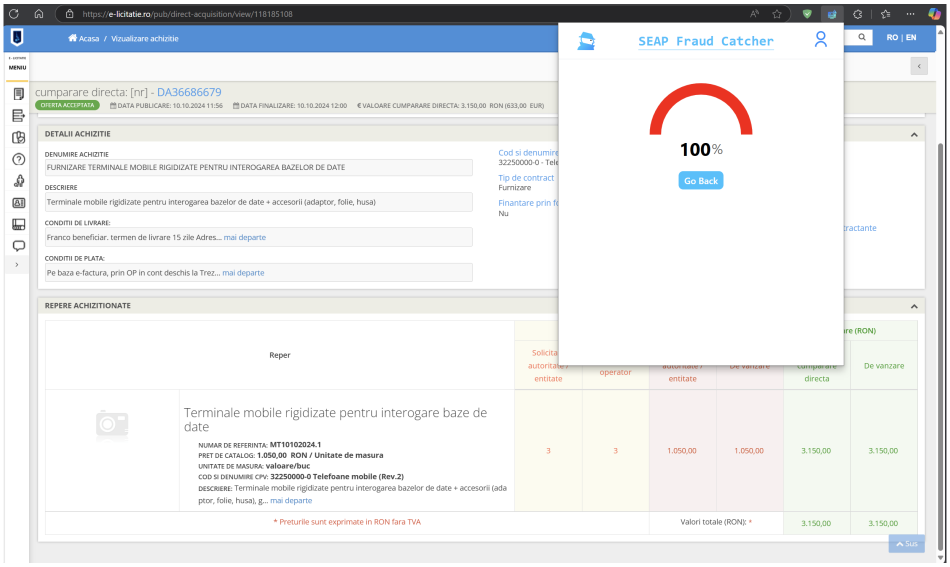
Description
Telefon mobil Motorola Moto g24, Dual SIM, 8GB RAM, 128GB, Matte Charcoal

Cluster Items

Total: 13 items

ID	Name	Description	Unit Type	Price	CPV Code
6784fb04b8ebefe2b53976a8	Telefon MOTOROLA Moto G24, 128GB, 8GB RAM, Dual SIM, Matte Charcoal	Telefon MOTOROLA Moto G24, 128GB, 8GB RAM, Dual SIM, ...	bucata	462.1	32250000-0 - Telefoane mobile (Rev.2)
6784fb04b8ebefe2b53976ac	Telefon mobil Motorola Moto g24, Dual SIM, 8GB RAM, 128GB, Ice Green	Telefon mobil Motorola Moto g24, Dual SIM, 8GB RAM, 128G...	bucata	489.26	32252000-4 - Telefoane GSM (Rev.2)
6784fb04b8ebefe2b53976ae	Telefon mobil Motorola Moto g24, Dual SIM, 8GB RAM, 128GB, Ice Green	Telefon mobil Motorola Moto g24, Dual SIM, 8GB RAM, 128G...	bucata	420.16	32252000-4 - Telefoane GSM (Rev.2)
6784fb28b8ebefe2b53976bc	Telefon mobil Motorola Moto g24, Dual SIM, 8GB RAM, 128GB, Ice Green	Telefon mobil Motorola Moto g24, Dual SIM, 8GB RAM, 128G...	bucata	420.16	32252000-4 - Telefoane GSM (Rev.2)
6784fbf7b8ebefe2b53976ec	Telefon mobil Motorola Moto g24, Dual SIM, 8GB RAM, 128GB, Ice Green	Telefon mobil Motorola Moto g24, Dual SIM, 8GB RAM, 128G...	bucata	460.58	32252000-4 - Telefoane GSM (Rev.2)
6784fea07043121573538a1c	Telefon mobil Motorola Moto g24, Dual SIM, 8GB RAM, 128GB, Matte Charcoal	Telefon mobil Motorola Moto g24, Dual SIM, 8GB RAM, 128G...	bucata	499.0	32252000-4 - Telefoane GSM (Rev.2)
6784feee7043121573538a6f	Telefon MOTOROLA Moto G24, 128GB, 8GB RAM, Dual SIM, Matte Charcoal	Telefon MOTOROLA Moto G24, 128GB, 8GB RAM, Dual SIM, ...	bucata	503.36	32250000-0 - Telefoane mobile (Rev.2)
6784feee7043121573538a70	Telefon MOTOROLA Moto G24, 128GB, 8GB	Telefon MOTOROLA Moto G24, 128GB, 8GB RAM, Dual SIM, P...	bucata	503.36	32250000-0 - Telefoane

Figure 5: Cluster containing the item with a medium fraud score. This cluster reveals moderate variability in pricing, highlighting the need for additional analysis.



https://e-licitatie.ro/pub/direct-acquisition/view/118185108

cumparare directa: [nr] - DA36686679

OFERTA ACCEPTATA DATA PUBLICARE: 10.10.2024 11:56 DATA FINALIZARE: 10.10.2024 12:00 € VALOARE CUMPARARE DIRECTA: 3.150,00 RON (633,00 EUR)

DETALII ACHIZITIE

DENUMIRE ACHIZITIE
FURNIZARE TERMINALE MOBILE RIGIDIZATE PENTRU INTEROGAREA BAZELOR DE DATE

DESCRIERE
Terminale mobile rigidizate pentru interogarea bazelor de date + accesorii (adaptor, folie, husa)

CONDITII DE LIVRARE:
Franco beneficiar. termen de livrare 15 zile Adres... mai departe

CONDITII DE PLATA:
Pe baza e-factura, prin OP in cont deschis la Trez... mai departe

Cod si denumire
32250000-0 - Tele

Tip de contract
Furnizare

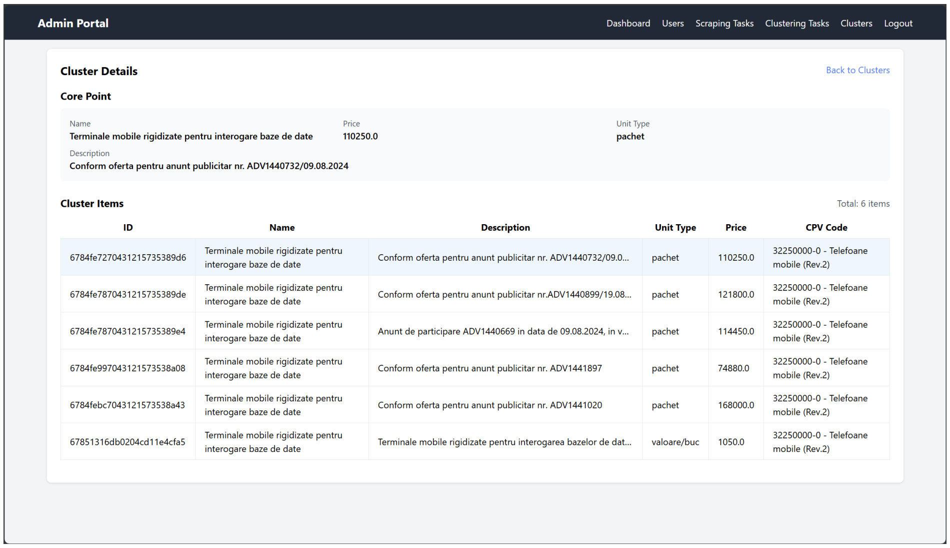
Finantare prin f
Nu

tractante

REPERE ACHIZITIONATE

Reper					re (RON)	
	Solicita autoritate / entitate	operator	autoritate / entitate	De vanzare	cumparare directa	De vanzare
Terminale mobile rigidizate pentru interogare baze de date NUMAR DE REFERINTA: MT10102024.1 PRET DE CATALOG: 1.050,00 RON / Unitate de masura UNITATE DE MASURA: valoare/buc COD SI DENUMIRE CPV: 32250000-0 Telefoane mobile (Rev.2) DESCRIERE: Terminale mobile rigidizate pentru interogarea bazelor de date + accesorii (adaptor, folie, husa), g... mai departe	3	3	1.050,00	1.050,00	3.150,00	3.150,00
* Preturile sunt exprimate in RON fara TVA			Valori totale (RON): *		3.150,00	3.150,00

Figure 6: Procurement item with a high fraud score. This item exhibits significant deviation from the expected price range, indicating potential fraudulent activity.



Cluster Details

Core Point

Name
Terminale mobile rigidizate pentru interogare baze de date

Price
110250.0

Unit Type
pachet

Description
Conform oferta pentru anunt publicitar nr. ADV1440732/09.08.2024

Cluster Items

Total: 6 items

ID	Name	Description	Unit Type	Price	CPV Code
6784fe7270431215735389d6	Terminale mobile rigidizate pentru interogare baze de date	Conform oferta pentru anunt publicitar nr. ADV1440732/09.0...	pachet	110250.0	32250000-0 - Telefoane mobile (Rev.2)
6784fe7870431215735389de	Terminale mobile rigidizate pentru interogare baze de date	Conform oferta pentru anunt publicitar nr.ADV1440899/19.08...	pachet	121800.0	32250000-0 - Telefoane mobile (Rev.2)
6784fe7870431215735389e4	Terminale mobile rigidizate pentru interogare baze de date	Anunt de participare ADV1440669 in data de 09.08.2024, in v...	pachet	114450.0	32250000-0 - Telefoane mobile (Rev.2)
6784fe997043121573538a08	Terminale mobile rigidizate pentru interogare baze de date	Conform oferta pentru anunt publicitar nr. ADV1441897	pachet	74880.0	32250000-0 - Telefoane mobile (Rev.2)
6784febc7043121573538a43	Terminale mobile rigidizate pentru interogare baze de date	Conform oferta pentru anunt publicitar nr. ADV1441020	pachet	168000.0	32250000-0 - Telefoane mobile (Rev.2)
67851316db0204cd11e4cfa5	Terminale mobile rigidizate pentru interogare baze de date	Terminale mobile rigidizate pentru interogarea bazelor de dat...	valoare/buc	1050.0	32250000-0 - Telefoane mobile (Rev.2)

Figure 7: Cluster containing the item with a high fraud score. The cluster shows substantial price discrepancies, strongly suggesting irregularities or fraud.

The examples above demonstrate how the fraud detection system identifies and clusters procurement items based on their characteristics. By associating fraud scores with clusters, the system ensures a comprehensive and accurate evaluation of procurement anomalies. They also show the system's capability to detect anomalies across a spectrum of fraud scores. Low fraud scores indicate that the procurement is consistent with expectations, medium scores highlight items requiring further review, and high scores suggest likely fraudulent activity. This tiered approach aids in prioritizing investigations and enhances transparency in public procurement processes.

5 Comparison with Existing Solutions

5.1 Comparative Analysis

The SEAP Fraud Catcher project distinguishes itself through its modular and hybrid clustering approach tailored specifically for Romanian public procurement data on the SEAP platform. In contrast to existing solutions like Fraud.net [4] and Onfido [5], the SEAP Fraud Catcher provides more flexibility and local specificity.

Fraud.net employs an advanced AI-powered platform that offers real-time fraud monitoring across various industries, but it is primarily focused on financial transactions and lacks the procurement-specific functionality required for detecting fraud in public tenders. Its primary use case revolves around detecting payment fraud, account takeovers, and synthetic identity fraud, with strong emphasis on compliance in regulated industries. SEAP Fraud Catcher, by contrast, focuses exclusively on public procurement data and integrates semantic analysis and clustering to handle textual descriptions in procurement documents, offering a significant advantage in identifying anomalies in product specifications or pricing. For example, it can detect cases where procurement descriptions contain inconsistencies or are repeatedly altered to favor specific bidders.

Similarly, Onfido focuses on fraud detection for identity verification and compliance in the financial and e-commerce sectors. While it leverages machine learning for fraud prevention, its functionality is centered around verifying personal identities and combating identity theft. It does not cater to the complexities of public procurement systems or offer tools for detecting collusion or irregular pricing. SEAP Fraud Catcher addresses this gap by employing clustering mechanisms that combine traditional algorithms, such as K-Means++, with rule-based refinements. These refinements allow it to detect irregularities such as overpriced tenders, recurring awards to the same bidders, or patterns indicative of collusion among contractors. The system is further enhanced by its ability to group tenders semantically, ensuring better accuracy in identifying outliers and fraudulent activities.

SEAP Fraud Catcher also stands out through its user-centric approach. While Fraud.net and Onfido are enterprise-focused tools, often requiring subscriptions or specialized integration, SEAP Fraud Catcher is openly accessible to the public without login barriers. This ensures that both public institutions and citizens can transparently access insights into potential procurement frauds. Additionally, SEAP Fraud Catcher's modular architecture allows it to be customized and scaled easily, ensuring it can evolve to handle increased data volumes or incorporate additional algorithms as needed. By focusing on transparency and adaptability, SEAP Fraud Catcher provides a tailored solution for Romania's procurement-specific challenges, standing as a more locally effective alternative to broader, less specialized systems.

5.2 Discussion

The SEAP Fraud Catcher stands out due to its hybrid clustering methodology, which effectively handles complex and non-standardized procurement data. By integrating rule-based checks alongside standard clustering algorithms such as K-Means or DBSCAN, it refines the clustering process, making it especially effective in detecting anomalies within the SEAP dataset. For instance, the system can identify and recluster tenders displaying textual irregularities, such as alphanumeric inconsistencies, which are often indicative of collusion or inflated pricing.

However, the effectiveness of SEAP Fraud Catcher depends heavily on the quality of the input data. Challenges like incomplete or inconsistent procurement records, common to public platforms, could impact its accuracy. Addressing this issue could involve employing more sophisticated data preprocessing techniques or enriching the dataset with external sources to ensure greater reliability.

Overall, SEAP Fraud Catcher addresses specific challenges in Romanian public procurement fraud detection with a tailored, transparent, and modular approach. Its focus on integrating advanced clustering methods with domain-specific rules positions it as a compelling alternative to broader, less specialized international systems.

6 Future Work

6.1 Current Limitations

Although the SEAP Fraud Detection Application demonstrates significant potential, it currently faces a few notable limitations that require attention for future development:

- **Domain-Specific Training:** The system is currently trained exclusively on procurement data related to mobile phones. This domain-specific approach arises from the challenges associated with generalizing item clustering methods across diverse product categories. Each category requires customized clustering approaches to accurately group similar items and subsequently calculate a fraud score. This limitation restricts the system's applicability to broader procurement domains and underscores the need to develop adaptive algorithms capable of handling diverse product categories with varying characteristics.
- **Limited Fraud Score Parameters:** The current methodology for fraud detection relies exclusively on price-based analysis to calculate fraud scores. Although price anomalies are a critical indicator of fraudulent activity, other essential parameters, such as procurement timelines, product descriptions, and supplier consistency, are not yet integrated into the scoring algorithm. This narrow focus may lead to the overlooking of complex fraudulent patterns that are influenced by temporal and contextual factors.

Addressing these limitations is essential for enhancing the system's robustness and versatility, enabling it to adapt to a wider range of procurement scenarios while providing a more comprehensive evaluation of potential fraud.

6.2 Opportunities for Improvement

To enhance the scope and effectiveness of the SEAP Fraud Detection Application, several key areas for improvement have been identified:

- **Expansion to Multiple Categories:** A significant opportunity lies in extending the system's applicability to a wider range of procurement categories beyond mobile phones. Developing a generalized framework for detecting fraud across diverse product categories will require adaptable clustering techniques and fraud detection models that can accommodate varying data characteristics. This expansion will significantly broaden the system's utility and impact.
- **Incorporation of Additional Parameters:** The current fraud detection model primarily relies on price as the sole parameter for scoring fraud. Integrating additional dimensions, such as the product's release period (e.g., year of manufacture, relevance over time), detailed descriptions (e.g., condition, specifications), and supplier history, can provide a more nuanced and comprehensive analysis. By leveraging these factors, the system can uncover more intricate fraud patterns and reduce the likelihood of false positives.
- **Adoption of Advanced Clustering Algorithms:** To improve the efficiency and accuracy of the system's clustering processes, adopting more advanced and scalable clustering algorithms presents a valuable opportunity. Although the current approach employs agglomerative clustering, k-means++, and OPTICS, exploring newer or hybrid algorithms could improve the speed and precision of data processing, particularly for large and complex datasets. Such improvements are invaluable for achieving real-time or near-real-time fraud detection.

These advancements will not only enhance the system's accuracy and performance but also increase its adaptability, paving the way for a more robust and versatile fraud detection platform.

7 Conclusions

The SEAP Fraud Detection Application is a innovative tool designed to elevate the standards of transparency and accountability in Romania's public procurement processes. By harnessing the power of artificial intelligence and advanced clustering algorithms, the application not only identifies fraudulent activities but also redefines how procurement data is analyzed and interpreted. Its ability to detect irregularities such as inflated prices and collusion, coupled with a user-centric interface, ensures accessibility for all stakeholders—from public officials to ordinary citizens.

While the system demonstrates exceptional performance in analyzing large datasets and uncovering price-related anomalies, its potential extends far beyond. Future enhancements, such as integrating additional analytical dimensions and refining clustering methods, will solidify its position as a robust, versatile, and adaptive solution.

By bridging the gap between technology and governance, the SEAP Fraud Detection Application sets a new benchmark for ethical public procurement, ensuring that resources are allocated fairly and efficiently while fostering greater trust in governmental institutions.

References

[1] Roberto Nai, Emilio Sulis, and Rosa Meo. "Public Procurement Fraud Detection and Artificial Intelligence Techniques: A Literature Review". In: *EKAW'22: Companion Proceedings of the 23rd International Conference on Knowledge Engineering and Knowledge Management*. Bozen-Bolzano, Italy: CEUR Workshop Proceedings, 2022, pp. 1–8. URL: http://ceur-ws.org/Vol-3222/.

[2] Nikola Modrušan, Kornelije Rabuzin, and Leo Mršić. "Review of Public Procurement Fraud Detection Techniques Powered by Emerging Technologies". In: *International Journal of Advanced Computer Science and Applications (IJACSA)* 12.2 (2021), pp. 1–10. DOI: 10.14569/IJACSA.2021.0120272. URL: https://thesai.org/Publications/ViewPaper?Volume=12&Issue=2&Code=IJACSA&SerialNo=1.

[3] Lothar Kuhl. "Implementation of Effective Measures against Fraud and Illegal Activities in Cohesion Policies". In: *eucrim* (2020). DOI: 10.30709/eucrim-2020-011. URL: https://doi.org/10.30709/eucrim-2020-011.

[4] Fraud.net. *Fraud.net: AI-powered fraud detection platform*. 2021. URL: https://fraud.net.

[5] Onfido. *Onfido: Fraud Detection Solution*. 2024. URL: https://onfido.com/solutions/fraud-detection.