Processing and Analysis of Data from Fiscal Electronic Cash Registers in the Context of IoT and Big Data

Dragoș-Cătălin BARBU[1]*, Adela BÂRA[2], Simona-Vasilica OPREA[2]

[1] National Institute for Research and Development in Informatics – ICI Bucharest,
8-10 Mareșal Averescu Avenue, Bucharest, 011455, Romania
dragos.barbu@ici.ro (**Corresponding author*)

[2] Bucharest University of Economic Studies, Department of Economic Informatics and Cybernetics,
6 Piața Romană, Bucharest, 010374, Romania
bara.adela@ie.ase.ro, simona.oprea@csie.ase.ro

Abstract: This paper proposes a framework for processing and analysing fiscal data from electronic cash registers using Big Data and IoT. The system is designed for handling large volumes of transactional data by integrating data collection, cleaning, transformation, and analysis through a modular architecture based on microservices, distributed messaging, and both relational and NoSQL databases. Before carrying out the data analysis, the missing or inconsistent values are addressed using regression models, which enhances data quality. In order to contribute to anomaly detection in fiscal activities, the proposed platform supports statistical analysis, time series analysis, and pattern recognition. Real-world data-based tests revealed that the proposed technological solution can help the tax authorities track data compliance and increase the effectiveness of fiscal data operations.

Keywords: Big Data, Cloud Computing, IoT, Data Mining, Data Aggregation, Big Data Analytics, Digital Fiscal Data, Electronic Cash Register.

1. Introduction

The inclusion of Big Data analytics into the field of taxation and fiscal data analysis represents a major change towards a more dynamic and sophisticated public financial management. Considering its potential to transform conventional tax systems and boost the economy, this developing field of research has attracted great interest among both the university academics and government finance professionals. According to Belahouaoui & Attak (2024), Big Data analytics not only enhances fiscal compliance and optimizes revenue but it also plays a crucial role in shaping fiscal policies.

In comparison with previous, limited-scope fiscal analyses, the use of large-scale fiscal datasets allows tax authorities to detect subtle differences in tax collection and evasion patterns that were previously undetectable, which represents a major methodological advancement (Zhang et al., 2021). By analysing these datasets, tax authorities can anticipate future trends and enact proactive measures to mitigate tax evasion, thus securing revenue streams that are vital for public spending. Furthermore, this analytical capacity allows for a more nuanced understanding of taxpayer behaviour, enabling targeted interventions that can improve compliance rates and the overall tax system efficiency.

Although prior research has investigated the application of big data in fiscal analysis (Slemrod & Weber, 2012; OECD, 2017), it frequently lacks a comprehensive technical framework that would encompass real-time data intake, preprocessing, and anomaly detection. This work addresses this gap by presenting a complete and flexible architecture designed for large-scale fiscal data produced by ECRs.

Fiscal data analysis involves the meticulous examination and interpretation of transactional data to extract insights relevant to taxation, economic analysis, and policy-making. This process provides critical insights that are instrumental in assessing tax compliance, understanding revenue generation mechanisms, and discerning consumer behaviour and broader economic trends (Slemrod, 2019; Slemrod & Weber, 2012). Such realisations are priceless for governments trying to strike a balance between economic development and fiscal responsibility in order to guarantee that tax laws are fair and efficient. Examining fiscal data goes beyond mere revenue collection or compliance. It affects everything from personal tax changes to more general macroeconomic policies, so it impacts a spectrum of economic decisions. Fiscal data analysis contributes to a more informed policy-making and a more solid economic planning by offering a clearer picture of the economic terrain (Barbu, 2020; Kleven et al., 2011).

As finances are getting more complicated and interconnected with global economies, the need for advanced data analysis tools has become more evident. Professionals point out that the analysis of fiscal data is required so that fact-based decisions

https://doi.org/10.24846/v34i2y202510

can be made, tax management can be improved, and economic growth can be stimulated (Advani et al., 2023; Pomeranz, 2015). Advanced data analytics techniques like data mining and machine learning, help uncover patterns, anomalies, and correlations. These revelations help financial authorities to refine their plans for an improved tax collection and compliance. They can thus create more sensible and flexible financial policies.

The remainder of this paper is structured as follows. Section 2 presents the state of the art regarding Big Data technologies applied to fiscal data analysis, highlighting the role of advanced analytics, machine learning, and data integration in improving tax compliance and identifying anomalies. Section 3 describes the preprocessing and cleaning steps applied to the raw fiscal data collected from electronic cash registers, including imputation techniques and error correction using regression models. Further on, Section 4 outlines the architecture and design specifications for the proposed platform, detailing the data acquisition process, the technological components, and analytical modules used for processing and analyzing fiscal data. Finally, Section 5 provides the main conclusions regarding the proposed platform's capabilities and potential for supporting real-time fiscal monitoring and decision-making.

2. State of the Art

The analysis of economic data has significantly changed due to a large extent to the Big Data technology. These instruments today enable tax authorities and legislators to gather thorough data, therefore revealing important insights that support the creation of more efficient tax laws and enhance the revenue collection methods.

Modern technical developments and fresh approaches have greatly improved the financial data analysis. By means of machine learning and data mining, researchers and authorities have been able to identify intricate patterns and odd outcomes which were undetectable using more traditional methods. By evaluating prior data patterns, machine learning algorithms can today identify tax fraud and evasion (OECD, 2017; Naritomi, 2019). These prediction tools enable fast and effective reactions by the early identification of probable tax inequalities.

The legal framework in charge of Big Data application has evolved to meet the technological development. The 2024 Strategic Plan of the Directorate-General for Taxation and Customs Union (DG TAXUD, 2020) outlines initiatives using Big Data to maximise tax systems inside the internal market, thereby improving tax compliance and minimising administrative burdens by means of the data acquired through the automatic exchange of tax information. This initiative involves the funding of European IT systems enabling real-time data processing and sharing across EU member states, therefore enhancing the member states` capacity to track cross-border transactions and so fight tax evasion.

Scholarly research constantly shapes the use of Big Data in fiscal policy by offering actual data and pragmatic concepts proving the effectiveness of these tools, which include advanced data analytics platforms, machine learning algorithms, data mining techniques, and statistical modeling methods, all of which are used for analyzing large-scale fiscal data and detecting patterns in taxpayer behavior. Supported by sophisticated data analytics, the papers of Slemrod & Weber (2012) and Barbu (2020) provide comprehensive analyses of taxpayer compliance behaviour and regarding the efficacy of several tax systems. These intellectual contributions not only support the useful applications of Big Data in taxation but they also widen the boundaries of possible developments via constant innovation.

And notwithstanding many innovations, it is imperative to address the continuous challenges and future directions in the industry, especially with regard to data privacy, security, and the ethical use of information. One major problem is balancing the preservation of personal privacy with maximising Big Data for commercial goals. Moreover, the difficulty of combining new technologies with the current tax systems features operational difficulties needing continuous attention and creative ideas.

Collaborations and global initiatives, such as OECD's efforts to standardize and enhance technological tools, have played a key role in the international fight against tax evasion and fraud (OECD, 2017), illustrate a global commitment to adopting Big Data strategies. These initiatives aim to advance a coherent plan which could solve the issues of a technologically linked global economy.

The increasing relevance of producing and responding to questions about corporate data has

been underlined in recent years by the expanding reach of data science in corporate environments encompassing subjects like Business Analytics and Business Intelligence. Emphasising the use of Big Data to include advancements in data science, these fields make use of several instruments from statistics, data management, data visualisation, and machine learning (Mikalef et al., 2018). Such an integration is crucial since it reflects their relevance by enabling their broader applicability across many sectors and disciplines, therefore improving the value and usefulness of these approaches.

The integration of modern analytical approaches has attracted the interest of researchers who underline how Big Data Analytics may enable an improved decision-making, strengthen corporate intelligence, and improve the general organisational performance. Sophisticated analytical approaches, patterns, correlations, and trends within large datasets reveal such supporting tasks including forecasting, risk analysis, and work optimisation (Fan et al., 2014). This feature is particularly useful in fiscal data analysis, in which merging external data sources - including socioeconomic data, demographic information, industry-specific data, or macroeconomic indicators - offers more insights (Wang et al., 2018).

The integration of external data sources into fiscal data analysis has gained relevance as it provides a better understanding of the elements influencing revenue patterns and economic trends. The techniques for data integration must ensure consistency and accuracy, with data preprocessing including cleaning and standardization processes for aligning external data with fiscal data (Batini et al., 1986; Halevy et al., 2006). This combination enhances the analysis and offers a broader perspective on fiscal administration and policy formulation (Khan et al., 2014; Einav & Levin, 2014).

The developments in data infrastructure for Big Data have made it much easier for the scientific community to manage and work with large-scale datasets, opening up new possibilities for analysis and discovery. For instance, Demchenko et al. (2013) presented a study on the development of the Big Data infrastructure, aiming to aid in the selection and adoption of suitable Big Data technologies aligned with technological needs and specific application requirements. This includes the development of models such the Scientific Data Lifecycle Management (SDLM), which addresses the particularities of data management in contemporary electronic science and combines all main phases of the scientific data management process, including data creation, storage, processing, analysis, sharing, and preservation.

An additional contribution is that of Wang et al. (2018), whose thorough review of the basic components of Big Data projects especially in sectors like healthcare offers a whole picture. These works highlight how the effective integration of data storage, extraction, analysis, and visualization can enhance organisational performance and support evidence-based decision-making.

Technological frameworks and their applications like Hadoop, Spark, and NoSQL databases have become essential tools in Big Data analysis, helping organizations process and make sense of massive amounts of information more efficiently (Sahal et al., 2020; Deng et al., 2020). In modern computing environments, particularly wherein IoT, edge computing, and artificial intelligence converge, Hadoop, for example, is absolutely vital in data management. Being fit for iterative algorithms and real-time analytics, Spark offers a quick and scalable framework for data processing and analysis (Zaharia et al., 2016). For managing unstructured and semi-structured data, NoSQL databases provide scalable and flexible storage options (Han et al., 2011).

3. Pre-processing and Cleaning of Fiscal Data

While gathering information and during the early assessment, the financial data for the study was obtained via electronic cash registers spread over several Romanian industries. Originally, the data consisted in millions of transaction records including information on the transaction value, tax rate applied, timestamp, and merchant category. Initially, the preprocessing consisted in confirming whether the data was complete and in spotting records with odd or absent values. Following file organisation for both structured and unorganised files, over 626,000 receipts were imported. These reflected roughly 4,600 daily closing fiscal receipts from 33 enterprises using a total of 37 electronic cash registers (ECRs). Using the electronic journal files produced by these 37 ECRs as the test data source, almost 2 million items were extracted at the product level from these files.

The data exhibited typical issues such as null values in the item description field and

inconsistent merchant IDs. To address this issue, multiple regression models were developed and assessed, including linear and logistic regression, using contextual variables such as the merchant type, transaction category, and temporal patterns as input features. Their performance was assessed using cross-validation, based on the mean squared error and classification accuracy metrics. For instance, a linear regression model projected the missing tax rates based on the average tax rate of the transactions within the same category and the total related amounts expected.

As the data was being cleaned, a descriptive statistical analysis found some oddities including irregular tax rates and transaction value outlays. Transactions which deviated too much from the norm such as by more than three standard deviations were deemed possible mistakes or anomalies. To better match the pattern noticed in most of the data, a strong regression model adjusted these values.

The timestamps of all transaction were changed to fix errors, therefore ensuring consistency in the full dataset. If various currencies came from data entry errors, all the transaction values were converted to Romanian Lei (RON). By using a logistic regression model, which scored every transaction depending on different variables including transaction value, location, and time, the errors in the merchant category field were also fixed.

Following data cleaning, it was found that approximately 12.3% of the data contained missing or erroneous fields; 9.6% of it was successfully recovered using imputation techniques, while 2.7% of it was discarded due to irreparability. The selected subsets were tested with regard to their accuracy and consistency, therefore verifying the success of the data cleaning operations. The performance of the regression models in imputing the missing values and error correction was evaluated using cross-valuation methods so that the models could forecast the missing or erroneous data in an unbiased manner.

The dataset prepared for analysis, resulting from a thorough preprocessing and cleaning, provided a coherent and reliable foundation for the subsequent analytical procedures. The cleaned data retained the integrity of the original records while significantly reducing noise and errors, thereby enhancing the quality and reliability of the subsequent fiscal analyses.

4 Architecture and Design Specifications

The work of Barbu et al. (2024) studied the impact of electronic cash registers on tax collection. This paper emphasizes the implementation of ECRs as an essential step in improving the accuracy of fiscal reporting and increasing the transparency of financial transactions. It also presents the challenges encountered in adopting ECR systems, including technical and operational aspects and it highlights the importance of monitoring data in real time in order to reduce tax evasion and optimize the tax collection process.

This study proposes a high-level architecture (Figure 1) for the collection and analysis of fiscal data. This architecture is intended to enhance the accuracy of fiscal reporting and increase transparency, thus contributing to the reduction of tax evasion and the optimization of the tax collection process.

The framework elaborated for the collection of fiscal data from electronic cash registers contains a large amount of varied information that captures all the sales operations conducted by economic agents. The data features accurate information on transactions such as prices, quantities, VAT rates, and the total number of receipts issued. The proposed architecture must be able to efficiently manage this large amount of information, ensuring the collection, storage, and analysis of data to support fiscal processes and prevent tax evasion.

A set of high-level functional requirements aims to guide the implementation process for data collection and the functionalities of the analysis platform.

In the following subsection, the types of information that can be collected from electronic cash registers are detailed.

4.1 Types Of Data from Cash Registers

This subsection analyses the specific details of the available data, recorded by the cash registers and, further on, this paper discusses how this information can be used to comply with fiscal regulations and to improve the transparency and efficiency of the data collection process.

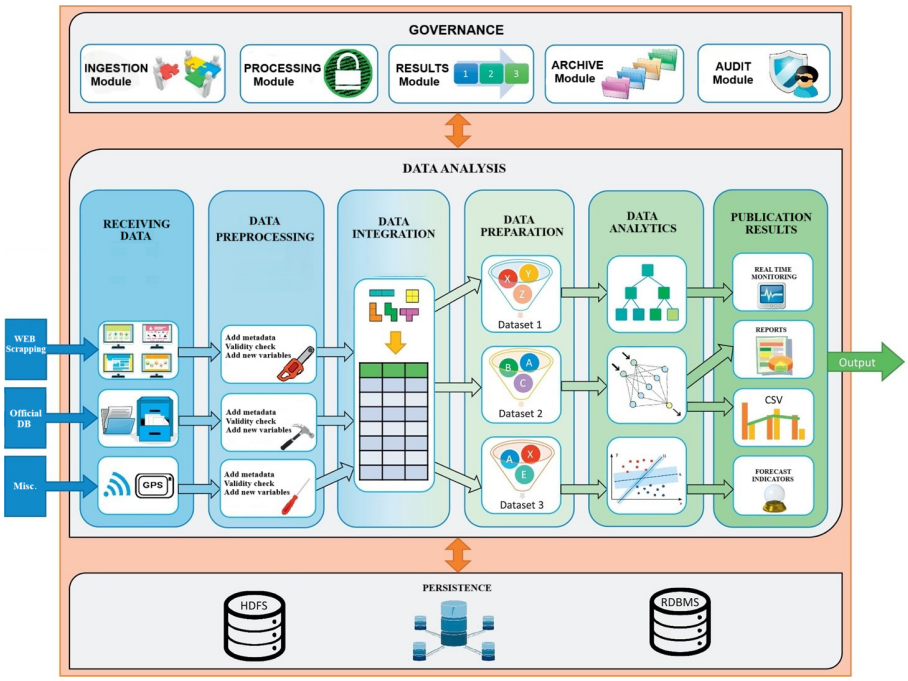

Figure 1. Big Data Architecture for Fiscal Data Analysis (Barbu et al., 2024)

The fiscal data that can be collected is in different formats:

- Unstructured – Within the fiscal module of the ECR, data is stored in TXT format, an electronic journal file that stores all the information from the fiscal receipts printed during a fiscal day;
- Semi-structured – At the communication module level of the ECR, XML files with synthetic data for the workday are generated and transmitted by the ECR (online or offline);
- Structured – Information held in the databases of the National Agency for Fiscal Administration (ANAF), including quarterly reports and historical data;
- Traditional retail sales data has been structured and largely derived from ECRs. This data captures information about sales, the number of sold items, prices, and the timestamps of transactions.

4.2 Digital Fiscal Data Collection

This subsection describes the use of advanced technologies and IT systems for automating the collection of fiscal data. A secure architecture based on digital certificates enables the real-time communication between reporting platforms and electronic fiscal devices. These mechanisms reflect broader trends in secure digital infrastructure design, similar to those employed in blockchain-based systems to ensure transparency and trust in electronic processes (Barbu et al., 2022). Once the connection is established, XML summary files are transmitted to specialized microservices within the data collection system. This architecture supports near real-time data acquisition from over 600,000 ECRs operating nationwide, significantly enhancing the efficiency and scope of fiscal data monitoring. For illustration purposes, Figure 2 shows the online data collection architecture based on microservices, highlighting the secure communication flow

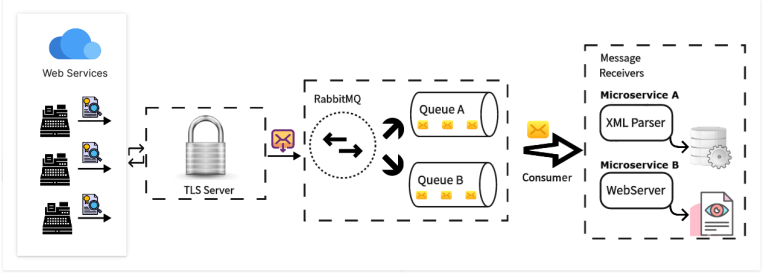

Figure 2. Online Collection Architecture via Microservices

between electronic fiscal devices and the central reporting platform. This modular system enables scalable, real-time data processing, ensuring an efficient and reliable fiscal information transfer.

4.3 Process Flow and Technology Overview

This subsection details the flow diagram and the employed technologies, structured according to the modules of the fiscal data lifecycle:

- Data Ingestion - The online communication component uses the RabbitMQ distributed message broker to receive messages from ECRs and forward them to other microservices in the system.

- Preprocessing - The received messages are processed by various microservices, and the resulting XML files are stored in a PostgreSQL database system. This modular, microservices-based design also supports the scalable deployment and tuning of machine learning components, consistent with the recent architectural frameworks designed for optimizing ML workflows (Oprea et al., 2023).

- Data Integration - At this stage, using the Python programming language, the platform retrieves both XML files received online and TXT files from offline-integrated electronic journals. These are parsed and stored in a MongoDB database.

- Data Processing - This module performs data transformations and validations using Python, operating on the data collections stored in MongoDB.

- Data Analysis - Leveraging Python libraries such as Scikit-learn, Pandas, and Plotly, the system analyses the data and prepares it for the publication module.

- Data Publication - The analysed data is published for result visualization and stored in MongoDB collections.

- Data Export/Reuse - This module enables the saving and exporting of analysed datasets for replication and archiving purposes, using Python packages such as PyMongo and Plotly.

The diagram in Figure 3 illustrates the technological tools used for the implementation of the fiscal data collection and analysis platform. This platform was designed to demonstrate the efficiency and functionality of the entire workflow from the initial data acquisition to the extraction of valuable insights for identifying fiscal trends and anomalies.

4.4 Data Analytics

Following the data preparation phase, the proposed platform enables a wide range of analytical operations aimed at extracting insights from the processed datasets, including predictive modelling using decision trees and support vector machines in order to forecast fiscal anomalies based on merchant categories, transaction frequency, and value patterns. These approaches are consistent with the prior findings emphasizing the role of data analysis and IT tools in combating tax evasion and enhancing fiscal transparency (Barbu, 2019). Other approaches include:

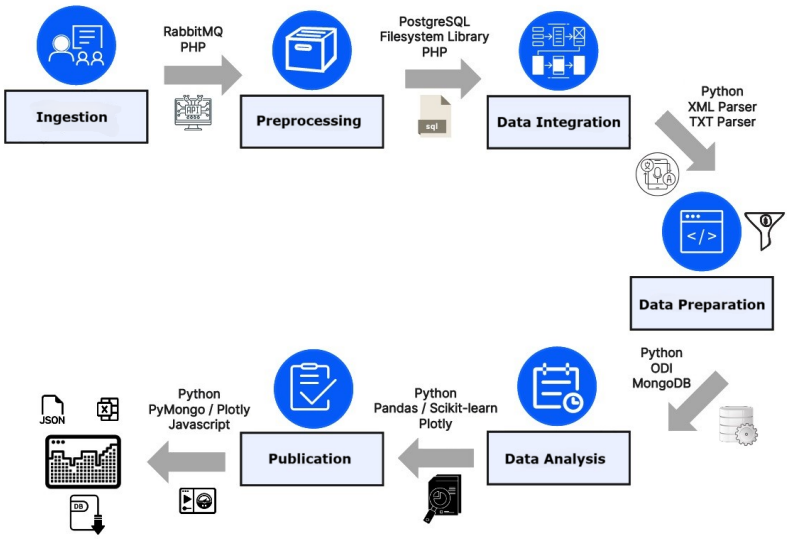

Figure 3. Process Flow and Technology Overview

- Statistical Analysis and Visualization: Tools are employed to explore data distributions and identify the underlying patterns. This helps users gain a comprehensive understanding of the structure and variability of the fiscal data.

- Time Series Analysis: A key application which focuses on the analysis of daily sales, allowing the identification of temporal trends, seasonal patterns, and anomalies related to transaction volumes.

- Detailed Record Analysis: Enables the in-depth examination of individual data entries, supporting the granular investigation of the fiscal activity.

- Transaction-Level Analysis: Users can analyse the frequency and characteristics of daily transactions. As illustrated in Figure 4, this subsection allows the selection of specific data collections and time intervals. Filters can be applied to extract and analyse the transaction records corresponding to the selected criteria, enabling targeted investigations and operational monitoring.

This module transforms raw fiscal data into actionable intelligence, supporting data-driven decision-making and the early detection of irregularities in commercial activities.

After analysing this information, it is possible to download the displayed data in JSON format. The downloaded file would contain the processed, filtered, and visualized data, which can be useful for further analysis or archival purposes. All the displayed charts are interactive, allowing the user to explore the data in detail.

5. CONCLUSIONS

This study presents the technical specifications and functional requirements of the proposed fiscal data collection and analysis platform, which includes modules for data acquisition, cleaning, preprocessing, and transformation. These components ensure data integrity and compliance prior to its analytical processing. A key contribution of this platform lies in the integration of advanced analytic functions, which enable the identification of fiscal behaviour patterns and the generation of predictive reports.

The platform's performance was evaluated using real-world test datasets, which confirmed its ability to efficiently and accurately process substantial volumes of fiscal data. The obtained results demonstrate the platform's robustness in supporting real-time data analysis and its practical applicability in operational tax environments.

In conclusion, this work sets forth a practical, scalable, and efficient platform for automated fiscal data analysis using a modular big data framework. It provides clear advantages with

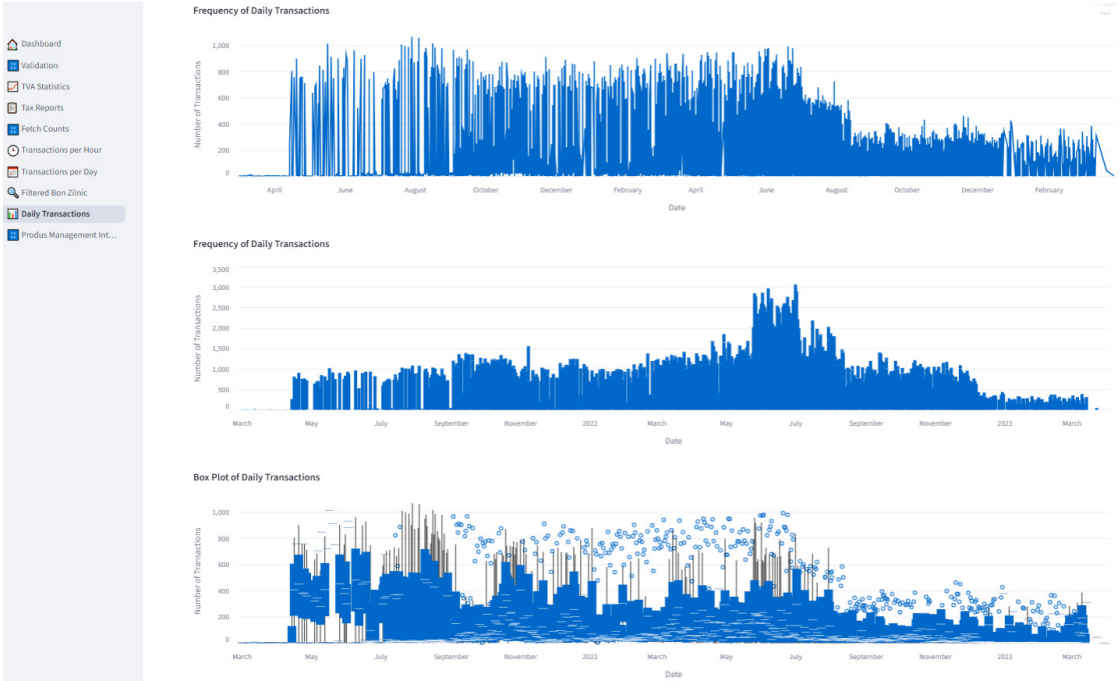

Figure 4. Daily Transactions analysis

regard to real-time anomaly detection, policy compliance support, and the flexibility of its integration with the emerging technologies such as AI and edge computing.

The integration of electronic cash registers in an architecture for an efficient data capturing mechanism requires the careful consideration of its processing and analytical components including data capture, transmission, storage, processing, and analysis. Regarding its performance, scalability, and security, the proposed architecture provides an ample support to users from fiscal data monitoring to its management and reporting. Nevertheless, a serious shortcoming lies in its dependency on high-quality data, as the proposed approaches are best suited for environments that are controlled, which is hardly the case in real life. Moreover, the developed framework could benefit from the emerging technologies such as artificial intelligence and edge computing with a view to optimizing its performance for real-time decision making.

REFERENCES

Advani, A., Elming, W. & Shaw, J. (2023) The Dynamic Effects of Tax Audits. *The Review of Economics and Statistics*. 105(3), 545–561. https://doi.org/10.1162/rest_a_01101.

Barbu D.-C. (2019) Soluții de prelucrare specifice Big Data [Big Data Processing Solutions]. *Revista Română de Informatică și Automatică* [*Romanian Journal of Information Technology and Automatic Control*]. 29(2), 35–48, https://doi.org/10.33436/v29i2y201903.

Barbu, D.-C. (2020) Utilizarea tehnicilor de analiză a datelor în procesul de identificare a evaziunii fiscale [Use of data analysis techniques in the process of identifying tax evasion]. *Revista Română de Informatică și Automatică* [*Romanian Journal of Information Technology and Automatic Control*]. 30(4), 39–50. https://doi.org/10.33436/v30i4y202003.

Barbu, D.-C., Dobrița, G., Oprea, S.-V. et al. (2022) Provocări și beneficii ale sistemului de vot electronic bazat pe Blockchain [Challenges and benefits of Blockchain-based Electronic Voting System]. *Revista Română de Informatică și Automatică* [*Romanian Journal of Information Technology and Automatic Control*]. 32(4), 117-128. https://doi.org/10.33436/v32i4y202210.

Barbu, D.-C., Bâra, A., Oprea, S.-V. (2024) Impact of Electronic Cash Registers on Tax Collection. In: Ciurea, C., Pocatilu, P. & Filip, F.G. (eds.) *Proceedings of 22nd International Conference on Informatics in Economy (IE 2023)* (*Smart Innovation, Systems and Technologies*, vol 367). Singapore, Springer, pp. 25-37.

Batini, C., Lenzerini, M. & Navathe, S.B. (1986) A Comparative Analysis of Methodologies for Database Schema Integration. *ACM Computing Surveys*. 18(4), 323–364. https://doi.org/10.1145/27633.27634.

Belahouaoui, R. & Attak, E.-H. (2024) Digital taxation, artificial intelligence and Tax Administration 3.0: improving tax compliance behavior – a systematic literature review using textometry (2016–2023). *Accounting Research Journal*. 37(2), 172–191. https://doi.org/10.1108/ARJ-12-2023-0372.

Demchenko, Y., Grosso, P., de Laat, C. et al. (2013) Addressing Big Data Issues in Scientific Data Infrastructure. In: 2013 International Conference on Collaboration Technologies and Systems (CTS), *20-24 May 2013, San Diego, CA, USA*. New York, USA, IEEE. pp. 48–55.

Deng, S., Zhao, H., Fang, W. et al. (2020) Edge Intelligence: The Confluence of Edge Computing and Artificial Intelligence. *IEEE Internet of Things Journal*. 7(8), 7457–7469. https://doi.org/10.1109/JIOT.2020.2984887.

Directorate-General for Taxation and Customs Union (DG TAXUD). (2020) *Strategic Plan 2020-2024 – DG Research and Innovation*. https://commission.europa.eu/system/files/2020-11/rtd_sp_2020_2024_en.pdf [Accessed 22nd June 2025]

Einav, L. & Levin, J. (2014) Economics in the age of big data. *Science*. 346(6210). https://doi.org/10.1126/science.1243089.

Fan, J., Han, F. & Liu, H. (2014) Challenges of Big Data analysis. *National Science Review*. 1(2), 293–314. https://doi.org/10.1093/nsr/nwt032.

Halevy, A., Rajaraman, A. & Ordille, J. (2006) Data integration: the teenage years. In: *Proceedings of the 32nd International Conference on Very large Data Bases (VLDB '06)*, *12 – 15 September 2006, Seoul, Korea*. Los Angeles, USA, VLDB Endowment. pp. 9–16.

Han, J., Haihong, E, Le, G. et al. (2011) Survey on NoSQL Database. In: *2011 6th International Conference on Pervasive Computing and Applications*, *26-28 October 2011*, *Port Elizabeth, South Africa*. New York, USA, IEEE. pp. 363-366.

Khan, N., Yaqoob, I., Hashem, I. A. T. et al. (2014) Big Data: Survey, Technologies, Opportunities, and

Challenges. *The Scientific World Journal*. 2014, Art. ID 712826. https://doi.org/10.1155/2014/712826.

Kleven, H.-J., Knudsen, M. B., Kreiner, K. T. et al. (2011) Unwilling or Unable to Cheat? Evidence From a Tax Audit Experiment in Denmark. *Econometrica*. 79(3), 651–692. https://doi.org/10.3982/ECTA9113.

Mikalef, P., Pappas, I. O., Krogstie, J. et al. (2018) Big data analytics capabilities: a systematic literature review and research agenda. *Information Systems and E-Business Management*. 16(3), 547–578. https://doi.org/10.1007/s10257-017-0362-y.

Naritomi, J. (2019) Consumers as Tax Auditors. *American Economic Review*. 109(9), 3031–3072. https://doi.org/10.1257/aer.20160658.

OECD. (2017) *Technology Tools to Tackle Tax Evasion and Tax Fraud*. OECD.

Oprea S.-V., Bâra, A., Dobrița, G. et al. (2023) A Horizontal Tuning Framework for Machine Learning Algorithms Using a Microservice-based Architecture. *Studies in Informatics and Control*. 32(3), 31-43. https://doi.org/10.24846/v32i3y202303.

Pomeranz, D. (2015) No Taxation without Information: Deterrence and Self-Enforcement in the Value Added Tax. *American Economic Review*. 105(8), 2539–2569. https://doi.org/10.1257/aer.20130393.

Sahal, R., Breslin, J. G. & Ali, M. I. (2020) Big Data and Stream Processing Platforms for Industry 4.0 Requirements Mapping for a Predictive Maintenance Use Case. *Journal of Manufacturing Systems*. 54, 138-151. https://doi.org/10.1016/j.jmsy.2019.11.004.

Slemrod, J. (2019) Tax Compliance and Enforcement. *Journal of Economic Literature*. 57(4), 904–954. https://doi.org/10.1257/jel.20181437.

Slemrod, J. & Weber, C. (2012) Evidence of the Invisible: Toward A Credibility Revolution in the Empirical Analysis of Tax Evasion and the Informal Economy. *International Tax and Public Finance*. 19(1), 25–53. https://doi.org/10.1007/s10797-011-9181-0.

Wang, Y., Kung, L. & Byrd, T. A. (2018) Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations. *Technological Forecasting and Social Change*. 126, 3–13. https://doi.org/10.1016/j.techfore.2015.12.019.

Zaharia, M., Xin, R. S., Wendell, P. et al. (2016) Apache Spark: A Unified Engine for Big Data Processing. *Communications of the ACM*. 59(11), 56-65. https://doi.org/10.1145/2934664.

Zhang, L., Nan, X., Huang, E., & Liu, S. (2021). Social E-commerce Tax Evasion Detection Using Multimodal Deep Neural Networks. In: *2021 Digital Image Computing: Techniques and Applications (DICTA)*, *29 November - 1 December 2021, Gold Coast, Australia*. New York, USA, IEEE. https://doi.org/10.1109/DICTA52665.2021.9647152.