Revista JURÍDICA PORTUCALENSE

www.upt.pt

UNIVERSIDADE PORTUCALENSE
IJP
Instituto Jurídico Portucalense

Fundação para a Ciência e a Tecnologia

Nº 37 | Universidade Portucalense | Porto | 2025

https://doi.org/10.34625/issn.2183-2705(37)2025

Anastasiia CHYSTIAKOVA, Kostiantyn PODLIEHAIEV, Oleg KHORONOVSKYI, Tetiana SOKUR, Ivan KUBARIEV

The Role of Big Data Analytics in the Investigation of Corruption Offences

DOI: https://doi.org/10.34625/issn.2183-2705(37)2025.ic-6

Secção

Investigação Científica / Scientific Research*

* Os artigos presentes nesta secção não foram sujeitos a processo de revisão segundo o método *blind peer review* / The articles in this section have not undergone a blind peer review process.

The Role of Big Data Analytics in the Investigation of Corruption Offences

O Papel da Análise de Big Data na Investigação de Crimes de Corrupção

Anastasiia CHYSTIAKOVA[1]
Kostiantyn PODLIEHAIEV[2]
Oleg KHORONOVSKYI[3]
Tetiana SOKUR[4]
Ivan KUBARIEV[5]

ABSTRACT: The article deals with the integration of big data (BD) analytics in the investigation of corruption offenses and assesses its potential to detect complex criminal schemes. The use of large data volumes, such as financial transactions, communication patterns, and public records, significantly improves the effectiveness of investigations. Big Data (BD) analysis makes it possible to reveal hidden connections and anomalies, which provides new tools for the fight against corruption. The article combines legal analysis with technological innovation and explores the legal and ethical issues that arise when using such technologies. It focuses on how Big Data (BD) analytics can be integrated into the legal process within existing legal systems. The academic novelty is the study of the changes that Big Data (BD) analytics bring to the corruption investigation practices. The study also identifies data protection and compliance challenges that require the creation of an appropriate legal framework and developing specialized skills.

KEYWORDS: data analytics; corruption offenses; criminal law; criminology; criminological principles; criminal offense.

---

[1] Candidate of Law Sciences, Department of Criminal Procedure and Criminalistics, Educational and Scientific Institute of Law, Taras Shevchenko National University of Kyiv, Kyiv, Ukraine. anastasiya.spaska@gmail.com
[2] Candidate of Law Sciences, Department of Anti-Corruption and Law Enforcement Educational and scientific, Interregional Academy of Personnel Management, Kyiv, Ukraine. kmar.design22@gmail.com
[3] Candidate of Legal Sciences, Recipient of the Scientific Degree of Doctor of Sciences, National Academy of the State Security Service of Ukraine, Kyiv, Ukraine. olegseremet45@gmail.com
[4] Doctor of Philosophy, Department of Criminal, Civil and International Law, Law Faculty, National University "Zaporizhzhia Polytechnic", Zaporizhzhia, Ukraine. tanya.val1505@gmail.com
[5] PhD in Law, Donetsk State University of Internal Affairs, Kropyvnytskyi, Ukraine. educ660u@gmail.com

RESUMO: O artigo trata da integração de análises de big data (BD) na investigação de crimes de corrupção e avalia seu potencial para detectar esquemas criminosos complexos. O uso de grandes volumes de dados, como transações financeiras, padrões de comunicação e registros públicos, melhora significativamente a eficácia das investigações. A análise de Big Data (BD) permite revelar conexões e anomalias ocultas, o que fornece novas ferramentas para o combate à corrupção. O artigo combina análise jurídica com inovação tecnológica e explora as questões legais e éticas que surgem ao usar tais tecnologias. Ele se concentra em como a análise de Big Data (BD) pode ser integrada ao processo legal dentro dos sistemas jurídicos existentes. A novidade acadêmica é o estudo das mudanças que a análise de Big Data (BD) traz para as práticas de investigação de corrupção. O estudo também identifica desafios de proteção de dados e conformidade que exigem a criação de uma estrutura legal apropriada e o desenvolvimento de habilidades especializadas.
PALAVRAS-CHAVE: análise de dados; crimes de corrupção; direito penal; criminologia; princípios criminológicos; infração penal.

1. Introduction

The integration of Big Data (BD) analytics into the corruption offense investigation processes is an important step in the development of modern legal systems that face increasingly complex criminal schemes. Given the rapid digitization of the economy, the use of digital tools in anti-crisis management demonstrates the potential of data analytics to prevent and respond to corruption. Big Data (BD) analytics can support regulatory authorities in detecting irregular financial flows and identifying areas vulnerable to corruption risks, providing effective implementation[6]. Corruption scandals become more complex and cover wide networks, so traditional investigative methods often cannot keep up with the volume and complexity of information[7]. Therefore, the role of Big Data (BD) analytics in detecting corruption links becomes indispensable in the conditions of modern legal and technological realities. Anti-corruption judicial initiatives show that electronic declaration systems significantly contribute to increased transparency and accountability through real-time monitoring. These systems provide

[6] KABACHENKO, Dmytro et al. Implementation of digitization for anti-crisis management of business entities. Economic Affairs, 2023, vol. 68, no. 01s, pp. 361-369. https://doi.org/10.46852/0424-2513.1s.2023.39
[7] UNITED NATIONS. Introduction to Role of Data Analytics in Anti-Corruption and Fraud. [online]. 2024. Available from https://www.unodc.org/roseap/en/what-we-do/anti-corruption/topics/2021/data-analytics-anti-corruption-fraud.html

access to important financial information, which allows for more effective use of Big Data (BD) analytics to detect and investigate cases of corruption[8].

Big Data (BD) analytics tools can process huge amounts of information from a variety of sources, revealing hidden patterns, connections, and financial flows that would otherwise go unnoticed[9]. This is especially important for increasing transparency, ensuring accountability, and strengthening the effectiveness of law enforcement agencies in the fight against corruption at various levels of government and business. Anti-corruption efforts can become more precise, proactive, and transparent through Big Data (BD) analytics[10]. The implementation of a tough anti-corruption policy indicates a growing need for complex investigative tools to track complex multi-level corruption. The integration of Big Data (BD) analytics enables law enforcement agencies to better understand the complex networks used for illicit enrichment and facilitates the detection of corruption schemes across sectors[11]. The practice of the European Union emphasizes the importance of using data-based anti-corruption mechanisms to detect and prevent financial crimes. Big Data (BD) analytics is becoming a valuable tool, especially when combined with international anti-corruption standards, to track illicit financial transactions and facilitate international coordination of anti-corruption efforts[12].

The integration of big data (BD) analytics into corruption investigations represents a significant advance in modern legal systems. Corruption schemes are becoming increasingly complex, spanning multiple jurisdictions and involving large volumes of financial transactions, contracts, and communications records. The use of Big Data (BD) analytics can improve the detection of such schemes by identifying patterns, anomalies, and hidden relationships. However, its application must be assessed within the framework of existing legal norms governing corruption investigations.

[8] MAZIYCHUK, Vitaliy et al. The effectiveness of electronic declaration monitoring in the detection and investigation of corruption offences in the judicial system. Pakistan Journal of Life and Social Sciences, 2024, vol. 22, no. 2, pp. 4833-4845. https://doi.org/10.57239/pjlss-2024-22.2.00358

[9] N-IX. Fraud detection with big data analytics and machine learning: How to make it work. [online]. 2021. Available from https://www.n-ix.com/fraud-detection-big-data-analytics-machine-learning/

[10] UDEH, Ezekiel Onyekachukwu et al. The role of big data in detecting and preventing financial fraud in digital transactions. World Journal of Advanced Research and Reviews, 2024, vol. 22, no. 2, pp. 1746–1760. https://doi.org/10.30574/wjarr.2024.22.2.1575

[11] SHCHOKIN, Rostyslav et al. Sport management in the context of criminal liability for corruption. Retos, 2023, vol. 48, pp. 708–719. https://doi.org/10.47197/retos.v48.96768

[12] MELNYK, Dmytro S. et al. Practice of the member states of the European Union in the field of anti-corruption regulation. Journal of Financial Crime, 2021b, vol. 29, no. 3, pp. 853–863. https://doi.org/10.1108/jfc-03-2021-0050

The legal framework at both the international and national levels plays a crucial role in determining the acceptability and ethical use of Big Data (BD) analytics in corruption investigations. The United Nations Convention against Corruption (UNCAC)[13] sets a global standard for anti-corruption efforts, calling for improved investigative techniques, including the analysis of digital evidence. The OECD Convention on Combating Bribery[14] also criminalizes corruption in international business transactions, emphasizing the importance of cross-border cooperation. National anti-corruption laws, such as the US Foreign Corrupt Practices Act (FCPA)[15] and the UK Bribery Act,[16] include digital forensics techniques to track illicit financial flows. This study explores the potential of Big Data (BD) analytics in improving corruption investigations by analyzing the legal issues surrounding its use, including due process, admissibility of evidence, and confidentiality issues.

This study analyzes the role and potential of Big Data (BD) analytics in improving the investigation and prosecution of corruption offenses. It focuses on ways to use Big Data (BD) analytics to identify, track, and prosecute those involved in corruption. The research also analyzes the challenges that arise from the application of such technologies, including legal, ethical, and practical issues. The purpose is to provide recommendations for the optimal use of Big Data (BD) analytics in future legal practices.

The aim involves the fulfillment of the following research objectives:

1. Analyse the existing methods and tools of Big Data (BD) analytics for investigating corruption, including their effectiveness and limitations.

2. Study ethical issues related to the use of Big Data (BD) analytics in corruption cases, particularly in the context of data privacy and security.

3. Determine the impact of using Big Data (BD) analytics on the effectiveness of corruption investigations in different legal systems, particularly through comparative studies.

---

[13] UNITED NATIONS OFFICE ON DRUGS AND CRIME (UNODC). United Nations Convention against Corruption. [online]. 2004. Available from https://www.unodc.org/documents/brussels/UN_Convention_Against_Corruption.pdf

[14] ORGANISATION FOR ECONOMIC CO-OPERATION ANF DEVELOPMENT (OECD). Convention on Combating Bribery of Foreign Public Officials in International Business Transactions and Related Documents. [online]. 2024. Available from https://www.oecd.org/content/dam/oecd/en/topics/policy-sub-issues/fighting-foreign-bribery/Convention%20and%20commentaries%20booklet%202024.pdf

[15] UNITED STATES GOVERNMENT. Compilation of the National Environmental Policy Act of 1969. [online]. 2024. Available from https://www.govinfo.gov/content/pkg/COMPS-9569/pdf/COMPS-9569.pdf

[16] UNITED KINGDOM. Equality Act 2010. [online]. 2010. Available from https://www.legislation.gov.uk/ukpga/2010/23/contents

2. Literature Review

The growing complexity of corruption offenses and their significant socio-economic consequences require new approaches to investigations. Big Data (BD) analytics is one of those approaches, which is becoming an important tool in the fight against corruption. Oatley et al.[17] examine forensic intelligence processes, particularly the role of Big Data (BD) analytics in criminal investigations. They emphasize that the integration of forensic intelligence techniques improves data processing and analysis, which contributes to the detection of corruption schemes. The authors also emphasize the need for accurate data analysis to develop a methodology for integrating Big Data (BD) analytics into investigations and the importance of improving forensic standards to improve interagency cooperation. Das et al.[18] explore predictive analytics in the context of corruption offenses. They note that predictive models help to identify high-risk areas, which is important for effective response. The authors apply a gradual learning model for solving the problems related to dynamic data sets. They emphasize that adaptive algorithms are necessary for corruption investigations where the flow of data is unpredictable.

Oatley[19] analyses the application of data mining to the study of crime, particularly emphasizing the importance of Big Data (BD) analytics for identifying patterns of corrupt activity. It combines various methods of data analysis, including clustering and neural networks, to detect corrupt practices. The author also indicates the disadvantages of working with unstructured data and suggests improvements in natural language processing (NLP) methods to improve the accuracy of corruption analytics.

Hou et al.[20] developed an integrated graph model for predicting crime using an attentional approach to improve spatiotemporal analysis. Although the study focuses on

---

[17] OATLEY, Giles, CHAPMAN, Brendan and SPEERS, James. Forensic intelligence and the analytical process. Wiley Interdisciplinary Reviews Data Mining and Knowledge Discovery, 2020, vol. 10, no. 3. https://doi.org/10.1002/widm.1354

[18] DAS, Priyanka et al. Incremental classifier in crime prediction using bi-objective Particle Swarm Optimization. Information Sciences, 2021, vol. 562, pp. 279-303. https://doi.org/10.1016/j.ins.2021.02.002

[19] OATLEY, Giles C. Themes in data mining, big data, and crime analytics. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2021, vol. 12, no. 2, pp. e1432. https://doi.org/10.1002/widm.1432

[20] HOU, Miaomiao et al. An Integrated graph Model for Spatial–Temporal Urban Crime Prediction based on attention mechanism. International Journal of Geo-Information, 2022, vol. 11, no. 5, pp. 294. https://doi.org/10.3390/ijgi11050294

urban crime, attentional methods may be useful for monitoring changes in corruption hotspots. The authors demonstrated how these models can identify new areas of risk, which can also be adapted to track sectors prone to corruption threats. Schuilenburg and Soudijn[21] investigate the use of Big Data (BD) analytics in the work of the Dutch police, in particular its impact on the effectiveness of anti-corruption initiatives. They argue that big data helps to optimize the allocation of resources and strengthen anti-corruption activities. However, they note the importance of creating a regulatory framework that will ensure a balance between the effectiveness of technologies and the protection of citizens' rights. Lutska[22] studies digital technologies in anti-corruption activities, focusing on e-governance as a means of prevention. She notes that e-government can increase transparency and reduce corruption risks. However, the successful implementation of these initiatives requires a developed technological infrastructure and public trust, which is a difficult task for developing countries.

Rudyk[23] analyses the role of e-government in Ukraine's anti-corruption strategy, in particular the introduction of online services to facilitate public transactions. The author notes that infrastructure limitations may become an obstacle to the implementation of similar initiatives in other countries, although positive results are observed in Ukraine. Bauhr et al.[24] examine the impact of transparency on reducing corruption in public procurement, suggesting that BD can enhance transparency by providing access to information. The work of the authors emphasizes the importance of open data for reducing corruption risks in this area. Taylor[25] reviews Lagunas's book on anti-corruption, focusing on the dynamics of corruption in America and the effectiveness of control mechanisms. The review highlights the importance of detailed data analysis to

---

[21] SCHUILENBURG, Marc and SOUDIJN, Melvin. Big data in het veiligheidsdomein: onderzoek naar big-datatoepassingen bij de Nederlandse politie en de positieve effecten hiervan voor de politieorganisatie. Tijdschrift Voor Veiligheid, 2021, vol. 20, no. 4, pp. 44–62. https://doi.org/10.5553/tvv/.000028

[22] LUTSKA, Kateryna. Prospects of the use of modern technologies in counteraction to corruption in public administration. Bulletin of Taras Shevchenko National University of Kyiv Public Administration, 2021, vol. 14, no. 2, pp. 14-18. https://doi.org/10.17721/2616-9193.2021/14-3/6

[23] RUDYK, T. O. E-government in Ukraine as a means of combating corruption. In Management of the XXI Century: Modern Models, Strategies, Technologies. VI All-Ukrainian Scientific and Practical Internet Conference Collection of Scientific Papers. Vinnytsia: Tsentr Pidhotovky Naukovykh ta Navchalno-Metodychnykh Vydan VTEI KNTEU, 2019, vol. 2, pp. 517-524. Available from http://vtei.com.ua/doc/2019/22_10_2.pdf#page=517

[24] BAUHR, Monika et al. Lights on the shadows of public procurement: Transparency as an antidote to corruption. Governance, 2019, vol. 33, no. 3, pp. 495-523. https://doi.org/10.1111/gove.12432

[25] TAYLOR, Matthew M. Paul Lagunes, The Eye and the Whip: Corruption Control in the Americas. Oxford: Oxford University Press, 2021. Figures, tables, illustrations, bibliography, index, 168 pp.; hardcover $74, ebook. Latin American Politics and Society, 2022, vol. 64, no. 4, pp. 170-173. https://doi.org/10.1017/lap.2022.44

track changes in corruption trends. Akinbowale et al.[26] investigate the use of Big Data (BD) analytics to improve the accuracy of forensic accounting in the analysis of complex financial statements. They emphasize the need to standardize practices in different legal systems to improve the effectiveness of corruption investigations.

Navarrete and Gallego[27] apply a qualitative approach to evaluate forensic accounting tools, emphasizing their ability to prevent fraud. The study demonstrates that combining forensic accounting with big data can create an effective system for detecting fraudulent schemes. Kaur et al.[28] conducted a systematic review of the impact of forensic accounting on fraud detection, noting that the integration of Big Data (BD) analytics improves forensic methods. They suggest using interdisciplinary approaches to improve the accuracy of corruption investigations. Blahuta and Movchan[29] review crime investigation technologies, including digital forensic tools. They point to the great potential of Big Data (BD) analytics to transform criminal investigations but note the existing limitations to the availability of these technologies.

Despite significant contributions to the field, there are still certain gaps in the academic literature, particularly regarding the integration of machine learning (ML) for corruption data analysis and cross-border data sharing. There are also controversies surrounding privacy issues and the use of Big Data (BD) analytics, requiring the development of mechanisms to ensure a balance between the effectiveness of investigations and the protection of privacy rights. Addressing these issues will improve the accuracy and ethical aspects of using Big Data (BD) analytics in anti-corruption activities.

2.1. Regulatory Framework and Regulatory Analysis

The use of Big Data (BD) analytics in corruption investigations must comply with

---

[26] AKINBOWALE, Oluwatoyin Esther, KLINGELHÖFER, Heinz Eckart and ZERIHUN, Mulatu Fikadu. An innovative approach in combating economic crime using forensic accounting techniques. Journal of Financial Crime, 2020, vol. 27, no. 4, pp. 1253-1271. https://doi.org/10.1108/jfc-04-2020-0053

[27] NAVARRETE, Alberto Clavería and GALLEGO, Amalia Carrasco. Forensic accounting tools for fraud deterrence: a qualitative approach. Journal of Financial Crime, 2022, vol. 30, no. 3, pp. 840–854. https://doi.org/10.1108/jfc-03-2022-0068

[28] KAUR, Baljinder, SOOD, Kiran and GRIMA, Simon. A systematic review on forensic accounting and its contribution towards fraud detection and prevention. Journal of Financial Regulation and Compliance, 2022, vol. 31, no. 1, pp. 60-95. https://doi.org/10.1108/jfrc-02-2022-0015

[29] BLAHUTA, R. I. and MOVCHAN, A. V. The Latest Technologies in the Investigation of Crimes: The Current State and Use Problems. Lviv: LvDUVS, 2020. Available from https://dspace.lvduvs.edu.ua/handle/1234567890/3337

established legal principles regarding the collection, processing, and use of data. The rights to due process, enshrined in the European Convention on Human Rights (ECHR)[30] and the International Covenant on Civil and Political Rights (ICCPR)[31], require that any evidence used in criminal proceedings be obtained lawfully and respect fundamental rights.

International and domestic legal standards United Nations Convention against Corruption (UNCAC)[32]: encourages states to take measures that facilitate the detection of corruption through digital and forensic analysis.

European Union Anti-Corruption Directives[33]: establish rules on cross-border data exchange and the admissibility of electronic evidence in judicial proceedings. National anti-corruption laws: Countries such as Germany (Corruption Crimes StGB[34]), France (Sapien II Act[35]) and Brazil (Clean Company Act[36]) have provisions regarding the use of electronic evidence in corruption cases.

The role of Big Data (BD) analytics in corruption investigations raises concerns regarding the principle of proportionality under Article 52 of the Charter of Fundamental Rights of the European Union (CFREU)[37]. While data analytics increases detection capabilities, it should not infringe on privacy rights beyond what is necessary for law enforcement purposes.

3. Methods

[30] EUROPEAN COURT OF HUMAN RIGHTS. Convention for the Protection of Human Rights and Fundamental Freedoms. [online]. 2025. Available from https://www.echr.coe.int/documents/d/echr/convention_ENG

[31] UN TREATY COLLECTION. International Covenant on Civil and Political Rights (ICCPR). United Nations, 1976. Available from https://treaties.un.org/doc/treaties/1976/03/19760323%2006-17%20am/ch_iv_04.pdf

[32] UNITED NATIONS OFFICE ON DRUGS AND CRIME (UNODC). United Nations Convention against Corruption. [online]. 2004. Available from https://www.unodc.org/documents/brussels/UN_Convention_Against_Corruption.pdf

[33] EUROPEAN COMMISSION. EU legislation on anti-corruption. [online]. 2025. Available from https://commission.europa.eu/strategy-and-policy/policies/justice-and-fundamental-rights/democracy-eu-citizenship-anti-corruption/anti-corruption/eu-legislation-anti-corruption_en

[34] FEDERAL MINISTRY OF JUSTICE. German Penal Code (StGB). [online]. 2025. Available from https://www.gesetze-im-internet.de/englisch_stgb/englisch_stgb.html

[35] FRENCH ANTI-CORRUOTION AGENCY. French Anti-Corruption Agency Guidelines. [online]. 2025. Available from https://www.agence-francaise-anticorruption.gouv.fr/files/files/French%20AC%20Agency%20Guidelines%20.pdf

[36] GLOBAL PRIVATE CAPITAL ASSOCIATION. Brazilian Clean Company Act. [online]. 2017. Available from https://www.globalprivatecapital.org/app/uploads/2017/03/Brazilian-Clean-Company-Act.pdf

[37] EU FUNDAMENTAL RIGHTS AGENCY. Article 52: Scope and interpretation of rights and principles. European Union Agency for Fundamental Rights. [online]. 2025. Available from https://fra.europa.eu/en/eu-charter/article/52-scope-and-interpretation-rights-and-principles

3.1. Research Design

Research stages (Figure 1):

1. Data collection (January - March 2023). The data was obtained from national anti-corruption agencies, law enforcement agencies, and open government sources.

2. Data pre-processing and cleansing (April 2023). All data were cleansed of duplicate, irrelevant, and incomplete information to ensure high-quality materials for further analysis.

3. Big Data (BD) analytics (May - July 2023). Modern algorithms and ML models were used to identify patterns and anomalies in data that could indicate corruption.

4. Data analysis and establishment of interrelationships of corruption offenses (August - September 2023). Identified regularities were compared with known cases of corruption, based on data from national and international databases.

5. Reporting and conclusions (October - December 2023). The analysis of the results made it possible to assess the effectiveness of using Big Data (BD) analytics for the investigation of corruption offences, and also proposed ways of integrating such methods into criminal investigations.

Figure 1. Research stages in the Integration of Big Data (BD) Analytics for Corruption Investigations

Source: developed by the author based on MiniTAB[38]

3.2. Sampling

[38] MINITAB. Data analysis, statistical & process improvement tools. [online]. 2024. Available from https://www.minitab.com/en-us/

A sample of 500 cases of corruption detected using Big Data (BD) analytics was formed for the study. The sample included criminal cases related to various types of corruption, from petty bribery to large-scale fraud. An important aspect was the collection of data from various sources: financial transactions, government contracts, whistleblower reports, and other records. The sample spanned different geographic locations, industries, and levels of corruption, which made it possible to examine the impact of Big Data (BD) analytics on crime detection. The data were selected using stratified random sampling to ensure a balanced representation of all types of corruption and to minimize bias.

3.3. Methods

The study includes a combination of methods for data collection and analysis:

1. Statistical analysis. Descriptive statistical methods such as mean, median, and standard deviation were used to assess the main characteristics of the data. Inferential statistics identified correlations between data anomalies and known corruption cases. Pearson’s correlation coefficient was used to measure the strength of the relationship between variables, and regression analysis was used to predict the probability of corruption offenses.

2. ML algorithms.

– *Random forest classifier* – used for classification tasks to distinguish between corrupt and non-corrupt organizations.

– *k-means clustering* – used to identify groups of suspicious activity that may not be obvious at first glance.

– *Isolation forest* – used to detect anomalies and help to detect outliers in financial transactions.

3. Network analysis. This method made it possible to study the relationships between individuals, organizations, and transactions involved in corruption cases. Network analysis uses social network analysis (SNA) to visualize the structure of corruption networks and graph theory to identify the central actors of the schemes. This approach reveals hidden patterns of cooperation and collusion that may be invisible without special analysis.

4. Results

4.1. Descriptive and Inferential Statistical Analysis

The descriptive statistics of the study provide an overview of the sample characteristics and main trends in the 500 corruption cases. Table 1 displays the mean, median, mode, and standard deviation of variables such as transaction amounts, frequencies of government contract irregularities, and whistleblower-reported incidents.

Table 1. Descriptive statistics of corruption indicators

Variable	Value	Median	Mode	Standard deviation
Transaction amount	$45,300	$42,100	$50,000	$15,000
Contract Irregularities	14 cases	12 cases	10 cases	8 cases
Whistleblower reports	4 reports	3 reports	3 reports	1.5 reports

Source: developed by the author based on World Bank Group[39], Our World in Data[40]

Inferential statistical analysis revealed an important relationship between anomalies in financial transactions and cases of corruption. The results showed a significant positive correlation between frequent anomalies and confirmed corruption cases (Pearson correlation coefficient $r = 0.73$, $p < 0.01$). Regression analysis also confirmed that the probability of corruption increases with the number of anomalies in financial transactions, with a prediction accuracy of 87% for cases where transaction figures exceed two standard deviations from the mean. The average transaction amount in corruption cases is $45,300, which is a typical indicator of the value of such transactions. While most transactions are close to this average, some amounts are significantly higher or lower. The most common transaction amount is $50,000,

[39] WORLD BANK GROUP. Corruption Perceptions Index Rank. Indicator Profile Prosperity Data 360. [online]. 2024. Available from https://prosperitydata360.worldbank.org/en/indicator/TI+CPI+Rank
[40] OUR WORLD IN DATA. Political corruption index. [online]. 2024. Available from https://ourworldindata.org/grapher/political-corruption-index

indicating possible thresholds for corrupt schemes. A standard deviation of $15,000 shows moderate variability in amounts, with both large outliers and transactions near the mean. An average of 14 violations were found per case in government and commercial contracts, indicating a high level of opacity, hidden interests, or violations of requirements. The most common number of violations is 10, which may be the standard for typical contract problems. A standard deviation of 8 cases indicates a significant variety in the number of violations, where their number is significantly higher than the average value in some cases. On average, there are 4 whistleblower reports per case, which is a low but stable level of reporting. The most common number of reports is 3, which may indicate organizational limitations or a standard level of awareness about official reports. A standard deviation of 1.5 indicates low variability in the number of reports, with a few extreme cases deviating from the mean.

4.2. ML Classification and Anomaly Detection

ML algorithms allow the classification and detection of anomalies, which is important for distinguishing between corrupt and non-corrupt organizations.

1. The Random Forest model achieved 91% accuracy in detecting signs of corruption. The results of the classification by industries and regions, shown in Figure 1 confirm the effectiveness of the model in various conditions.

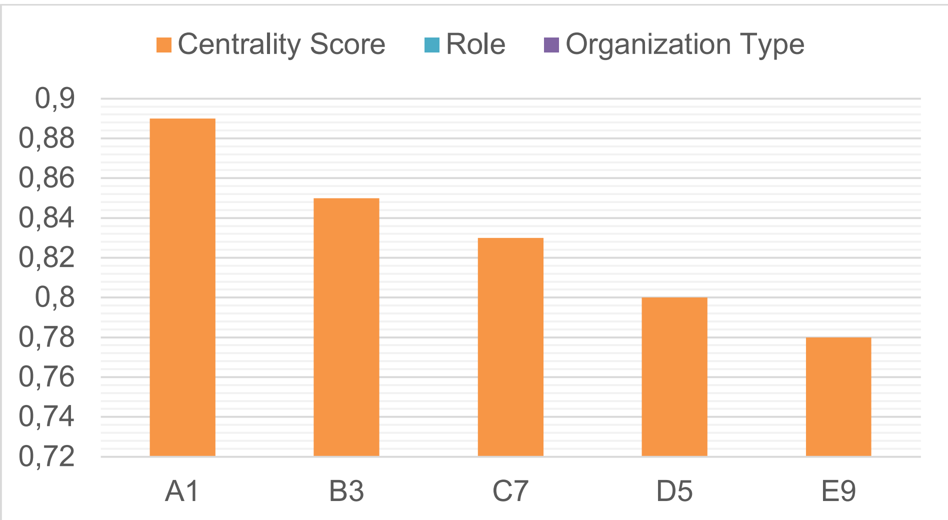

Figure 2. Classification of organizations using a random forest model

Source: Developed by the author based on Simplilearn[41], Donges[42]

Construction and energy are the sectors with the highest corruption rates: 70% of organizations in the construction sector and 65% in the energy sector are classified as corrupt. This may indicate an increased vulnerability of these industries to corruption risks, which is probably related to the significant financial transactions and frequent contracts with the state that are common in these sectors. In the public health sector, approximately 40% of organizations were classified as corrupt, while 60% were found to be non-corrupt. The relatively low level of corruption in this sector may be determined by a stronger regulatory framework or more complex corruption schemes that are difficult to detect using standard analytical methods. The public sector shows an almost even distribution, with about 50% of institutions classified as corrupt. Government organizations are often prone to corruption because of their vast resources and the influence they have on policymaking. The lowest corruption rate was recorded in the financial sector, approximately 25%. This may indicate a lower propensity for corruption in the industry or more effective internal control and compliance mechanisms. The overall indicators emphasize varying corruption rates across sectors, indicating increased vulnerability of industries with significant contracts or regulatory influence. The analysis also takes into account interregional differences, as the represented organizations come from different geographical areas, which increases the reliability of the study and allows for a broader assessment of corruption risks.

2. Figure 2 presents three key groups of suspicious activity patterns revealed by K-means clustering analysis. The first cluster covers transactions with high frequency and significant value, which are often characteristic of large-scale fraud schemes. The second cluster contains frequent transactions of small value, which usually indicate cases of petty bribery. The third cluster includes medium-sized irregular transactions that may indicate possible false reporting or procedural violations. These groups form the basis for targeted investigative measures in areas with high-risk activity patterns.

[41] SIMPLILEARN. Random Forest Algorithm. [online]. 2023. Available from https://www.simplilearn.com/tutorials/machine-learning-tutorial/random-forest-algorithm

[42] DONGES, Niklas. Random Forest: A complete guide for machine learning. Built In, 2024. Available from https://builtin.com/data-science/random-forest-algorithm

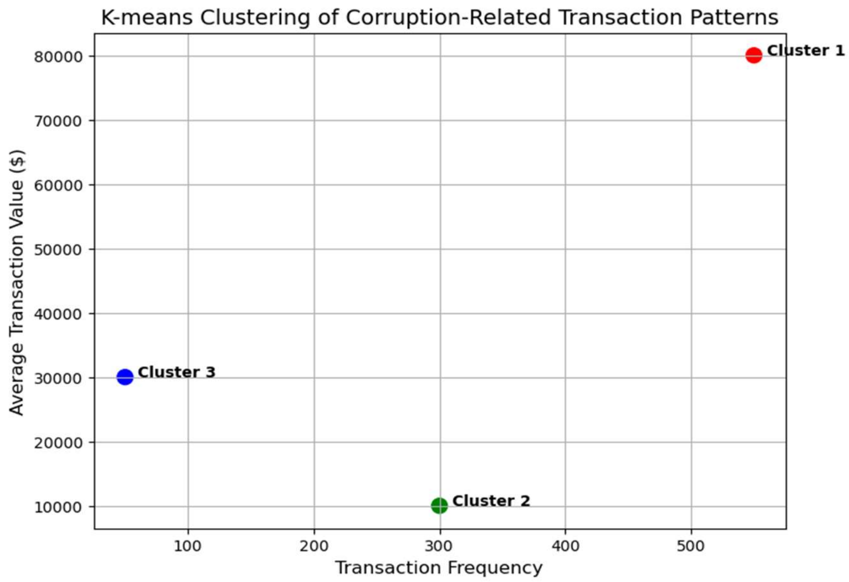

Figure 3. K-means clustering of corruption-related transaction patterns

Source: Developed by the author based on Riswanto[43]

3. The Isolation Forest algorithm revealed 37 anomalous cases of financial transactions that deviated significantly from typical patterns. These anomalies presented in Table 2 can serve as important indicators for further detailed analysis.

Table 2. Summary detected by the Isolated Forest algorithm

Transaction ID	Anomaly score	Transaction cost ($)	Industry sector	Country
10234	0.97	$98,500	Construction	Country A
10567	0.95	$120,000	Energy	Country B
10789	0.92	$150,000	Health care	Country C

Source: Developed by the author based on Scikit-Learn[44]

Transaction identifiers (10234, 10567, 10789) are used to detect unusual financial transactions. An anomaly score indicates how different a transaction is from others in the data set. It is calculated using the Isolated Forest algorithm: the higher the score (closer to 1), the more abnormal the transaction. For example, transaction ID 10234 has a score of 0.97, indicating a significant difference from other transactions.

[43] RISWANTO, Ujang. K-Means Clustering for Anomaly Detection. Medium, 2023. Available from https://ujangriswanto08.medium.com/k-means-clustering-for-anomaly-detection-1bbbb0b20b52

[44] SCIKIT-LEARN. IsolationForest. [online]. 2024. Available from https://scikit-learn.org/1.5/modules/generated/sklearn.ensemble.IsolationForest.html

Transactions 10567 and 10789 have scores of 0.95 and 0.92, which also confirms their anomaly. The transaction cost reflects its amount in US dollars. Outliers typically have high amounts that do not fit typical financial models for certain industries. For example, transaction 10234 at $98,500 is well above average amounts in the construction sector. Transactions 10567 and 10789 of $120,000 and $150,000 in the energy and healthcare sectors are also large, which may indicate possible corruption, including bribery, kickbacks, or fraud. Corruption is often associated with certain industries where there is a greater likelihood of improper financial transactions. Transaction 10234 in the construction sector may be related to bribery in public procurement. Transaction 10567 in the energy sector may indicate fraud, and 10789 in health care may indicate kickbacks or misuse of public funds. Anomalies can be a sign of international corruption because they cover different countries.

4.3. Network Analysis of Corruption Relations

The study of the network revealed hidden patterns of interaction between individuals and organizations in corruption cases.

1. Social Network Analysis (SNA) visualizes the structure of corruption networks, indicating key nodes, connections, intermediary organizations, and clusters of coordinated activity shown in Figure 3.

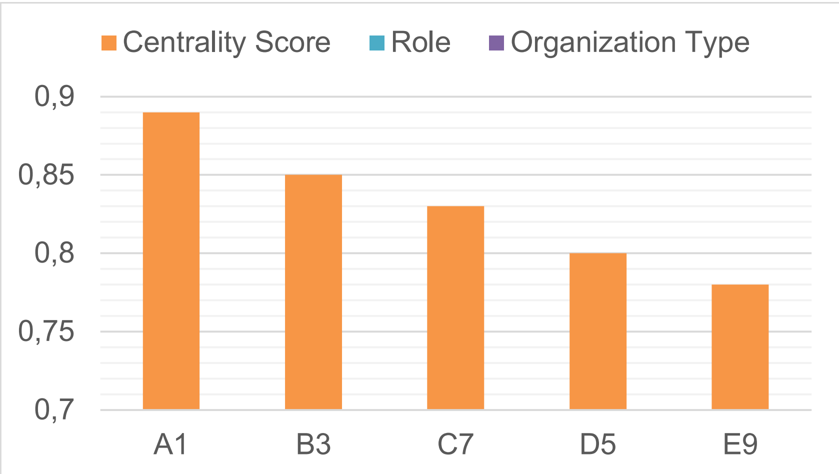

Figure 4. Visualization of subjects of corruption in social networks

Source: Developed by the author based on Ortiz-Ospina and Roser[45], Nash et al.[46]

[45] ORTIZ-OSPINA, Esteban and ROSER, Max. Corruption. Our World in Data, 2024. Available from https://ourworldindata.org/corruption

[46] NASH, Richard et al. Understanding corruption and social norms: a case study in Natural resource management. RTI Press, 2023. https://doi.org/10.3768/rtipress.2023.op.0089.2309

Nodes in a diagram are individual units that can represent individuals, organizations, or government agencies. Nodes with high centrality are key network participants actively interacting in corruption schemes. Centrality is measured by several criteria: degree (number of connections), betweenness (how often a node connects other nodes), and proximity (how quickly a node can reach other participants). Nodes with high centrality usually play an important role in spreading corrupt activities. Clusters of nodes formed due to close ties between participants may indicate groups of individuals or organizations prone to joint corrupt practices. For example, government officials regularly working with private contractors may be part of organized bribery schemes. Identifying such key participants through the visualization of diagrams allows investigators to focus on important elements of corruption networks. The method of social network analysis (SNA) is an effective tool for identifying hidden connections that may not be obvious with traditional methods of investigation. For example, a financial institution that appears isolated in a diagram may turn out to be an important element of a corruption scheme because of its connections with other participants. Visualization of the network highlights isolated nodes that indicate individual participants, as well as densely connected areas that indicate a high level of cooperation.

2. Analysis of graph theory. Analysis of centrality in the network revealed influential nodes where organizations with a high level of centrality contributed to the spread of corruption. Table 3 presents the 10 most central nodes and their role in corruption schemes.

Table 3. Top-10 central nodes of corruption networks

Node ID	Centrality score	Role	Organization type
A1	0.89	Facilitator	Financial institution
B3	0.85	Key mediator	Consulting company
C7	0.83	Primary beneficiary	Public sector
D5	0.81	Co-author	Private contractor
E2	0.79	Coordinator	State agency
F6	0.77	Provider	Construction company

G4	0.75	Investor	Private equity firm
H8	0.73	Performer	Law firm
I9	0.72	Bribery recipient	Civil servant
J10	0.71	Scheme organizer	State enterprise

Source: Developed by the author based on Council of Europe[47], OECD[48]

Identification of network nodes determines specific participants, for example, A1, B3, and C7, which may be individuals or organizations involved in corruption schemes. Each node has a unique identifier to track its role in the network. The centrality score, which is measured through graph theory metrics, specifically degree centrality, reflects the importance of a node in the network. A higher score (from 0.71 to 0.89) indicates a greater influence of the node. For example, node A1, with a score of 0.89 has significant influence on the corruption network, while node J10, with a score of 0.71, has less influence. The role of a node indicates the function of a person or organization within a corruption scheme. Node A1, marked Intermediary, coordinates financial transactions such as money laundering or illegal payments. Node B3, the Key Intermediary from a consulting firm, provides communication between parties involved in corruption, and node C7 – the Primary Beneficiary – from the public sector, benefits from corrupt deals. The organization type helps to understand the context of corruption. For example, financial institutions (A1) may carry out illegal transactions, consulting firms (B3) help to legalize schemes, and government organizations (C7) may engage in bribery. The nodes with high centrality, such as A1, B3, and C7, are key players in the corruption network, playing an important role in coordinating and spreading corrupt activities through multiple connections. The analysis of these nodes reveals which participants have the greatest influence on corruption processes and where attention should be focused during investigations. A high level of centrality indicates the importance of these actors for revealing the scale of corruption schemes and their interrelationships between different sectors and actors.

4.4. Due Process and Admissibility of Evidence

[47] COUNCIL OF EUROPE. Network of Corruption Prevention Authorities. [online]. 2023. Available from https://www.coe.int/en/web/corruption/ncpa-network

[48] OECD. Anti-Corruption Network for Eastern Europe and Central Asia. [online]. 2024. Available from https://www.oecd.org/en/networks/anti-corruption-network-for-eastern-europe-and-central-asia.html

The use of Big Data (BD) analytics as evidence in legal proceedings varies across legal systems.

Common law systems (US, UK): Courts assess the admissibility of Big Data (BD) analytics evidence based on chain of custody, reliability, and exclusionary rules (e.g., Fourth Amendment limitations).

Civil law systems (Germany, France): Electronic evidence must be obtained through legal authorisation that ensures due process.

A key legal challenge is the potential for algorithmic bias in Big Data (BD) analytics, which could affect the neutrality of investigations. Courts should establish clear standards of proof, ensuring that machine-generated results are subject to rigorous review.

5. Discussion

The study emphasizes the importance of using Big Data (BD) analytics to investigate corruption offenses and reveals new opportunities for identifying patterns and predicting risks. The results show that the integration of Big Data (BD) analytics with advanced analytical tools provides law enforcement agencies with more effective methods of detecting anomalies in financial transactions, procurement, and other sectors exposed to corruption risks. Analysis of large volumes of structured and unstructured data identifies individual cases of corruption but also reveals systemic vulnerabilities in state and corporate institutions. Several assumptions were made during the research process, including data availability and quality. As data from state archives, financial transactions and databases are considered accurate and complete, their manipulation can reduce the reliability of the obtained results. The effectiveness of Big Data (BD) analytics also depends on the technological infrastructure and the qualifications of researchers. These assumptions affect research results, especially when data are limited or of poor quality.

Akinbowale et al.[49] investigated the use of forensic accounting to combat economic crime. Their work focuses on traditional methods, but our study extends their

[49] AKINBOWALE, Oluwatoyin Esther, KLINGELHÖFER, Heinz Eckart and ZERIHUN, Mulatu Fikadu. An innovative approach in combating economic crime using forensic accounting techniques. Journal of Financial Crime, 2020, vol. 27, no. 4, pp. 1253-1271. https://doi.org/10.1108/jfc-04-2020-0053

findings by adding the role of Big Data (BD) analytics as a complementary tool in forensic investigations. The integration of Big Data (BD) analytics enables the use of a more dynamic approach to analyze financial transactions in real-time and predict future risks. This improves judicial investigations, in particular when detecting complex corruption schemes. Research by Ali and Jasim[50] highlights the importance of transparency in public procurement to reduce corruption. Our research supports their findings but adds that big data analytics can further increase transparency by revealing suspicious patterns and anomalies. The use of such tools automates the monitoring process and ensures its updating in real time.

Blahuta and Movchan[51] examine modern technologies used in the investigation of crimes, noting their limitations in anti-corruption activities. Our research supports their findings, but identifies the unique capabilities of BD analytics to monitor complex and multidimensional corruption offences, particularly cross-border ones. The use of BD makes it possible to achieve effective solutions that exceed the capabilities of traditional methods. The study by Das et al.[52] on crime prediction using Particle Swarm Optimization (PSO) points to the potential of predictive analytics in criminal investigations. Although their research also focuses on forecasting, our work aims to apply Big Data (BD) analytics to detect corrupt practices, allowing for the prediction of potential violations before they occur. The integrated graph model of Hou et al.[53] for predicting crime has common features with our study, especially in the use of complex models to predict crime. However, our approach is more focused on corruption, in particular the detection of illegal financial flows and inefficient management of public resources.

Kaur et al.[54] emphasize the role of accounting tools in detecting fraud. We extend

---

50 ALI, Sufyan Lateef and JASIM, Zeyad Tareq. The Effectiveness of International Rules in the Fight against Corruption. Revista Jurídica Portucalense, 2024, no. 36, pp. 135-161. https://doi.org/10.34625/issn.2183-2705(36)2024.ic-6

51 BLAHUTA, R. I. and MOVCHAN, A. V. The Latest Technologies in the Investigation of Crimes: The Current State and Use Problems. Lviv: LvDUVS, 2020. Available from https://dspace.lvduvs.edu.ua/handle/1234567890/3337

52 DAS, Priyanka et al. Incremental classifier in crime prediction using bi-objective Particle Swarm Optimization. Information Sciences, 2021, vol. 562, pp. 279-303. https://doi.org/10.1016/j.ins.2021.02.002

53 HOU, Miaomiao et al. An Integrated graph Model for Spatial–Temporal Urban Crime Prediction based on attention mechanism. International Journal of Geo-Information, 2022, vol. 11, no. 5, pp. 294. https://doi.org/10.3390/ijgi11050294

54 KAUR, Baljinder, SOOD, Kiran and GRIMA, Simon. A systematic review on forensic accounting and its contribution towards fraud detection and prevention. Journal of Financial Regulation and Compliance, 2022, vol. 31, no. 1, pp. 60-95. https://doi.org/10.1108/jfrc-02-2022-0015

this vision by considering Big Data (BD) analytics as a more scalable real-time alternative. Traditional methods can detect fraud, but Big Data (BD) analytics enables the analysis of multiple data sets from different sources. Rudyk[55] emphasizes the importance of e-governance for reducing the corruption rate in Ukraine. We support his view, but point out that the integration of Big Data (BD) analytics with e-government platforms can significantly improve the process of detecting and preventing corruption. Automation of Big Data (BD) analytics provides a deeper understanding of the activities of state institutions. Lutska[56] examines the role of modern technologies in the anti-corruption activities. Our results are consistent with her findings, but they provide a clearer picture of the application of Big Data (BD) analytics specifically for the detection of corrupt practices. We offer tools that can not only respond to breaches but also anticipate them before they escalate. Navarrete and Gallego[57] propose a forensic accounting approach to fraud prevention. However, our research adds emphasis to the role of big data in the real-time detection of fraud and corruption, as well as the application of predictive analytics to predict breaches. Oatley[58] examines data mining and criminal analytics, emphasizing the importance of Big Data (BD) analytics for improving criminal investigations. Our study extends this approach by showing how big data can be effectively applied to investigate corruption offenses, particularly in sectors where traditional methods are ineffective. Schuilenburg and Soudijn[59] confirm the importance of Big Data (BD) analytics in policing, but our study focuses on the specific needs of corruption investigations. We demonstrate how Big Data (BD) analytics can help to solve complex problems of corruption in public and corporate institutions.

---

55 RUDYK, T. O. E-government in Ukraine as a means of combating corruption. In Management of the XXI Century: Modern Models, Strategies, Technologies. VI All-Ukrainian Scientific and Practical Internet Conference Collection of Scientific Papers. Vinnytsia: Tsentr Pidhotovky Naukovykh ta Navchalno-Metodychnykh Vydan VTEI KNTEU, 2019, vol. 2, pp. 517-524. Available from http://vtei.com.ua/doc/2019/22_10_2.pdf#page=517

56 LUTSKA, Kateryna. Prospects of the use of modern technologies in counteraction to corruption in public administration. Bulletin of Taras Shevchenko National University of Kyiv Public Administration, 2021, vol. 14, no. 2, pp. 14-18. https://doi.org/10.17721/2616-9193.2021/14-3/6

57 NAVARRETE, Alberto Clavería and GALLEGO, Amalia Carrasco. Forensic accounting tools for fraud deterrence: a qualitative approach. Journal of Financial Crime, 2022, vol. 30, no. 3, pp. 840–854. https://doi.org/10.1108/jfc-03-2022-0068

58 OATLEY, Giles C. Themes in data mining, big data, and crime analytics. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 2021, vol. 12, no. 2, pp. e1432. https://doi.org/10.1002/widm.1432

59 SCHUILENBURG, Marc and SOUDIJN, Melvin. Big data in het veiligheidsdomein: onderzoek naar big-datatoepassingen bij de Nederlandse politie en de positieve effecten hiervan voor de politieorganisatie. Tijdschrift Voor Veiligheid, 2021, vol. 20, no. 4, pp. 44–62. https://doi.org/10.5553/tvv/.000028

Taylor[60] analyses the fight against corruption in America, offering a regional approach. Instead, our study demonstrates the global applicability of Big Data (BD) analytics to corruption investigations, with universal applicability beyond specific regions.

5.1. Privacy and Data Protection Considerations

A fundamental challenge in investigating corruption using Big Data (BD) analytics is compliance with data protection laws, in particular the General Data Protection Regulation (GDPR)[61] and national privacy laws.

Lawful processing: Investigators must establish a lawful basis for data collection, ensuring compliance with Article 6 of the GDPR.

Data minimisation: Big Data (BD) analytics should only process data that is necessary to avoid excessive surveillance.

Proportionality and necessity: Courts must balance the needs of the investigation with the right to privacy, as required by Article 8 of the ECHR.

One critical risk is the potential for mass surveillance of Big Data (BD) analytics, which could violate privacy rights if not properly regulated. Providing mechanisms for judicial oversight and independent monitoring can minimize these risks.

Despite the studies indicating the importance of using modern technologies, in particular Big Data (BD) analytics, to fight corruption, there are discrepancies between our results and the conclusions of previous studies. Some researchers emphasize the importance of transparency in anti-corruption initiatives, but our research shows that transparency without the use of analytical tools is not enough. Differences may be due to factors such as access to data, technological infrastructure, and political will.

5.2. Comparative Legal Analysis

A comparative analysis of different jurisdictions highlights different approaches to regulating Big Data (BD) analytics in corruption investigations.

[60] TAYLOR, Matthew M. Paul Lagunes, The Eye and the Whip: Corruption Control in the Americas. Oxford: Oxford University Press, 2021. Figures, tables, illustrations, bibliography, index, 168 pp.; hardcover $74, ebook. Latin American Politics and Society, 2022, vol. 64, no. 4, pp. 170-173. https://doi.org/10.1017/lap.2022.44

[61] GDPR-INFO.EU. General Data Protection Regulation (GDPR) EU Law. [online]. 2025. Available from https://gdpr-info.eu

United States: The Foreign Corrupt Practices Act (FCPA) allows electronic evidence in corruption cases provided it meets the Daubert standard of scientific reliability.

European Union: EU Directive 2016/680 regulates the processing of data by law enforcement agencies, requiring safeguards to prevent misuse.

China: The National Intelligence Law raises concerns about the government's power to oversee corruption investigations.

A comparison of these approaches highlights the need for a balanced regulatory framework that ensures effective investigations while protecting fundamental rights.

5.3. Doctrinal and Jurisprudential Support

The inclusion of legal science strengthens the case for the legitimate use of Big Data (BD) analytics in corruption investigations. Legal doctrines such as proportionality, necessity, and legitimate interest guide the application of Big Data (BD) analytics in international law.

Case law references: ECtHR case Weber and Saravia v. Germany[62]: examining the legality of mass surveillance in criminal investigations.

United States v. Microsoft[63]: examined the issue of cross-border access to data and privacy rights in law enforcement.

Digital Rights Ireland v. Minister for Communications[64]: held that mass data storage must be compatible with fundamental rights. These cases highlight the evolving nature of digital evidence and the need for strong legal safeguards.

5.4. Limitations

The main drawback of using Big Data (BD) analytics in corruption investigations is the threat of breach of confidentiality through the analysis of sensitive data. Besides,

---

[62] EUROPEAN COURT OF HUMAN RIGHTS. Case of (name of case) v. (name of respondent). [online]. 2024. Available from https://hudoc.echr.coe.int/fre#{%22itemid%22:[%22001-76586%22]}

[63] U.S. DEPARTMENT OF JUSTICE. Complaint in United States v. Microsoft Corporation. [online]. 1998. Available from https://www.justice.gov/atr/complaint-us-v-microsoft-corp

[64] EUROPEAN COURT OF JUSTICE (ECJ). Digital Rights Ireland Ltd v Minister for Communications, Marine and Natural Resources, C-293/12 and C-594/12. [online]. 2014. Available from https://globalfreedomofexpression.columbia.edu/cases/ecj-digital-rights-ireland-ltd-v-minister-for-communications-marine-and-natural-resources-c%E2%80%9129312-and-c%E2%80%9159412-2014/

the volume of information can lead to errors and the loss of important evidence, which complicates the investigation.

5.5. Recommendations

The effective use of Big Data (BD) analytics in corruption investigations requires the introduction of changes in legislation that will allow the use of the latest technologies, taking into account the confidentiality of information and the citizens' rights. International cooperation is important to create uniform legal standards for data exchange, which will ensure a coherent strategy to fight corruption without violating human rights.

6. Conclusions

The integration of Big Data (BD) analytics into the investigation of corruption offenses represents a significant advancement in modern legal and law enforcement frameworks. This study demonstrates that Big Data (BD) analytics enables the identification of complex financial patterns, facilitates anomaly detection, and enhances the predictive capabilities of law enforcement agencies. By processing vast quantities of data, including financial transactions and communication records, Big Data (BD) analytics reveals hidden links and systemic irregularities that traditional methods may overlook.

The research findings underscore that the application of Big Data (BD) analytics substantially improves investigative accuracy and accelerates the detection of corruption schemes. Machine learning algorithms, such as the Random Forest classifier and k-means clustering, effectively identify high-risk areas and highlight suspicious transactions. Furthermore, network analysis uncovers key actors and central nodes, shedding light on the organizational structures behind corrupt activities.

A crucial aspect of Big Data (BD) analytics in corruption investigations is its ability to process and analyze diverse datasets across multiple sectors and jurisdictions. This study reveals that industries with significant public sector interactions, such as construction and energy, exhibit higher rates of detected corruption. Moreover, the use of anomaly detection algorithms, including the Isolation Forest method, provides actionable insights for preemptive intervention in corruption-prone environments.

Legal and ethical considerations remain paramount in the use of Big Data (BD) analytics. The research emphasizes the need for a robust legal framework that upholds due process, protects individual privacy, and ensures the admissibility of machine-generated evidence. Compliance with international legal standards, including the General Data Protection Regulation (GDPR) and the European Convention on Human Rights (ECHR), is essential to maintaining the legitimacy of corruption investigations.

The study also highlights discrepancies in legal approaches across different jurisdictions. For instance, the United States relies on scientific reliability standards under the Foreign Corrupt Practices Act (FCPA), while the European Union mandates proportionality and privacy safeguards under Directive 2016/680. Harmonizing these regulatory frameworks is critical for effective cross-border cooperation and ensuring the integrity of Big Data (BD) analytics in legal proceedings.

To advance the practical implementation of Big Data (BD) analytics, the study recommends the following measures:

Development of a Unified Legal Framework: Establishing clear and standardized legal protocols for the collection, processing, and use of Big Data (BD) analytics in corruption investigations.

Enhanced International Collaboration: Promoting data-sharing agreements and aligning national regulations to facilitate cross-border investigations while respecting privacy and human rights standards.

Judicial Oversight and Transparency: Implementing independent monitoring mechanisms to ensure Big Data (BD) analytics are used proportionately and do not infringe on fundamental rights.

Technological Integration and Capacity Building: Investing in advanced machine learning models and providing specialized training for law enforcement personnel to improve the accuracy and efficiency of corruption investigations.

In conclusion, Big Data (BD) analytics offers a transformative approach to the investigation of corruption offenses. Future research should focus on refining predictive algorithms, addressing emerging privacy challenges, and developing best practices for the ethical use of Big Data (BD) analytics in the fight against corruption.

REFERENCES

AKINBOWALE, Oluwatoyin Esther, KLINGELHÖFER, Heinz Eckart and ZERIHUN, Mulatu Fikadu. An innovative approach in combating economic crime using forensic accounting techniques. *Journal of Financial Crime,* 2020, vol. 27, no. 4, pp. 1253-1271. https://doi.org/10.1108/jfc-04-2020-0053

ALI, Sufyan Lateef and JASIM, Zeyad Tareq. The Effectiveness of International Rules In The Fight Against Corruption. *Revista Jurídica Portucalense,* 2024, no. 36, pp. 135-161. https://doi.org/10.34625/issn.2183-2705(36)2024.ic-6

BAUHR, Monika et al. Lights on the shadows of public procurement: *Transparency as an antidote to corruption. Governance,* 2019, vol. 33, no. 3, pp. 495-523. https://doi.org/10.1111/gove.12432

BLAHUTA, R. I. and MOVCHAN, A. V. *The Latest Technologies in the Investigation of Crimes: The Current State and Use Problems*. Lviv: LvDUVS, 2020. Available from: https://dspace.lvduvs.edu.ua/handle/1234567890/3337

COUNCIL OF EUROPE. *Network of Corruption Prevention Authorities.* [online]. 2023. Available from: https://www.coe.int/en/web/corruption/ncpa-network

DAS, Priyanka et al. Incremental classifier in crime prediction using bi-objective Particle Swarm Optimization. *Information Sciences,* 2021, vol. 562, pp. 279-303. https://doi.org/10.1016/j.ins.2021.02.002

DONGES, Niklas. *Random Forest: A complete guide for machine learning.* Built In, 2024. Available from: https://builtin.com/data-science/random-forest-algorithm

EU FUNDAMENTAL RIGHTS AGENCY. *Article 52: Scope and interpretation of rights and principles.* European Union Agency for Fundamental Rights. [online]. 2025. Available from: https://fra.europa.eu/en/eu-charter/article/52-scope-and-interpretation-rights-and-principles

EUROPEAN COMMISSION. *EU legislation on anti-corruption.* [online]. 2025. Available from: https://commission.europa.eu/strategy-and-policy/policies/justice-and-fundamental-rights/democracy-eu-citizenship-anti-corruption/anti-corruption/eu-legislation-anti-corruption_en

EUROPEAN COURT OF HUMAN RIGHTS. *Case of (name of case) v. (name of respondent).* [online]. 2024. Available from: https://hudoc.echr.coe.int/fre#{%22itemid%22:[%22001-76586%22]}

EUROPEAN COURT OF HUMAN RIGHTS. *Convention for the Protection of Human Rights and Fundamental Freedoms.* [online]. 2025. Available from: https://www.echr.coe.int/documents/d/echr/convention_ENG

EUROPEAN COURT OF JUSTICE (ECJ). *Digital Rights Ireland Ltd v Minister for Communications, Marine and Natural Resources, C-293/12 and C-594/12.* [online]. 2014. Available from: https://globalfreedomofexpression.columbia.edu/cases/ecj-digital-rights-ireland-ltd-v-minister-for-communications-marine-and-natural-resources-c%E2%80%9129312-and-c%E2%80%9159412-2014/

FEDERAL MINISTRY OF JUSTICE. *German Penal Code (StGB).* [online]. 2025. Available at: https://www.gesetze-im-internet.de/englisch_stgb/englisch_stgb.html

FRENCH ANTI-CORRUOTION AGENCY. *French Anti-Corruption Agency Guidelines.* [online]. 2025. Available from: https://www.agence-francaise-anticorruption.gouv.fr/files/files/French%20AC%20Agency%20Guidelines%20.pdf

GDPR-INFO.EU. *General Data Protection Regulation (GDPR) EU Law.* [online]. 2025. Available from: https://gdpr-info.eu

GLOBAL PRIVATE CAPITAL ASSOCIATION. *Brazilian Clean Company Act.* [online]. 2017. Available from: https://www.globalprivatecapital.org/app/uploads/2017/03/Brazilian-Clean-Company-Act.pdf

HOU, Miaomiao et al. An Integrated graph Model for Spatial–Temporal Urban Crime Prediction based on attention mechanism. *International Journal of Geo-Information,* 2022, vol. 11, no. 5, pp. 294. https://doi.org/10.3390/ijgi11050294

KABACHENKO, Dmytro et al. Implementation of digitization for anti-crisis management of business entities. *Economic Affairs,* 2023, vol. 68, no. 01s, pp. 361-369. https://doi.org/10.46852/0424-2513.1s.2023.39

KAUR, Baljinder, SOOD, Kiran and GRIMA, Simon. A systematic review on forensic accounting and its contribution towards fraud detection and prevention. *Journal of Financial Regulation and Compliance,* 2022, vol. 31, no. 1, pp. 60-95. https://doi.org/10.1108/jfrc-02-2022-0015

LEXTEGRITY. *Data Analytics for Anti-Corruption Compliance.* [online]. 2024. Available from: https://www.lextegrity.com/resources/use-cases/data-analytics-for-anti-corruption-compliance

LUTSKA, Kateryna. Prospects of the use of modern technologies in counteraction to corruption in public administration. *Bulletin of Taras Shevchenko National University of Kyiv Public Administration,* 2021, vol. 14, no. 2, pp. 14-18. https://doi.org/10.17721/2616-9193.2021/14-3/6

MAZIYCHUK, Vitaliy et al. The effectiveness of electronic declaration monitoring in the detection and investigation of corruption offences in the judicial system. *Pakistan Journal of Life and Social Sciences,* 2024, vol. 22, no. 2, pp. 4833-4845. https://doi.org/10.57239/pjlss-2024-22.2.00358

MELNYK, Dmytro S. et al. Practice of the member states of the European Union in the field of anti-corruption regulation. *Journal of Financial Crime,* 2021, vol. 29, no. 3, pp. 853–863. https://doi.org/10.1108/jfc-03-2021-0050

MINITAB. *Data analysis, statistical & process improvement tools.* [online]. 2024. Available from: https://www.minitab.com/en-us/

MOHR, Tim and PADILLA, Victor. *Leveraging Data Analytics to Uncover FCPA Violations and Strengthen Compliance.* Weaver, 2024. Available from: https://weaver.com/resources/leveraging-data-analytics-to-uncover-fcpa-violations-and-strengthen-compliance/

N-IX. *Fraud detection with big data analytics and machine learning: How to make it work.* [online]. 2021. Available from: https://www.n-ix.com/fraud-detection-big-data-analytics-machine-learning/

NASH, Richard et al. *Understanding corruption and social norms: a case study in Natural resource management.* RTI Press, 2023. https://doi.org/10.3768/rtipress.2023.op.0089.2309

NAVARRETE, Alberto Clavería and GALLEGO, Amalia Carrasco. Forensic accounting tools for fraud deterrence: a qualitative approach. *Journal of Financial Crime,* 2022, vol. 30, no. 3, pp. 840–854. https://doi.org/10.1108/jfc-03-2022-0068

OATLEY, Giles C. Themes in data mining, big data, and crime analytics. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery,* 2021, vol. 12, no. 2, pp. e1432. https://doi.org/10.1002/widm.1432

OATLEY, Giles, CHAPMAN, Brendan and SPEERS, James. Forensic intelligence and the analytical process. Wiley Interdisciplinary Reviews Data Mining and Knowledge Discovery, 2020, vol. 10, no. 3. https://doi.org/10.1002/widm.1354

OECD. *Anti-Corruption Network for Eastern Europe and Central Asia.* [online]. 2024. Available from: https://www.oecd.org/en/networks/anti-corruption-network-for-eastern-europe-and-central-asia.html

ORGANISATION FOR ECONOMIC CO-OPERATION ANF DEVELOPMENT (OECD). *Convention on Combating Bribery of Foreign Public Officials in International Business Transactions and Related Documents.* [online]. 2024. Available from: https://www.oecd.org/content/dam/oecd/en/topics/policy-sub-issues/fighting-foreign-bribery/Convention%20and%20commentaries%20booklet%202024.pdf

ORTIZ-OSPINA, Esteban and ROSER, Max. *Corruption.* Our World in Data, 2024. Available from: https://ourworldindata.org/corruption

OUR WORLD IN DATA. Political corruption index. [online]. 2024. Available from: https://ourworldindata.org/grapher/political-corruption-index

RISWANTO, Ujang. *K-Means Clustering for Anomaly Detection.* Medium, 2023. Available from: https://ujangriswanto08.medium.com/k-means-clustering-for-anomaly-detection-1bbbb0b20b52

RUDYK, T. O. E-government in Ukraine as a means of combating corruption. In *Management of the XXI Century: Modern Models, Strategies, Technologies. VI All-Ukrainian Scientific and Practical Internet Conference Collection of Scientific Papers.* Vinnytsia: Tsentr Pidhotovky Naukovykh ta Navchalno-Metodychnykh Vydan VTEI KNTEU, 2019, vol. 2, pp. 517-524. Available from: http://vtei.com.ua/doc/2019/22_10_2.pdf#page=517

SCHUILENBURG, Marc and SOUDIJN, Melvin. Big data in het veiligheidsdomein: onderzoek naar big-datatoepassingen bij de Nederlandse politie en de positieve effecten hiervan voor de politieorganisatie. *Tijdschrift Voor Veiligheid,* 2021, vol. 20, no. 4, pp. 44–62. https://doi.org/10.5553/tvv/.000028

SCIKIT-LEARN. *IsolationForest.* [online]. 2024. Available from: https://scikit-learn.org/1.5/modules/generated/sklearn.ensemble.IsolationForest.html

SHCHOKIN, Rostyslav et al. Sport management in the context of criminal liability for corruption. *Retos,* 2023, vol. 48, pp. 708–719. https://doi.org/10.47197/retos.v48.96768

SIMPLILEARN. *Random Forest Algorithm.* [online]. 2023. Available from: https://www.simplilearn.com/tutorials/machine-learning-tutorial/random-forest-algorithm

TAYLOR, Matthew M. Paul Lagunes, The Eye and the Whip: Corruption Control in the Americas. Oxford: Oxford University Press, 2021. Figures, tables, illustrations, bibliography, index, 168 pp.; hardcover $74, ebook. *Latin American Politics and Society,* 2022, vol. 64, no. 4, pp. 170-173. https://doi.org/10.1017/lap.2022.44

U.S. DEPARTMENT OF JUSTICE. *Complaint in United States v. Microsoft Corporation.* [online]. 1998. Available from: https://www.justice.gov/atr/complaint-us-v-microsoft-corp

UN TREATY COLLECTION. *International Covenant on Civil and Political Rights (ICCPR).* United Nations, 1976. Available from: https://treaties.un.org/doc/treaties/1976/03/19760323%2006-17%20am/ch_iv_04.pdf

UDEH, Ezekiel Onyekachukwu et al. The role of big data in detecting and preventing financial fraud in digital transactions. *World Journal of Advanced Research and Reviews,* 2024. vol. 22, no. 2, pp. 1746–1760. https://doi.org/10.30574/wjarr.2024.22.2.1575

UNITED KINGDOM. *Equality Act 2010.* [online]. 2010. Available from: https://www.legislation.gov.uk/ukpga/2010/23/contents.

UNITED NATIONS. *Introduction to Role of Data Analytics in Anti-Corruption and Fraud.* [online]. 2024. Available from: https://www.unodc.org/roseap/en/what-we-do/anti-corruption/topics/2021/data-analytics-anti-corruption-fraud.html

UNITED NATIONS OFFICE ON DRUGS AND CRIME (UNODC). *United Nations Convention against Corruption.* [online]. 2004. Available from: https://www.unodc.org/documents/brussels/UN_Convention_Against_Corruption.pdf

UNITED STATES GOVERNMENT. *Compilation of the National Environmental Policy Act of 1969.* [online]. 2024. Available from: https://www.govinfo.gov/content/pkg/COMPS-9569/pdf/COMPS-9569.pdf

WORLD BANK GROUP. *Corruption Perceptions Index Rank. Indicator Profile Prosperity Data 360.* [online]. 2024. Available from: https://prosperitydata360.worldbank.org/en/indicator/TI+CPI+Rank

Data de submissão do artigo: 13/01/2025

Data de aprovação do artigo: 10/03/2025

Edição e propriedade:

Universidade Portucalense Cooperativa de Ensino Superior, CRL

Rua Dr. António Bernardino de Almeida, 541 - 4200-072 Porto

Email: upt@upt.pt