INNOVATIVE ANALYTICAL AND STATISTICAL TECHNOLOGY IN ELECTION FORENSICS

Yuliia Yatsyna,[1] Igor Kudinov[2]

Zaporizhzhia National University (Zaporizhzhia, Ukraine)

ABSTRACT
The article examines the crucial role of innovative analytical and statistical technology in electoral forensics, which are increasingly used for detecting and preventing electoral corruption and fraud. By analysing vast amounts of data and detecting anomalies, electoral forensic investigations can contribute to fair and transparent democratic processes. The research aims to explore the effectiveness of these technologies and their potential impact on improving the transparency and fairness of electoral processes, using a multi-method approach that includes analysing relevant documents, media coverage, public opinion, and recent fraud cases. The authors divide the implementation of innovative analytical and statistical technologies for combating election corruption into four groups. The first is the analysis of statistical data and research on corruption, including election processes, which can be called secondary data analysis. The second is the analysis of documentary data containing information on corrupt actions and offences, including election processes. The third is the development of mathematical methods and algorithms using cutting-edge technologies such as artificial intelligence and machine learning for detecting anomalies and hidden patterns. The fourth is experimental developments in information technologies as a means of ensuring proper governance and combating corruption. While the use of algorithms for detecting anomalies in electoral statistics data can be an important tool, it should be used with caution, and in combination with other sources of information, to avoid the consequences of delegitimising the election results.
KEY WORDS: *election fraud, election corruption, forensics, Benford's Law, artificial intelligence, machine learning*.

JEL CODE: D73.
DOI: https://doi.org/10.15181/rfds.v40i2.2528

Introduction

Fraud and corruption are a barrier to the progressive socio-economic development of any state. They distort moral standards, reduce the level of trust by citizens in government institutions, and are rather difficult and elusive phenomena for researchers. The blurriness of the manifestations of fraud and corruption causes a difference in the opinions of researchers regarding the essence of these phenomena, as well as endless discussions regarding the definition of their essence, causes, consequences, and ways of prevention. One of the most significant manifestations of fraud and corruption in the political sphere is electoral fraud, or electoral corruption, which refers to the manipulation of the electoral process for personal gain. Election corruption and election fraud occur during the electoral process. Election corruption or election fraud is an activity or behaviour that is intended to manipulate or influence the outcome of an election. It is a serious threat to the democratic process, as it undermines the principle of free and fair elections.

In recent years, there has been growing recognition of the potential for information technology to play a critical role in addressing this challenge (*Computational Social Science: Discovery and Prediction*, 2016).

[1] Yuliia Yatsyna – head of the Union of Social Engineers of Ukraine NGO (Zaporizhzhia, Ukraine)
Scientific interests: civil society organisations, political corruption
yatsyna.yuliia@gmail.com

[2] Igor Kudinov PhD – associate professor in the Department of Sociology, Zaporizhzhia National University (Zaporizhzhia, Ukraine)
Scientific interests: business intelligence, social audit, life actor
ikudinov@cisr.org.ua

By leveraging innovative technologies, such as blockchain, artificial intelligence, and big data analytics, governments, civil society groups and other stakeholders explore new ways to detect and prevent electoral fraud, improve transparency and accountability, and enhance public confidence in the integrity of electoral processes. The digitisation of the state is considered one of the solutions in this matter. The digital transformation of society has a serious anti-corruption meaning and dimension. The development of fundamentally new mechanisms of public administration based on information and communication technologies allows the development of previously unknown and completely unexpected means of countering corruption in the system of public administration, including election procedures. At the same time, the development of information technologies can give rise to new corruption and bureaucratic schemes, which in its potential can be reduced to electronic bureaucracy, electronic corruption, or even a digital concentration camp.

This is why the formation of an anti-corruption policy with the use of innovative technologies requires serious scientific analysis and the formulation of such directions of state development on the way to a digital format of work to exclude the possibility of the development of the above-mentioned negative trends and consequences of digitalisation, and become the basis of fundamentally new mechanisms for regulating state administration, public authorities, law enforcement activities, and, of course, election processes. By highlighting the benefits and challenges of utilising technology in the fight against election corruption, this article aims to contribute to the ongoing conversation surrounding the role of innovation in election forensics, such as the process of examining and analysing election data, processes and outcomes, to identify patterns, detect anomalies, and assess the integrity of an election (Klimek, Jimenez, Hidalgo, Hinteregger, Thurner, 2018). It involves the use of various statistical, computational and social science methods to evaluate the quality of the electoral process, ensure transparency, and detect possible instances of fraud or manipulation. Election forensics can include examining voter registration data, voting patterns, turn-out rates, and the distribution of votes across different demographics and geographic areas. It can also involve analysing the procedures and mechanisms used in the election process, such as voter identification, ballot design, vote counting, and result reporting, to ensure they are fair and unbiased. The goal of election forensics is to help maintain public trust in the electoral system, by ensuring that elections are conducted fairly, transparently and accurately, and to identify any issues that may compromise the integrity of the process. One of these directions of research can be considered the analysis of scientific and practical projects implemented, the results of which are directly related to, or based on, the active involvement of the latest information, including analytical and statistical technologies to detect election corruption.

The article will focus on innovative analytical and statistical technologies used in electoral forensics (the purpose of the research). The use of these advanced techniques has become increasingly important in identifying and preventing electoral corruption and fraud. By analysing vast amounts of data and detecting anomalies, electoral forensic investigations can provide crucial evidence for legal action and contribute to fair and transparent democratic processes. The main tasks of the research are to explore the effectiveness of innovative analytical and statistical technologies in detecting and preventing electoral fraud, and to examine the potential impact of these technologies on improving the transparency and the fairness of electoral processes. To achieve these goals, a multi-method approach is used, including the analysis of relevant documents, content analysis of media coverage and public opinion, and case studies of recent electoral fraud cases. This methodology will provide a comprehensive understanding of the current state of electoral corruption and the role of technology in addressing this critical issue.

1. Analysis of terms

To address the issue of electoral corruption and the use of innovative analytical and statistical technologies for forensic investigations effectively, it is essential to begin with a clear understanding of the key concepts and terms used in this field.

By corruption, we understand ‘the misuse of public office for private gain’ (Rose-Ackerman, Palifka, 2016), which can have both a material and a non-material form. At the same time, misuse is a violation of

both formal regulatory and legal institutions, including norms of official behaviour and ethics, as well as informalised norms of behaviour, ethics and morality.

Fraud is 'the use of one's occupation for personal enrichment through the deliberate misuse or misapplication of the employing organisation's resources or assets' (Occupational Fraud 2022: A Report to the Nations, 2022, 6). According to J. Wells (2017), corruption is one of the forms of corporate fraud.

For the research aim, it is essential to understand the distinction between the two, in order to determine when one can be considered primary and the other secondary.

Electoral fraud refers to illegal interference in the process of an election. It can involve various activities, such as vote rigging, ballot stuffing, voter intimidation, or vote buying (Dechert, 2013). Electoral corruption, on the other hand, is a form of political corruption in which politicians, parties or other actors use illegal means to influence election outcomes in their favour (Birch, 2011). This can involve bribery, the abuse of power, or the misuse of public resources.

In cases where electoral fraud is the primary phenomenon and electoral corruption is secondary, the primary objective is to manipulate the electoral process for personal or political gain. The fraudsters aim to affect the election results directly, without any pre-existing corrupt agreements or intent. The subsequent electoral corruption may arise as a by-product, as individuals or parties take advantage of fraudulent practices to further their interests. For example, a political party might engage in electoral fraud by tampering with the vote count to secure a victory. In doing so, they create an environment conducive to corruption, as elected officials may feel emboldened to abuse their power or engage in bribery, knowing that they won through illicit means.

Conversely, in situations where electoral corruption is the primary phenomenon and electoral fraud is secondary, the main objective is to engage in corrupt activities to secure power, wealth, or other benefits. Electoral fraud may be employed as a means to achieve these corrupt ends. For instance, a group of politicians might enter into a corrupt agreement to exchange favours or resources. To ensure that the agreement is fulfilled, they may resort to electoral fraud, such as vote-buying or intimidation, to secure the necessary electoral outcome.

The main difference between fraud and corruption is that fraud can exist at an individual level, whereas corruption exists at a group level only, because this activity involves collusion between participants (Kratcoski, 2018).

In summary, while electoral fraud and electoral corruption can be interconnected, their relationship depends on the specific circumstances and objectives involved. When electoral fraud is the primary focus, with electoral corruption emerging as a by-product, the primary goal is to manipulate the electoral process. In contrast, when electoral corruption is the main objective, with electoral fraud employed as a means to achieve it, the central aim is to engage in corrupt activities for personal or political gain. Either way, both election fraud and election corruption are detrimental to the democratic process, as they can lead to unfair elections and undermine public trust in the political system.

Fraud or corruption detection tools are defined as 'data processing systems driven by tasks or problems designed to, with a degree of autonomy, identify, predict, summarize, and/or communicate actions related to the misuse of position, information and/or resources aimed at private gain at the expense of the collective good' (Odilla, 2023).

Election fraud and election corruption are complex phenomena that can be difficult to detect and investigate. This is where electoral criminology or election forensics come into play. 'Election forensics' is a multidisciplinary field that utilises scientific methods, statistical analyses and investigative techniques to examine and detect electoral irregularities, fraud and manipulation. It involves the application of forensic principles to the electoral process, with the goal of ensuring fair and transparent elections. Election forensics covers a wide range of activities, including analysing election data and records, examining voting equipment and technologies, monitoring election administration and procedures, and investigating allegations of electoral misconduct (Lacasa, Fernandez-Gracia, 2019).

Election forensics may involve various methods and techniques, including:

1. Statistical analysis – examining election data, such as voter turn-out, vote shares, and margins of victory, to identify unusual patterns or anomalies that might suggest fraud or manipulation. This can include techniques like Benford's Law, which looks at the distribution of digits in the data, or regression analysis to detect deviations from expected patterns (Nigrini, 2012; Zhang, Alvarez, Levin, 2019).
2. Geographical information systems (GIS) – mapping and analysing the spatial distribution of election data, such as polling locations, voter registration and election results, to identify patterns or discrepancies that could indicate irregularities (Centre for Research Transparency and Accountability, 2020; Shalaev, 2016).
3. Qualitative analysis – reviewing election procedures, voter registration processes, and ballot counting methods to identify weaknesses or potential areas for manipulation (Fisher, Hamilton, 2020).
4. Digital forensics – examining electronic voting systems, digital records and communication channels to detect potential hacking, tampering, or other forms of cyber interference (Lawless, 2022).
5. Auditing – verifying the accuracy and integrity of election results by comparing paper records or voter-verified paper audit trails with electronic vote tallies (Sridhar, Rivest, 2018).

Innovative analytical and statistical technologies are defined as:

- in a broad sense, as a set of methods and tools based on the use of mathematical and statistical methods of data analysis in order to identify useful dependencies and regularities in data, increase the efficiency of decision-making, and identify anomalies in various spheres of activity;
- in a narrow sense, as a process of using the most advanced methods and technologies of data analysis, such as machine learning, deep learning, neural networks, natural language processing, graph analysis, etc, in order to identify complex dependencies and useful patterns in data. Such technologies also include methods of data analysis in real time, which allow for obtaining quick and accurate results of the analysis of large volumes of data.

The process of the implementation of innovative analytical and statistical technologies in the field of anti-fraud or anti-corruption can conventionally be divided into several stages (Artificial Intelligence in International Development: A Discussion Paper, 2019):

1. Digitalisation – information technologies are used in the field of anti-fraud/corruption policy as means in the automation of processes. For example, the deployment of electronic databases, electronic document flow and electronic reporting.
2. Open data – the creation of web or application platforms for public access to data on public procurement, budget expenditure, the income and property of officials, etc. This makes it easier for interested stakeholders (citizens as taxpayers, NGO representatives, businesses) to track and identify cases of fraud or corruption.
3. Digital identification – implementing technologies for the electronic identification of Internet users as a specific natural or legal person in the state, which allows for an increase in the transparency and traceability of state processes.
4. Artificial intelligence and analytics – the widespread implementation of data analytics and artificial intelligence systems and technologies, which allows for the more effective detection and prevention of fraud/corruption cases, for example, due to the analysis of large volumes of data and the detection of anomalies in the activities of officials and state structures.
5. Digital economy – the active development and implementation of blockchain technologies, which are used for reliable protection against forgery, manipulation and falsification, including document circulation, electronic voting, and other processes relating to anti-fraud/corruption activity.

The analysis of scientific thought concerning the problem of innovative analytical and statistical technologies as a tool for the detection of fraud and corruption allows us to characterise it as a scientific direction focused on the practical implementation of cutting-edge technologies (firstly artificial intelligence and ma-

chine learning) in the field of state governance quality control (Managing Machine Learning Projects in International Development: a Practical Guide, 2022; Paul, Jolley, Anthony, 2020).

Machine learning and artificial intelligence are two closely related but different concepts in computer science. In a general sense, artificial intelligence is a field of computer science that seeks to create machines that can operate with human-like intelligence. Machine learning is one of the technologies used to create such machines (Machine Learning Applications for Accounting Disclosure and Fraud Detection, 2021).

More specifically, machine learning is a methodology that allows computers to learn from existing data without using explicit programming. Instead of a person writing a program that solves a particular task, machine learning algorithms are used to teach the computer about certain patterns in the data. The computer can use this information to make decisions or solve problems that it has not seen before.

Artificial intelligence, on the other hand, is a more general concept that encompasses all technologies aimed at creating computer systems that can act intelligently, that is, perceive, process, and use knowledge and solve tasks that require human intelligence. Artificial intelligence technologies may use machine-learning techniques, but may include other approaches, such as expert knowledge systems, knowledge-based problem-solving, and neural networks (Russell, Norvig, 2022).

Both concepts are used in various fields, including the search for fraud or corruption, where machine learning can be used to analyse large amounts of data, and artificial intelligence can be used to develop systems that can make decisions based on this analysis.

In addition to the technologies of artificial intelligence and machine learning, the use of the following information technologies is necessary for the successful fight against fraud and corruption (Artificial Intelligence in International Development: A Discussion Paper, 2019; Artificial Intelligence Technology, 2023):

1. Blockchain is a technology that can ensure transparency and the irreplaceability of information. It can be used to provide electronic voting systems, agreements and contracts, and to protect them from falsification.
2. Decentralised data storage systems allow data to be stored without centralised management, which provides additional protection against unauthorised access and data changes.
3. Big Data data can be used to create predictive models of fraud/corruption schemes, identify key factors contributing to fraudulent behaviour, and monitor the dynamics of fraud/corruption processes.
4. GIS technologies allow the use of spatial data for the identification of connections between edges on the map.
5. Internet of Things (IoT) can be used for data collection, monitoring and control of key facilities, such as public buildings, roads, transport, etc.
6. Voice technologies and speech recognition can be used to create voice recognition systems and further automate government processes.

Despite all the advantages of implementing innovative analytical and statistical technologies as a tool for combating fraud and corruption in the state, it is important to note that the use of these technologies must be accompanied by a legal framework, ethical norms, and strong political and civic support. Otherwise, these technologies can become the basis for creating a ‘digital concentration camp’.

2. An analysis of the implementation of innovative analytical and statistical technology tools

In order to get an idea of the current trends in using innovative analytical and statistical technologies to detect election corruption or fraud, publications going back ten years were chosen, since the research does not aim to cover all publications on problematic topics, but is guided by the principle of increasing the number of analysed sources until reaching the threshold value when the amount of new information about methods and approaches obtained from each successive source does not decrease so much that further annotation becomes impractical. Therefore, after analysing at least 45 thematic publications, it was determined that 15

to 20 works are sufficient for our analysis in connection with the repetition of the methods used, unclearly defined methodology, etc.

The implementation of innovative analytical and statistical technologies as a tool for combating election corruption can be divided into four groups:

1. Secondary data analysis – in which the authors try to develop their own indicators of the level of corruption, or describe the state of the development of corruption relations in a certain period of time based on existing indicators (see, for instance, Adam, Fazekas, 2018, 2021; Berru, Batista, Torres-Carrión, Jimenez, 2020; Davenport, Mittal, 2023; De Francesco, Trein, 2020; Lima, Delen, 2020; Mansour, Taha, Taha, 2023; Norris, 2020; Odilla, 2023; Zemankova, 2019).
2. Automated text data analysis systems – reflect the results of implementing NLP technologies (natural language processing) in combination with spatial data visualisation technologies (Artemova, Maksimenko, Ohrimenko, 2022; Gawthorpe, 2018; López-Iturriaga, Sanz, 2017; Mamun, Azad, Pramanik, 2023; Noerlina *et al*., 2018).
3. Methodological direction – in which researchers develop universal methods (algorithms) of data analysis (detection of anomalies) with the determination of possible areas of their practical application, including in the field of election corruption (Aggarwal, 2017; Chatera, Borgib, Slamaa, Sfar-Gandouraa, Landoulsi, 2022; Chen, Zhang, Qian, Yuan, Ren, 2023; Dou *et al*., 2020; Eswar, Kannan, Vuduc, Park, 2021; Goglev, Migalin, Kasatkina, 2022; Han, Hu, Huang, Jiang, Zhao, 2022; Hojjati, Ho, Armanfard, 2022; Isson, 2018; Kaplan, 2023; Lawless, 2022; Liu *et al*., 2022; Lu *et al*., 2022; Ma *et al*., 2022; Pang, Shen, Cao, Hengel, 2021; Pinheiro, McNeill, 2014; Salehi *et al*., 2022; Sehwag, Chiang, Mittal, 2021; Shao, Du, Yu, Chen, 2022; Vaughan, 2020; Vincent *et al*., 2021; Zhao, Chen, Jia, 2022; Zhao *et al*., 2021).
4. The results of the practical implementation of machine learning and artificial intelligence technologies for the identification of subjects of corrupt relations and/or obtaining statistically substantiated confirmation of the presence/absence of corruption (Alvarez, Levin, Li, 2018; Chan, Hogaboam, Cao, 2022; Hassan, Passing, Goméz, 2023; Hicken, Mebane, 2017; Klimek *et al*., 2018; Kobak, Shpilkin, Pshenichnikov, 2020; Lacasa, Fernandez-Gracia, 2019; Li *et al*., 2020; Mebane, 2015; Mebane, Klaver, 2015; Mebane, Wall, 2015; *Next-generation AML: 6 Tips to Modernize Your Fight Against Money Laundering*, 2023; Podlazov, 2020; Ringsholm, 2022a; Rozenas, 2017; Steif, 2022; Using Machine Learning for Anti-Corruption Risk and Compliance, 2021; Zhang *et al*., 2019).

In our case, the methodological and practice-oriented directions are of the greatest importance, because it is at this level that the relevant theoretical models are tested, and precedents are created for using the results of analytical and statistical research as a component of the evidence base for certain facts of election fraud or corruption.

Let us look at the last two directions.

1. Methodological direction

This category of research differs significantly from other categories in several parameters.

Firstly, the purpose of such papers is to present one's own information system, method or algorithm when solving anomaly detection problems. Therefore, probably deliberately, in order to increase the audience of potential readers, the authors try to reach the maximum number of interested readers, and declare that their methods, algorithms or technologies developed and presented in the work are widely used, starting from biology and ending in the field of combating fraud or corruption.

Secondly, research in this category is presented in two formats: first, as a scientific publication (in most cases in the form of an article or a conference thesis); secondly, as a set of codes in social media for developers (e.g. github or kaggle) that allow to combine the efforts of different specialists, communicate, comment on or edit each other's codes with the function of tracking versions of the code, and the ability to reproduce the proposed methods independently using training or own data sets.

In this respect, it will be enough for us to review the most popular repository on www.github.com AD-Bench (Han *et al*., 2022). ADBench is a joint project by researchers from Shanghai University of Finance and Economics (SUFE) and Carnegie Mellon University (CMU). The project was developed by the authors of the most popular anomaly detection libraries, including anomaly detection for tabular data or databases (PyOD), time series (TODS), and graphs (PyGOD).

According to the authors, the performance of 30 algorithms for detecting anomalies in data arrays was evaluated using 57 data sets to the amount of 98,436 experiments according to three parameters:

- type of machine learning method (supervision): algorithm performance tests include 14 controlled, seven semi-supervised, and nine unsupervised learning algorithms;
- anomaly nature: local, global, cluster, dependent;
- stability of algorithm: behaviour in the presence of information noise or incomplete data.

The project combined developments in three directions: the detection of anomalies in tabular, time, and graph data.

1. PyOD is a Python library for detecting anomalous objects in multi-dimensional data. The original PyOD includes more than 40 detection algorithms, ranging from the classical LOF to the latest ECOD (Python Outlier Detection (PyOD), 2023; Zhao, Nasrullah, Li, 2019).
2. TODS is a universal automatic machine learning system for detecting outliers (anomalies) in multi-dimensional time series data. TODS includes modules for building anomaly detection systems based on machine learning, including data processing, time series processing, feature analysis (mining), detection algorithms, and an enhancement module. The functionalities provided through these modules include general purpose data pre-processing, time series data smoothing/transformation, feature extraction from time/frequency blocks of data, and various detection algorithms, including expert (human expertise) algorithms for system calibration (Lai *et al*., 2021; TODS: Automated Time-series Outlier Detection System, 2023).
3. PyGOD is a Python library for detecting outliers (anomalies) in graphs. PyGOD includes more than ten graph anomaly detection algorithms, such as DOMINANT or GUIDE (Liu *et al*., 2022; PyGOD, 2023). One of the advantages of this complex of algorithms is their ease of application in the sense of the amount of code lines used: five lines of code are enough to run most algorithms.

2. Practical direction

Statistical methods designed to solve problems of election fraud or corruption are called ‘electoral forensics’. At the same time, electoral forensics can be understood in two ways. In a broad sense, electoral forensics includes methods such as the parallel counting of votes, watching the voting process, or recounting a sample of ballots after voting, which are difficult to call purely analytical or statistical methods. In a narrow sense, electoral forensics involves focusing on analytical and statistical methods, with the minimisation of the human factor and operating with all data in general. The use of analytical and statistical methods in electoral forensics is based on the use of the principle of normality (Hicken, Mebane, 2017).

Within the framework of the narrow meaning of electoral forensics, two groups of methods are distinguished. The first group derives its origin from number theory, and refers to the frequency characteristics of numerical data in electoral statistics. The second group of methods is based on the search for anomalies in the relationship between various parameters of the electoral process, for example, the level of turn-out and the level of support for candidates. The main criterion used to detect falsifications is the discrepancy between real (documented) election results and normative (model) ones.

In the first case, a certain distribution of numbers expected from the ‘spontaneous’ recording of the will of voters acts as a normative model; in the second case, certain relations between the general parameters of elections (usually turn-out) and private ones (usually shares of votes for candidates or parties).

The first group of methods was based on attempts to apply Benford’s Law to the analysis of electoral data (Mebane, 2009). This law applies to the distribution of the first few digits of large numbers. Based on this law, there are always more numbers starting with one than numbers starting with two.

The number of numbers starting with the number two is always more than the numbers starting with the number three, etc: 3>4, 4>5, 5>6, 6>7, 7>8, 8>9. Since humans cannot efficiently generate random numbers, there is a significant chance that the numbers that a person generates on their own will not obey Benford's Law. The logic of Benford's Law is that when numbers appear in protocols chosen by humans as 'random', the probability of assigning a 'round' last digit, such as 5 or 0, will be higher than an 'awkward' one, such as 8. And similarly, when electoral data is artificially obtained, the probability of encountering even numbers in the lower ranks (11, 22, 33...) will differ from the expected probability of 1/10. This method is usually called the Beber and Scacco method (Beber, Scacco, 2012).

In this case, A. Podlazov's study of the problem of election falsification is indicative. 'If the numbers in the election protocols are not related to the contents of the ballot boxes, then the array of these numbers acquire properties that are not typical for real results. This can be demonstrated using statistical hypothesis testing tools' (Podlazov, 2019, 3). Analytical and statistical technology for detecting falsification is based on the following provisions.

First of all, the proposed method of identifying falsification does not allow for detecting their forms based on various physical manipulations with ballot papers: their deliberate misreading, throwing, carousels, etc. Such actions may be fraudulent, but they are procedures that result in numbers determined by the logic of these procedures. Therefore, the method is relevant only to the detection of numbers invented directly by counterfeiters.

Secondly, there are four simple statistical tests that differ in the degree of validity, reliability, analysed characteristics, and level of selectivity. A combination of these tests gives a versatile idea of the forms and scale of falsification: 1) test for the predominance of round numbers; 2) clot test; 3) greedy test; and 4) invalid ballot test (analysis). The test for the predominance of round numbers is completely well grounded, and at the same time quite simple for anyone to use. The test for clots is somewhat more difficult to implement, allows for parameter variations, and makes it difficult to interpret borderline situations. The greedy test demonstrates clearly the extremely wide spread of falsifications, and allows us to estimate their total volume. Finally, the analysis of the share of invalid ballots shows that far from all signs of mass falsifications have been described so far, and there is still room for further research in this field.

Thirdly, these electoral characteristics are the object of falsification most often: 1) the number of voters who took part in the elections; 2) voter turnout, the share of registered voters who took part in the elections (who received a ballot); and 3) the result of power, the proportion of voters who supported the party/candidate in power, measured by the number of those who participated in the vote.

Fourthly, the most difficult form of falsification of the results is their fabrication, when numbers in the election protocols are not in any way related to the contents of the ballot boxes. With such a crude approach, the 'results' are dominated by psychologically attractive numbers. For whole electoral indicators, these are round numbers, and for percentages, values without tenths.

In an integer random variable with a spread of many tens and even hundreds of units, the last digit should almost equally probably take all possible values. If the probability of encountering the digit 0 at the end of a number exceeds 10%, we can assume the presence of fabricated results. The verification of this assumption is reduced to the verification of the statistical hypothesis about the occurrence of excess. And if its level of significance (the probability of rejecting this hypothesis when it is true) turns out to be small, then the presence of falsification should be considered confirmed.

The registration at a polling station of some electoral characteristic is a Bernoulli test with probabilities of success (a round number) and failures (non-round number) . The number of successes is described by the binomial distribution, for which the probability of registering at least k successes in n trials . This is the probability that its rise above the p level is only the result of a coincidence, although the probability of meeting k / n may seem abnormally high. In specific elections for any electoral value starting from about two should be considered suspicious, with three exceptional, and with four improbable.

In addition to making up numbers, there is another mechanism of falsification that contributes to the emergence of a round number of voters who participated in the elections. The number of ballots received by

the precinct election commission usually ends with 0, as they are counted in tens. If the falsifiers equate the number of voters who came to the number of ballots available, this can also lead to the appearance of round numbers. The effect of greedy voting appears: a situation where the number of voters who participated in the elections matches exactly the number of ballots received by the precinct election commission, but at the same time turns out to be less than the number of voters registered in its lists.

Despite the fact that greedy voting in a specific precinct almost certainly means the falsification of election results, its scale is usually relatively small. And it is precisely the lack of ballot papers that limits it. In the case of the sufficient number of ballots, having achieved the desired results, counterfeiters may stop. But the shortage of ballots forces them to use all of them. Greedy voting in some precincts most likely indicates falsification in many other precincts where there are more ballots. And the higher the prevalence of greedy voting, the stronger this relationship.

A very simple and reliable indicator of unreliable election results is the low proportion of invalid ballots. In the case of falsifications, it can only decrease, since there is no political actor that protects the will of those who spoiled the ballot. The deliberate spoiling of ballots as a form of protest voting began to take place in the Russian Federation after 2006, when the option 'against all' was abolished.

Finally, A. Podlazov notes the situation that has developed in the Russian Federation since 2001: 'Falsifications, in which the content of election protocols corresponds to the content of ballot boxes, are elementary confirmed by an independent counting of ballots. And the fact that such a recount is not carried out for those cases where there is reason to suspect that the results were fabricated, indicates the fundamental unsuitability of the existing system of election commissions regarding the behavior of voting results. And the fact that even a mathematically rigorously proven fact of falsification does not automatically receive the appropriate legal registration indicates the fundamental unsuitability of the judicial system for resolving conflicts in the electoral sphere' (Podlazov, 2020, 189).

The second group of methods searches for anomalous dependencies between general (turn-out) and private (success of a specific party list or candidate) indicators of electoral statistics using the modelling method. If we turn to the results of the research work by the University of Michigan (Hicken, Mebane, 2017), there are three kinds of development of models for evaluating election results, which allow us to obtain statistically justified conclusions about electoral fraud.

1. The first can be considered a model of multimodal fraud (Klimek, Yegorov, Hanel, Thurner, 2012). According to the authors' concept, the basic assumption is that votes in non-fraudulent elections are formed by interacting processes, the effects of which can be summarised by two normal distributions: one distribution for turn-out shares and another independent distribution for vote shares in favour of the 'winner' (i.e. parties or candidates with the largest number of votes).

The authors suggest that electoral fraud is a situation in which the number of votes for the winner is increased in violation of official voting procedures. Some votes are transferred to the winner from the opposition, and some from those who did not show up at the polling station. At the same time, the authors distinguish two types of electoral fraud: moderate ('incremental') and greedy ('extreme'). The first means that the transfer of votes is carried out carefully. The second means the transfer is carried out for the entire total number of polling stations without additional calculations for the votes that actually took place.

The critical values of the parameters that determine the probability of one or another variant of fraud were calculated: f_i is the probability of moderate fraud, and f_e is the probability of greedy fraud. Other parameters fully describe bimodal and trimodal distributions, which the model characterises as consequences of electoral fraud. A derivative of this model is its modification with the determination of the fraud probability indicator at the level of the voting station. In this case, the focus is on: 1) statistical tests for the presence of fraud; and 2) estimates of the probability that each observed unit of vote aggregation, for example, each precinct, is fraudulent (Mebane, Wall, 2015).

2. To solve the shortcomings of the previous model, a model of geographic clustering emerged (Hicken, Mebane, 2017; Mebane, Wall, 2015; Shalaev, 2016). Indicators or phenomena that are geographically clustered deserve special attention. Geographical clustering can reveal where cooperation or collusion occurs

during the electoral process. Geographical clustering can also hint to those with the relevant expertise about other factors that may contribute to observed patterns in electoral outcomes. These other factors may or may not be related to the possibility of fraud. For example, a cluster may coincide with a political leader's home base, or with an area dominated by the leader's (or minority) political party or ethnic group. Based on this method, the use of geographical coordinates allows for obtaining additional confirmation of interference in the electoral process. In its simplest form, this function is assumed to depend on the distance between geographical points: the closer the polling stations are (and therefore the closer the voters live to each other), the smaller the difference in the voting results at these stations should be.

Unlike previous models, dealing with geographical data involves much greater labour costs, with unclear research prospects. Without a clear idea of the geographical characteristics of the electorate, it is difficult to justify the costs of getting the addresses of the polling stations and comparing them with geographical coordinates. Taking into account the possibilities of the latest information technologies, this method can have a significant development in the case of the introduction of electronic voting technologies, which is not far off.

3. The third model is a combination of the previous two, and is presented not only in the form of publications, but in the form of the online service 'Election Forensics Toolkit' (Mebane, 2015; Mebane, Kalinin, 2023), which allows every interested person to carry out a statistical evaluation of their own electoral datasets, or to look at the work with the data available in the system.

The information system evaluates arrays according to the following tests: 2BL, LastC, C05s, P05s, Skew, Kurt, DipT. 2BL is a test of the average value of the second digit. 'Second digit' refers to the second significant digit in each count to which the test is applied (for example, if the count is '1234', then '2' is the second significant digit). LastC is a test of the average value of the last digit. 'Last digit' refers to the last digit in each count to which the test is applied (for example, if the count is '1234', then '4' is the last digit). C05s is a test of the mean of a binary variable that indicates whether the last digit of the vote count for a given party or candidate is a zero or a five. P05s is a test of the mean of a binary variable indicating whether the last digit of the rounded percentage of the vote for the respective party or candidate is zero or five. Skew is a test for asymmetry. Kurt is a kurtosis test. DipT is a unimodality test.

Nevertheless, there are precedents when the use of anomaly detection algorithms in elections has led to the detection of falsification not only in the Russian Federation and Uganda, but, for example, in 2009 in Iran. Experts who conducted an analytical-statistical analysis of the election results in Iran concluded a relative victory for Mahmoud Ahmadinejad, although they did so very cautiously. Silver compared the results for Mahmoud Ahmadinejad in 2009 with the results from the first round in 2005 for candidates from the conservative camp (Ahmadinejad, Larijani, Ghalibaf), and found some discrepancies. For example, in the province of Lorestan, in 2005, conservative candidates received only 20% of the votes; while in 2009, 71% of voters voted for Mahmoud Ahmadinejad. In Tehran, on the contrary, support for representatives of the right-wing camp in 2009 decreased compared to 2005. Overall, across the country, there was a correlation between the results of 2005 and 2009, but in a weak form. Therefore, Silver did not make a definitive conclusion about falsification, as voter preferences could change over time (Silver, 2009).

Some researchers have analysed the elections in Iran using Benford's Law (Mebane, 2009). Mebane concluded that there were statistical distortions in the vote count for Mehdi Karroubi and Mohsen Rezai, in the direction of reduction. Also, statistical discrepancies were found in Mahmoud Ahmadinejad's results in favour of increasing the number of ballots cast for him. However, he believed that Mir-Hossein Mousavi's data was accurate. Mebane concluded that if there was any falsification of the voting results, it consisted of taking votes from Karroubi and Rezai and transferring them to Ahmadinejad. In this case, there was a possibility of a second round of voting, although Mebane did not have any data to confirm this claim. He also found a pattern that the more votes Mahmoud Ahmadinejad received at polling stations, the fewer spoiled or invalid ballots there were. This dynamic was not observed for Mir-Hossein Mousavi. The simplest explanation for this pattern could be the stuffing of ballots for Ahmadinejad, or some other action that artificially adds votes to his results. Mebane concluded that in this way, the incumbent president's share in the official voting results could have increased by 5%.

Another example of statistical data analysis is based on the 2022 Hungarian election (Ringsholm, 2022a; 2022b). The author explores the use of Benford's Law to test the Hungarian election for potential fraud with the use of Python. To determine if the election data conforms to Benford's Law, the author first tests a legitimate election, using the 2019 Danish election as an example. The results indicate that the Danish election data conforms to Benford's Law, as expected. Ringsholm tests the Hungarian election data: the reelection of Prime Minister Viktor Orban. While the initial results indicate a deviation from Benford's Law, the author discovers that using the law to detect election fraud has been criticised by experts as being inconclusive. They recommend using second digit analysis, which is less sensitive to constituency size. After testing the Hungarian election data using second digit analysis, the results conform to the expected distribution. The author concludes that although Benford's Law is an interesting concept, it should be used with caution when detecting election fraud. Deviations from Benford's Law could have reasons other than manipulation, and even when deviations are found, they can only be considered as red flags, not as proof of fraud.

Conclusion

The implementation of innovative analytical and statistical technologies as a tool for combating corruption in the state can be divided into four groups.

Firstly, analysis, the visualisation of statistical data and data of sociological or related research in the field of corruption, including elections (Heritage Foundation, Transparency International, UN, International Bank, etc), which can be called secondary data analysis.

Secondly, analysis and visualisation of documentary data (news, reports, official appeals) containing information about corrupt (fraudulent) actions and offences including election processes: document analysis, content analysis.

Thirdly, creating and development of pure maths (statistical) methods (algorithms) with the application of cutting-edge technologies such as artificial intelligence and machine learning applied to the detection of anomalies in hidden pattern discovery that has a wide range of applications, including in the field of combating (identifying) corruption as an anomalous event.

Finally, experimental developments in the field of the implementation of information technologies as a means of ensuring proper governance and combating corruption. This direction also includes publications on the application of analytical and statistical technologies (methods, algorithms) to search for anomalies in the political sphere or the sphere of public administration (elections, public procurement, etc), in order to obtain statistical confirmation of the presence of corruption.

In this case, the last two directions are of the greatest importance, because it is at this level that the relevant theoretical models and methodological tools are directly tested, and precedents are created for using the results of analytical and statistical research as a component of the evidence base for certain facts of fraud or corruption.

It should be noted that none of the illustrated methods can unequivocally confirm the fact of election fraud or election corruption. They can only indicate the presence of data anomalies, and raise the issue of further investigation. The decision about the fact of falsification of elections is made by the relevant authorities, and should be based on additional research and evidence, of course, if the falsification is not the actions of power structures, because in such cases, the country has only two likely options for the development of events: Maidan (the Ukrainian scenario) or state capture (the Russian scenario).

In general, the use of algorithms for finding anomalies in electoral statistics data can be an important tool for detecting real facts of both election fraud and election corruption. However, as the researchers note, it is necessary to use them with caution, and analyse the results obtained in combination with other sources of information, since, as dubious practice shows, statistical calculations that are not supported by additional evidence can be considered by interested parties, for example, in the situation of national elections, not as a tool to combat corruption, but only as a means of delegitimising the election results, with the corresponding consequences and reaction of the government/opposition.

References

Adam, I., Fazekas, M. (2018). Are Emerging Technologies Helping Win the Fight Against Corruption in Developing Countries? *Pathways for Prosperity Commission Background Paper Series 21*. Doi: https://doi.org/10.13140/RG.2.2.17930.52162.

Adam, I., Fazekas, M. (2021). Are Emerging Technologies Helping Win the Fight Against Corruption? A Review of the State of Evidence. *Information Economics and Policy*, *57*. Doi: https://doi.org/10.1016/j.infoecopol.2021.100950.

Aggarwal, C. C. (2017). *Outlier Analysis*. Cham, Switzerland: Springer International Publishing AG.

Alvarez, M. R., Levin, I., Li, Y. (2018). Fraud, Convenience, and e-voting: How Voting Experience Shapes Opinions About Voting Technology. *Journal of Information Technology & Politics*, *15* (2), 1–33. Doi: https://doi.org/10.1080/19331681.2018.1460288.

Artemova, E., Maksimenko, A., Ohrimenko, D. (2022). Application of machine learning methods in the classification of corruption related content in Russian-speaking and English-speaking Internet media. *Sociology: methodology, methods, mathematical modeling (Sociology: 4M)*, *27* (52), 131–157. Doi: https://doi.org/10.19181/4m.2021.52.5.

Artificial Intelligence in International Development: A Discussion Paper. (2019). https://www.idiainnovation.org/s/AIandinternationalDevelopment.pdf.

Artificial Intelligence Technology. (2023). Beijing: Posts & Telecom Press.

Beber, B., Scacco, A. (2012). What the Numbers Say: A Digit-Based Test for Election Fraud. *Political Analysis*, *20* (2), 211–234. Doi: https://doi.org/10.1093/pan/mps003.

Berru, Y. T., Batista, V. F. L., Torres-Carrión, P., Jimenez, M. G. (2020). Artificial Intelligence Techniques to Detect and Prevent Corruption in Procurement: A Systematic Literature Review. *ICAT 2019, CCIS*, *1194*, 254–268. Doi: https://doi.org/10.1007/978-3-030-42520-3_21.

Birch, S. (2011). *Briefing Paper: Electoral Corruption*. https://repository.essex.ac.uk/4484/1/05_11.pdf.

Center for Research Transparency and Accountability. (2020). https://vismo.com/case-studies/crta/.

Chan, L., Hogaboam, L., Cao, R. (2022). *Applied Artificial Intelligence in Business: Concepts and Cases*. Cham, Switzerland: Springer Nature Switzerland AG.

Chatera, M., Borgib, A., Slamaa, M. T., Sfar-Gandouraa, K., Landoulsi, M. I. (2022). Fuzzy Isolation Forest for Anomaly Detection. *Procedia Computer Science*, *207*, 916–925. Doi: https://doi.org/10.1016/j.procs.2022.09.147.

Chen, J., Zhang, J., Qian, R., Yuan, J., Ren, Y. (2023). An Anomaly Detection Method for Wireless Sensor Networks Based on the Improved Isolation Forest. *Applied Sciences*, *13* (2). Doi: https://doi.org/10.3390/app13020702.

Computational Social Science: Discovery and Prediction. (2016). Ed. M. R. Alvarez. New York, NY: Cambridge University Pres.

Davenport, T. H., Mittal, N. (2023). *All in on AI: How Smart Companies Win Big with Artificial Intelligence*. Boston, Massachusetts: Harvard Business Review Press.

De Francesco, F., Trein, P. (2020). How Does Corruption Affect the Adoption of Lobby Registers? A Comparative Analysis. *Politics and Governance*, *8* (2), 116–127. Doi: https://doi.org/10.17645/pag.v8i2.2708.

Dechert, J. (2013). *Patterns of Fraud: Tools for Election Forensics: dissertation*. Oregon.

Dou, Y., Liu, Z., Sun, L., Deng, Y., Peng, H., Yu, P. S. (2020). Enhancing Graph Neural Network-based Fraud Detectors against Camouflaged Fraudsters. *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, 315–324. Doi: https://doi.org/10.1145/3340531.3411903.

Eswar, S., Kannan, R., Vuduc, R., Park, H. (2021). ORCA: Outlier detection and Robust Clustering for Attributed graphs. *Journal of Global Optimization*, *81* (4), 967–989. Doi: https://doi.org/10.1007/s10898-021-01024-z.

Fisher, O., Hamilton, D. V. (2020). *Sint Eustatius 2020 Island Council Elections: Technical Observation and Assessment Report*. https://www.ifes.org/sites/default/files/migrate/ifes_electoral_assessment_st_eustatius_2020_island_council_elections_final_edits_03.12.21.pdf.

Gawthorpe, S. (2018). *Rethinking Corruption in the Czech Republic: A Mixed-Methods Approach to a Systemic Problem.* Prague: Charles University, Faculty of Social Sciences, Institute of Sociological Studies.

Goglev, N. N., Migalin, S. A., Kasatkina, E. V. (2022). Risk Identification Approach using Artificial Intelligence and Big Data Analysis. *International Journal of Open Information Technologies*, *10* (10), 111–119.

Han, S., Hu, X., Huang, H., Jiang, M., Zhao, Y. (2022). ADBench: Anomaly Detection Benchmark. *NeurIPS*, *45*. Doi: https://doi.org/10.48550/arXiv.2206.09426.

Hassan, P., Passing, F., Goméz, J. M. (2023). ESG Fingerprint: How Big Data and Artificial Intelligence Can Support Investors, Companies, and Stakeholders? In R. Schmidpeter, R. Altenburger (eds.). *Responsible Artificial Intelligence: Challenges for Sustainable Management*, 219–234. Cham, Switzerland: Springer Nature Switzerland AG.

Hicken, A., Mebane, W. R. J. (2017). *A Guide to Elections Forensics: Research and Innovation Grants Working Papers Series*. https://pdf.usaid.gov/pdf_docs/PA00MXR7.pdf.

Hojjati, H., Ho, T. K. K., Armanfard, N. (2022). Self-Supervised Anomaly Detection: A Survey and Outlook. *arXiv.2205.05173*, *18*. Doi: https://doi.org/10.48550/arXiv.2205.05173.

Isson, J. P. (2018). *Unstructured Data Analytics: How to Improve Customer Acquisition, Customer Retention, and Fraud Detection and Prevention*. Hoboken, New Jersey: Wiley.

Kaplan, J. (2023). *Crime by the Numbers: A Criminologist's Guide to R*. Boca Raton, FL: CRC Press.

Klimek, P., Jimenez, R., Hidalgo, M., Hinteregger, A., Thurner, S. (2018). Forensic analysis of Turkish elections in 2017–2018. *PLoS One*, *13* (10), e0204975. Doi: https://doi.org/10.1371/journal.pone.0204975.

Klimek, P., Yegorov, Y., Hanel, R., Thurner, S. (2012). Statistical detection of systematic election irregularities. *PNAS*, *109* (41), 16469–16473. Doi: https://doi.org/10.1073/pnas.1210722109.

Kobak, D., Shpilkin, S., Pshenichnikov, M. S. (2020). Suspect Peaks in Russia's "Referendum" Results. *Significance*, *17* (5), 8–9. Doi: https://doi.org/10.1111/1740-9713.01438.

Kratcoski, P. C. (2018). Introduction: Overview of Major Types of Fraud and Corruption. In P. C. Kratcoski, M. Edelbacher (eds.). *Fraud and Corruption. Major Types, Prevention, and Control*, 3–20. Cham, Switzerland: Springer International Publishing AG.

Lacasa, L., Fernandez-Gracia, J. (2019). Election Forensics: Quantitative Methods for Electoral Fraud Detection. *Forensic Sci Int*, *294*, e19-e22. Doi: https://doi.org/10.1016/j.forsciint.2018.11.010.

Lai, K.-H., Zha, D., Wang, G., Xu, J., Zhao, Y., Kumar, D. (...), Hu, X. (2021). TODS: An Automated Time Series Outlier Detection System. Doi: https://doi.org/10.48550/arXiv.2009.09822.

Lawless, C. (2022). *Forensic Science: A Sociological Introduction*. Oxon: Routledge.

Li, X., Liu, S., Li, Z., Han, X., Shi, C., Hooi, B. (…), Cheng, X. (2020). FlowScope: Spotting Money Laundering Based on Graphs. *Proceedings of the AAAI Conference on Artificial Intelligence*, *34* (4), 4731–4738. Doi: https://doi.org/10.1609/aaai.v34i04.5906.

Lima, M. S. M., Delen, D. (2020). Predicting and explaining corruption across countries: A machine learning approach. *Government Information Quarterly*, *37*(1). Doi: https://doi.org/10.1016/j.giq.2019.101407.

Liu, K., Dou, Y., Zhao, Y., Ding, X., Hu, X., Ding, R. Z. K. (…), Yu, P. S. (2022). BOND: Benchmarking Unsupervised Outlier Node Detection on Static Attributed Graphs. *36th Conference on Neural Information Processing Systems (NeurIPS 2022) Track on Datasets and Benchmarks*, 25. Doi: https://doi.org/10.48550/arXiv.2206.10071.

López-Iturriaga, F. J., Sanz, I. P. (2017). Predicting Public Corruption with Neural Networks: An Analysis of Spanish Provinces. *Social Indicators Research*, *140* (3), 975–998. Doi: https://doi.org/10.1007/s11205-017-1802-2.

Lu, M., Han, Z., Rao, S. X., Zhang, Z., Zhao, Y., Shan, Y. (…), Jiang, J. (2022). BRIGHT – Graph Neural Networks in Real-time Fraud Detection. *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*, 3342–3351. Doi: https://doi.org/10.1145/3511808.3557136.

Ma, X., Wu, J., Xue, S., Yang, J., Zhou, C., Sheng, Q. Z. (…), Akoglu, L. (2022). A Comprehensive Survey on Graph Anomaly Detection with Deep Learning. *IEEE Transactions on Knowledge and Data Engineering*, 32. Doi: https://doi.org/10.1109/TKDE.2021.3118815.

Machine Learning Applications for Accounting Disclosure and Fraud Detection. (2021). Hershey, PA: IGI Global.

Mamun, A. A., Azad, A. K., Pramanik, I. (2023). Crime-Finder: A System for Extraction and Visualization of Crime Data from Bengali Online Newspaper Articles. In N. Siddique, M. Shamsul, J. Wall, S. M. Kaiser (eds.). *Applied Informatics for Industry 4.0*, 97–108. Boca Raton, FL: CRC Press.

Managing Machine Learning Projects in International Development: a Practical Guide. (2022). https://www.usaid.gov/sites/default/files/2022-05/Vital_Wave_USAID-AIML-FieldGuide_FINAL_VERSION_1.pdf.

Mansour, E., Taha, R., Taha, N. (2023). The Impact of Internet of Things on the Quality of Financial Reporting. In A. M. A. M. Al-Sartawi, A. Razzaque, M. M. Kamal (eds.). *From the Internet of Things to the Internet of Ideas: The Role of Artificial Intelligence: Proceedings of EAMMIS 2022*, 367–374. Cham, Switzerland: Springer Nature Switzerland AG.

Mebane, W. R. J. (2009). *Note on the Presidential Election in Iran, June 2009*. http://websites.umich.edu/~wmebane/note29jun2009.pdf.

Mebane, W. R. J. (2015). *Election Forensics Toolkit DRG Center Working Paper*. https://electionforensics.cps.isr.umich.edu/pdf/report.pdf.

Mebane, W. R. J., Kalinin, K. (2023). *Guide to Election Forensics Toolkit.* https://electionforensics.cps.isr.umich.edu/election.

Mebane, W. R. J., Klaver, J. (2015). *Election Forensics: Strategies versus Election Frauds in Germany*. http://www.umich.edu/~wmebane/epsa15.pdf.

Mebane, W. R. J., Wall, J. (2015). *Election Frauds, Postelection Legal Challenges and Geography in Mexico* [Press release]. http://www.umich.edu/~wmebane/apsa15.pdf.

Next-generation AML: 6 Tips to Modernize Your Fight Against Money Laundering. (2023). https://www.sas.com/en/whitepapers/next-generation-aml-110644.html.

Nigrini, M. J. (2012). *Benford's law: applications for forensic accounting, auditing, and fraud detection*. Hoboken, New Jersey: John Wiley & Sons, Inc.

Noerlina, Dewanti, R., Mursitama, T. N., Fairianti, S. P., Kristin, D. M., Sasmoko (…), Makalew, B. A. (2018). Development of a Web Based Corruption Case Mapping using Machine Learning with Artificial Neural Network. *2018 International Conference on Information Management and Technology (ICIMTech)*, 400–405. Doi: https://doi.org/10.1109/ICIMTECH.2018.8528150.

Norris, P. (2020). *Rusty Guillotines: Electoral Accountability and Government Corruption*. https://www.pippanorris.com/new-research-papers.

Occupational Fraud 2022: A Report to the Nations. (2022). https://acfepublic.s3.us-west-2.amazonaws.com/2022+Report+to+the+Nations.pdf.

Odilla, F. (2023). Bots against corruption: Exploring the benefits and limitations of AI-based anti-corruption technology. *Crime, Law and Social Change*. Doi: https://doi.org/10.1007/s10611-023-10091-0.

Pang, G., Shen, C., Cao, L., Hengel, A. V. D. (2021). Deep Learning for Anomaly Detection: A Review. *ACM Computing Surveys*, *54* (2), 1–38. Doi: https://doi.org/10.1145/3439950.

Paul, A., Jolley, C., Anthony, A. (2020). *Reflecting the Past, Shaping the Future: Making AI Work for International Development*. https://www.usaid.gov/sites/default/files/2022-05/AI-ML-in-Development.pdf.

Pinheiro, C. A. R., McNeill, F. (2014). *Heuristics in Analytics : a Practical Perspective of What Influences Our Analytical World*. Hoboken, New Jersey: Wiley.

Podlazov, A. V. (2019). Issledovanie statisticheskih metodov vyyavleniya vydumannyh rezul'tatov vyborov: CHast' 1. Kruglye chisla. *Preprinty IPM im. M. V. Keldysha*, *147*, 1–28. Doi: https://doi.org/10.20948/prepr-2019-147.

Podlazov, A. V. (2020). Formal'noe vyyavlenie vydumannyh rezul'tatov vyborov. *Proektirovanie budushchego. Problemy cifrovoj real'nosti: trudy 3-j Mezhdunarodnoj konferencii*, 176–190. https://keldysh.ru/future/2020/15.pdf doi: https://doi.org/10.20948/future-2020

PyGOD. (2023). *Github.com.* https://github.com/pygod-team/pygod.

Python Outlier Detection (PyOD). (2023). *Github.com.* https://github.com/yzhao062/pyod.

Ringsholm, J. F. (2022a). *How I Tested the Hungarian Election for Fraud Using Benford's Law.* https://towardsdatascience.com/how-i-tested-the-hungarian-election-for-fraud-using-benfords-law-2d32ea92fe7c.

Ringsholm, J. F. (2022b). *Hungarian Election Fraud*. https://github.com/JensFugl/Hungarian-election-fraud.

Rose-Ackerman, S., Palifka, B. J. (2016). *Corruption and Government: Causes, Consequences, and Reform*. Cambridge: Cambridge University Press.

Rozenas, A. (2017). Detecting Election Fraud from Irregularities in Vote-Share Distributions. *Political Analysis*, *25* (1), 41–56. Doi: https://doi.org/10.1017/pan.2016.9.

Russell, S. J., Norvig, P. (2022). *Artificial Intelligence: A Modern Approach*: Pearson Education Limited.

Salehi, M., Mirzaei, H., Hendrycks, D., Li, Y., Rohban, M. H., Sabokrou, M. (2022). A Unified Survey on Anomaly, Novelty, Open-Set, and Outof-Distribution Detection: Solutions and Future Challenges, *81*. Doi: https://doi.org/10.48550/arXiv.2110.14051.

Sehwag, V., Chiang, M., Mittal, P. (2021). SSD: A Unified Framework for Self-Supervised Outlier Detection. *ICLR 2021*. Doi: https://doi.org/10.48550/arXiv.2103.12051.

Shalaev, N. E. (2016). *Elektoral'nye anomalii v postsocialisticheskom prostranstve: opyt statisticheskogo analiza: dis. ... kandidata politicheskih nauk: 23.00.02*. Sankt-Peterburg.

Shao, C., Du, X., Yu, J., Chen, J. (2022). Cluster-Based Improved Isolation Forest. *Entropy (Basel)*, *24* (5), 17. Doi: https://doi.org/10.3390/e24050611.

Silver, N. (2009). Iranian Election Results by Province [UPDATED]. *Five Thirty Eight Politics*.

Sridhar, M., Rivest, R. L. (2018). k-Cut: A Simple Approximately-Uniform Method for Sampling Ballots in Post-Election Audits. 20. Doi: https://doi.org/10.48550/arXiv.1811.08811.

Steif, K. (2022). *Public Policy Analytics: Code and Context for Data Science in Government*. Boca Raton, FL CRC Press.

TODS: Automated Time-series Outlier Detection System. (2023). *Github.com.* https://github.com/datamllab/tods.

Using Machine Learning for Anti-Corruption Risk and Compliance. (2021). https://www.coalitionforintegrity.org/wp-content/uploads/2021/04/Using-Machine-Learning-for-Anti-Corruption-Risk-and-Compliance.pdf.

Vaughan, G. (2020). Efficient Big Data Model Selection with Applications to Fraud Detection. *International Journal of Forecasting*, *36* (3), 1116–1127. Doi: https://doi.org/10.1016/j.ijforecast.2018.03.002.

Vincent, J., Fei, S., Arnaud, S., Bijan, R., Yanlei, D., Nesime, T. (2021). Exathlon: A Benchmark for Explainable Anomaly Detection over Time Series. *Proceedings of the VLDB Endowment*, *14* (11), 2613–2626. Doi: https://doi.org/10.14778/3476249.3476307.

Wells, J. T. (2017). *Corporate Fraud Handbook. Prevention and Detection*. Hoboken, New Jersey: Wiley.

Zemankova, A. (2019). Artificial Intelligence in Audit and Accounting: Development, Current Trends, Opportunities and Threats – Literature Review. *2019 International Conference on Control, Artificial Intelligence, Robotics & Optimization (ICCAIRO)*, 148–154. Doi: https://doi.org/10.1109/iccairo47923.2019.00031.

Zhang, M., Alvarez, R. M., Levin, I. (2019). Election Forensics: Using Machine Learning and Synthetic Data for Possible Election Anomaly Detection. *PLoS One*, *14* (10), e0223950. Doi: https://doi.org/10.1371/journal.pone.0223950.
Zhao, Y., Chen, G. H., Jia, Z. (2022). TOD: GPU-accelerated Outlier Detection via Tensor Operations. *Proceedings of the VLDB Endowment, 16*(1). Doi: https://doi.org/10.48550/arXiv.2110.14007.
Zhao, Y., Hu, X., Cheng, C., Wang, C., Wan, C., Wang, W. (…), Akoglu, L. (2021). SUOD: Accelerating Large-Scale Unsupervised Heterogeneous Outlier Detection. *Proceedings ofthe 4th MLSys Conference*. Doi: https://doi.org/10.48550/arXiv.2003.05731
Zhao, Y., Nasrullah, Z., Li, Z. (2019). PyOD: A Python Toolbox for Scalable Outlier Detection. *Journal of Machine Learning Research*, *20*, 1–7. Doi: https://doi.org/10.48550/arXiv.1901.01588.

INOVATYVIOS ANALITINĖS IR STATISTINĖS TECHNOLOGIJOS ORGANIZUOJANT RINKIMUS

Yuliia Yatsyna, Igor Kudinov
Zaporižios nacionalinis universitetas (Ukraina), Ukrainos socialinių inžinierių sąjunga (Ukraina)

Santrauka

Straipsnyje nagrinėjamas svarbus novatoriškų analitinių ir statistinių technologijų, kurios vis dažniau taikomos, siekiant nustatyti rinkimų korupcijos ir sukčiavimo atvejus bei užkirsti tam kelią, organizuojant rinkimus vaidmuo. Didelių duomenų kiekių analizė ir anomalijų nustatymas bei rinkimų teismo ekspertizė gali užtikrinti sąžiningus ir skaidrius demokratinius procesus.

Atliekant tyrimą siekiama ištirti šių technologijų efektyvumą ir galimą jų poveikį didinant rinkimų procesų skaidrumą bei sąžiningumą, taikant kelis metodus: atitinkamų dokumentų, žiniasklaidos nušvietimo, viešosios nuomonės ir naujausių sukčiavimo atvejų analizę. Inovatyvių analitinių ir statistinių technologijų kovai su rinkimų korupcija diegimą autoriai skirsto į keturias kryptis. Pirmoji – statistinių duomenų analizė ir korupcijos, įskaitant rinkimų procesus, tyrimai, kuriuos galima pavadinti antrine duomenų analize. Antrasis – dokumentinių duomenų, kuriuose pateikiama informacija apie korupcinius veiksmus ir nusikaltimus, įskaitant rinkimų procesus, analizė. Trečiasis yra matematinių metodų ir algoritmų kūrimas taikant pažangiausias technologijas, tokias kaip dirbtinis intelektas ir mašinų mokymasis, siekiant aptikti anomalijas ir paslėptus modelius. Galiausiai eksperimentinė informacinių technologijų plėtra, kaip tinkamo valdymo ir kovos su korupcija užtikrinimo priemonė. Nors rinkimų statistikos duomenų anomalijų nustatymo algoritmai gali būti svarbi priemonė, juos reikėtų taikyti atsargiai ir kartu su kitais informacijos šaltiniais, siekiant išvengti rinkimų rezultatų delegitimizacijos pasekmių.

PAGRINDINIAI ŽODŽIAI: *sukčiavimas vykdant rinkimus, rinkimų korupcija, teismo ekspertizė, Benfordo įstatymai, mašinų mokymasis.*

JEL KLASIFIKACIJA: D73.

Received: 2023-04-03

Revised: 2023-04-23

Accepted: 2023-05-12