

Dmytro Diachenko, Mykhailo Prokopchyk, Vladyslav Rovenchak, Andrii Frolov

Kharkiv National University of Radio Electronics, Kharkiv, Ukraine

## METHODS OF DATA MINING USING MACHINE LEARNING

**Abstract. Relevance.** In the context of the continuous growth of information volumes across various fields of the economy, science, and technology, the problem of effective processing and analysis of large-scale data has become increasingly urgent. Traditional analytical methods are no longer capable of providing fast and accurate extraction of useful knowledge and patterns from massive information flows. The response to this challenge lies in the methods of data mining, which are based on modern machine learning technologies. These methods enable the automatic discovery of hidden patterns, the generation of accurate predictions, and the support of data-driven decision-making in real time. Given the rapid development of digitalization, artificial intelligence, and the need for prompt decision-making in a competitive environment, the relevance of developing and improving data mining methods is growing steadily. **The object of research** is process of data mining using machine learning methods, namely the set of algorithms, models, tools, and approaches that ensure the detection of hidden patterns, anomalies, and structures in large volumes of heterogeneous information. **Purpose of the article.** This study explores contemporary approaches to intelligent data analysis based on machine learning techniques and assesses their effectiveness across a range of application domains. The article aims to provide a structured overview of state-of-the-art algorithms and to evaluate their respective advantages and limitations in processing large-scale and high-dimensional datasets. **Research results.** A systematic analysis of key data mining methods based on machine learning algorithms was carried out. It was found that the most effective approaches for processing large and heterogeneous datasets include classification, clustering, regression analysis, and dimensionality reduction techniques. Deep neural networks demonstrated effectiveness when applied to unstructured data such as text, images, and time series. The study revealed that the appropriate choice of algorithm depends not only on the data type but also on the specific nature of the task. A comparative assessment of tools showed that the Python ecosystem offers the greatest flexibility, while AutoML platforms simplify model deployment for users with limited programming experience. The research also included a review of recent publications that confirm the practical value of machine learning in real-world use cases. Overall, the findings indicate that machine learning is a driving force behind the evolution of data mining methods, enabling accurate, scalable, and adaptive data processing in the context of modern digital transformation. **Conclusions.** Machine learning has significantly expanded the capabilities of intelligent data analysis by enabling the automatic detection of patterns, forecasting, and decision-making based on large volumes of information. The study demonstrates the effectiveness of various algorithms in tasks such as classification, clustering, regression, and deep learning. Python-based tools and cloud platforms have been identified as the most convenient environments for implementing analytical models. A promising direction lies in the development of explainable AI and hybrid approaches that combine algorithmic precision with domain-specific expertise.

**Keywords:** intelligent data analysis, machine learning, classification, clustering, regression, deep learning, neural networks, analytical tools, Data Mining, CRISP-DM, AutoML, Big Data, Python.

### Introduction

Intelligent data analysis (Data Mining) [1] is a significant area within modern information technologies, aimed at the automated extraction of meaningful patterns, knowledge, and deep insights from large volumes of both structured and unstructured data. In today's world, data is generated at an unprecedented rate – from online platforms and social networks to financial transactions, medical records, and industrial sensors. In this context of information overload, traditional analytical methods are losing their effectiveness and are increasingly being replaced by intelligent approaches.

Machine learning plays a central role as the driving force behind intelligent data analysis. Its ability to autonomously detect patterns, adapt to new data, and improve performance without human intervention makes these techniques exceptionally powerful analytical tools. The primary tasks addressed through machine learning include classification, regression, clustering, anomaly detection, forecasting, and dimensionality reduction.

Historically, data mining emerged in the 1990s from the convergence of statistics, artificial intelligence, and database systems. With the rapid advancement of computational capabilities and the exponential growth

of digital data, the need for more complex, self-learning algorithms has become apparent. Machine learning has evolved from simple models into advanced deep neural networks capable of performing multi-level information processing with high accuracy.

Unlike traditional statistical methods, which are oriented toward hypothesis testing and formal modeling, machine learning can work with large-scale datasets without the need for predefined assumptions. It can identify patterns within chaotic, heterogeneous, and high-dimensional data – a crucial feature in the era of Big Data. Moreover, intelligent data analysis based on machine learning is inherently interdisciplinary, integrating knowledge from computer science, statistics, mathematics, linguistics, medicine, economics, psychology, and the social sciences. This synergy enables the development of new analytical techniques that are applicable across a broad range of domains, including manufacturing management, logistics, energy, bioinformatics, environmental monitoring, and digital humanities.

The objective of this article is to investigate the key methods of intelligent data analysis based on machine learning, to examine relevant algorithms, tools, and technologies, and to explore practical application scenarios. The article also discusses the benefits, limita-

tions, and prospects of employing machine learning in data analytics tasks.

**Review of Recent Studies and Publication.** In recent years, the topic of intelligent data analysis using machine learning has become highly relevant, as evidenced by the growing number of specialized reviews and systematic studies across various fields. This section highlights the most influential publications in finance, education, healthcare, and multimedia, as well as methodological papers focused on imbalanced data processing and open science initiatives.

A significant contribution in this area is presented in [2], a comprehensive review on fraud detection methods based on data mining and machine learning techniques. The paper covers multiple applied domains, such as banking, insurance, telecommunications, and e-commerce. The authors emphasize that traditional fraud detection algorithms, despite their past effectiveness, are increasingly incapable of addressing the challenges posed by today's high-speed and dynamic environments. The review proposes a generalized classification of approaches, including both classical techniques and modern ensemble or hybrid architectures that combine multiple models to improve accuracy when working with imbalanced datasets. Special attention is paid to model interpretability, which is critical in financial sectors where decision-making must be transparent and accountable. In this context, the researchers highlight the role of explainable AI as a direction capable of integrating the computational power of modern algorithms with the need for human oversight and auditing. The paper also stresses the growing importance of adaptive and context-aware models capable of operating with limited, fragmented, or unreliable data in the era of increasingly complex fraud schemes.

Another influential work is the updated systematic review [3], focused on the use of machine learning in education. The authors trace the evolution of Educational Data Mining and Learning Analytics over the past decade, analyzing both the growth in publication metrics and the methodological maturity of the field. The paper presents educational analytics as an emerging scientific discipline that integrates pedagogy, artificial intelligence, psychology, sociology, and statistics. It explores the technological pipeline for knowledge discovery in educational data – from data collection and preprocessing to interpretation and practical implementation. Numerous examples are provided, including the use of classification models for learning outcome prediction, learning style analysis, and recommendation systems in MOOCs. The paper also examines the application of deep learning to identify behavioral patterns among students.

Emphasis is placed on ethical challenges such as data privacy, model transparency, and adherence to the principles of open science. The authors argue that education demands a more delicate approach to analytics than commercial or financial systems, as misguided modeling could distort learners' trajectories. They recommend practices such as preregistration, open dataset publication, and replication studies to foster an ethically responsible research environment.

The study in [4] explores the application of machine learning to financial anomaly detection, a domain crucial for maintaining market order and protecting investor interests. Through comparative analysis, the paper evaluates the performance of traditional statistical methods and modern machine learning algorithms in identifying anomalies. It covers supervised, unsupervised, and deep learning methods, assessing their ability to handle high-dimensional financial data and detect complex fraud patterns. Ensemble models are found to offer a strong balance between detection accuracy and interpretability. However, challenges remain, particularly regarding class imbalance and model generalization. The study proposes hybrid approaches that combine domain knowledge with data-driven analysis and emphasizes the potential of explainable AI to enhance transparency and trust in automated financial anomaly detection systems. The results significantly expand the understanding of machine learning in financial analytics and suggest promising directions for improving anomaly detection efficiency and reliability.

A notable interdisciplinary example is a study dedicated to the use of machine learning in cardiology. The publication [5] presents an extensive meta-analysis of over forty studies focused on automated prediction of coronary artery disease using clinical and demographic data. Ensemble algorithms – combining logistic regression, decision trees, gradient boosting, and neural networks – emerge as the most effective models. The study emphasizes evaluation using medically relevant metrics such as sensitivity, specificity, and accuracy, highlighting the importance of not only statistical performance but also the applicability of models within medical systems, where false positives or negatives can have serious consequences. The authors draw attention to the limitations of publicly available medical datasets and recommend local model validation across demographic groups. Interpretable algorithms and hybrid systems that integrate medical expertise with the flexibility of machine learning are proposed as future directions.

In [6], the role of data mining methods – particularly text mining – in digital forensics is examined. The authors emphasize the growing complexity of digital sources and devices involved in investigations, which renders manual analysis ineffective. Machine learning models and pattern recognition algorithms are shown to be essential for uncovering hidden digital evidence that traditional methods might overlook. The paper demonstrates that text analysis techniques, as a subset of data mining, are critical to improving the precision and efficiency of digital forensic workflows. The review outlines the main data mining-based approaches to digital forensics and highlights their potential to address the challenges associated with modern data volume and complexity.

Across all reviewed studies, a common trend is observed: the shift from traditional statistical techniques to more flexible, interpretable, scalable, and adaptive machine learning solutions. At the same time, researchers stress the necessity of an interdisciplinary approach that combines algorithmic rigor with ethical, legal, and domain-specific considerations. Collectively, these publi-

cations offer a comprehensive view of the current state of data mining research, showcasing both theoretical advances and practical achievements, while outlining the promising avenues for future development.

**The purpose of this work** is to investigate modern intelligent data analysis methods based on machine learning and examine their application across various domains to uncover hidden patterns, anomalies, and key insights in large, complex datasets. Special emphasis is placed on comparing traditional statistical approaches with machine learning algorithms, analyzing the advantages and limitations of current models, and identifying future trends such as explainable AI, hybrid modeling, and ethical considerations. The article also aims to consolidate findings from recent literature to present a coherent picture of machine learning as a pivotal tool in contemporary data analysis.

### Main part

Today, data mining is regarded not merely as a component of information technologies, but as a multidisciplinary field that integrates computer science, statistics, linguistics, sociology, medicine, and economics. The application of intelligent data analysis has become a strategic tool for decision-making, process optimization, risk identification, and forecasting in a wide array of domains – from industry and finance to healthcare and cybersecurity. Data mining encompasses a set of tasks aimed at uncovering hidden patterns, knowledge, and models in large volumes of information. These tasks are conventionally divided into descriptive and predictive categories. The primary types of analytical operations within data mining include classification, regression, clustering, association rule mining, anomaly detection, dimensionality reduction, and sequence analysis. These tasks are often interrelated and frequently combined in real-world analytical projects. For instance, clustering or dimensionality reduction may precede classification, while anomaly detection may be integrated with a predictive model.

The process of intelligent data analysis involves more than just applying individual algorithms; it represents a full multi-stage cycle that encompasses all stages of working with data – from acquisition to the interpretation of results and the implementation of knowledge in practical activities. This process is commonly modeled as a knowledge discovery lifecycle or structured according to the CRISP-DM framework [7], which has become an industry standard for analytical projects. The first stage is domain understanding, during which the objectives of the analysis are defined, research questions are formulated, and business or strategic requirements are identified. This stage is critically important because it establishes the context for analysis and sets the success criteria for the models.

The second stage involves data collection and understanding. It includes identifying information sources, aggregating data, verifying quality, and conducting initial assessments. Often, this stage reveals that the available data is incomplete, noisy, or heterogeneous, requiring further processing. The third stage – data preparation – covers data cleaning, normalization, encoding of categorical variables, handling of missing values, fea-

ture engineering, and dimensionality reduction. This is the stage where data is transformed into a format suitable for applying machine learning algorithms.

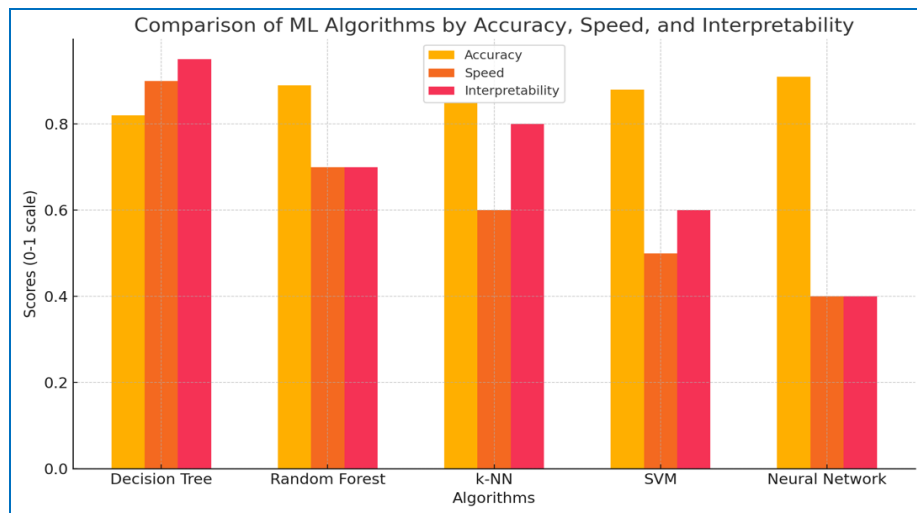
At the next, modeling stage, appropriate algorithms are selected, hyperparameters are tuned, and models are created and trained. Depending on the objectives, these may include classifiers, regression models, clustering algorithms, or anomaly detection techniques. Common practices at this stage include data partitioning into training and test sets, cross-validation, and various performance evaluation techniques.

Subsequently, model evaluation takes place, during which the modeling results are analyzed in terms of accuracy, completeness, stability, interpretability, and relevance to the original objectives. Both general metrics (e.g., accuracy, F1-score, ROC-AUC) and specialized ones tailored to the specific domain may be used.

The final stage involves the deployment and interpretation of knowledge, where the obtained results are adapted for practical use: they are integrated into business processes, decision-making systems, or visualized for further analysis. At this stage, explainable artificial intelligence plays a key role, enabling model behavior to be made understandable to experts without deep technical backgrounds.

In an era of exponential data growth, deep learning gains particular importance as it enables the extraction of complex, abstract representations from unstructured information, making it indispensable in computer vision, natural language processing, audio analysis, and big data analytics in finance, medicine, industry, and cybersecurity. One of the most significant achievements of deep learning is the development of convolutional neural networks (CNNs), which are highly effective for visual data. These networks can automatically extract features from images at various levels of abstraction – from edges and textures to complex objects. Their application in medical imaging, MRI scan analysis, and radiograph diagnostics ensures high accuracy in identifying pathologies, often surpassing or augmenting expert performance. In industry, CNN models are used for real-time quality control, defect detection, and monitoring of production processes.

Fig. 1 illustrates a comparison of popular machine learning algorithms across three key criteria: accuracy, speed, and interpretability. The diagram reflects how effectively each approach meets the basic demands placed on analytical models in real-world conditions. The analysis includes five models: decision tree, random forest, k-nearest neighbors, support vector machine, and neural network. For each, the performance is visualized on a scale from zero to one, enabling a visual assessment of the strengths and weaknesses of each algorithm. Neural networks provide the highest accuracy but face significant limitations in interpretability and speed. Conversely, decision trees demonstrate the most transparent structure, making them convenient for explaining results while maintaining acceptable performance. Random forests show high accuracy but lower transparency. K-nearest neighbors and SVMs offer relatively balanced results, though each has its limitations in terms of computational complexity or the explainability of predictions.



**Fig. 1.** Comparative Analysis of Machine Learning Methods

In tasks where the sequence of data is critical – such as in text analysis, time series processing, or tracking user actions – recurrent neural networks (RNNs) are widely applied. They enable the retention of information from previous states, allowing for contextual event processing. Modifications such as LSTM and GRU have overcome classic RNN limitations like gradient vanishing, paving the way for long-term forecasting models, for instance, in financial analytics or consumer behavior prediction.

Another important component of deep learning is autoencoders, used for dimensionality reduction, anomaly detection, data generation, and pretraining of other models. Thanks to their ability to reveal hidden structures in large datasets, autoencoders are actively used in noise filtering, data reconstruction, and the creation of compressed vector representations for further analysis.

Among the most innovative and complex technologies are generative adversarial networks (GANs), consisting of two components – a generator and a discriminator – that learn by competing. These models can generate new, realistic images, texts, or other data types, opening new possibilities in information synthesis, privacy protection, data augmentation, and handling rare cases. Despite their computational intensity, GANs are actively used in fields that require the generation of new but statistically reliable examples.

The effective implementation of data mining methods is impossible without the appropriate toolkit – both software and hardware. In the current environment of growing data volumes, rising demands for model performance and accuracy, and the need to integrate analytics into real systems, universal, flexible, and scalable platforms are of particular importance. These platforms not only provide access to a wide range of machine learning algorithms but also support the full analytics lifecycle – from data collection and processing to result visualization and model deployment in production environments.

Among the most popular programming languages for implementing data mining solutions, Python leads. Its popularity stems from its simple syntax and the availability of powerful libraries covering the full range of tasks – from preprocessing to complex modeling. For example, the Scikit-learn library offers implementations of classic

machine learning algorithms such as regression, classification, clustering, dimensionality reduction, and ensemble models. It is user-friendly for education and research projects while being stable enough for production use. TensorFlow and PyTorch libraries are used for building neural networks and implementing deep learning. The former is more oriented toward industrial solutions, offering wide scalability and GPU optimization, while the latter is valued for its flexibility, transparency of internal processes, and ease of use in academic research.

The R programming language occupies a separate niche, popular especially among statisticians and researchers. It has a rich set of packages for data processing, analysis, and visualization, such as caret, random Forest, e1071, and nnet, which enable the implementation of machine learning methods in an intuitive manner. Due to its high statistical accuracy and graphical capabilities, R is frequently used in scientific research, bioinformatics, and social analytics.

In cases where data volumes exceed local computing capacity or rapid scalability is required, cloud platforms become especially relevant. Amazon Web Services, Microsoft Azure, and Google Cloud Platform offer comprehensive services for storing, processing, and analyzing big data, as well as ready-to-use environments for training and deploying AI models. These platforms enable the creation of end-to-end solutions with data stream integration, automated model training (AutoML), built-in security tools, and support for distributed computing. This opens the door to scalable and reliable analytics systems accessible even to small and medium-sized organizations without the need for costly infrastructure.

Interactive analytics environments such as Jupyter Notebook, Google Colab, and RStudio also gain particular significance. They allow integration of code, visualization, and explanatory text in a single interface, which is extremely convenient for research, report preparation, and education. Using such environments promotes transparency of the analytical process, ensures experiment reproducibility, and supports collaborative work on projects.

Fig. 2 presents a comparative analysis of the most widely used software tools and platforms employed for

implementing machine learning methods in data mining tasks. The evaluation covers key characteristics such as performance in classification, clustering, regression, and deep learning tasks, capacity to handle large datasets, level of model interpretability, visualization capabilities,

and accessibility for cloud deployment. The table enables comparison of tools such as Scikit-learn, TensorFlow, PyTorch, R, Jupyter Notebook, as well as cloud services offered by AWS, Azure, and Google Cloud, in the context of their suitability for specific analytical objectives.

Tool / Platform	Classification	Clustering	Regression	Deep Learning
Python (Scikit-learn)	very good	good	very good	basic (via Keras)
Python (TensorFlow)	good	limited	good	very good
Python (PyTorch)	good	limited	good	very good
R (caret, randomForest)	very good	good	very good	weak
Jupyter Notebook / Colab	depends on libraries	depends	depends	good (via integration)
AWS / Azure / GCP (AutoML)	very good	good	very good	good
Google Cloud BigQuery ML	good	weak	good	weak
Tool / Platform	Big Data Handling	Model Interpretability	Data Visualization	Cloud Support
Python (Scikit-learn)	limited	high	high	possible via API
Python (TensorFlow)	good	medium	limited	fully supported
Python (PyTorch)	average	medium	basic	fully supported
R (caret, randomForest)	limited	high	very high	partially via RStudio Cloud
Jupyter Notebook / Colab	limited	high	very high	fully supported (in Colab)
AWS / Azure / GCP (AutoML)	very good	high	high	native support
Google Cloud BigQuery ML	very good	limited	basic	fully

Fig. 2. Comparative Analysis of Development Software

In conclusion, machine learning serves as a powerful instrument for data mining, capable of effectively uncovering patterns, forecasting events, and supporting decision-making across various domains. With the advancement of algorithms, tools, and computational platforms, the ability to analyze large-scale and complex datasets has significantly expanded. The integration of classical approaches with modern deep learning models, the use of cloud-based solutions, and the incorporation of explainable AI provide a foundation for further enhancement of analytical systems and the extension of their practical applications.

### Conclusions

As a result of the conducted research, a comprehensive analysis was performed on fundamental approaches and modern practices in data mining based on machine learning methods. It was established that traditional statistical techniques lack the scalability, adaptability, and precision required to handle the growing volume, velocity, and complexity of data characteristic of the digital era. Consequently, machine learning – serving as a core subset of artificial intelligence – has become the foundation for building contemporary analytical systems.

The study examined the core concepts of data mining, including classification, clustering, regression analysis, and dimensionality reduction. It was demonstrated how algorithms such as decision trees, k-NN methods, neural networks, Bayesian classifiers, and ensemble models enable adaptive processing of heterogeneous and high-dimensional datasets. Particular attention was paid to deep learning, which opens new possibilities for

the automatic processing of complex structured and unstructured data (images, text, signals, etc.) and has proven highly effective in solving tasks related to visual recognition, time series forecasting, and generative modeling.

An essential component of the research involved reviewing tools and platforms that enable practical implementation of these methods. It was determined that the Python ecosystem, complemented by libraries such as TensorFlow, PyTorch, and Scikit-learn, provides the flexibility and scalability required for the full model development lifecycle – from data preprocessing to cloud deployment. Other environments, such as R and AutoML platforms on AWS, Azure, and Google Cloud, facilitate rapid development and effective model testing, particularly for users without deep programming expertise.

The analyzed publications and real-world case studies confirm that machine learning is successfully integrated across a wide range of domains: medicine, finance, cybersecurity, education, forensics, and industry. However, the findings also highlight persistent challenges, such as the interpretability of complex models, class imbalance, overfitting risks, and the ethical handling of personal data.

In conclusion, machine learning not only broadens the scope of data mining but also shapes a new paradigm of analytics – flexible, adaptive, and capable of self-improvement. Future research in this field should focus on integrating explainable AI, domain-aware hybrid systems, and automated model design tools, making data mining even more accessible and reliable for real-world applications.

### REFERENCES

1. Flach P. A. Machine Learning: The Art and Science of Algorithms that Makes Sense of Data. Cambridge: Cambridge University Press, 2012. 291 p. <https://doi.org/10.1017/CBO9780511973000>.
2. Phua, C. et al. A Comprehensive Survey of Data Mining-Based Fraud Detection Research. Artificial Intelligence Review, 55, 2021. P. 1985-2033.
3. C. Romero, S. Ventura. Educational data mining and learning analytics: An updated survey. Cornell University. Computer Science, 2024. 30 p. <https://doi.org/10.48550/arXiv.2402.07956>.

4. Q. Wang. Research on the Application of Machine Learning in Financial Anomaly Detection. iBusiness, 16, 2024. 173-183. <https://doi.org/10.4236/ib.2024.164012>.
5. О.Д. Земляний, О.Г. Байбуз. Огляд методів аналізу даних та методів машинного навчання при прогнозуванні ішемічної хвороби серця. Актуальні проблеми автоматизації та інформаційних технологій, т.27, 2023. С. 109-129.
6. E. Abdallah, Esraa A. Eloud, A. Abdallah. A Survey of Data Mining Techniques for Digital Forensic Analysis. Procedia Computer Science, Volume 257, 2025. P. 731-736. <https://doi.org/10.1016/j.procs.2025.03.094>
7. A.Rotty, T.Dewayana, A.Habyba. Cross-Industry Standard Process for Data Mining (CRISP-DM) Approach in Determining the Most Significant Employee Engagement Drivers to Sales at X Car Dealership. IEOM Society International, 2023. P. 3368-2279. <https://doi.org/10.46254/AP03.20220552>.

Received (Надійшла) 18.06.2025

Accepted for publication (Прийнята до друку) 22.10.2025

#### ВІДОМОСТІ ПРО АВТОРІВ / ABOUT THE AUTHORS

**Дяченко Дмитро Олександрович** – аспірант кафедри електронних обчислювальних машин, Харківський національний університет радіоелектроніки, Харків, Україна;

**Dmytro Diachenko** – PhD student, Department of Electronic Computers, Kharkiv National University of Radio Electronics Kharkiv, Ukraine;

e-mail: [dmytro.diachenko2@nure.ua](mailto:dmytro.diachenko2@nure.ua); ORCID Author ID: <http://orcid.org/0009-0006-5751-3511>.

**Прокопчик Михайло Володимирович** – студент кафедри електронних обчислювальних машин, Харківський національний університет радіоелектроніки, Харків, Україна;

**Mykhailo Prokopchuk** – student, Department of Electronic Computers, Kharkiv National University of Radio Electronics Kharkiv, Ukraine;

e-mail: [mykhailo.prokopchuk@nure.ua](mailto:mykhailo.prokopchuk@nure.ua); ORCID Author ID: <http://orcid.org/0009-0003-4217-1326>.

**Ровенчак Владислав Миколайович** – студент кафедри електронних обчислювальних машин, Харківський національний університет радіоелектроніки, Харків, Україна;

**Vladyslav Rovenchak** – student, Department of Electronic Computers, Kharkiv National University of Radio Electronics Kharkiv, Ukraine;

e-mail: [vladyslav.rovenchak@nure.ua](mailto:vladyslav.rovenchak@nure.ua); ORCID Author ID: <http://orcid.org/0009-0007-7847-2411>.

**Фролов Андрій Віталійович** – кандидат технічних наук, доцент кафедри комп'ютерно-інтегрованих технологій, автоматизації та робототехніки, Харківський національний університет радіоелектроніки, Харків, Україна;

**Andrii Frolov** – Candidate of Technical Sciences, Associate Professor, Associate Professor of Department of Computer-Integrated Technologies, Automation and Robotics, Kharkiv National University of Radio Electronics Kharkiv, Ukraine;

e-mail: [andrii.frolov@nure.ua](mailto:andrii.frolov@nure.ua); ORCID Author ID: <https://orcid.org/0000-0001-7335-0712>.

Scopus Author ID: <https://www.scopus.com/authid/detail.uri?authorId=57193455909>.

#### Методи інтелектуального аналізу даних з використанням машинного навчання

Д. О. Дяченко, М. В. Прокопчик, В. М. Ровенчак, А. В. Фролов

**Анотація. Актуальність.** В умовах постійного збільшення обсягів інформації в різних галузях економіки, науки та техніки, гостро постає проблема ефективної обробки та аналізу великих масивів даних. Традиційні методи аналітики вже не здатні забезпечити швидке й точне вилучення корисних знань та закономірностей з величезних інформаційних потоків. Відповіддю на цей виклик стають методи інтелектуального аналізу даних, які базуються на сучасних технологіях машинного навчання. Ці методи дозволяють автоматично виявляти приховані закономірності, формувати точні прогнози та приймати обґрунтовані рішення в реальному часі. Враховуючи стрімкий розвиток цифровізації, штучного інтелекту і необхідність швидкого прийняття рішень у конкурентному середовищі, актуальність розробки та удосконалення методів інтелектуального аналізу даних зростає з кожним днем. **Об'єкт дослідження:** процес інтелектуального аналізу даних із застосуванням методів машинного навчання, а саме сукупність алгоритмів, моделей, інструментів та підходів, які забезпечують виявлення прихованих закономірностей, аномалій та структур у великих обсягах різномірної інформації. **Мета статті:** дослідження підходів до інтелектуального аналізу даних із використанням методів машинного навчання, а також виявлення їх ефективності в різних галузях застосування. Стаття спрямована на систематизацію сучасних алгоритмів, аналіз їх переваг і недоліків у контексті обробки великих та високовимірних даних. **Результати дослідження.** здійснено систематичний аналіз ключових методів інтелектуального аналізу даних, які базуються на алгоритмах машинного навчання. Було встановлено, що найбільш ефективними підходами до обробки великих і різномірних обсягів даних є методи класифікації, кластеризації, регресійного аналізу та зменшення розмірності. Особливої ефективності досягають глибокі нейронні мережі при роботі з неструктурованими даними, такими як текст, зображення та часові ряди. Виявлено, що обґрунтований вибір алгоритму залежить не лише від типу даних, а й від задачі. Порівняльний аналіз інструментів показав, що екосистема Python пропонує найбільшу гнучкість, тоді як AutoML-платформи спрощують впровадження моделей для користувачів без глибоких знань програмування. Також проведено огляд сучасних публікацій, які підтверджують практичну цінність машинного навчання в реальних кейсах. Загалом дослідження засвідчило, що машинне навчання є ключовим рушієм еволюції методів Data Mining, дозволяючи здійснювати точну, масштабовану та адаптивну обробку даних в умовах сучасної цифрової трансформації. **Висновки.** Машинне навчання значно розширило можливості інтелектуального аналізу даних, забезпечуючи автоматичне виявлення закономірностей, прогнозування та прийняття рішень на основі великих обсягів інформації. У роботі показано ефективність різних алгоритмів у задачах класифікації, кластеризації, регресії та глибокого навчання. Інструменти на основі Python та хмарні платформи визнані найзручнішими для реалізації аналітичних моделей. Перспективним напрямом є розвиток пояснювального ШІ та гібридних підходів, що поєднують алгоритмічну точність з галузеву експертизою.

**Ключові слова:** інтелектуальний аналіз даних, машинне навчання, класифікація, кластеризація, регресія, глибоке навчання, нейронні мережі, інструменти аналітики, Data Mining, CRISP-DM, AutoML, великі дані, Python.