DOI: https://doi.org/10.36910/6775-2524-0560-2024-54-19
UDC 004.7:004.032.4:006.3:681.3:006.6
Nesterov Vasyl, Data Analyst
https://orcid.org/0009-0000-3204-1382
Itel inc, Jacksonville, Florida, USA

OPTIMIZATION OF BIG DATA PROCESSING AND ANALYSIS PROCESSES IN THE FIELD OF DATA ANALYTICS THROUGH THE INTEGRATION OF DATA ENGINEERING AND ARTIFICIAL INTELLIGENCE

Nesterov V. Optimization of Big Data Processing and Analysis Processes in the Field of Data Analytics Through the Integration of Data Engineering and Artificial Intelligence. In this article, we provide an overview of the current state of artificial intelligence (AI) and big data analytics, which are the most needed needs of our age. This report is designed to provide an overview of the latest research and achievements in this rapidly growing field and to highlight, check and complement the many ways in which AI can be combined to improve data. The most important developments in the field of data analysis demonstrate the development of machine learning models, the importance of explaining AI, and the use of AI in computing. **Methodology:** for this research we have analyse the latest trend in the related field of AI and data engineering. We have short listed nine main papers for this report. Meta-analysis requires the use of descriptive and/or statistical methods to collect data from multiple studies on a particular topic. These ideas help build knowledge from many studies in both qualitative and quantitative ways**. Results:** The article provides an overview of the complexities of integrating large amounts of data with intelligence, exploring topics such as data integrity, social and ethical issues. Additionally, it introduces new applications across multiple industries and reveals opportunities to transform environmental monitoring, supply chain operations, cybersecurity, and health monitoring. Overall, this article provides the reader with an in-depth understanding of the latest advances and challenges in data mining and intelligence-based analysis. **Conclusion:** It contributes to the ongoing debate on how AI can be used to gain meaningful insights from massive data sets, and provides a forward-looking perspective on the potential impacts on multiple sectors of the economy and society in general. Another new concept that expresses the openness of AI algorithms is descriptive intelligence (XAI). Understanding and interpreting intelligence decisions is important because it plays an important role in information processing, especially in areas that require compliance with ethical and legal standards.
Keywords: Artificial Intelligence, Big Data Analytics, Data Processing Optimization, Machine Learning Models, Ethical AI in Data Analytics.

Нестеров В.Ф. Оптимізація процесів обробки та аналізу великих даних у сфері аналітики даних шляхом інтеграції інженерії даних та штучного інтелекту. У цій статті ми надаємо огляд сучасного стану штучного інтелекту (ШІ) та аналітики великих даних, які є найбільш затребуваними потребами нашого часу. Цей звіт має на меті надати огляд останніх досліджень і досягнень у цій швидкозростаючій галузі, а також висвітлити, перевірити і доповнити численні способи поєднання штучного інтелекту для покращення даних. Найважливіші події в галузі аналізу даних демонструють розробку моделей машинного навчання, важливість пояснення ШІ та використання ШІ в обчисленнях. Методологія: для цього дослідження ми проаналізували останні тенденції в суміжних галузях ШІ та інженерії даних. Ми відібрали дев'ять основних статей для цього звіту. Мета-аналіз вимагає використання описових та/або статистичних методів для збору даних з декількох досліджень на певну тему. Ці ідеї допомагають побудувати знання з багатьох досліджень як в якісному, так і в кількісному плані. Результати: У статті представлено огляд складнощів інтеграції великих обсягів даних з аналітикою, досліджено такі теми, як цілісність даних, соціальні та етичні питання. Крім того, вона знайомить з новими застосуваннями в різних галузях і розкриває можливості для трансформації моніторингу навколишнього середовища, операцій в ланцюгах поставок, кібербезпеки і моніторингу здоров'я. Загалом, ця стаття дає читачеві глибоке розуміння останніх досягнень і викликів у галузі інтелектуального аналізу даних та аналізу на основі розвідданих. Висновок: Ця стаття робить свій внесок у триваючі дебати про те, як можна використовувати ШІ для отримання значущих висновків з величезних масивів даних, і дає перспективний погляд на потенційний вплив на різні сектори економіки і суспільства в цілому. Ще одна нова концепція, яка виражає відкритість алгоритмів ШІ, - це дескриптивний інтелект (XAI). Розуміння та інтерпретація рішень інтелекту має важливе значення, оскільки він відіграє важливу роль в обробці інформації, особливо в сферах, які вимагають дотримання етичних та правових норм.
Ключові слова: штучний інтелект, аналіз великих даних, оптимізація обробки даних, моделі машинного навчання, етичний ШІ в аналітиці даних.

Scientific problem statement. As big data analytics and operations merge with the data analytics world, the paper faces a questionable task as organizations face ever-increasing data volume, spread, and diversity. Due to the complexity of this environment, optimization must use an effective method that combines data mining and artificial intelligence (AI). In this case, the statement of the problem focuses on the need to coordinate these two factors to increase the efficiency, speed and accuracy of big data and analysis.

Unparalleled diversity and volume of data: Traditional operating systems cannot handle the large volume and variety of data that companies collect from different sources; because they store many forms of information and do not have a structure. This issue goes past the boundaries of conventional information

handling and capacity and requires superior approaches to oversee the integration of distinctive information sorts.

Inefficient information processing workflow: Current information handling workflows frequently include profound and long information upkeep, change, and integration forms. These wasteful aspects ruin the method of extricating valuable data from the information and constrain the adaptability of the investigation team.

These biases make it impossible to extract useful information from the data and limit the search team's flexibility. Organizations that need to quickly respond to business changes and embrace automation need to address these limitations.

Inadequate

Many of the large-scale data systems in use today have difficulty scaling to meet data development needs. Larger files cannot be displayed due to this limitation. It can reduce the success rate and make it impossible to evaluate the data search plan.

Low Processing Intelligence

Traditional data planning techniques cannot identify patterns, anomalies, and connections in comprehensive data. Advanced data science and machine learning require intelligence directed at data processing pipelines.

The evolution of data and AI integration is different

There is a separation between the data research, analysis and AI departments of the organization. These disciplines influence collaboration and therefore enhance joint activities. To close this gap, technology is needed to connect expertise in data science, research and big data analysis.

Privacy and Data Security Issues

As data analysis and processing becomes more difficult, the potential for security breaches and threats will also increase. Privacy concerns are increasing. While businesses gain access to sensitive data, they strive to protect the availability, confidentiality and integrity of that data. Finding the best balance between security and data availability is a key challenge in the optimization process.

Latest research and publications. Popularity and evolution of big data analytics using AI: AI is revolutionizing data analysis through the development of machine learning models, development of explanatory AI (XAI), and integration of Edge computing for time analysis. These advancements aim to improve accuracy and efficiency in processing large datasets, particularly in sensitive domains like healthcare and finance. Edge computing allows for faster decision-making, reduced latency, and more efficient handling of streaming data, crucial for applications like IoT devices [1, 2].

Automated Machine Learning (AutoML) tools and platforms have gained popularity, allowing organizations to automate the process of building, training, and deploying machine learning models. This democratization empowers users with varying levels of technical expertise to leverage the benefits of machine learning in their data analysis tasks.

Natural Language Processing (NLP) is increasingly integrated into big data analytics, allowing systems to understand and derive insights from unstructured textual data, making it easier to analyze vast amounts of text-based information. AI-driven predictive analytics is a focal point in industries like finance, healthcare, and marketing, enabling proactive risk management and strategic planning [3].

Data security and ethics are important as awareness of the need for data security measures and responsible AI practices increases. The integration of big data and AI technology has become unprecedented, allowing organizations to manage big data and high-performance AI-driven analytics in the network together.

According to the research, the following topics are addressed: AI-powered analytics faces many challenges, including data quality and importance, orchestration and design, ethically fair AI and bias reduction, scalability and resource optimization, and humans and machines. partnership. Ensuring data quality is critical to the success of AI-driven analytics, and organizations are investing in processes that improve data quality. Standardization efforts are important to promote interoperability and facilitate the integration of different AI technologies into the big data ecosystem [4]. Ethical AI and bias mitigation are a priority, with researchers exploring techniques to ensure fairness, transparency, and accountability in AI-driven analytics. Scalability and resource optimization are key considerations, with AI models needing to scale efficiently to handle the increasing volume of data. Human-AI collaboration is evolving, and designing interfaces and workflows that facilitate effective collaboration is crucial for harnessing the complementary strengths of human expertise and AI capabilities in data analysis tasks[4, 5].

Latest applications using AI in field of data science: With AI playing a critical role in illness prediction, therapy enhancement, and supply chain management, healthcare analytics is revolutionizing the way medical professionals manage their resources. Predictive analytics powered by AI helps businesses estimate demand, manage inventory, and increase productivity. AI is also being used to analyze the climate and environment, offering insights about changes in the environment and patterns in the climate. In order to address the effects of climate change and the management of natural resources, this data-driven strategy is essential. Furthermore, through real-time analysis of user behavior and network traffic patterns, AI-driven cybersecurity analytics assists in the identification and mitigation of threats[3, 6-9].

The research purpose: This article's goal is to give readers a thorough understanding of the most recent developments and analyses in the big data analytics and AI fields. The purpose of this article is to gain an understanding of how organizations can improve their data analysis and performance by identifying the strengths that exist at the intersection of these two transitions. The work explores several significant themes, such as the development of models for machine learning, the emergence of explainable AI, the use of AI in edge computing, including the increasing significance of ethical approaches in data analytics.

The article also discusses the difficulties of integrating AI with big data analytics, focusing on problems like data quality, compatibility, and the moral ramifications of AI-driven decision-making. It also looks at new uses in a variety of industries, showing how AI is changing cybersecurity, supply chain optimization, healthcare analytics, and environmental monitoring.

Essentially, the goal of this article is to provide readers with a thorough grasp of the most recent advancements in AI-based big data analytics by acting as a guide. The article aims to add to the larger conversation on using AI to extract valuable insights from big and complex databases while emphasizing the possible effects on many companies as well as society at large by looking at trends, obstacles, and new applications.

Main research material presentation: In the present era of information explosion, enterprises have a complex problem optimizing big data handling and analysis within the field of data analytics. As the amount, velocity, and diversity of data increase, so does the need for creative solutions to guarantee that the enormous datasets can effectively be converted into insights that can be put to use. Given this, combining data engineering with AI becomes a potent tactic to improve large data processing and analysis's overall effectiveness[1].

The Big Data Processing Environment: The emergence of big data has revolutionized the information management practices of enterprises. The sheer amount and variety of data being created, including both structured and unstructured information from several sources, is too much for traditional databases and processing techniques to handle. The exponential growth of information has created a pressing need for optimization solutions to enable enterprises to quickly extract useful insights.

The Function of Data Engineering: Throughout the optimization process, data engineering is essential. It entails creating and overseeing methods, tools, and structures for gathering, storing, and evaluating data. Data engineering is the basis for downstream analytics and includes operations like data input, cleansing, transformation, and integration. Workflows for processing data manually frequently create bottlenecks, which limit the adaptability of analytics groups and lengthen the amount of time it takes to provide insights that are actionable. In data engineering, automation and sophisticated technologies can lead to efficiency improvements. By decreasing manual involvement and increasing the pace at which information is made accessible for analysis, computerized information pipelines can optimize the procedure from data gathering to transformation. Optimizing data retrieval and storage methods also improves the overall efficiency of big data processing[2, 10].

Utilizing AI's Power: Through the introduction of intelligent data processing capabilities, AI expands the scope of big data analytics. A subset of AI known as machine learning algorithms is capable of pattern recognition, data analysis, and prediction without the need for explicit programming. By combining AI with data analysis, businesses can leverage insights hidden in complex data beyond the capabilities of traditional analytical methods [4]. AI's ability to perform predictive analysis is one of its strengths. Machine learning models can use historical data to predict future trends, allowing businesses to make informed decisions. Predictive capability is especially useful in situations where predicting user behavior, physical activity, or work performance is important [6].

Also, AI improves data processing results using cumbersome procedures. Accuracy and speed in tasks such as data classification, anomaly detection, and natural language processing can enable human analysts to focus more on research and management [4].

Integrated Alignment: True optimization can be found in the combination of AI and knowledge engineering. Although the two sometimes appear distinct from each other, their integration is necessary to realize the benefits of big data and success [11].

Automation in Data Science: Data engineering jobs will be greatly impacted by AI-driven automation. For example, using machine learning algorithms to streamline data cleaning processes can improve the accuracy of data preparation. Similarly, AI-driven automatic data integration solutions can adapt to changing data sources and formats to have diverse and dynamic analyzes [10].

Using Intelligent Data Processing for Analytics: Smart Features Like Adaptive Learning AI features instant pattern recognition when integrated with analytics. This allows analysis to change and advance over time in response to changing business needs and data trends. By combining data engineering methods with the complex analysis of AI, a strategy for the optimization of big data is created [3].

Scalability solution: One of the most important factors in processing big data is scalability. As information proceeds to grow, organizations must guarantee that their information operations can suit information development through flat scaling. Adaptability issues in conventional frameworks can lead to diminished execution and expanded uptime. Utilizing AI and data designing together opens up better approaches to unravel scaling issues. An AI-driven auto-scaling framework can distribute assets as required. By doing this, it is guaranteed that the data processing framework can be effortlessly scaled to prepare bigger data and at the same time give great execution [5].

About Safety and Security: As the investigation of huge data gets to be more troublesome, security and privacy issues ended up more imperative. To ensure delicate information, organizations must compromise between simple get to and rigid security methods. Manufactured intelligence can offer assistance illuminate this issue by taking successful security measures and utilizing prescient analytics to identify security dangers. Calculations utilizing machine learning can analyze client activity patterns to recognize peculiarities which will demonstrate vulnerabilities. Moreover, data can be secured utilizing privacy-oriented encryption strategies and mental property models amid examination and handling. A great methodology has been made to fathom these problems by combining information designing with AI-driven security measures and information management [7].

The crevice between AI and information designing: In spite of their benefits, data engineering and AI departments in companies frequently need communication. This gap has to be closed so that these locales can coordinated consistently. Through collaboration, data engineers and AI engineers can make coordinates arrangements that back each other's capabilities [11]. Collaborative teams combine specialized knowledge, skills, and knowledge engineering to achieve a more efficient approach. By establishing effective communication and consistent processes, we ensure that engineering data analysis informs the creation and development of intelligence models, generates feedback, and improves the overall process and analysis performance of big data [10,12].

Talent A lack and Skill Gap: The quick development of AI and big data analytics technologies has resulted in a skills shortage and skill gap that is widening. Finding and keeping experts in both data engineering and AI is a common challenge for organizations. It will take money, upskilling efforts, and deliberate recruiting to close this gap and create a workforce capable of navigating the intricacies of integrated big data optimization [13].

Practical Uses and Success Stories: To maximize big data processing and analysis, a number of businesses have effectively deployed combined data engineering and AI solutions. For instance, AI-driven algorithms are employed in the finance industry to instantly identify fraudulent transactions, improving security and guarding against financial losses. AI-powered predictive analytics in healthcare helps to maximize resource allocation and spot possible disease outbreaks [14].

Medical Data Analytics - AI is being used more and more in healthcare analytics to help with drug development, customized treatment, and diagnostics. In order to provide insights that support illness prediction and therapy, machine learning models examine patient data, medical records, and genetic data [3].

Supply Chain Optimization: - AI is a key component in supply chain management optimization. AI-powered predictive analytics aids businesses in demand forecasting, inventory optimization, and overall supply chain efficiency improvement [8].

Climate and Environmental Analysis: AI is being used to analyze environmental data in order to address natural resource management and climate change. Large datasets from sensors, satellites, and climate models are processed by machine learning algorithms to reveal patterns in the climate and changes in the environment [9].

Intelligence on Cybersecurity: AI-driven analytics plays a critical role in locating and reducing cybersecurity risks. Algorithms that employ machine learning examine trends in user activity and network traffic to instantly identify anomalies and possible security breaches [9].

Conclusions: The field at the nexus of AI and big data analytics is still dynamic and changing. The most recent evaluations draw attention to developments in models for machine learning, the significance of AI that can be explained, the incorporation of AI into edge computing, and the persistent problems with data quality, ethics, and scalability. The future of analytics for data across a range of sectors is anticipated to be shaped by the collaboration between big data and AI, as firms endeavor to extract significant insights from ever growing databases.

In conclusion, a revolutionary strategy for addressing the problems presented by the current data environment is the optimization of big data processing and analysis through the integration of data engineering and AI. Intelligent analysis and electronic data processing working together can increase efficiency and create new opportunities to collect information from large and diverse data sources.

The combination of data engineering and AI solves important problems such as capacity building, security and complexities. As business transitions into this field, it is important to promote collaboration between knowledge engineering and enterprise intelligence, connecting intelligence, keeping up with changes and technology. The interaction between these departments not only accelerates the current process, but also lays the foundation for future innovation in the dynamic field of big data analytics.

Future Directions: Future developments should leverage big data and analytics through the combination of AI and data engineering. Edge computing has gained importance because it processes data closer to its location rather than to the data center. Edge AI integration reduces the need to send large amounts of data and enables decisions to be made faster. Another new concept that expresses the openness of AI algorithms is descriptive intelligence (XAI). Understanding and interpreting intelligence decisions is important because it plays an important role in information processing, especially in areas that require compliance with ethical and legal standards**.** Furthermore, as AI models and algorithms continue to progress and hardware capabilities rise, new avenues for big data processing optimization are anticipated. If implemented at scale, quantum computing might improve data processing rates and make it possible to analyze extraordinarily complex information quickly.

References

1. C. Shang and F. You, "Data analytics and machine learning for smart process manufacturing: Recent advances and perspectives in the big data era," *Engineering,* vol. 5, pp. 1010-1016, 2019.
2. J. Li, M. S. Herdem, J. Nathwani, and J. Z. Wen, "Methods and applications for Artificial Intelligence, Big Data, Internet of Things, and Blockchain in smart energy management," *Energy and AI,* vol. 11, p. 100208, 2023.
3. Z. F. Khan and S. R. Alotaibi, "Applications of artificial intelligence and big data analytics in m-health: a healthcare system perspective," *Journal of healthcare engineering,* vol. 2020, pp. 1-15, 2020.
4. O. M. Araz, T. M. Choi, D. L. Olson, and F. S. Salman, "Role of analytics for operational risk management in the era of big data," *Decision Sciences,* vol. 51, pp. 1320-1346, 2020.
5. R. H. Hariri, E. M. Fredericks, and K. M. Bowers, "Uncertainty in big data analytics: survey, opportunities, and challenges," *Journal of Big Data,* vol. 6, pp. 1-16, 2019.
6. S. Benzidia, N. Makaoui, and O. Bentahar, "The impact of big data analytics and artificial intelligence on green supply chain process integration and hospital environmental performance," *Technological forecasting and social change,* vol. 165, p. 120557, 2021.
7. M. Andronie, G. Lăzăroiu, M. Iatagan, C. Uță, R. Ștefănescu, and M. Cocoșatu, "Artificial intelligence-based decision-making algorithms, internet of things sensing networks, and deep learning-assisted smart process management in cyber-physical production systems," *Electronics,* vol. 10, p. 2497, 2021.
8. R. Dash, M. McMurtrey, C. Rebman, and U. K. Kar, "Application of artificial intelligence in automation of supply chain management," *Journal of Strategic Innovation and Sustainability,* vol. 14, pp. 43-53, 2019.
9. A. Y. Sun and B. R. Scanlon, "How can Big Data and machine learning benefit environment and water management: a survey of methods, applications, and future directions," *Environmental Research Letters,* vol. 14, p. 073001, 2019.
10. S. Raschka, J. Patterson, and C. Nolet, "Machine learning in python: Main developments and technology trends in data science, machine learning, and artificial intelligence," *Information,* vol. 11, p. 193, 2020.
11. M. Naeem, T. Jamal, J. Diaz-Martinez, S. A. Butt, N. Montesano, M. I. Tariq, *et al.*, "Trends and future perspective challenges in big data," in *Advances in Intelligent Data Analysis and Applications: Proceeding of the Sixth Euro-China Conference on Intelligent Data Analysis and Applications, 15–18 October 2019, Arad, Romania,* 2022, pp. 309-325.
12. J. Wang, C. Xu, J. Zhang, and R. Zhong, "Big data analytics for intelligent manufacturing systems: A review," *Journal of Manufacturing Systems,* vol. 62, pp. 738-752, 2022.
13. Z. Allam and Z. A. Dhunny, "On big data, artificial intelligence and smart cities," *Cities,* vol. 89, pp. 80-91, 2019.
14. A. Belhadi, K. Zkik, A. Cherrafi, and M. Y. Sha'ri, "Understanding big data analytics for manufacturing processes: insights from literature review and multiple case studies," *Computers & Industrial Engineering,* vol. 137, p. 106099, 2019.