IJSES
International Journal of Scientific Engineering and Science
Volume 8, Issue 12, pp. 54-58, 2024. ISSN (Online): 2456-7361

Automation of Data Analysis Processes Using Modern Technologies

Aleksandr Vasilchenko
Data Analyst
Moscow, Russia

Abstract— *The article discusses the problems of automating data analysis processes using the latest technological developments. The relevance of this topic is justified by the increasing need for efficient processing of large amounts of information against the background of digitalization of various fields of activity. It is emphasized that automation today is becoming not only a tool for increasing productivity, but also a key factor in the competitiveness of business entities. However, the choice of methods and specific solutions is associated with several contradictions related to their adaptation to the specifics of the tasks, limitations of computing resources, and the complexity of interpreting the results. The study aims to explore the conceptual foundations, opportunities, advantages, and "bottlenecks" of innovative developments that allow automating data analysis. The results obtained make it possible to assert that the latest technologies for automation have great potential; at the same time, their successful implementation requires careful preparation, taking into account the specifics of the tasks. It is important to combine the capabilities of various tools to overcome their limitations, as well as invest in the development of specialist competencies to ensure maximum return on the steps taken. The information presented in the article will be useful to specialists in the field of analytics, developers of intelligent systems, as well as researchers dealing with the problems of digitalization.*

Keywords— *Automation, data analysis, big data, intelligent analysis, machine learning, cloud technologies, processing processes.*

I. Introduction

The era of digitalization has created a necessity for processing significant volumes of information, which serves as the "foundation" for decision-making across various fields of activity. Advances in computational power, along with the enhancement of artificial intelligence (AI) algorithms, have driven increased demand for the automation of data analysis. However, the effectiveness of these approaches is determined not only by their technical parameters but also by their adaptation to specific application conditions. In this regard, many researchers today are interested in exploring contemporary directions and trends in automation that leverage advanced technologies and methodologies.

The research problem lies in the need to develop effective approaches for integrating new technologies for data analysis automation to improve productivity and accuracy in information processing. Despite significant progress in the development of intelligent algorithms, their integration and practical implementation are complicated by limitations in computational resources and the complexity of interpreting results.

II. Methods and Materials

The preparation of this article involved comparative analysis, systematization, synthesis, and generalization. The literature on the topic demonstrates a variety of approaches. It is appropriate to outline the key directions of research presented in the materials.

The works of K.S. Andryushchenko and A.S. Zlobin [1], as well as S.A. Berezyuk [2], focus on the application of cloud solutions for data analysis. The authors emphasize the importance of flexibility and scalability in such platforms, particularly for streaming video processing. General advantages, including integration with analytical tools, are discussed.

Significant contributions to the study of machine learning have been made by D.V. Grigoriev [4] and A.V. Dagaev [5]. These researchers focus on the application of natural language processing for text information analysis, highlighting its potential for working with unstructured data. Methods are examined in detail, with an evaluation of their effectiveness and applicability across various fields. K.A. Kazakov [6] discusses intelligent analytical techniques and their role in automation.

The publications of A.O. Butenko [3] and E.V. Khrapova [9] address the industry-specific applications of these technologies, such as their use in the construction sector. They underscore improvements in project management efficiency and specific features related to the real estate market, emphasizing the unique nature of informational content in these areas.

Issues related to the analysis of specific data types are explored in the works of E.N. Prokofieva [7] and R.I. Tamarova [8]. The authors examine the automation of geospatial information processing, focusing on current trends.

The study by V.I. Chumkin and S.B. Kaidanov [10] is devoted to evaluating artificial intelligence tools for automation. The authors consider the versatility of these approaches and their adaptability to various tasks.

The review of sources demonstrates that contemporary developments in this field encompass a wide range of directions, from industry-specific applications to the processing of complex data types. However, contradictions exist regarding the most effective methods for specific tasks, such as choosing between machine learning and cloud platforms. Additionally, questions about integrating automation tools under conditions of limited computational resources remain insufficiently addressed. These aspects require further research to expand the potential for applying these technologies across diverse fields.

III. Results and Discussion

The automation of data analysis is a complex process involving the systematization of approaches, technologies, and methodologies aimed at minimizing human involvement in the processing, interpretation, and presentation of information. Research on the effectiveness and efficiency of automated data analysis has been conducted since 2021. Data analysis, including its automated aspects, represents an interdisciplinary endeavor. The conceptual foundations of this field are formed at the intersection of data science, information technology, and cognitive research. Figure 1 outlines the foundational principles.

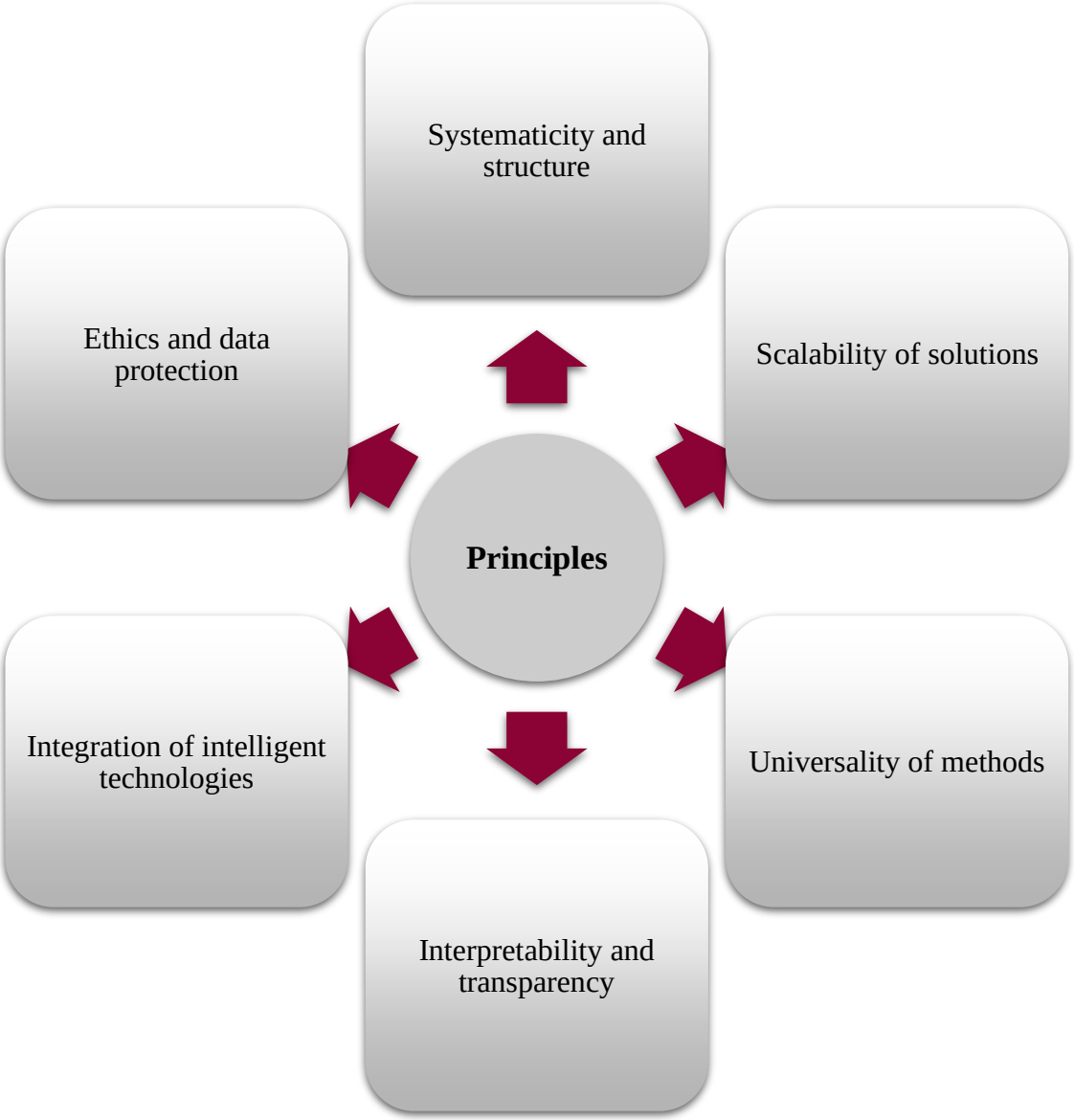

Fig. 1. Systematization of the principles of automation of data analysis processes [3, 6-9]

The process follows a sequential series of stages: data collection, processing, interpretation, and visualization. Automation relies on integrating these steps into a unified system capable of handling various data formats (structured and unstructured) while ensuring interaction between system modules.

Regarding scalability, modern technologies must adapt to the increasing volumes of data streams, necessitating the use of high-performance platforms.

Another principle is applicability across various industries, including finance, healthcare, education, and manufacturing. This is achieved by developing flexible algorithms that can be tailored to specific tasks.

Interpretability and transparency are also critical guiding principles. For instance, deep neural networks are often criticized for their lack of explainability. The conceptual basis of automation includes the development of explainable AI methods that clarify the factors influencing the system's conclusions.

Machine learning algorithms and AI enable the detection of patterns, forecasting, and decision-making without human intervention. A crucial component is the consideration of

privacy aspects, particularly when processing personal information. This requires strict compliance with information security standards.

Based on the conceptual principles and the characteristics of fundamental tenets discussed above, the author proposes the following definition: the automation of data analysis processes involves the implementation of intelligent algorithms and software solutions for systematically and autonomously performing tasks related to data processing, interpretation, and presentation with minimal human involvement. This approach aims to enhance the accuracy, speed, and overall efficiency of information handling. The proposed definition emphasizes the following key aspects:

- Intelligent algorithms and software solutions: Highlighting the use of advanced technologies to perform analytical tasks.
- Systematic and autonomous operations: Indicating that processes should be executed within a structured framework, minimizing human intervention.
- Target-oriented approach: Stating that the goal is not merely to simplify work but also to improve the accuracy, speed, and productivity of data processing.

This definition encompasses both technical and organizational dimensions.

The diagram in Figure 2 illustrates the time efficiency of data analysis processes: manual analysis exhibits linear time growth, whereas automated analysis maintains a fixed duration.

Data analysis processes require consideration of numerous factors, ranging from the quality of the initial data to the complexity of result interpretation. Traditional processing methods often prove insufficient when working with large data volumes. Automation becomes a critical component, reducing human involvement, minimizing errors, and significantly accelerating analytical tasks.

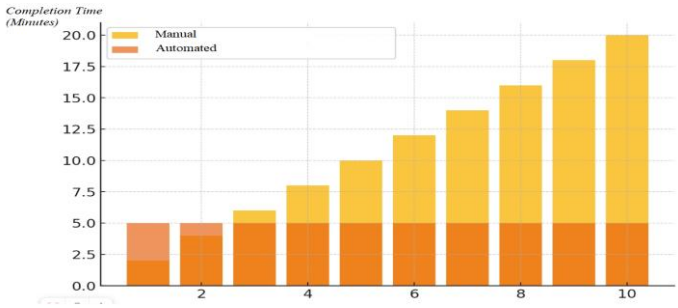

Fig. 2. Graphs of the effectiveness of manual and automated data analysis in comparison [11]

Modern solutions in this field rely on machine learning (ML) algorithms, deep learning (DL) methods, and data mining tools. These approaches extend beyond the capabilities of manual processing, enabling the detection of hidden patterns and the development of complex forecasts.

Machine learning holds a central position in the automation of analytical processes. Its applications (Figure 3) range from linear regression for simple models to neural networks for handling nonlinear data. For instance, gradient boosting demonstrates high accuracy in classification tasks, while clustering methods such as k-means are widely used for data segmentation.

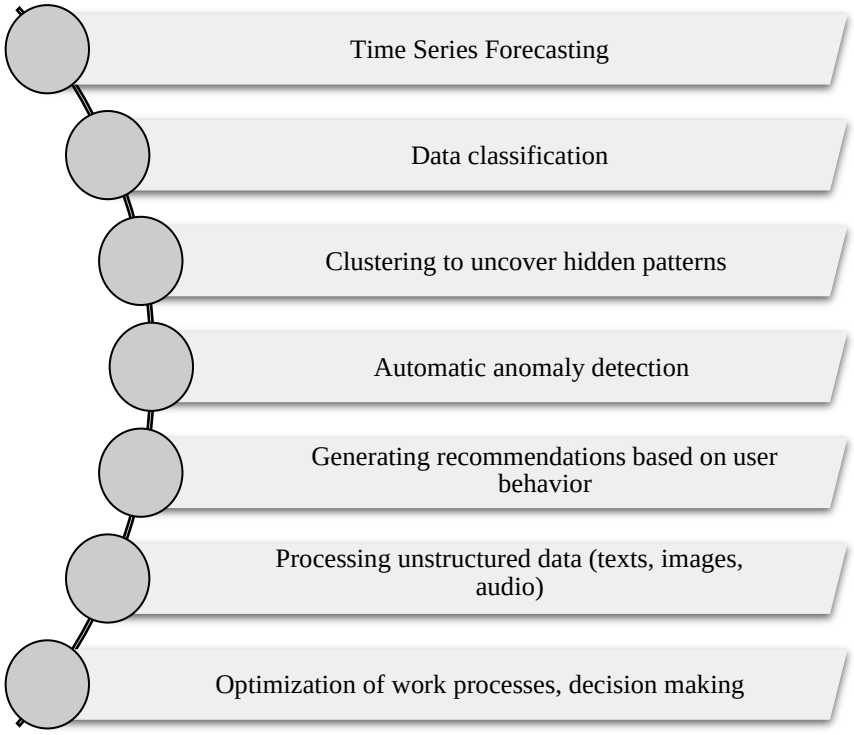

Fig. 3. The possibilities of machine learning in automating data analysis processes [4, 5]

International Journal of Scientific Engineering and Science
Volume 8, Issue 12, pp. 54-58, 2024. ISSN (Online): 2456-7361

Particular attention should be given to reinforcement learning algorithms, which are applied in optimizing complex processes such as dynamic pricing and inventory management.

Cloud platforms, including Google Cloud, AWS, Azure, and others, provide infrastructure for storing and processing information, significantly accelerating the implementation of automated solutions. These platforms offer built-in tools such as BigQuery and SageMaker, facilitating automation at all stages, from preprocessing to the visualization of results [2].

The primary advantage of cloud technologies lies in their scalability and accessibility. Users can adjust computational resources to meet current tasks, thereby reducing the costs associated with maintaining server infrastructure [1].

AI methods expand possibilities by enabling the processing of unstructured data, including images, text, and audio. Computer vision technologies are actively used in monitoring industrial operations. For example, AI-enabled video surveillance systems can detect anomalies in real-time, reducing the costs associated with preventing emergencies [10].

Robotic Process Automation (RPA) is another tool designed to perform routine tasks. Software robots integrate with various information systems, streamlining data uploads, report generation, and document verification.

A notable example of successful RPA implementation is the automation of application processing in the banking sector, which accelerates service delivery and significantly enhances customer service quality. In banking, fraud detection and customer behavior analysis have become more efficient. For instance, E. Sh. Akhmadullina and K. V. Fedorova analyzed machine learning methods, including logistic regression, decision trees, and random forests, to detect fraud in banking card systems. The results demonstrated that the random forest method performed the best (F1 = 0.82) (Figure 4).

Thus, modern technologies for automating data analysis processes offer a rich arsenal of tools that simplify and accelerate information processing. However, despite their evident advantages, each technology has specific limitations that must be considered during practical implementation. Based on research into the efficiency of various data analysis methods, Table 1 summarizes the key characteristics of the most widely used solutions.

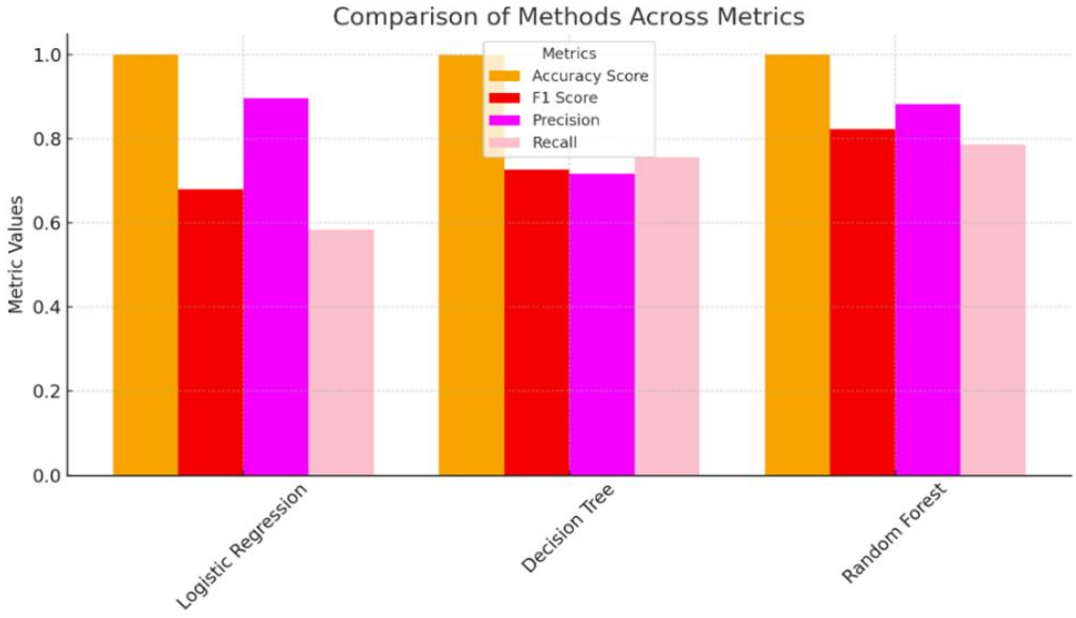

Fig. 4. Accuracy of teaching methods [12]

TABLE 1. Systematization of the advantages and limitations of modern technologies for automating data analysis processes (compiled by the author)

Technology	Advantages	Limitations
Machine Learning (ML)	High accuracy in prediction and classification; adaptability to new data.	Requires large datasets for training; complexity in interpreting some models.
Deep Learning (DL)	Effective with unstructured data; identifies complex dependencies.	High computational complexity; requires large volumes of data.
Robotic Process Automation (RPA)	Automates routine tasks; easy integration with existing systems.	Limited flexibility; unsuitable for complex analytical tasks.
Cloud Technologies	Scalability; access to computational resources without purchasing infrastructure.	Dependency on internet connection; potential risks of data breaches.
Big Data Platforms	Efficient processing of large data streams; integration with analytical tools.	High infrastructure requirements; complexity in configuration and management.
Artificial Intelligence (AI)	Self-learning capabilities; versatility across various industries.	Lack of transparency in decision-making; requires highly skilled specialists for development.

The advantages of modern technologies significantly enhance analytical processes through the automation of routine operations, handling large data volumes, and improving result accuracy.

However, the substantial number of limitations demonstrates that adopting these innovations requires considerable resources, including data, computational power, and professional expertise. For example, while deep learning offers high efficiency, its application is challenging for small organizations due to the high cost of equipment.

IV. Conclusions

The automation of data analysis processes using modern technologies provides extensive opportunities to enhance the efficiency of information management. However, success in this area depends not only on the implementation of advanced developments but also on the careful consideration of factors such as data quality, specialist training, and confidentiality.

Machine learning, artificial intelligence, cloud computing, and related technologies already demonstrate impressive results. Nevertheless, their further development and integration will make data analysis even more accessible, reliable, and accurate, significantly impacting various fields of activity.

The future appears to be tied to the integration of Internet of Things (IoT) technologies and artificial intelligence. IoT generates vast volumes of information, while AI algorithms enable real-time processing. Another promising area is the development of systems incorporating elements of explainable AI. Such solutions facilitate the interpretation of analytical results, increasing user trust and simplifying implementation processes.

Another critical direction is the advancement of hybrid systems that combine the strengths of RPA, ML, and cloud technologies. This development will enable the creation of universal platforms capable of automating a wide range of tasks.

References

[1]. Andryushchenko K.S. Review of current methods of streaming data processing in a cloud environment for video information analysis / K.S. Andryushchenko, A.S. Zlobin // Actual issues of modern science. Collection of articles of the XII International Scientific and Practical Conference. – Penza: 2024. – pp. 46-49.

[2]. Berezyuk S.A. Cloud technologies in storage and analysis of big data / S.A. Berezyuk // Digitalization of society: status, problems, prospects. Materials of the X Annual All-Russian Scientific and Practical Conference. – Moscow: 2023. – pp. 19-27.

[3]. Butenko A.O. Automation of the data analysis process in construction companies: improving efficiency and decision-making / A.O. Butenko // Collection of scientific papers of Russian universities "Problems of economics, finance, and production management". – 2023. – No. 53. – pp. 273-276.

[4]. Grigoriev D.V. Application of machine learning and natural language processing in text data analysis: a review of methods, effectiveness and prospects / D.V. Grigoriev // Scientific research of young scientists. Collection of articles of the XXVIII International Scientific and Practical Conference. – Penza: 2024. – pp. 9-11.

[5]. Dagaev A.V. Methods of data analysis in machine learning / A.V. Dagaev // Economic potential and corporate social responsibility. Materials of the International Scientific and Practical Conference. – St. Petersburg: 2024. – pp. 31-39.

[6]. Kazakov K.A. Research of data processing automation processes using intelligent analysis technologies / K.A. Kazakov // Development of inland waterway transport infrastructure: traditions, innovations (RIVVT-2023). Proceedings of the scientific and practical conference. – St. Petersburg: 2024. – pp. 176-179.

[7]. Prokofieva E.N. Modern trends and trends in automation of geospatial data analysis processes / E.N. Prokofieva, N.V. Tekutyeva, I.I. Tsvetkovskaya // System Administrator. – 2020. – No. 10 (215). – Pp. 86-88.

[8]. Tamarova R.I. Automation of the process of data collection and analysis in science: new opportunities and solutions in the process of professional training / R.I. Tamarova, Yu.V. Medyanik, N.V. Bogdanova // Problems of modern pedagogical education. – 2024. – No. 82-3. – pp. 346-349.

[9]. Khrapova E.V. Automation of the process of analyzing data on the real estate market / E.V. Khrapova, N.M. Kalinina // Scientific and methodological electronic journal "Concept". – 2019. – No. 11. – Pp. 145-153.

[10]. Shumkin V.I. Automation of data analysis using Artificial Intelligence methods / V.I. Shumkin, S.B. Kaysanov // Bulletin of Shakarim University. Technical Sciences. – 2024. – No. 3 (15). – Pp. 37-42.

[11]. Manual and automated testing: advantages and limitations // URL: https://kurshub.ru/journal/blog/ruchnoe-i-avtomatizirovannoe-testirovanie-preimushhestva-i-ogranicheniya/? (date of application: 12/12/2024).

[12]. Akhmadullina E.S. Research of machine learning methods for detecting fraud with bank cards / E.S. Akhmadullina, K.V. Fedorova // Applied mathematics and Computer science: modern research in the field of natural and technical sciences. Collection of materials of the IX International Scientific and Practical Conference. – Tolyatti: 2023. – pp. 217-221.