

Erik Askarovich Aktayev – undergraduate student of the specialty @Computer Engineering@; Astana IT University, Astana, Republic of Kazakhstan; e-mail: aktaeverikwork@gmail.com.

Pavel Aleksandrovich Pak – undergraduate student of the specialty @Computer Engineering@; Astana IT University, Astana, Republic of Kazakhstan; e-mail: park.pavel04@gmail.com.

Поступила в редакцию 07.08.2024

Принята к публикации 12.09.2024

[https://doi.org/10.53360/2788-7995-2024-3\(15\)-6](https://doi.org/10.53360/2788-7995-2024-3(15)-6)

MPHTI: 20.15.05



V. Shumkin*, S. Kaysanov

Shakarim University of Semey,
071412, Republic of Kazakhstan, Semey, 20 A Glinka Street
e-mail: shumkin1999@list.ru

AUTOMATION OF DATA ANALYSIS USING ARTIFICIAL INTELLIGENCE METHODS

Annotation: The article discusses modern approaches to automating data analysis using artificial intelligence (AI) methods. With the rapid growth of data volumes entering various systems, their analysis and processing are becoming a complex task. Automating these processes with AI allows us to increase the efficiency and accuracy of data analysis, minimize the human factor, and speed up decision-making. The article discusses machine learning and deep learning methods used to automate data analysis, as well as examples of their application in various industries, such as finance, medicine, industry, and marketing. Particular attention is paid to the advantages and limitations of existing approaches, as well as prospects for their further development. The article discusses in detail the conditions and methods of research aimed at studying and evaluating the effectiveness of various AI models in automating data analysis. The obtained results are analyzed and prospects for further development of AI technologies in this area are discussed. The study emphasizes the importance of interpretability of AI models, the need to develop new methods that can effectively work with limited and noisy data, as well as reducing the computational costs associated with their use.

Key words: automation of data analysis, artificial intelligence, machine learning, deep learning, big data, intelligent systems.

Introduction

With the development of digital technologies, the volume of data generated by various devices and systems continues to grow exponentially. According to experts, by 2025, the amount of data in the world will reach 175 zettabytes (ZB) [1]. In this regard, the need for effective methods of analyzing and processing information increases. Traditional approaches to data analysis do not always cope with such volumes, which leads to the need to automate the data analysis process.

Artificial intelligence methods, especially machine learning (ML) and deep learning (DL), open up new opportunities for automating data analysis. They allow you to create models that can independently learn from data and make predictions with high accuracy [2]. These methods have already found wide application in various industries, including finance, healthcare, industry and marketing.

The purpose of this article is to review modern methods of automating data analysis using AI, as well as discuss their advantages, limitations and development prospects.

Research Setting and Methods

For the analysis, various approaches and methods for automating data analysis were studied, including classical machine learning algorithms such as linear regression, decision trees, and random forests [3], as well as more modern approaches using neural networks and deep learning [4]. To evaluate the effectiveness of these methods, data from various sources were used, including open datasets from the fields of finance, medicine, and industry [5].

The Python and R programming languages were used for data analysis, as well as specialized libraries such as TensorFlow and PyTorch for working with neural networks [6]. The metrics for assessing the quality of the models included accuracy, recall, F-measure, and ROC curves.

The study was conducted based on the following conditions:

Study Objects: The study objects were diverse datasets from various industries such as finance, medicine, manufacturing, and marketing. These datasets included structured and unstructured data, images, texts, and numeric data. Examples of such datasets include financial transactions, medical images (e.g., MRI or X-ray), industrial sensory data, and customer reviews.

Selection of AI methods: The study used both traditional machine learning methods (e.g., linear regression, decision trees, random forests) and more modern deep learning methods (e.g., convolutional neural networks (CNNs), recurrent neural networks (RNNs), transformers).

Technical resources: The study was conducted using high-performance computing resources, including graphics processing units (GPUs) to accelerate the training of deep learning models. This allowed for experiments with large amounts of data and complex models that require significant computing power.

Software tools: Modern software tools and libraries such as Python and R, as well as specialized libraries for working with AI and machine learning: TensorFlow, Keras, PyTorch, Scikit-Learn and others were used to develop and train the models.

The study included several key steps related to the automation of data analysis using AI methods:

Data collection and preprocessing: This step involved collecting data from various sources (e.g. Kaggle, UCI Machine Learning Repository) and preprocessing it, including cleaning, normalizing, and transforming the data into formats suitable for use in machine and deep learning models. This included handling missing values, removing outliers, categorizing, and scaling features.

Model development and training: For each type of data, appropriate models were developed and trained. For example, recurrent neural networks (RNNs) were used for time series analysis, convolutional neural networks (CNNs) for image classification, and transformer-based models (e.g. BERT) for text analysis. The models were trained on training datasets, and their hyperparameters were selected using grid search or Bayesian optimization methods.

Model Quality Assessment: Trained models were evaluated on test datasets using various metrics such as accuracy, recall, F1-score, and area under the ROC curve (AUC-ROC). Cross-validation methods were used to assess the models' resistance to overfitting.

Analysis and Interpretation of Results: At this stage, the obtained results were analyzed to identify patterns and test hypotheses. Interpretable AI methods (e.g., LIME, SHAP) were used to interpret the models and explain their predictions, which made it possible to understand which features most strongly influence the model's decisions.

Testing and Implementation: At the final stage, the models were tested in conditions close to real ones in order to assess their performance and reliability. For models that showed high results, prototypes were developed for implementation in production systems.

Efficiency Analysis and Optimization: After the models were implemented, their efficiency was analyzed in real conditions. To achieve this, A/B testing and monitoring methods were used to assess the impact of data analysis automation on processes and identify potential areas for further improvement and optimization.

Research Results

The results of the study showed that AI methods, especially deep learning, can significantly improve the accuracy and speed of data analysis compared to traditional methods. For example, in medicine, neural networks have shown excellent results in image recognition and diagnosis of diseases such as cancer. In finance, machine learning methods have improved risk prediction and fraud detection [7].

In addition, it has been found that the automation of data analysis using AI can significantly reduce the time it takes to process and analyze large amounts of data. In industry, this leads to increased efficiency of production processes and reduced costs.

In industry, automated data analysis using AI has found application in the tasks of predicting equipment failures and optimizing production processes. A study by Lee et al. (2013) describes the use of machine learning to predict equipment failures on production lines, which allows for preventive maintenance before problems occur [8]. The use of predictive analysis algorithms, such as regression models and neural networks, has increased the accuracy of failure prediction to 90%. This significantly reduces equipment downtime and repair costs.

In addition, automated data analysis has helped improve quality control processes. The use of CNN for automatic inspection and defect analysis on production lines has reduced the number of

defects by 25% and improved product quality. Such improvements were achieved due to the ability to analyze high-resolution images and detect small defects that are difficult to notice with the naked eye.

Discussion of scientific results

The results obtained confirm the effectiveness of using AI to automate data analysis in various industries. However, despite the obvious advantages, there are also certain limitations. Firstly, the implementation of AI requires significant computing resources and highly qualified specialists. Secondly, AI algorithms can be subject to errors and distortions in data, which requires constant monitoring and adjustment of models [9].

In the future, we can expect further development of AI methods, including improved learning algorithms, increased interpretability of models, and reduced computational costs. It is also important to consider the ethical aspects of AI application, including privacy and data protection issues.

The results of this study confirm the significant potential of artificial intelligence (AI) methods in automating data analysis. The discussion of these results allows us not only to evaluate current achievements, but also to identify the prospects, challenges, and limitations faced by the use of AI in various industries. Below is a detailed discussion of the key aspects.

1. Accuracy and performance of AI methods

The main results show that AI methods, such as machine and deep learning, significantly improve the accuracy and speed of data analysis. In particular, the use of convolutional neural networks (CNN) in medicine has demonstrated impressive accuracy in diagnosing diseases based on images, comparable to the results obtained by qualified dermatologists [10]. This opens up new opportunities for automating diagnostics and early detection of diseases, which is especially important in the context of a shortage of medical personnel and the need to process large volumes of data.

On the other hand, despite the high accuracy, the use of AI in medicine requires strict control and validation to avoid errors that can lead to negative consequences for patients. This highlights the importance of developments in interpretable AI, where methods such as LIME and SHAP make it possible to explain model predictions and make them more transparent to users .

2. Application of AI in Finance

In the financial sector, the use of machine learning algorithms for fraud detection and risk management has also proven to be effective. Algorithms such as random forests and gradient boosting can significantly improve the accuracy of fraudulent transaction detection and reduce false positives. This not only helps financial institutions protect themselves from losses, but also improves customer experience by reducing unnecessary transaction blocking.

However, the implementation of AI in finance also faces challenges related to model interpretation and regulatory compliance. Financial regulators require companies to explain automated decisions, especially in cases of loan or other financial product denials. This requires additional research and development in the field of creating transparent and interpretable models that are not only effective but also compliant with legal requirements and ethical standards.

3. Optimization of production processes

In industry, AI has demonstrated significant success in predicting equipment failures and optimizing production processes. The use of predictive analysis methods such as regression models and neural networks allows for early detection of potential faults and maintenance planning, which reduces equipment downtime and repair costs. In addition, automation of quality control using CNNs enables more accurate detection of defects on production lines, which leads to improved product quality and reduced losses [11].

However, the implementation of AI in production processes requires significant investments in infrastructure and personnel training. It is important to consider that the successful application of AI depends not only on the quality of models and data, but also on the company's readiness for organizational change and the adaptation of new technologies. This highlights the importance of strategic planning and change management when implementing AI in manufacturing.

4. Limitations and Challenges

Despite significant advances, there are certain limitations and challenges in applying AI to automate data analysis:

Need for high-quality data: Training effective AI models requires a large amount of high-quality data. Insufficient data or poor quality data can result in models with low accuracy and

reliability. This is especially true for industries where data collection is difficult or the data contains a lot of noise and outliers.

High computational costs: Training complex deep learning models requires significant computing resources, such as graphics processing units (GPUs) or tensor processing units (TPUs). This creates barriers for small and medium-sized companies that may not have access to such resources.

Ethical and legal aspects: The implementation of AI raises questions of ethics and data privacy, especially in sensitive areas such as medicine and finance. The need to protect data and ensure the fairness of algorithms is becoming an important aspect in the development and application of AI [12].

5. Prospects and directions for further research

The results of this study highlight the need for further developments in the field of interpretable AI, which will improve the transparency and explainability of model decisions, which is especially important in critical industries. In addition, an important direction is the development of methods that can effectively work with.

Conclusion

Automation of data analysis using artificial intelligence methods opens up new opportunities for efficient and accurate analysis of large volumes of information. This article has reviewed the main AI methods, their advantages and limitations, as well as application examples in various industries. Despite the existing challenges, further development of AI technologies promises significant improvements in the field of data analysis automation.

Future developments in AI technologies will focus on improving learning algorithms, reducing computational costs, and increasing the interpretability of models. One of the key areas of research is to create models that are more robust to noise and data distortions, which will allow AI to be applied in real-world settings with even greater efficiency [13]. We should also expect the emergence of more advanced methods for processing and analyzing unstructured data, such as texts and images, which will open up new opportunities for automating data analysis in various industries.

In conclusion, the automation of data analysis using artificial intelligence methods is a powerful tool that is already having a significant impact on various industries and fields of activity. However, to maximize the potential of AI, it is necessary to consider both the technical and ethical aspects of its application, ensuring sustainable and responsible development of technologies in the future.

References

1. Reinsel D. The Digitization of the World From Edge to Core / D. Reinsel, J. Gantz, J. Rydning. – 2018. – 28 p.
2. Dermatologist-level classification of skin cancer with deep neural networks / A. Esteva et al // Nature. – 2017. – № 542(7639). – P. 115-118.
3. Wang F. AI in Health: State of the Art, Challenges, and Future Directions / F. Wang, A. Preininger // Yearbook of Medical Informatics. – 2019. – № 28(01). – P. 16-26.
5. Kshetri N. Blockchain's roles in meeting key supply chain management objectives. / N. Kshetri // International Journal of Information Management. – 2017. – № 39. – P. 80-89.
6. LeCun Y. Deep learning / Y. LeCun, Y. Bengio, G. Hinton // Nature. – 2015. – № 521(7553). – P. 436-444.
7. UCI Machine Learning Repository. (n.d.). <https://archive.ics.uci.edu/ml/index.php>, (date accessed: 27.08.2024)
8. TensorFlow: A System for Large-Scale Machine Learning / M. Abadi et al // OSDI. – 2016. – № 16. – P. 265-283.
9. Mastering the game of Go with deep neural networks and tree search / D. Silver et al // Nature. – 2016. – № 529(7587). – P. 484-489.
10. Dermatologist-level classification of skin cancer with deep neural networks / A. Esteva et al // Nature. – 2017. – № 542(7639). – P. 115-118.
11. Bolton, R.J. Statistical fraud detection: A review. / R.J. Bolton, D.J. Hand // Statistical Science. – 2002. – № 17(3). – P. 235-255.
12. Recent advances and trends in predictive manufacturing systems in big data environment / J. Lee et al // Manufacturing Letters. – 2013. – № 1(1). – P. 38-41.
13. Jordan M.I. Machine learning: Trends, perspectives, and prospects / M.I. Jordan, T.M. Mitchell // Science. – 2015. – № 349(6245). – P. 255-260.

В.И. Шумкин*, С.Б. Кайсанов
Университет имени Шакарима города Семей,
071412, Республика Казахстан, г. Семей, ул. Глинки, 20 А
*e-mail: shumkin1999@list.ru

АВТОМАТИЗАЦИЯ АНАЛИЗА ДАННЫХ С ИСПОЛЬЗОВАНИЕМ МЕТОДОВ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

В статье рассматриваются современные подходы к автоматизации анализа данных с использованием методов искусственного интеллекта (ИИ). В условиях стремительного роста объемов данных, поступающих в различные системы, их анализ и обработка становятся сложной задачей. Автоматизация этих процессов с помощью ИИ позволяет повысить эффективность и точность анализа данных, минимизировать человеческий фактор и ускорить принятие решений. В статье обсуждаются методы машинного обучения и глубинного обучения, используемые для автоматизации анализа данных, а также примеры их применения в различных отраслях, таких как финансы, медицина, промышленность и маркетинг. Особое внимание уделяется преимуществам и ограничениям существующих подходов, а также перспективам их дальнейшего развития. В статье подробно рассматриваются условия и методы исследования, направленные на изучение и оценку эффективности различных моделей ИИ в автоматизации анализа данных. Проводится анализ полученных результатов и обсуждаются перспективы дальнейшего развития технологий ИИ в этой области. Исследование подчеркивает важность интерпретируемости моделей ИИ, необходимости разработки новых методов, способных эффективно работать с ограниченными и шумными данными, а также снижения вычислительных затрат, связанных с их применением.

Ключевые слова: автоматизация анализа данных, искусственный интеллект, машинное обучение, глубинное обучение, большие данные, интеллектуальные системы.

В.И. Шумкин*, С.Б. Кайсанов
Семей қаласының Шәкәрім атындағы университеті,
071412, Қазақстан Республикасы, Семей қ., Глинки к-сі, 20 А
*e-mail: shumkin1999@list.ru

ЖАСАНДЫ ИНТЕЛЛЕКТТІК ӘДІСТЕРДІ ПАЙДАЛАНАТЫН ДЕРЕКТЕРДІ ТАЛДАУДЫ АВТОМАТТАНДЫРУ

Мақалада жасанды интеллект (AI) әдістерін пайдалана отырып, деректерді талдауды автоматтандырудың заманауи тәсілдері талқыланады. Әртүрлі жүйелерге түсетін деректер көлемінің жылдам өсуімен оларды талдау және өңдеу күрделі міндетке айналады. Бұл процестерді AI көмегімен автоматтандыру деректерді талдаудың тиімділігі мен дәлдігін жақсартады, адам қатесін азайтады және шешім қабылдауды жылдамдатады. Мақалада деректерді талдауды автоматтандыру үшін қолданылатын машиналық оқыту және терең оқыту әдістері, сондай-ақ оларды қаржы, медицина, өндіріс және маркетинг сияқты әртүрлі салаларда қолдану мысалдары қарастырылады. Қолданыстағы тәсілдердің артықшылықтары мен шектеулеріне, сондай-ақ оларды одан әрі дамыту перспективаларына ерекше назар аударылады. Мақалада деректерді талдауды автоматтандыруда әртүрлі AI модельдерінің тиімділігін зерттеуге және бағалауға бағытталған шарттар мен зерттеу әдістері егжей-тегжейлі қарастырылады. Алынған нәтижелер талданады және осы саладағы AI технологияларын одан әрі дамыту перспективалары талқыланады. Зерттеу AI үлгілерінің интерпретациялануының маңыздылығын, шектеулі және шулы деректермен тиімді жұмыс істей алатын жаңа әдістерді өзірлеу қажеттілігін және оларды пайдаланумен байланысты есептеу шығындарын азайтуды көрсетеді.

Түйін сөздер: деректерді талдауды автоматтандыру, жасанды интеллект, машиналық оқыту, терең оқыту, үлкен деректер, интеллектуалды жүйелер.

Information about the authors

Vladislav Shumkin* – Master of Technical Sciences, lecturer of the Department «IT Technology»; Shakarim University of Semey, Republic of Kazakhstan; e-mail: shumkin1999@list.ru. ORCID: <https://orcid.org/0009-0005-0652-5603>.

Sovetkazy Kaysanov – lecturer of the Department «IT Technology»; Shakarim University of Semey, Republic of Kazakhstan; e-mail: kaisanov@mail.ru.

Сведения об авторах

Владислав Игоревич Шумкин* – магистр технических наук, преподаватель кафедры «IT-технологий»; Университет имени Шакарима города Семей, Республика Казахстан; e-mail: shumkin1999@list.ru. ORCID: <https://orcid.org/0009-0005-0652-5603>.

Советказы Бекенович Кайсанов – преподаватель кафедры «IT-технологий»; Университет имени Шакарима города Семей, Республика Казахстан; e-mail: kaisanov@mail.ru.

Авторлар туралы мәліметтер

Владислав Игоревич Шумкин* – техника ғылымдарының магистрі, «IT технологиялар» кафедрасының оқытушысы; Семей қаласының Шәкәрім атындағы университеті, Қазақстан Республикасы; e-mail: shumkin1999@list.ru. ORCID: <https://orcid.org/0009-0005-0652-5603>.

Советказы Бекенович Кайсанов – «IT технологиялар» кафедрасының оқытушысы; Семей қаласының Шәкәрім атындағы университеті, Қазақстан Республикасы; e-mail: kaisanov@mail.ru.

Received 01.08.2024

Revised 28.08.2024

Accepted 12.09.2024

[https://doi.org/10.53360/2788-7995-2024-3\(15\)-7](https://doi.org/10.53360/2788-7995-2024-3(15)-7)

FTAXP: 81.93.29



А. Бимырзақызы*, Ж.М. Алимжанова

Әл-Фараби атындағы Қазақ Ұлттық Университеті,
050040, Қазақстан Республикасы, Алматы қаласы, Әл-Фараби даңғылы, 71
*e-mail: akerkebeimirzakizi@gmail.com

ЗЕРТТЕУ АРҚЫЛЫ КИБЕРҚАУІПТЕРДІ АНЫҚТАУ

Аңдатпа: Жұмыста киберқауіптерді анықтау үшін қолданылатын әлеуметтік желілерді талдаудың негізгі әдістеріне шолу берілген. Әлеуметтік желілердегі қауіптердің негізгі түрлері көрсетілген және олардың алдын алудың кейбір қорғау әдістері сипатталған. Киберқауіптерді анықтауға бағытталған әлеуметтік желіні талдаудың типтік міндеттері, мысалы, желідегі қауымдастықтарды анықтау, қауымдастықтардағы көшбасшылар мен сарапшыларды анықтау, қауымдастықтардың тұрақтылығын талдау, мәтіндік ақпаратты кластерлеу және т.б. Цифрландыруды ұлғайту және әлеуметтік желілерді коммуникация құралы ретінде белсенді пайдалану жағдайында киберқауіпсіздікті қамтамасыз ету үшін тиімді мониторинг пен деректерді талдаудың маңыздылығы барған сайын өзекті бола түсуде. Әлеуметтік желілердегі деректердің үлкен көлемін өңдеуге байланысты күрделіліктер мен қиындықтар, соның ішінде құпиялылық мәселелері мен пайдаланушы деректерін бақылаудың этикалық аспектілері де талданады. Зерттеу әлеуметтік желіні киберқауіптерді белсенді түрде анықтау және алдын алу құралы ретінде пайдаланудың тұтас көрінісін ұсынады, аналитикалық жүйелерді ұйымдар мен жеке тұлғалардың жалпы киберқауіпсіздік шеңберіне біріктірудің маңыздылығын көрсетеді. Зерттеу әлеуметтік желіні киберқауіптерді белсенді анықтау және алдын алу құралы ретінде пайдаланудың тұтас көрінісін береді, аналитикалық жүйелерді ұйымдар мен жеке тұлғалардың жалпы киберқауіпсіздік ұстанымына біріктірудің маңыздылығын көрсетеді, сонымен бірге машиналық оқытуды жақсарту қажеттілігіне назар аударады және сандық ортадағы қауіптердің алдын алу мен бейтараптандырудың тиімдірек болуына ықпал ете алатын пайдаланушы мінез-құлқын және қауымдастық динамикасын тереңірек және дәлірек талдауға арналған жасанды интеллект әдістері.

Түйін сөздер: әлеуметтік желіні талдау, фишинг, киберқауіптер, көшбасшыны анықтау әдістері, ақпараттың шығуы, Advanced Persistent Threat (APT) шабуылдары.

Кіріспе

Әлеуметтік желілердің қарқынды дамуы және олардың мәліметтерді жинақтау қабілеті олардың аналитикасына деген қызығушылықты арттырды және таланттарды иемдену, кәсіби топтастыру, әлеуметтік ұсыныстар, маркетинг, қоғаммен байланыс және жарнама сияқты көптеген салаларда қолданылатын жаңа әдістерді жасауға ықпал етті. Әлеуметтік желіні талдау қазіргі уақытта экономикалық және басқарушылық процестер мен құбылыстарды зерттеу үшін белсенді түрде қолданылады. Сондай-ақ ол жеке басын ұрлаумен, интернет-алаяқтықпен, киберқауіптермен, бағалы қағаздармен айла-шарғы жасау және инвестициялық алаяқтықпен күресу үшін, сондай-ақ қылмыстың алдын алу және басқа да осыған ұқсас мәселелерді шешу үшін қолданылады [1].

Осыған байланысты әлеуметтік желілер ақпараттық өріске және адамдардың психологиясына әсер ету үшін көбірек қолданылады. Олар қоғамдық пікірді қалыптастыруда,