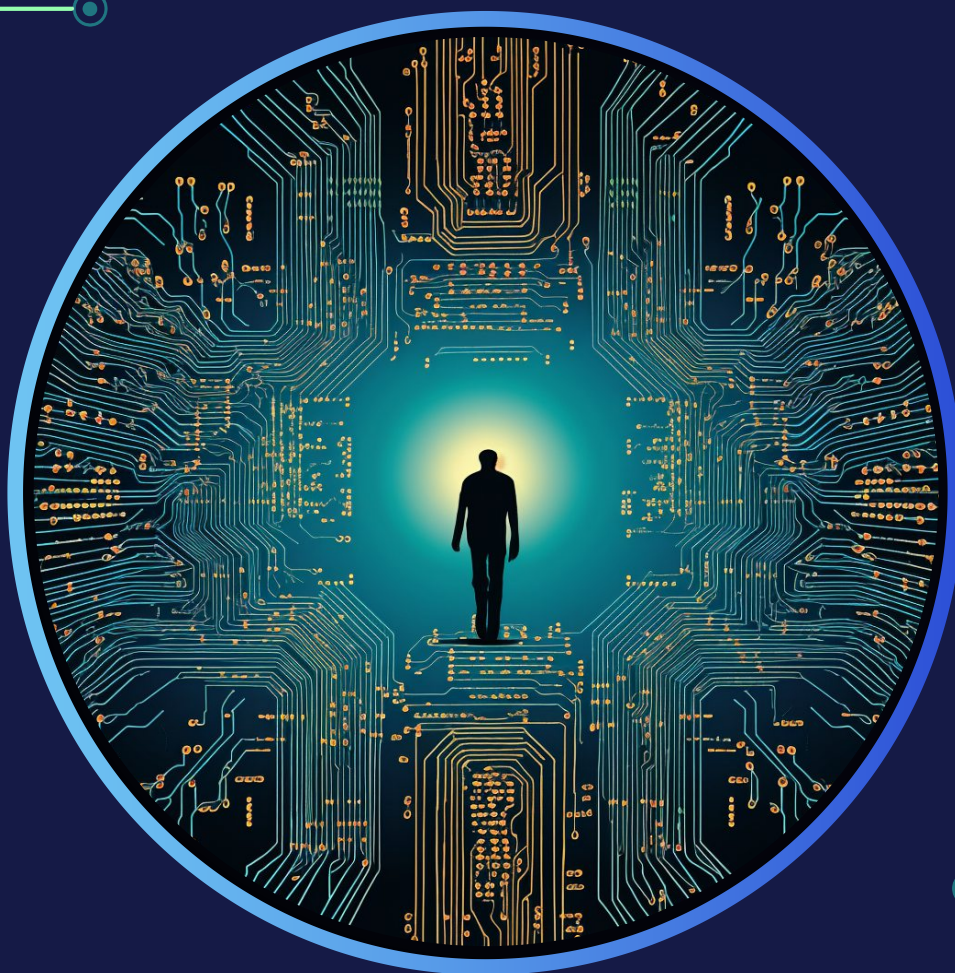




SATBAYEV
UNIVERSITY

COMPUTING AND ENGINEERING



1(1)
March 2023

EDITOR-IN-CHIEF

Raissa Uskenbayeva, doctor of technical sciences, professor, Institute of Automation and Information Technologies of Satbayev University, Kazakhstan

MANAGING EDITOR

Assem Ayapbergenova, master, Institute of Automation and Information Technologies of Satbayev University, Kazakhstan

MEMBERS OF THE EDITORIAL BOARD

Vladimir Barakhnin, doctor of technical sciences, professor, Novosibirsk State University, Russia

Akambai Beisembayev, candidate of technical sciences, associate professor, Institute of Automation and Information Technologies of Satbayev University, Kazakhstan

Oleg Bodin, doctor of technical sciences, professor, Penza State Technological University, Russia

Young-Im CHO, PhD, Gachon University, South Korea

Sergey Gnatyuk, doctor of technical sciences, associate professor, National Aviation University, Ukraine

Zhuldyz Kalpeeva, PhD, associate professor, Institute of Automation and Information Technologies of Satbayev University, Kazakhstan

Erol Kurt, PhD, professor, Gazi University, Turkey

Vyacheslav Kosnikov, candidate of technical sciences, associate professor, Institute of Automation and Information Technologies of Satbayev University, Kazakhstan

Vitaly Levashenko, doctor of technical sciences, professor, Žilinská univerzita, Slovakia

Ayman Moldagulova, candidate of physical and mathematical sciences, professor, Institute of Automation and Information Technologies of Satbayev University, Kazakhstan

Ravil Mukhamediev, doctor of technical sciences, professor, Institute of Automation and Information Technologies of Satbayev University, Kazakhstan

Kasymbek Ozhikenov, candidate of technical sciences, professor, Institute of Automation and Information Technologies of Satbayev University, Kazakhstan

Ryshan Satybaldieva, candidate of technical sciences, associate professor, Institute of Automation and Information Technologies of Satbayev University, Kazakhstan

Batyrbek Suleimenov, doctor of technical sciences, professor, Institute of Automation and Information Technologies of Satbayev University, Kazakhstan

Olga Shiryaeva, candidate of technical sciences, associate professor, Institute of Automation and Information Technologies of Satbayev University, Kazakhstan

Yerlan Tashtai, candidate of technical sciences, associate professor, Institute of Automation and Information Technologies of Satbayev University, Kazakhstan

Waldemar Wojcik, doctor of technical sciences, professor, Lublin Polytechnic University, Poland

Gulbakhar Yussupova, PhD, associate professor, Institute of Automation and Information Technologies of Satbayev University, Kazakhstan

Kanat Zhunussov, candidate of physical and mathematical sciences, professor, Institute of Automation and Information Technologies of Satbayev University, Kazakhstan

БАС ҒЫЛЫМИ РЕДАКТОР

Раиса Өскенбаева, техника ғылымдарының докторы, профессор, Satbayev University Автоматика және ақпараттық технологиялар институты, Қазақстан

ЖАУАПТЫ ХАТШЫ

Әсем Аяпбергенова, магистр, Satbayev University Автоматика және ақпараттық технологиялар институты, Қазақстан

РЕДАКЦИЯЛЫҚ АЛҚА МҮШЕЛЕРІ

Владимир Барахнин, техника ғылымдарының докторы, профессор, Новосібір мемлекеттік университеті, Ресей

Акамбай Бейсембаев, техника ғылымдарының кандидаты, қауымдастырылған профессор Satbayev University Автоматика және ақпараттық технологиялар институты, Қазақстан

Олег Бодин, техника ғылымдарының докторы, профессор, Пенза мемлекеттік технологиялық университеті, Ресей

Young-Im CHO, PhD, Гачон Университеті, Оңтүстік Корея

Сергей Гнатюк, д.т.н., доцент, Ұлттық авиация университеті, Украина

Жұлдыз Қальпеева, PhD, қауымдастырылған профессор, Satbayev University Автоматика және ақпараттық технологиялар институты, Қазақстан

Erol Kurt, PhD, профессор, Гази Университеті, Түркия

Вячеслав Косников, техника ғылымдарының кандидаты, қауымдастырылған профессор Satbayev University Автоматика және ақпараттық технологиялар институты, Қазақстан

Виталий Левашенко, техника ғылымдарының докторы, профессор, Жилин университеті, Словакия

Айман Молдағұлова, физика-математика ғылымдарының кандидаты, профессор Satbayev University Автоматика және ақпараттық технологиялар институты, Қазақстан

Равиль Мұхамедиев, техника ғылымдарының докторы, профессор, Satbayev University Автоматика және ақпараттық технологиялар институты, Қазақстан

Қасымбек Өжікенов, техника ғылымдарының кандидаты, профессор, Satbayev University Автоматика және ақпараттық технологиялар институты, Қазақстан

Рысхан Сатыбалдиева, техника ғылымдарының кандидаты, қауымдастырылған профессор Satbayev University Автоматика және ақпараттық технологиялар институты, Қазақстан

Батырбек Сүлейменов, техника ғылымдарының докторы, профессор, Satbayev University Автоматика және ақпараттық технологиялар институты, Қазақстан

Ольга Ширяева, техника ғылымдарының кандидаты, қауымдастырылған профессор Satbayev University Автоматика және ақпараттық технологиялар институты, Қазақстан

Ерлан Таштай, техника ғылымдарының кандидаты, қауымдастырылған профессор Satbayev University Автоматика және ақпараттық технологиялар институты, Қазақстан

Waldemar Wojcik, техника ғылымдарының докторы, профессор, Люблин политехникалық университеті, Польша

Гүльбахар Юсупова, PhD, қауымдастырылған профессор, Satbayev University Автоматика және ақпараттық технологиялар институты, Қазақстан

Канат Жунусов, физика-математика ғылымдарының кандидаты, қауымдастырылған профессор, Satbayev University Автоматика және ақпараттық технологиялар институты, Қазақстан

ГЛАВНЫЙ НАУЧНЫЙ РЕДАКТОР

Раиса Ускенбаева, д.т.н., профессор, Институт автоматике и информационных технологий Satbayev University, Казахстан

ОТВЕТСТВЕННЫЙ СЕКРЕТАРЬ

Асем Аяпбергенова, магистр, Институт автоматике и информационных технологий Satbayev University, Казахстан

ЧЛЕНЫ РЕДАКЦИОННОЙ КОЛЛЕГИИ

Владимир Барахнин, д.т.н., профессор, Новосибирский государственный университет, Россия

Акамбай Бейсембаев, к.т.н., ассоциированный профессор, Институт автоматике и информационных технологий Satbayev University, Казахстан

Олег Бодин, д.т.н., профессор, Пензенский государственный технологический университет, Россия

Young-Im CHO, PhD, Университет Гачон, Южная Корея

Сергей Гнатюк, д.т.н., доцент, Национальный авиационный университет, Украина

Жулдыз Кальпеева, PhD, ассоциированный профессор, Институт автоматике и информационных технологий Satbayev University, Казахстан

Erol Kurt, PhD, профессор, Университет Гази, Турция

Вячеслав Косников, к.т.н., ассоциированный профессор, Институт автоматике и информационных технологий Satbayev University, Казахстан

Виталий Левашенко, д.т.н., профессор, Жилинский университет, Словакия

Айман Молдагулова, к.ф.-м.н., профессор, Институт автоматике и информационных технологий Satbayev University, Казахстан

Равиль Мухамедиев, д.т.н., профессор, Институт автоматике и информационных технологий Satbayev University, Казахстан

Касымбек Ожикенов, к.т.н., профессор, Институт автоматике и информационных технологий Satbayev University, Казахстан

Рысхан Сатыбалдиева, к.т.н., ассоциированный профессор, Институт автоматике и информационных технологий Satbayev University, Казахстан

Батырбек Сулейменов, д.т.н., профессор, Институт автоматике и информационных технологий Satbayev University, Казахстан

Ольга Ширяева, к.т.н., ассоциированный профессор, Институт автоматике и информационных технологий Satbayev University, Казахстан

Ерлан Таштай, к.т.н., ассоциированный профессор, Институт автоматике и информационных технологий Satbayev University, Казахстан

Waldemar Wojcik, д.т.н., профессор, Люблинский политехнический университет, Польша

Гульбахар Юсупова, PhD, ассоциированный профессор, Институт автоматике и информационных технологий Satbayev University, Казахстан

Канат Жунусов, к.ф.-м.н., ассоциированный профессор, Институт автоматике и информационных технологий Satbayev University, Казахстан

Blockchain-based voting system: a systematic literature review

A. Omar*, Zh. Kalpeyeva

Satbayev University, Almaty, Kazakhstan

*Corresponding author: abylay.omar@gmail.com

Abstract. Democratic elections are a cornerstone of modern society, enabling citizens to exercise their right to vote and express their preferences for political leaders and policies. However, traditional voting systems have faced numerous challenges in recent years, including allegations of fraud, hacking, and misinformation. To address these challenges, many countries have started exploring new technologies that can help secure and modernize the voting process. One such technology is blockchain, a decentralized and tamper-proof database that allows multiple parties to maintain a shared ledger without the need for a central authority. By providing a transparent and immutable record of all transactions, blockchain technology has the potential to revolutionize the way we conduct elections, making them more secure, transparent, and efficient. In this paper, we will explore the benefits and challenges of using blockchain technology in voting systems.

Keywords: *blockchain, voting, e-voting, voting system, decentralized.*

1. Introduction

Blockchain technology has emerged as a promising solution for secure and transparent voting systems. By providing a decentralized and tamper-proof database, blockchain systems can help ensure the integrity of voting processes and increase public trust in democratic institutions. In this paper, we will explore the benefits and challenges of using blockchain technology in voting systems, drawing on case studies from around the world. We will also discuss some of the criticisms and concerns that have been raised about blockchain-based voting systems, and provide recommendations for future research and implementation. The topic of using blockchain technology in voting systems is important and worth studying for several reasons:

1. Ensuring the integrity of democratic processes: Voting is a fundamental component of democratic processes, and any attempt to manipulate or influence the outcome of an election can undermine the legitimacy of the democratic system. Using blockchain technology in voting systems can help ensure the integrity of the voting process by providing a transparent, tamper-proof, and auditable record of all transactions.

2. Increasing trust in the voting process: Trust is essential for the functioning of democratic systems, and any perceived lack of trust in the voting process can undermine public confidence in the democratic system. By using blockchain technology, voting systems can provide a high degree of transparency and security, increasing trust in the voting process.

3. Improving efficiency and reducing costs: Traditional voting systems can be time-consuming and expensive to administer, requiring significant resources and infrastructure. By using blockchain technology, voting systems can be designed to be more efficient and cost-effective, reducing the burden on election officials and taxpayers.

4. Facilitating more inclusive and accessible voting: Traditional voting systems can present barriers to participation

for certain groups, such as people with disabilities or those living in remote areas. By using blockchain technology, voting systems can be designed to be more inclusive and accessible, allowing more people to participate in the democratic process.

5. Advancing the development and implementation of blockchain technology: Blockchain technology is a rapidly developing area with many potential applications, and the study of blockchain-based voting systems can contribute to the advancement of this technology by identifying challenges and opportunities for further development and implementation. Overall, the study of blockchain-based voting systems is important and worth pursuing because it has the potential to improve the integrity, efficiency, accessibility, and inclusivity of democratic processes, while also contributing to the development and implementation of blockchain technology.

2. Background

Blockchain is a type of distributed ledger technology (DLT) that allows multiple parties to maintain a shared database without the need for a central authority. In a blockchain system, each block in the chain contains a cryptographic hash of the previous block, making it difficult to tamper with past transactions. This makes blockchain technology well-suited for applications where transparency, security, and immutability are important.

One of the key features of blockchain technology is decentralization, which means that no single entity controls the database. Instead, all parties in the network have a copy of the database, and any changes to the database must be approved by consensus among the parties. This makes blockchain systems resistant to tampering and hacking, as any attempt to change the data in one copy of the database will be rejected by the other copies.

Another important feature of blockchain technology is transparency. In a blockchain system, all transactions are

recorded in a public ledger that can be accessed and verified by anyone in the network. This makes it possible to trace the history of any transaction and ensure that it has not been tampered with. Additionally, some blockchain systems can be designed to allow users to verify the integrity of their own transactions, further increasing transparency and accountability.

Blockchain technology has a wide range of applications, including cryptocurrency, supply chain management, digital identity, and voting systems, among others. By providing a secure and transparent way to record and verify transactions, blockchain technology has the potential to revolutionize many industries and enable new forms of collaboration and innovation.

2.1. Evolution of voting systems

The evolution of voting systems has been a continuous process, adapting to societal changes and technological advancements. From paper-based ballots to electronic voting machines, each iteration has aimed to streamline the voting process and address inherent challenges. However, these advancements have brought about their own set of concerns, particularly related to the security and transparency of the electoral process. Instances of hacking, manipulation, and doubts surrounding the accuracy of results have underscored the need for a more robust and secure voting infrastructure.

The origins of modern voting systems can be traced back to ancient civilizations where rudimentary forms of voting, often conducted in public forums, laid the groundwork for democratic principles. However, it wasn't until the 17th and 18th centuries that structured voting methods began to emerge.

1. Voice Voting and Paper Ballots

a. Early democratic practices involved voice voting, where citizens verbally expressed their choices. This method, while simple, lacked privacy and was susceptible to external influences.

b. The introduction of paper ballots marked a significant step forward. Voters could now cast their votes in writing, providing a level of secrecy and reducing the potential for coercion. This approach became widespread during the 19th century.

2. Lever Machines and Mechanical Voting

a. The 19th and early 20th centuries witnessed the introduction of lever machines and mechanical voting systems. These innovations aimed to streamline the voting process and eliminate errors associated with manual vote counting.

b. Lever machines, for instance, allowed voters to pull levers corresponding to their chosen candidates, automatically recording and tallying the votes. While these systems expedited the counting process, they posed challenges related to maintenance and reliability.

3. Punch Card and Optical Scan Systems

a. The mid-20th century brought about the adoption of punch card voting systems. Voters would use a punch tool to indicate their choices on a card, which was then tabulated by machines. This automated counting but was prone to inaccuracies and logistical issues, as highlighted by the infamous "hanging chads" in the 2000 United States presidential election.

b. Optical scan systems, introduced later, used technology to read marked paper ballots, addressing some of the issues

associated with punch cards. These systems provided a more accurate and efficient means of tabulating votes.

4. Electronic Voting Machines

a. The late 20th century saw the rise of electronic voting machines, offering a departure from paper-based systems. These machines allowed voters to cast their votes electronically, aiming to reduce errors and streamline the counting process.

b. While electronic voting machines improved efficiency, they raised concerns about security vulnerabilities and the potential for tampering. Instances of hacking and doubts about the integrity of electronic voting systems prompted a reevaluation of their use in some regions.

5. Challenges and the Need for Innovation

a. Throughout this evolutionary journey, voting systems have faced persistent challenges, including issues of accessibility, security, and transparency. The desire for more inclusive, secure, and transparent elections has driven the exploration of innovative solutions, leading to the intersection of voting systems with emerging technologies, such as blockchain.

In this context, the emergence of blockchain technology represents a potential paradigm shift in the evolution of voting systems. By leveraging the principles of decentralization, immutability, and transparency, blockchain offers a unique opportunity to address longstanding challenges and pave the way for a more robust and trustworthy electoral process. The subsequent sections of this literature review will delve into the specific ways in which blockchain has been explored and implemented to enhance the security and transparency of voting systems.

2.2. The promise of blockchain in voting systems

Blockchain technology, originally conceived as the foundational architecture for cryptocurrencies, has garnered increasing attention for its potential to revolutionize various industries, with voting systems standing out as a domain ripe for innovation. The promises of integrating blockchain into voting systems are manifold and address some of the persistent challenges that traditional voting methods face:

1. Decentralization and Security

a. One of the key promises of blockchain in voting systems is the principle of decentralization. Traditional voting systems often rely on central authorities for oversight, which can be vulnerable to manipulation or hacking. Blockchain, being inherently decentralized, distributes the record of votes across a network of nodes, making it extremely challenging for any single entity to control or compromise the system.

b. The decentralized nature of blockchain enhances the security of the voting process, providing a safeguard against unauthorized access, tampering, or fraud. Each block in the chain is linked cryptographically to the previous one, creating a chain of blocks that is resistant to alteration.

2. Immutability and Transparency

a. Immutability, a fundamental characteristic of blockchain, ensures that once a vote is recorded, it cannot be altered or deleted. This feature instills confidence in the integrity of the electoral process, as voters and election officials can trust that the recorded votes remain unchanged.

b. Transparency is another crucial aspect facilitated by blockchain. Every participant in the network has access to a transparent and immutable record of the votes cast. This transparency not only engenders trust but also allows for

independent verification of the election results, fostering a more open and accountable electoral process.

3. Elimination of Fraud and Double Voting

a. Blockchain's cryptographic principles and consensus mechanisms significantly reduce the risk of fraudulent activities. Votes are securely recorded, and the transparency of the system makes it easier to identify and eliminate fraudulent attempts.

b. The use of cryptographic keys ensures that each voter can cast only one vote, preventing the possibility of double voting. This enhances the accuracy and fairness of the electoral process, addressing a common concern in traditional voting systems.

4. Accessibility and Inclusivity

a. Blockchain-based voting systems have the potential to enhance accessibility and inclusivity in the electoral process. The technology allows for remote and online voting, enabling individuals who face physical barriers or are geographically distant to participate in elections.

b. By leveraging blockchain, voting systems can potentially reach a broader demographic, including those with mobility challenges, expatriates, and individuals residing in remote areas.

5. Trust in the Electoral Process

a. The transparency, security, and immutability offered by blockchain contribute to building trust in the electoral process. Trust is a cornerstone of any democratic system, and blockchain's features help mitigate doubts and concerns related to the accuracy and legitimacy of election outcomes.

While the promises of blockchain in voting systems are compelling, it's essential to acknowledge that implementing such a transformative technology comes with its own set of challenges and considerations. Issues like scalability, user adoption, and the need for standardized protocols must be carefully navigated to fully realize the potential of blockchain in revolutionizing the way societies conduct their elections. The subsequent sections of this literature review will delve into the existing research and insights regarding the practical implementations and challenges associated with blockchain-based voting systems.

2.3. Challenges and Criticisms

While the promise of blockchain in voting systems is enticing, the integration of this transformative technology is not without its share of challenges and criticisms. Addressing these concerns is crucial for ensuring the viability, security, and widespread adoption of blockchain-based voting systems:

- Scalability Issues

Blockchain systems, particularly public blockchains, often face scalability challenges. As the number of transactions (votes, in this context) increases, the scalability of the network becomes a critical consideration. The time taken to reach consensus and add a new block to the chain can impact the speed and efficiency of the voting process. Scalability concerns become more pronounced in large-scale elections where millions of votes must be processed within a short timeframe. Researchers and developers are actively exploring solutions, such as sharding and layer-2 protocols, to address scalability issues in blockchain-based voting systems.

- User Experience and Accessibility

Blockchain technology, with its cryptographic keys and complex structures, can be intimidating for non-technical

users. Ensuring a user-friendly interface and a seamless voting experience is essential for the widespread adoption of blockchain-based voting systems. Accessibility is another concern, especially for populations with limited access to technology. Implementing blockchain in a way that does not disenfranchise individuals without access to smartphones or reliable internet connectivity is a challenge that needs to be carefully navigated.

- Privacy Concerns

While blockchain ensures the security and immutability of votes, it also raises concerns about voter privacy. The transparent nature of the technology means that all transactions are visible on the blockchain. Striking a balance between transparency and the anonymity of individual votes is a delicate task that requires robust cryptographic techniques. Researchers are actively exploring privacy-preserving technologies, such as zero-knowledge proofs, to allow voters to prove the validity of their votes without revealing the specific details of their choices.

- Centralization Risks

Paradoxically, the decentralization touted as a strength of blockchain can face challenges that lead to unintended centralization. Issues such as the concentration of mining power or the dominance of a few key players in the blockchain network can compromise the distributed nature of the technology.

In the context of voting systems, a high degree of centralization can undermine the security and integrity of the process, potentially leading to manipulation or collusion. Designing blockchain protocols that mitigate centralization risks is a critical consideration.

- Cybersecurity Threats

As with any technology, blockchain-based voting systems are susceptible to cybersecurity threats. The decentralized nature of blockchain doesn't make it immune to attacks, and novel threats may emerge as the technology evolves. Ensuring the resilience of the voting system against cyber threats, including hacking attempts and denial-of-service attacks, requires ongoing research and the implementation of robust security measures.

- Regulatory and Legal Challenges

The integration of blockchain in voting systems may face regulatory and legal challenges. The legal frameworks surrounding elections vary across jurisdictions, and introducing a novel technology like blockchain may require updates and adjustments to existing laws. Regulatory challenges also extend to issues such as the legal status of blockchain transactions, the enforceability of smart contracts, and the handling of disputes in a blockchain-based voting system.

- Public Trust and Acceptance

Trust in the electoral process is paramount, and introducing a new and unfamiliar technology can raise skepticism among the public. Building confidence in blockchain-based voting systems requires transparent communication, education, and a demonstration of the technology's reliability. Public acceptance is crucial for the success of blockchain in elections, and addressing concerns about security, privacy, and usability is key to garnering widespread support.

Navigating these challenges and criticisms requires a multidisciplinary approach, involving not only technologists but also policymakers, legal experts, and the public. The subsequent sections of this literature review will delve into

the existing research that explores potential solutions and mitigations for these challenges, providing insights into the current state of knowledge in the field of blockchain-based voting systems.

3. Case studies

A blockchain-based voting system has both advantages and disadvantages. Advantages:

1. Increased transparency: Decentralized voting systems can provide a transparent and auditable record of all transactions, making it easier to detect and prevent fraud and manipulation.
2. Improved security: Decentralized voting systems are resistant to hacking and tampering, as they rely on a consensus mechanism among multiple parties to validate and approve transactions.
3. Reduced costs: Decentralized voting systems can reduce the costs of conducting elections, as they eliminate the need for central authorities and intermediaries.
4. Increased accessibility: Decentralized voting systems can be designed to be more accessible and inclusive, allowing a broader range of participants to engage in the voting process.
5. Enhanced voter privacy: Decentralized voting systems can protect voter privacy by allowing voters to cast their vote without revealing their identity.

Disadvantages:

1. Technical challenges and complexity: Implementing decentralized voting systems can be technically challenging and requires expertise in blockchain technology and cryptography [1].
2. Limited scalability: Decentralized voting systems may have limitations in terms of scalability, as the number of transactions that can be processed at any given time may be limited by the capacity of the network [2].
3. Potential for unequal participation: Decentralized voting systems may not be accessible to all voters, particularly those who lack access to technology or have limited technical skills. And there can be several DDoS attacks while voting process [3].
4. Difficulty in ensuring the accuracy of vote counting: Decentralized voting systems may face challenges in ensuring the accuracy of vote counting, as errors or discrepancies may be difficult to detect and correct.
5. Lack of legal and regulatory frameworks: Decentralized voting systems may face legal and regulatory challenges, as they may not fit within existing legal frameworks and regulations for voting systems.

6. Security: There can be a bunch of an unpredictable attacks. Ddos, TLS, MM (man in the middle) attacks [4]. The client devices can have the viruses or some other malware software.

In recent years, two major e-voting applications have been developed, but they have also been found to have significant security risks. Following the 2015 election, the Virginia Information Technologies Agency (VITA) conducted security tests on several aspects of their e-voting system, including physical security, network security, operating system security, data security, and the vote tally process. VITA discovered that the system had used unsafe security protocols and weak passwords, and that an attacker could compromise the confidentiality and integrity of the voting data. Due to

these issues, VITA recommended discontinuing the use of the Advanced Voting System [5].

In addition, the Swiss government had been working on implementing an e-voting system for many years. Swiss Post was also involved in this effort and opened its applications for safety testing to the public in 2019 [6], believing in the transparency of the applications. However, international IT experts discovered a critical error in the source code of the Swiss Post application, which could not detect voting manipulation in the shuffle method. This error allowed hackers to replace valid votes with fraudulent ones. The IT experts noted that the codes were not standardized [7]. As a result of these critical issues, the Swiss government canceled the use of the system until a new appointment [8].

Over last there are several points of view based on blockchain based voting systems. Most of them bring the idea that it is hard to develop safe e-voting system itself [9,10]. The other part says that the blockchain based architecture gives an opportunity to design safe voting system.

4. Conclusions

In conclusion, the use of blockchain-based voting systems has the potential to provide numerous benefits, such as improved transparency, security, and efficiency. However, there are also significant challenges and limitations to consider, such as the need for widespread adoption, potential technical issues, and the risk of centralization. It is clear that further research and development are necessary to overcome these challenges and ensure the successful implementation of blockchain-based voting systems. As such, it is important for policymakers and researchers to carefully evaluate the pros and cons of these systems and work towards developing robust solutions that can effectively address the needs and concerns of all stakeholders. Ultimately, the adoption of blockchain-based voting systems could pave the way for more democratic and secure electoral processes in the future.

References

- [1] Taş, R. & Tanrıöver, Ö.Ö. (2020). A systematic review of challenges and opportunities of blockchain for E-voting. *Symmetry*, 12(8), 1328. <https://doi.org/10.3390/sym12081328>
- [2] Lauer, T.W. (2004). The risk of e-voting. *Electronic Journal of e-Government*, 2, 177–186
- [3] Bokslag, W. & de Vries, M. (2016). Evaluating e-voting: Theory and practice. *arXiv:160202509*
- [4] Cardillo, A. & Essex, A. (2018). The Threat of SSL/TLS Stripping to Online Voting. In *E-Vote-ID 2018. Electronic Voting*; 11143, 35–50. https://doi.org/10.1007/978-3-030-00419-4_3.
- [5] Manpearl, E. (2018). Securing US election systems: Designating US election systems as critical infrastructure and instituting election security reforms. *BUJ Sci. & Tech. L.*, 24, 168
- [6] Zetter, K. (2019). Experts Find Serious Problems with Switzerland's Online Voting System Before Public Penetration Test Even Begins. Retrieved from: <https://www.vice.com/en/article/vbwz94/experts-find-serious-problems-with-switzerlands-online-voting-system-before-public-penetration-test-even-begins>
- [7] Lewis, S.J., Pereira, O. & Teague, V. (2019). The use of trapdoor commitments in bayer-groth proofs and the implications for the verifiability of the scytl-swisspost internet voting system. Retrieved from: <https://cva.unifr.ch/content/ceci-n%E2%80%99est-pas-une-preuve-use-trapdoor-commitments-bayer-groth-proofs-and-implications>

- [8] Clarke, D. & Ali, S.T. (2017). End to end security is not enough. In Security Protocols XXV. 25th International Workshop, Cambridge, UK. https://doi.org/10.1007/978-3-319-71075-4_29
- [9] Bollinger, L.C. & McRobbie, M.A. (2018). Ensuring the Integrity of Elections. In Securing the Vote: Protecting American Democracy. National Academies of Sciences: Washington, DC, USA
- [10] Specter, M.A., Koppel, J. & Weitzner, D. (2020). The ballot is busted before the blockchain: A security analysis of voatz, the first internet voting application used in us federal elections. In Proceedings of the 29th USENIX Conference on Security Symposium

Блокчейнге негізделген дауыс беру жүйесі: әдебиеттерге жүйелі шолу

А. Омар*, Ж. Қалпеева

Satbayev University, Алматы, Қазақстан

*Корреспонденция үшін автор: abylay.omar@gmail.com

Аңдатпа. Демократиялық сайлау азаматтарға сайлау құқығын жүзеге асыруға және саяси көшбасшылар мен саясаттарға өз қалауларын білдіруге мүмкіндік беретін заманауи қоғамның ірге тасы болып табылады. Дегенмен, дәстүрлі дауыс беру жүйелері соңғы жылдары көптеген қиындықтарға тап болды, соның ішінде алаяқтық, бұзу және жалған ақпарат бар. Осы міндеттерді шешу үшін көптеген елдер дауыс беру процесін қауіпсіздендіруге және жаңғыртуға көмектесетін жаңа технологияларды зерттей бастады. Осындай технологиялардың бірі блокчейн болып табылады, орталықтандырылмаған және бұрмаланбайтын дерекқор, ол бірнеше тараптарға орталық органның қажеттілігіне ортақ кітапты жүргізуге мүмкіндік береді. Блокчейн технологиясы барлық транзакциялардың мөлдір және өзгермейтін жазбасын қамтамасыз ете отырып, сайлауды қауіпсіз, ашық және тиімді етіп өткізу тәсілін өзгертуге әлеуеті бар. Бұл мақалада біз блокчейн технологиясын дауыс беру жүйесінде қолданудың артықшылықтары мен қиындықтарын зерттейміз.

Негізгі сөздер: блокчейн, дауыс беру, электронды дауыс беру, дауыс беру жүйесі, орталықтандырылмаған.

Система голосования на основе блокчейна: систематический обзор литературы

А. Омар*, Ж. Қалпеева

Satbayev University, Алматы, Казахстан

*Автор для корреспонденции: abylay.omar@gmail.com

Аннотация. Демократические выборы являются краеугольным камнем современного общества, позволяя гражданам реализовать свое право голоса и выразить свои предпочтения в отношении политических лидеров и политики. Однако в последние годы традиционные системы голосования столкнулись с многочисленными проблемами, включая обвинения в мошенничестве, хакерстве и дезинформации. Чтобы решить эти проблемы, многие страны начали изучать новые технологии, которые могут помочь защитить и модернизировать процесс голосования. Одной из таких технологий является блокчейн - децентрализованная и защищенная от взлома база данных, которая позволяет нескольким сторонам вести общую бухгалтерскую книгу без участия центрального органа. Обеспечивая прозрачную и неизменяемую запись всех транзакций, технология блокчейн способна произвести революцию в проведении выборов, сделав их более безопасными, прозрачными и эффективными. В этой статье мы рассмотрим преимущества и проблемы использования технологии блокчейн в системах голосования.

Ключевые слова: блокчейн, голосование, электронное голосование, система голосования, децентрализованная.

Received: 10 December 2022

Accepted: 16 March 2023

Available online: 31 March 2023

Matrix combination in strip conversion for implementing hidden messages in the image

A. Yerimbetova¹, E. Daiyrbayeva^{1*}, I. Nechta², L. Lukpanova¹

¹Satbayev University, Almaty, Kazakhstan

²Siberian state university of telecommunications and information sciences, Novosibirsk, Russia

*Corresponding author: nurbekkyzy_e@mail.ru

Abstract. Steganography is a methodology for transmitting sensitive information while remaining undetected by an outside observer. One of the widely used approaches in this field is embedding hidden messages in various files, with special emphasis on the secrecy of the transmission process and the possibility of increasing the amount of data carried. The modification of the Least Significant Bit (LSB) method for embedding hidden messages in a graphical image proposed in this paper represents a step forward in improving steganographic techniques. This method is based on an innovative bandpass transform, where the embedded message is perceived as visible «noise» or interference added to the image. A key step in this process is the professional isolation of said noise from the signal, which allows for the extraction of the transmitted secret message with high accuracy. The uniqueness of the new method is manifested in the use of a complex combination of several matrices to «mix» the image fragments. Through experiments, different combinations of matrices were selected to provide a higher signal-to-noise ratio. The results obtained confirm that the new method, compared to conventional bandpass transform, shows a significant improvement in signal-to-noise ratio. This, in turn, enhances the ability to embed longer secret messages, improving the overall efficiency of the steganographic process.

Keywords: *LSB steganography, Hadamard, Slant, Haar, strip-method, secret message.*

1. Introduction

To ensure the security of the communication channel, messages transmitted between two subscribers are transformed so that their interception by a third party is useless. Usually such tasks are solved using cryptographic methods. In the general case, the cryptographic transformation of a message occurs with the participation of some secret key, available only to the sender and the recipient. Getting the original message from the transformed one is almost impossible without knowing the secret key. Accordingly, the analysis of data transmitted over an open communication channel does not allow a third party to freely read the original message.

When a message is received, for example, in the form of a file, then the problem of its further protection is also relevant. Thus, a graphic file created by one person can be copied by another person or slightly altered and further unlawfully issued as copyright property. Then it becomes necessary to create means that allow to uniquely identify the author when it comes to copyright, or to identify the end user when it comes to finding the source of unlicensed copies of a file. Such tools are being developed and researched within the framework of the science of steganography. Steganography studies techniques for creating a covert communication channel by embedding secret messages in digital data objects called containers. In cryptography, access to a message is limited if the secret key is unknown, and in steganography, the very existence of a secret message is hidden.

Depending on the task, such a container is chosen that allows you to get higher secrecy and a larger volume of the

transmitted message. Stealth is understood as the probability of not detecting the fact of the introduction. Obviously, the very fact of transferring a container without a message is not something suspicious.

Currently, text, audio, video, images and executable files (programs) are used as containers. Embedding into the text can be carried out by replacing synonyms. [1]. The text consists of words, some of them have synonyms. Such words in the source text are replaced with the synonym corresponding to the embedded message. The received text has the same meaning, but already contains a hidden message.

For embedding in an audio file, it is possible to add noise or slightly vary the signal value in accordance with the secret message. The resulting audio file does not differ from the original from the point of view of human perception, but it already has hidden data.

When shooting video, it is possible to use two cameras, standing next to each other. At the stage of editing, the final film is created by inserting short fragments (mise-en-scène) both from one camera and from the second, according to the embedded data. Variations in the shooting position are not detected by the viewer and do not affect the overall perception of the picture.

In steganography, the most widely used methods of embedding in the image. Any image is represented as a matrix of pixels, each of which has a color (represented in binary form). LSB implementation methods, for example [2], suggest replacing the least significant bit of the pixel color with a secret message. The use of this approach does not distort the visual perception of the image.

The rapid development of LSB implementation methods gave rise to the emergence of steganalysis methods for images, i.e. methods for detecting the fact of transmitting a secret message. As a result, in order to ensure an acceptable level of secrecy, the implementation is carried out not in all pixels, but only in some part (now it is several percent), and these pixels are selected in a pseudo-random way.

Let us briefly review the methods of steganography of images presented in the scientific literature. L.A. Mironovsky, V.A. Slaev [3] described the main matrix methods for processing continuous signals and images using strip transformation. The problem of estimating the potential noise immunity and synthesizing the optimal filter for the case of pulse interference is solved. The possibilities of two-dimensional strip transformation for storing and noise-resistant image transmission are investigated. Examples of image strip transformations are given and classes of images that are invariant with respect to symmetric orthogonal transformations are described.

A.P. Alekseev [4] concluded that steganography is a rapidly and dynamically developing science that uses the methods and achievements of cryptography, digital signal processing, communication theory and computer science. The monograph by B. Ya. Ryabko., A. N. Fionov [5] describes the main approaches and methods of modern cryptography and steganography for solving problems arising during the processing, storage and transmission of information.

O.I. Shelukhin et al., [6] described the issues of hiding information in text documents, network steganography, methods and algorithms for hiding data in audio signals of WAVE and MP3 formats, which are important for practical use. Methods and algorithms for hiding data in the spatial and frequency domains of still images are analyzed; software implementation of the introduction of a digital watermark into a video container in BMP and JPEG formats. The issues of introducing watermarks based on wavelet transformations are highlighted. In addition to the theoretical sections, the manual contains extensive practical material – a large number of software-implemented hiding algorithms using modern application software packages Matlab and Mathcad, as well as programming languages Python, C++, C#, etc.

In the article [7], F.A. Murzin et al., investigated the variants of the strip method. Namely, the variants based on the use of different matrices are considered: Hadamard, Haar, Conference, C-matrix, etc. Various types of matrices and signals were tested. A theoretical estimate is proposed in terms of spectral expansion coefficients for the error rate for a bandpass transformation based on the Hadamard matrix in the case of pulse interference. These variants of the strip method were implemented in our work.

Rosziati Ibrahim et al [8] propose a new algorithm for hiding data inside an image using the steganography technique. The proposed algorithm uses binary codes and pixels inside the image. Various data sizes are stored inside the images, and the PSNR (peak signal-to-noise ratio) is also recorded for each of the tested images. Based on the PSNR value of each image, the stego image has a higher PSNR value. Therefore, this new steganography algorithm is very effective for hiding data inside an image.

E.A. Bela [9] focused on Haar cascades and is based on the article by Viola P., Jones M. «Rapid Object Detection using a Boosted Cascade of Simple Features». Some subtleties of cascade training that were not described in the original article

are described here. In particular, this is a method for iterating through the thresholds of weak classifiers, as well as an optimized method for constructing a cascade of classifiers.

R. Sonic et al [10] investigated an efficient algorithm for calculating parametric Slant-Hadamard transformations. The authors presented the Slant-Hadamard matrix of order n^2 as a product of sparse matrices, developed the corresponding fast Slant-Hadamard transformation and its complexity.

According to A. Alharbi et al [11], one of the possible ways to conceal classified information is the use of images. Images are the most common type of payload in terms of their availability and use in steganographic applications. They are able to hide secret information, because the human eye is less sensitive to minor changes in the image. In this paper, they propose a steganographic method using a discrete Haar wavelet transform, in which the data is hidden in the frequency domain.

In this paper, the purpose of the study is to transmit the hidden data using an image object. This experiment is conducted by researchers in two ways. The sender's side sends encrypted and unencrypted hidden data in graphic formats with security concerns. The extracted hidden data will be decrypted on the receiver side in the case of encrypted hidden data, and as part of security, the goal of ensuring image quality remains unchanged for human eyes (Figure 1).

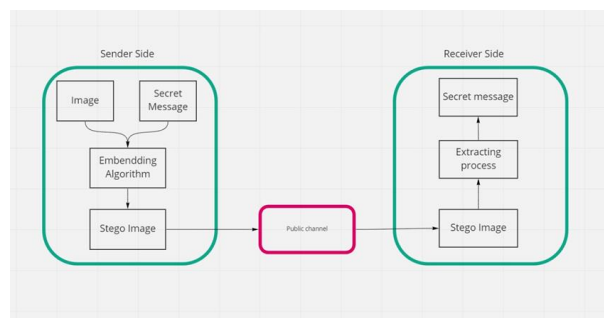


Figure 1. Scheme of transmission of hidden data using an image object

2. Materials and methods

Many tasks of information transformation and data analysis are related to image processing and transmission. As mentioned earlier, images are the most popular there are many different image file formats, most of which are designed for specific applications. Various steganographic algorithms exist for these different image file formats [11,12].

Image definition. For a computer, an image is a set of colored pixels (dots), the color is represented in the RGB palette in the computer's memory as triples of numbers [13]. Currently, there are quite a lot of different methods (and their variants) for embedding messages. All the methods used are fairly well described in special literature [6-9]. The list of steganographic methods is updated annually, more reliable and original ways of hiding information are being invented, which will require new, effective methods of analysis to neutralize.

This article explores the possibility of using the strip method for storing and noise-immune transmission of images. In this case, matrix transformations of the original image are used before transmission, during which the image fragments are mixed and superimposed on each other. The converted image is transmitted over a communication channel, where it is distorted by impulse noise. Its action can lead, for example,

to the complete loss of individual fragments of the image. When a signal is received at the receiving end, an inverse transformation is performed, as a result of which the image is restored. If we ensure a uniform distribution of impulse noise over the entire area of the image (without changing its energy), then a significant attenuation of the amplitude of the noise will occur and an acceptable quality of all areas of the reconstructed image will be achieved.

The first stage of strip conversion of two-dimensional signals consists in splitting the original image

$$P = \begin{pmatrix} p_{11} & \cdots & p_{1T} \\ \vdots & \ddots & \vdots \\ p_{s1} & \cdots & p_{sT} \end{pmatrix}$$

where T and S are the horizontal and vertical widths of the image. P is divided into N rectangular fragments of the same size [14]. Let's designate the number of horizontal and vertical stripes into which the image is cut through m and n; then $N=m \times n$. We represent the image obtained after splitting as a block matrix

$$X = \begin{pmatrix} x_{11} & \cdots & x_{1n} \\ \vdots & \ddots & \vdots \\ x_{m1} & \cdots & x_{mn} \end{pmatrix},$$

where

$$x_{ij} = \begin{pmatrix} p(i-1) \times m + 1, (j-1) \times n + 1 & \cdots & p(i-1) \times m + 1, (j-1) \times n + n \\ \vdots & \ddots & \vdots \\ p(i-1) \times m + m, (j-1) \times n + 1 & \cdots & p(i-1) \times m + m, (j-1) \times n + n \end{pmatrix}.$$

The image is presented in a gray palette, i.e. RGB components of the palette are equal for each pixel and $p_{st} \in \{0 \dots 255\}$.

Next, the fragments are linearly combined. In this paper, a matrix approach is considered. In this case, two approaches are possible - vector and matrix. There are three options for isometric transformation of this matrix in order to «mix» its fragments:

A) multiplication by the orthogonal $m \times m$ matrix B on the left: $Z_1 = BX$ (left-side matrix transformation)

B) multiplying by the orthogonal $n \times n$ matrix A on the right $Z_2 = XA$ (right-handed matrix transformation).

C) Simultaneous multiplication by the matrix B on the left and by the matrix on the right: $Z = BXA$ (two-sided matrix transformation).

The transmission and reception of images using a two-way strip transform is shown in Figure 2. Here Δ - denotes a secret message.

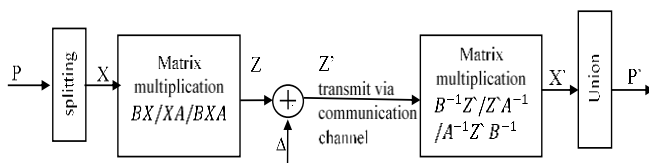


Figure 2. Scheme of transmission-reception

After receiving P' , the transmitted secret message is calculated $\Delta' = Z' - P'$. Due to the mathematical properties of the applied transformations $\Delta \approx \Delta'$.

Transformation of image fragments during transmission is carried out according to the formula

$$Z = A^T \times X \times A \quad (1)$$

Reconstruction on the receiving side of an image formed from fragments Y and transmitted over a communication network is performed through an inverse two-way transformation in the form

$$X = (A^T)^{-1} \times Z \times A^{-1} \quad (2)$$

The resulting image P is the result of defragmentation from fragments of the X type of the received image.

3. Results and discussion

This paper discusses a special steganographic method for hiding information in digital images. This steganographic method uses the Hadamard transform and works on grayscale images. This method has been extensively tested on a variety of images with various textures and is robust enough to avoid various attacks such as adding noise or squeezing. The experimental results show that the considered system successfully preserves the image quality and remains unnoticed by the known methods of steganalysis. The Hadamard transform is an example of a generalized class of Fourier transforms. It performs an orthogonal, symmetric, linear operation on real numbers (or complex numbers, although the Hadamard matrices themselves are purely real). The Hadamard transform has a significant computational advantage over other methods. Their unitary matrices and transformations consist of and are calculated only with the help of additions and subtractions, but they do not involve multiplication. Consequently, for processors for which multiplication is a laborious operation, sustainable savings are achieved.

During the experiments based on the Hadamard transformation, the following matrices were also tested: Slant, Discrete Cosine Transform, Haar, Conference, S-Matrix.

Choice of transformation matrix. Strip transform matrices are selected in order to achieve the most uniform distribution of interference in the signal or image as a result of decoding at the receiving end of the communication channel. This will allow the most accurate recovery of information about distorted or lost fragments.

The Hadamard matrix H is a given square matrix satisfying

$$HH' = nI_n \quad (3)$$

in which all records in the first row and first column are +1, and the rest of the elements are +1 or -1. The inner product of any two rows (columns) is 0. This is referred to as the orthogonal property. He assumed that the Hadamard matrix exists if and only if $n = 0 \pmod{4}$. Despite the efforts of several mathematicians, this hypothesis remains unproven, although it is widely believed to be true. This condition is necessary and the sufficiency of the part is still an open problem. These Hadamard matrices were systematically investigated by Paley in 1933. There are other orthogonal matrices such as Slant and Haar matrices, Discrete Fourier transform (DFT), discrete cosine transform (DCT), shell matrix, etc. [15].

Algorithm for implementing the Hadamard transform:

Block strip method for transforming images. A matrix of size M by N is divided into m into n blocks, to each of 3 of which a strip transformation is applied 4. Input parameters: file with an image, m is the number of blocks horizontally, n is the number of blocks vertically. We tested images with dimensions: 1600×1200 , 600×600 , 512×512 , 256×256 , 600×600 , 1024×1024 , 1024×1024 . For testing, the pictures were taken from the Internet.

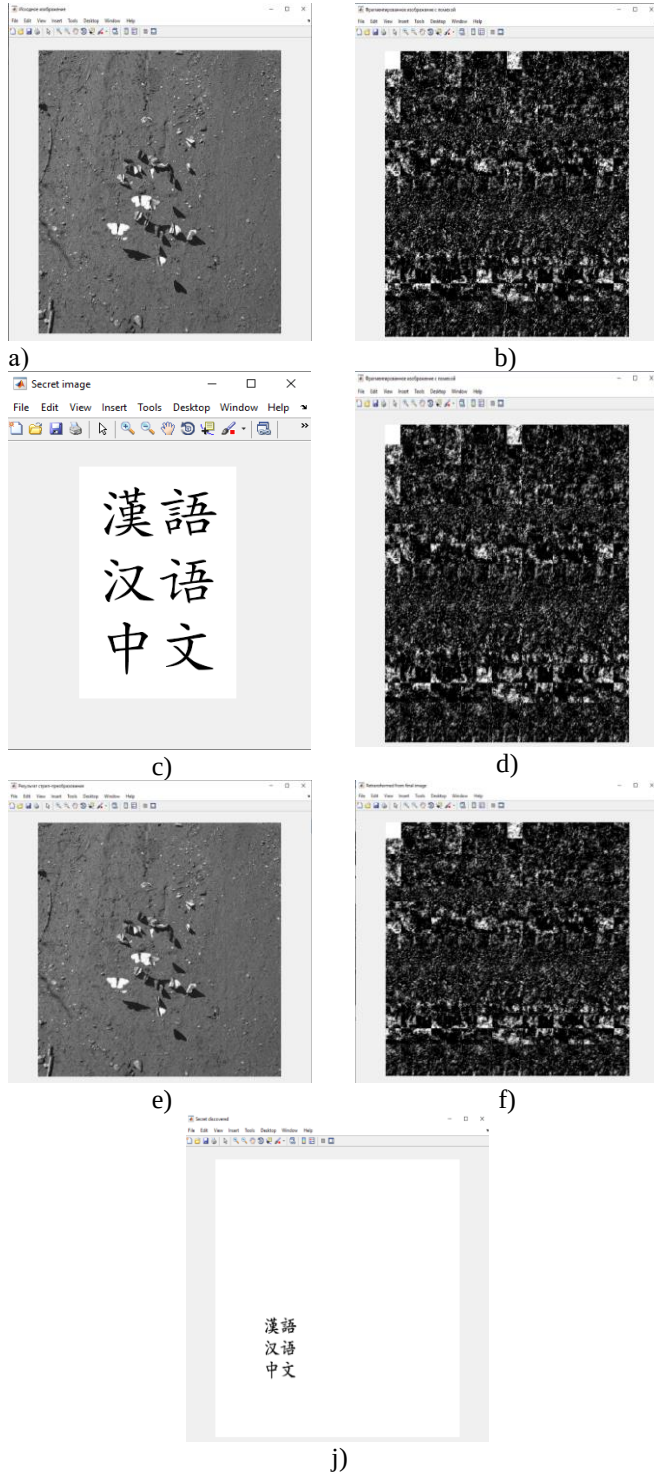


Figure 3. (a) is the original image, (b) is a strip-transformed image, (c) is a reconstructed image after transmission in a channel with impulse noise using a strip transform, (d) Noisy fragmented image, (e) is a reconstructed image after transmission in a channel with interference without using strip transforms, (f) Converted final image, (g) detected message

Results:

Two-sided (two-dimensional) strip transformation of images is shown in Figure 3. Here (a) is the original image, (b) is a strip-transformed image, (c) is a reconstructed image after transmission in a channel with impulse noise using a strip transform, (e) is a reconstructed image after transmission in a channel with interference without using strip transforms. The above example of a strip transform is implemented using a 4×4 Hadamard matrix, and a single noise corresponds to a 30×30 pixel rectangle. The weakening of the influence of impulse noise is realized by the fact that during inverse transformations performed on the receiving side with a transposed orthogonal matrix, the noise is evenly distributed over the entire reconstructed image. For maximum attenuation of the noise amplitude in the strip transform, as a rule, extremal orthogonal matrices of orders 4 and 8 are used. Hadamard matrices exist on orders $4t$, where t is a natural number. There are more than two such matrices on each order. However, the existing algorithms for calculating Hadamard matrices ensure the inheritance of only the structures of the matrices corresponding to these methods. There is no their structural and quantitative diversity in the orders of existence.

The result of the inverse transformation coincides with the original image, up to the transposition of the masking matrix: there is no need to separately store not only the inverse matrix itself, but also its half due to its symmetry. This saves memory when directly implementing the method on the system. It is the presented simplified version of the strip transformation (without using Kronecker multiplication and cutting the image into strips), but implemented with unique matrices, that was called masking.

To carry out calculations and mathematical operations, the Matlab application is used. The functionality of this application made it possible to perform orthogonal transformation with a digital image, as well as calculate the desired metrics.

For the comparison, the following orthogonal transformations were chosen: Hadamard, Haar, oblique, discrete cosine. To compare the effect of noise on the final result, the following indicators were selected [13]

average modulus of deviation

$$L1 = \sum_{i,j=1}^n \frac{|x_{ij} - y_{ij}|}{n^2} \quad (4)$$

root-mean-square deviation modulus

$$L2 = \sqrt{\sum_{i,j=1}^n \frac{|x_{ij} - y_{ij}|^2}{n^2}} \quad (5)$$

maximum deviation modulus

$$\max_{1 \leq i, j \leq n} |x_{ij} - y_{ij}| \quad (6)$$

signal-to-noise ratio PSNR

$$10 \log_{10} \frac{255^2 n^2}{\sum_{i,j=1}^n (x_{ij} - y_{ij})^2} \quad (7)$$

For the experiment, images with light and dark areas of different frequency characteristics were selected. Interference to the image was set in the form of a black area of various sizes in the center of the image.

For the experiments, images of sizes 400x400, 512x512, 600x600, 800x800, 1024x1024 were used. For each size, 25 images were tested. The images were “cut” into fragments, after which they were subjected to two-sided transformation using the matrices described above.

Table 1. Results of experiments using transformation matrices (size 1024×1024)

Methods	Max	L1	L2	PSNR
Haar+Haar	72,99699	2,56507	0,47514	40,8345
Haar+Hadamard(2^n)	39,2723	2,58196	0,67347	40,56836
Haar+Slant	42,32732	2,56632	0,656235	40,64133
Haar+Hadamard(p+1)	242,15	8,4986	1,490525	29,70511
Haar+CreatingLegendre	39,55325	2,560085	0,667305	40,7044
Hadamard(2^n)+Hadamard(2^n)	17,71095	4,1032	3,19386	36,20309
Hadamard(2^n)+Haar	39,2723	2,58196	0,67347	40,56836
Hadamard(2^n)+Slant	19,28821	4,40384	3,36667	35,67784
Hadamard(2^n)+Hadamard(p+1)	582,9547	146,0302	110,0094	5,039475
Hadamard(2^n)+creatingLegendre	16,33816	3,577025	2,761904	37,51159
CreatingLegendre+Haar	43,88584	4,203045	2,14059	35,97084
CreatingLegendre+Hadamard(2^n)	20,90581	3,598735	2,694915	37,43103
CreatingLegendre+Slant	23,3122	3,78383	2,767745	37,17978
CreatingLegendre+Hadamard(p+1)	242,15	13,74597	8,611805	25,87514
CreatingLegendre+CreatingLegendre	19,40275	3,21828	2,392725	38,40855
Hadamard(p+1)+Haar	238,94	13,63	31,56	18,14
Hadamard(p+1)+Hadamard(2^n)	80,53	13,11	16,49	23,78
Hadamard(p+1)+Slant	209,98	24,36	32,91	17,78
Hadamard(p+1)+Hadamard(p+1)	188,35	24,75	31,4	18,19
Hadamard(p+1)+slmtx	234,46	19,06	40,19	16,04
Slant +Haar	170,77	9,12	21,53	21,46
Slant+Hadamard(2^n)	89,48	9,06	13,15	25,74
slmtx+Slant	188,04	14,69	21,73	21,38
slmtx+Hadamard(p+1)	178,88	19,35	27,73	19,26
slmtx+slmtx	180,21	15,76	30,56	18,42

A visual comparison (Table 1.) of the images clearly shows that at any percentage of information loss, discrete cosine and Hadamard matrices, it is best to save the image, which allows it to be brought closer to the original using third-party software and other algorithms. Using an oblique matrix or Haar matrix allows you to keep some parts of the image exactly as in the original image (which may be due to the fact that they were not affected by the interference), but significantly spoils the image. At the same time, other areas of the image are seriously distorted or cannot be restored.

In the course of the work, new experiments were carried out with several types of strip transformation matrices. Those of them have been identified that are most successful in minimizing the negative impact of interference, evenly distributing it over the entire image, thereby making it possible to restore the lost areas. The dependence of the degree of reconstruction on the size of the interference was revealed for various matrices and images with different frequency characteristics.

Below is a comparative analysis of standard methods for suppressing noise in images (table 2). Figure 4 shows the diagrams for different values of N and M difference PSNR.

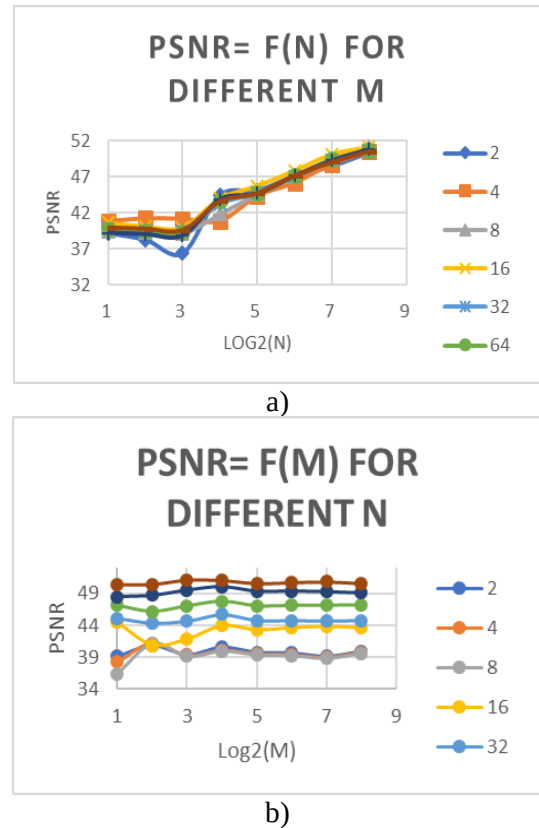


Figure 4. Schematic representation of PSNR results for different values of a) N and b) M

4. Conclusions

The purpose of this article is to study algorithms for digital image processing based on orthogonal transformations for image restoration. This goal was achieved, the tasks were completed. The program is implemented in the Matlab environment. As a result of the research, the method of introducing hidden messages, based on strip conversion, was improved. In the course of experiments, it was shown that the new method, in comparison with the conventional strip conversion, allows one to obtain a higher signal-to-noise ratio, which obviously allows embedding a longer secret message. It is assumed that the embedded message should be a regular image, and not any encrypted sequence.

References

- [1] Chang C.Y. & Clark S. (2014). Practical linguistic steganography using contextual synonym substitution and a novel vertex coding method. *Computational linguistics*, 40(2), 403-448. https://doi.org/10.1162/coli_a_00176

- [2] Chatterjee A., Ghosal S. K. & Sarkar R. (2020). LSB based steganography with OCR: an intelligent amalgamation. *Multimedia Tools and Applications*, 1-19. <https://doi.org/10.1007/s11042-019-08472-6>
- [3] Mironovsky, L.A. & Slaev, V.A. (2011). Strip method for transforming images and signals: Monograph. SPb.: Polytechnic. <https://doi.org/10.1515/9783110252569>
- [4] Alekseeva, A.P. (2010). Steganographic and cryptographic methods of information protection: a textbook on the discipline «Informatics» for students of PSUTI. Samara: PSUTI
- [5] Ryabko, B.Ya. & Fionova, A.N. (2016). Fundamentals of modern cryptography and steganography. M: Gorjachaja linija-Telekom
- [6] Shelukhin, O.I. & Kanaev, S.D. (2017). Steganography. Algorithms and software implementation. Moscow
- [7] Murzin, A. & Ryaskina, N.A. (2017). Analysis of noise stability of strip-transformation. *Bulletin of the Novosibirsk Computing Center*, (41), 41-54
- [8] Rosziati Ibrahim and Teoh Suk Kuan. (2011). Steganography Algorithm to Hide Secret Message inside an Image. *Computer Technology and Application*, (2), 102-108
- [9] Belyh, E.A. (2017). HAARA cascade training. *Syktvykar University Bulletin. Series 1: Mathematics. Mechanics. Computer science*, 1(22), 42-54
- [10] Hakobyan, S.R. (2014). Fast Slant-Hadamard Transform Algorithm. *Mathematical Problems of Computer Science*, (42), 113-120
- [11] Alharbi, A. & Kechadi, M.-T. (2017). A Steganography Technique for Images Based on Wavelet Transform. *Lecture Notes in Computer Science*, (10646), 273-281. https://doi.org/10.1007/978-3-319-70004-5_19
- [12] Setiadi, D. (2021). PSNR vs SSIM: imperceptibility quality assessment for image steganography. *Multimedia Tools and Applications*, 80(6), 8423-8444. <https://doi.org/10.1007/s11042-020-10035-z>
- [13] Jiao, S. & Feng, J. (2021). Image steganography with visual illusion. *Optics Express*, 29(10), 14282-14292. <https://doi.org/10.1364/OE.421398>
- [14] Zuo, X., Hu, H., Zhang, W. & Yu, N. (2018). Text Semantic Steganalysis Based on Word Embedding. *Lecture Notes in Computer Science*, (11066), 485-495. https://doi.org/10.1007/978-3-030-00015-8_42
- [15] Daiyrbayeva, E., Yerimbetova, A. & Toigozhinova, A. (2021). Comparative analysis of the results of image recovery based on the strip method using various matrices. *News of the National Academy of Sciences of the Republic of Kazakhstan, Physico – mathematical series*, (4), 29-34. <https://doi.org/10.32014/2021.2518-1726.70>

Стрип-түрлендіруде матрицаларды кескінге құпия хабарламаларды енгізу үшін біріктіру

А. Еримбетова¹, Э. Дайырбаева^{1*}, И. Нечта², Л. Лукпанова¹

¹Satbayev University, Алматы, Қазақстан

²Сібір Мемлекеттік телекоммуникация және информатика университеті, Новосибирск, Ресей

*Корреспонденция үшін автор: nurbekkyzy_e@mail.ru

Андатпа. Стеганография – сыртқы бақылаушыларға көрінбейтін болып қалуға тырысатын құпия ақпаратты берудің инновациялық әдістемесі. Бұл салада кеңінен қолданылатын тәсілдер әлемінде жасырын хабарламаларды әртүрлі файлдарға кірістіру ерекшеленеді, бірақ тек тасымалдау процесінің құпиялылығына ғана назар аударылмайды, сонымен бірге тасымалданатын деректер көлемін ұлғайту мүмкіндігіне көңіл бөлінеді. Осы мақалада ұсынылған графикалық кескінге жасырын хабарламаларды кірістіру үшін ең кіші маңызды бит (LSB) әдісінің модификациясы стеганографиялық технологияларды жетілдірудегі алға қадам болып табылады. Бұл әдіс инновациялық жолақты түрлендіруге негізделген, мұнда кірістірілген хабарлама суретке енгізілген «шу» немесе кедергі ретінде қабылданады. Бұл процестің негізгі кезеңі – берілген шуды сигналдан кәсіби түрде оқшаулау, бұл жіберілген құпия хабарламаны жоғары дәлдікпен шығаруға мүмкіндік береді. Жаңа әдістің бірегейлігі сурет фрагменттерін «араластыру» үшін бірнеше матрицалардың күрделі комбинациясын қолданудан көрінеді. Эксперименттер арқылы жоғары сигнал/шу қатынасын қамтамасыз ететін матрицалардың әртүрлі комбинациялары таңдалды. Нәтижелер жаңа әдіс әдеттегі жолақты түрлендірумен салыстырғанда сигнал/шу қатынасының айтарлықтай жақсарғанын көрсетеді. Бұл өз кезегінде стеганографиялық процестің тиімділігін арттыра отырып, ұзағырақ құпия хабарламаларды ендіру мүмкіндіктерін кеңейтеді. Осылайша, ұсынылған әдіс жасырын ақпаратты беру кезінде қауіпсіздік жолағын көтеріп қана қоймайды, сонымен қатар құпия деректерді енгізу үшін сыйымдылықта айтарлықтай өсуді қамтамасыз етеді, бұл стеганографияның дамуына маңызды үлес қосады.

Негізгі сөздер: LSB-стеганография, Адамар, Slant, Хаар, стрип-әдіс, құпия хабарлама.

Комбинирование матриц в стрип-преобразовании для внедрения скрытых сообщений в изображение

А. Еримбетова¹, Э. Дайырбаева^{1*}, И. Нечта², Л. Лукпанова¹

¹Satbayev University, Алматы, Казахстан

²Сибирский Государственный университет телекоммуникации и информатики, Новосибирск, Россия

*Автор для корреспонденции: nurbekkyzy_e@mail.ru

Аннотация. Стеганография представляет собой инновационную методологию передачи секретной информации, стремясь оставаться невидимой для внешних наблюдателей. В мире широко используемых подходов к этой области выделяется встраивание скрытых сообщений в различные файлы, при этом особое внимание уделяется не только секретности самого процесса передачи, но и возможности увеличения объема переносимых данных. Предложенная в данной статье модификация метода наименьших значащих бит (LSB) для встраивания скрытых сообщений в графическое изображение представляет собой шаг вперед в совершенствовании стеганографических технологий. Этот метод базируется на инновационном полосовом преобразовании, где встроенное сообщение воспринимается как видимый «шум» или помеха, добавленная к изображению. Ключевым этапом данного процесса является профессиональная изоляция указанного шума от сигнала, что позволяет извлекать передаваемое секретное сообщение с высокой точностью. Уникальность нового метода проявляется в применении сложной комбинации нескольких матриц для «смешивания» фрагментов изображения. Путем экспериментов были подобраны различные комбинации матриц, обеспечивающих более высокое отношение сигнал/шум. Полученные результаты подтверждают, что новый метод, по сравнению с обычным полосовым преобразованием, демонстрирует значительное улучшение отношения сигнал/шум. Это, в свою очередь, расширяет возможности встраивания более длинных секретных сообщений, улучшая в целом эффективность стеганографического процесса.

Ключевые слова: LSB-стеганография, Адамар, Slant, Хаар, стрип – метод, секретное сообщение.

Received: 11 November 2022

Accepted: 16 March 2023

Available online: 31 March 2023

Using machine learning algorithms for processing medical data

G. Mukazhanova¹, Zh. Alibiyeva^{2*}, A. Kassenkhan², N. Mukazhanov²

¹Innovative Eurasian University, Pavlodar, Kazakhstan

²Satbayev University, Almaty, Kazakhstan

*Corresponding author: zh.alibiyeva@satbayev.university

Abstract. The paper considers the comparative analyses of machine learning algorithms for dataset: cardio_train.csv from kaggle.com (link: <https://www.kaggle.com/sulianova/cardiovascular-disease-dataset>). Moreover, using machine learning algorithms there will be discovered the best accuracy algorithms for the cardio_train.csv. Considering procedures have done in Python 3.0 programming language, which represents confusion matrix and classification report, in order to see precision score, recall, f1-score, and support. Furthermore, in this paper you are able to see following classification models: KNN algorithm, Logistic Regression, Decision Tree, Random Forest, Naïve Bayes and SVM. As a result, it will be defined the superior accuracy for processing medical dataset.

Keywords: medical dataset, machine learning, algorithms, KNN algorithm, Logistic Regression, Decision Tree, Random Forest, SVM, Naïve Bayes.

1. Introduction

It is clear, that processing medical data plays significant role in our century, especially in period of worldwide pandemic which has changed and affected to world health organization and whole economy of the countries. The mission of this paper is to help and to analyze medical data via modern technology such as machine learning and to process cardiovascular diseases via finding out methods and models in order to atomize data and compare methods, search the best-adapted models and method using Python programming language and Machine learning algorithms. There is a good medicine in Kazakhstan, although it needs several methods in order to make it better. The purpose of the following work is to improve processing medical data, especially cardiovascular diseases, which is the top problem in Kazakhstan.

In this article, will be presented machine learning classification algorithms such as: KNN algorithm, Logistic Regression, SVM, Decision Tree, Random Forest, Naïve Bayes. All algorithms and procedures have done in Jupyter Notebook (Anaconda), Python 3.0 programming language. After considering machine learning algorithms will provided comparative analysis in tables of all procedures.

2. Materials and methods

2.1. Overview of dataset

There are 3 types of input features: objective, examination and subjective.

Objective: factual information;

Examination: results of medical examination;

Subjective: information given by the patient.

Features correspond to 12 columns: age, height, weight, gender, systolic blood pressure, diastolic blood pressure, cholesterol, glucose, smoking, alcohol intake, physical activity and presence or absence of cardiovascular disease.

Below are the column details:

Age | Objective Feature | age | int (days)

Height | Objective Feature | height | int (cm) |

Weight | Objective Feature | weight | float (kg) |

Gender | Objective Feature | gender | categorical code |

Systolic blood pressure | Examination Feature | ap_hi | int |

Diastolic blood pressure | Examination Feature | ap_lo | int |

Cholesterol | Examination Feature | cholesterol | 1: normal, 2: above normal, 3: well above normal |

Glucose | Examination Feature | gluc | 1: normal, 2: above normal, 3: well above normal |

Smoking | Subjective Feature | smoke | binary |

Alcohol intake | Subjective Feature | alco | binary |

Physical activity | Subjective Feature | active | binary |

Presence or absence of cardiovascular disease | Target Variable | cardio | binary | All of the dataset values were collected at the moment of medical examination.

All the following attributes will help to analyze and process the best adapted method and models in order to achieve in the article goal.

Cardiovascular disease (CVD) continue to be the most pressing health problem most countries of the world, including the Kazakhstan. According to the World Health Organization, every year in the world from cardiovascular disease (CVD) dies more than 17 million people, including more than 7 million from coronary heart disease (IHD) [1].

Predicting CVD risk is becoming increasingly more important in clinical decision making since their introduction at the international level in the latest guidelines. At the same time, predicting the risk of coronary artery disease at based on the analysis of traditional risk factors is fraught with a number of problems. In the FI, during observation for 26 years, a significant coincidence of groups of persons without established ischemic heart disease and people who develop coronary artery disease. By level traditional FRs, coincidence

was noted the level of total cholesterol (3.9-7.8 mmol / l) between the groups [2].

There was a significant overlap of groups of patients IHD and healthy men according to the level of traditional RF (TC, LDL cholesterol, smoking, AH, BMI) and significant difference in HDL cholesterol, TG and ratio LDL cholesterol / HDL cholesterol. The prospective NPHS2 study compared the predictive ability of algorithms for cardiovascular risk assessment by Framingham and Procarn. Both of these algorithms had a false negative result > 85%.

The low accuracy of predicting cardiovascular events has a number of reasons. First, the assessment the total risk must be adapted depending on national and regional features. Secondly, considering the design of the research scales included in the development, in them often not considered significant for the offensive cardiovascular event's clinical conditions (type I and II diabetes mellitus, chronic kidney disease or very high levels of certain risk factors). Third, the data that were used to compile scales, were received 30-50 years ago and may not correspond to modern realities. Fourthly, mathematical methods for calculating risks also have errors and limitations of applicability. This way we can speak with confidence about the problem of insufficient accuracy of results calculating cardiovascular risk based on generally accepted scales.

Machine learning provides good opportunities to solve this problem and significantly improve accuracy in predicting cardiovascular diseases and their complications in comparison with the use of existing methods, due to the nonlinear relationships of their fine tuning between cardiovascular risk factors and the manifestation of diseases. Recently calculation of the number of research and development in these areas.

In the Figure 1 is shown dataset attributes all columns and rows.

```
dataset = pd.read_csv('cardio_train.csv', ';')
```

```
dataset.sample(10)
```

	id	age	gender	height	weight	ap_hi	ap_lo	cholesterol	gluc	smoke	alco	active	cardio
47702	68105	15955	2	173	85.0	130	80	1	1	0	1	1	0
42749	61085	21912	2	167	80.0	140	90	1	1	0	0	0	1
8773	12518	23248	1	160	64.0	120	80	1	1	0	0	1	0
6340	9025	18299	1	156	104.0	120	80	1	1	0	0	1	1
56707	80954	15300	2	160	62.0	140	90	1	1	0	0	1	1
69834	99739	19776	1	156	59.0	120	80	2	1	0	0	0	0
16906	24167	17272	2	170	31.0	150	90	2	2	0	0	1	1
18950	27061	21158	1	168	64.0	120	80	1	1	0	0	0	0
36449	52065	20302	1	168	95.0	180	80	1	1	0	0	1	1
28611	40910	23455	2	176	85.0	140	90	2	2	0	0	1	1

Figure 1. Dataset description in Python

After dataset has been read in Python there is a method `dataset.nunique()` to check how many unique values are there in the each row. `dataset.isna().sum()` following method is used to verify are there any null or empty rows. If yes, then we need to fill all null values by average sum of row and normalize the dataset. Then there is prepared dataset for further procedures.

`X = dataset.iloc[:, :-1].values`

`y = dataset.iloc[:, -1].values`

Following code above is used to divide dataset into train and test. Furthermore, for X has chosen all columns except the last one, for y vice versa. It is important to separate them because our target to find people from dataset who has cardi-

ovascular disease and doesn't have. There is a reason why we have 2 option only 1 and 0. 1 – yes, 0 –no.

There is used following piece of code to import `train_test_split` for separating dataset into test and train:

`from sklearn.model_selection import train_test_split`

From Figure 2 you are able to visualize it more clearly.

```
from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state = 1, stratify = y)

from sklearn.neighbors import KNeighborsClassifier
```

Figure 2. Importing train test split

After all these steps our dataset is ready to be modified and proceed.

2.2. Using machine learning algorithms for processing medical data

Machine learning is a process used by companies to turn raw data into useful information. By using software to look for patterns in large batches of data, businesses can learn more about their customers to develop more effective marketing strategies, increase sales and decrease costs. Machine learning depends on effective data collection, warehousing, and computer processing [1].

Machine learning involves exploring and analyzing large blocks of information to glean meaningful patterns and trends. It can be used in a variety of ways, such as database marketing, credit risk management, fraud detection, spam Email filtering, or even to discern the sentiment or opinion of users [2].

The most popular algorithms of machine learning are represented in Figure 3.

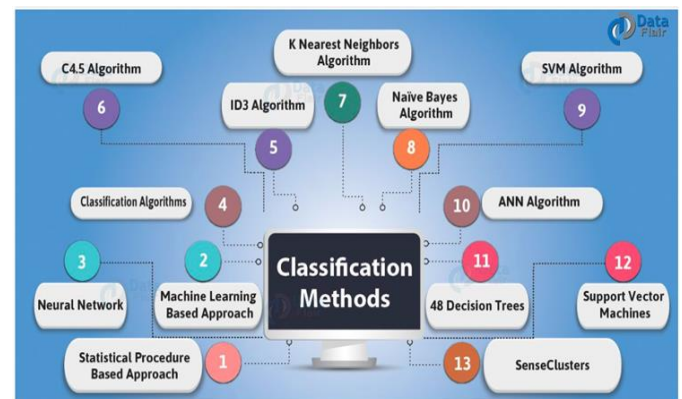


Figure 3. Classification Methods

Machine learning algorithms are able to be whether supervised or unsupervised.

Supervised learning: Algorithms that need a 'training' set of data to learn.

Unsupervised learning: Algorithms that don't need any training data to work properly.

Here are the main types of algorithm that is going to be used.

Classification: These algorithms put the existing data (or past data) into various 'classes' (hence classification) based on their attributes (properties) and use that classified data to make predictions.

Accuracy = $\frac{TP+TN}{TP+FP+FN+TN}$

Precision = $\frac{TP}{TP+FP}$

Recall = $\frac{TP}{TP+FN}$

F1 Score = $2 * (\text{Recall} * \text{Precision}) / (\text{Recall} + \text{Precision})$.

K Nearest Neighbor Algorithm. K-Nearest Neighbors, or KNN for short, is one of the simplest machine learning algorithms and is used in a wide array of institutions. KNN is a non-parametric, lazy learning algorithm. When we say a technique is non-parametric, it means that it does not make any assumptions about the underlying data. In other words, it

makes its selection based off of the proximity to other data points regardless of what feature the numerical values represent. Being a lazy learning algorithm implies that there is little to no training phase. Therefore, we can immediately classify new data points as they present themselves [3].

Implementation in Python KNN algorithm. Let's explore our cardiovascular dataset in KNN algorithm in Figure 4.

```
X_train, X_test, y_train, y_test = train_test_split(X,y, test_size=0.2, random_state = 1,stratify =y)

from sklearn.neighbors import KNeighborsClassifier

knn = KNeighborsClassifier(51)

re is no big difference between 11th and 33th k neighbours only 57% is predicted

X_test[1:2]

array([[5.3647e+04, 1.9730e+04, 1.0000e+00, 1.4700e+02, 5.8000e+01,
        1.6500e+02, 1.0000e+02, 1.0000e+00, 1.0000e+00, 0.0000e+00,
        0.0000e+00, 1.0000e+00]])

knn.fit(X_train, y_train)

KNeighborsClassifier(algorithm='auto', leaf_size=30, metric='minkowski',
                    metric_params=None, n_jobs=None, n_neighbors=51, p=2,
                    weights='uniform')

y_preds = knn.predict(X_test)

for i,j in zip(y_test, y_preds):
    print('real: ', i, ' predicted: ',j)
```

Figure 4. KNN

The first step to import KNeighborsClassifier in order to analyze dataset and find optimal accuracy for dataset. At the beginning it was tried to use 11 and 33 neighbors. However, an accuracy was not as it was expected. As a result, it has been noticed that there is no big difference between 11th and 33th k neighbors only 57% is predicted.

	precision	recall	f1-score	support
Not cardio	0.59	0.59	0.59	7004
cardio	0.59	0.58	0.59	6996
accuracy			0.59	14000
macro avg	0.59	0.59	0.59	14000
weighted avg	0.59	0.59	0.59	14000

Figure 5. KNN, Confusion matrix and classification report

Figure 5 is shown accuracy score better predicted in the nearest neighbors 51 but in 55 it is getting worse. In order to predict better result, we need find the best option here. Before 51 it was tried all classifiers but accuracy was only between 56-57, the best result was shown only here, anyway 58 it is not tending to be a good prediction. As a result, you are able to see here accuracy result is about 59% which is not bad, but still need some good options to increase result. Result: Accuracy of KNN is 59%.

Logistic Regression. Logistic regression is a classification algorithm used to assign observations to a discrete set of classes. Some of the examples of classification problems are Email spam or not spam, Online transactions Fraud or not Fraud, Tumor Malignant or Benign. Logistic regression transforms its output using the logistic sigmoid function to return a probability value [4].

Logistic regression defined as the «Sigmoid function» or also known as the «logistic function» instead of a linear function. Figure 6 is represented Sigmoid function, which defines Logistic regression.

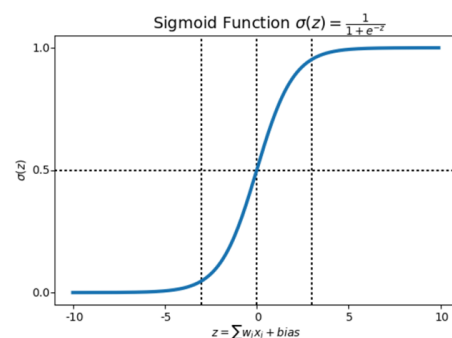


Figure 6. Sigmoid function

Implementation of Logistic Regression algorithm in Python. To explore our dataset by logistic Regression algorithm, there is been added method import LogisticRegression. Figure 7 is shown source code of importing methods and clusters in Python. Accuracy of logistic regression classifier on test set: 0.71(71%). This accuracy is better than KNN which has shown only 57%.

In the next Figure 8 we able to see confusion matrix and classification reports.

The result of confusion matrix comparing KNN and Logistic Regression it is definitely better confusion matrix in Logistic regression 5255+1749>2323+1749 in KNN we are able to see following result in confusion matrix 4154+4082>2914+2850, as a result confusion matrix of Logistic Regression is found as the best. In Classification report

of Logistic Regression there is precision 69%, recall 75%, f1 score 72% higher than in KNN here following classifiers pre-

cision 59%, recall 59%, f1 score 59% constantly. Precision better predict for 10% defense, recall 16%, f1 score 13%.

```
from sklearn.linear_model import LogisticRegression
from sklearn import metrics

logreg = LogisticRegression()
logreg.fit(X_train, y_train)

C:\Users\user\Anaconda3\lib\site-packages\sklearn\linear_model\logistic.py:432: FutureWarning: Default solver will be change
d to 'lbfgs' in 0.22. Specify a solver to silence this warning.
  FutureWarning)

LogisticRegression(C=1.0, class_weight=None, dual=False, fit_intercept=True,
  intercept_scaling=1, l1_ratio=None, max_iter=100,
  multi_class='warn', n_jobs=None, penalty='l2',
  random_state=None, solver='warn', tol=0.0001, verbose=0,
  warm_start=False)

y_pred = logreg.predict(X_test)
print('Accuracy of logistic regression classifier on test set: {:.2f}'.format(logreg.score(X_test, y_test)))

Accuracy of logistic regression classifier on test set: 0.71
```

Figure 7. Logistic regression

	precision	recall	f1-score	support
0	0.68	0.73	0.70	7004
1	0.71	0.66	0.68	6996
accuracy			0.69	14000
macro avg	0.69	0.69	0.69	14000
weighted avg	0.69	0.69	0.69	14000

Figure 8. Logistic Regression, Confusion matrix and classification report

Let's visualize True and Positive Rate False Negative Rate as well in Figure 9.

```
from sklearn.metrics import roc_auc_score
from sklearn.metrics import roc_curve
logit_roc_auc = roc_auc_score(y_test, logreg.predict(X_test))
fpr, tpr, thresholds = roc_curve(y_test, logreg.predict_proba(X_test)[:,1])
plt.figure()
plt.plot(fpr, tpr, label='Logistic Regression (area = %0.2f)' % logit_roc_auc)
plt.plot([0, 1], [0, 1], 'r--')
plt.xlim([0.0, 1.0])
plt.ylim([0.0, 1.05])
plt.xlabel('False Positive Rate')
plt.ylabel('True Positive Rate')
plt.title('Receiver operating characteristic')
plt.legend(loc='lower right')
plt.savefig('Log_ROC')
plt.show()
```

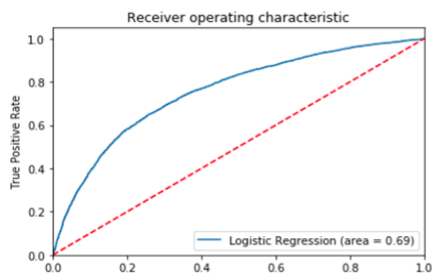


Figure 9. True Positive Rate and False Negative Rate of Logistic Regression

To sum up, I would probably choose in my case logistic regression, as a result it seems to make better prediction from 59% to 69%.

Result: Accuracy of Logistic Regression is 69%.

Support Vector Machine. The objective of the support vector machine algorithm is to find a hyperplane in an N-dimensional space (N — the number of features) that distinctly classifies the data points. To separate the two classes of data points, there are many possible hyperplanes that

could be chosen. Our objective is to find a plane that has the maximum margin, i.e. the maximum distance between data points of both classes. Maximizing the margin distance provides some reinforcement so that future data points can be classified with more confidence [5].

In SVM, we take the output of the linear function and if that output is greater than 1, we identify it with one class and if the output is -1, we identify it with another class. Since the threshold values are changed to 1 and -1 in SVM, we obtain this reinforcement range of values ([-1,1]) which acts as margin. In Formula 1 there is Hinge loss function.

$$c(x, y, f(x)) = \begin{cases} 0, & f y * f(x) \geq 1 \\ 1 - y * f(x), & \text{else} \end{cases} \quad (1)$$

Implementation in Python SVM algorithm. The earliest step we need to start with, is to import SVM. This method is used to call classifier SVC `svclassifier = SVC()`.

`plt.scatter(X_train[:, 0], X_train[:, 4], c=y_train, cmap = 'spring')`. Following code represents scatter plot that will show visualization of data, which is shown in Figure 10.

```
In [98]: plt.scatter(X_train[:, 0], X_train[:, 4], c=y_train, cmap = 'spring')
Out[98]: <matplotlib.collections.PathCollection at 0x20594c93848>
```

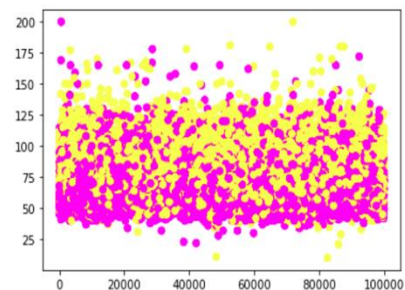


Figure 10. Scatter plot of dataset

For SVM is used 2 types of kernel sigmoid and rbf. Furthermore, it has got 2 accuracy results, confusion matrix and classifier report as well. Figure 11 is shown report of SVM rbf classifier report.

	precision	recall	f1-score	support
0	0.61	0.54	0.57	7004
1	0.59	0.66	0.62	6996
accuracy			0.60	14000
macro avg	0.60	0.60	0.60	14000
weighted avg	0.60	0.60	0.60	14000

Figure 11. SVM, rbf, Confusion matrix and classification report

As a result, it is clear that accuracy of rbf kernel for SVM is about 60% macro and weighted average the same 60% which better than in KNN which was 59% but worse than Logistic Regression which was 71%. Precision is 61% recall seems worse only 54% and support there are 7004 for 0 target. For target 1 result in precision worse 59% but recall 66% and f1-score 62%, support only 6996.

Now Figure 12 is presented Confusion matrix and classification report of SVM for sigmoid classifier. `svclassifier1 = SVC(kernel='sigmoid')` this following code is used to identify result for sigmoid.

	precision	recall	f1-score	support
0	0.51	0.50	0.51	7004
1	0.51	0.51	0.51	6996
accuracy			0.51	14000
macro avg	0.51	0.51	0.51	14000
weighted avg	0.51	0.51	0.51	14000

Figure 12. SVM, sigmoid, Confusion matrix and classification report

For cardiovascular dataset rbf classifier is better than sigmoid accuracy in rbf 0.508 accuracy in sigmoid 0.5995 comparing confusion matrix (sigmoid) [3419 3577] correct predictions 7112>6888 negative predictions comparing confusion matrix (rbf) [3782 3222] [2385 4611] correct predictions 8393>5607 negative predictions precision 51% in sigmoid, precision 61% in rbf recall no big difference 50% and 54% in rbf f1 51% f1 57% in svm kernel rbf.

Result: Accuracy of SVM kernel = rbf is 60%.

Naïve Bayes. A Naïve Bayes classifier is a probabilistic machine learning model that's used for classification task. The crux of the classifier is based on the Bayes theorem. In the following formula 2 you are able to see Bayes theorem about Using Bayes theorem, we can find the probability of A happening, given that B has occurred. Here, B is the evidence and A is the hypothesis. The assumption made here is that the predictors/features are independent. That is presence of one particular feature does not affect the other. Hence it is called naïve [6].

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \quad (2)$$

Implementation in Python Naïve Bayes algorithm. In this Naïve Bayes algorithms are used three types of classifiers, such as Bernoulli Naïve Bayes with following source code:

1. `from sklearn.naive_bayes import BernoulliNB`
2. `from sklearn.model_selection import train_test_split`
3. `bnb = BernoulliNB(binrize=0.0)`
4. After that has been used Multinomial Naïve Bayes with following source code:
5. `from sklearn.naive_bayes import MultinomialNB`
6. `from sklearn.model_selection import train_test_split`
7. `mnb = MultinomialNB(alpha=0.01)`

8. The last one Gaussian Naïve Bayes with following source code:

9. `from sklearn.naive_bayes import GaussianNB`
10. `gnb = GaussianNB()`

In Figure 12 is demonstrated accuracy 57% for Gaussian Naïve Bayes. Moreover, there is confusion matrix with not bad result and recall with 95% in classification result which the best comparing with others.

Accuracy:	precision	recall	f1-score	support
0	0.54	0.95	0.69	7004
1	0.78	0.19	0.30	6996
accuracy			0.57	14000
macro avg	0.66	0.57	0.49	14000
weighted avg	0.66	0.57	0.49	14000

Figure 13. Gaussian Naïve Bayes, Confusion matrix and classification report

In Figure 14 is demonstrated Bernoulli Naïve Bayes accuracy 52% with the same precision and f1-score as well. It seems Gaussian is better for the following dataset than Bernoulli.

Accuracy:	precision	recall	f1-score	support
0	0.51	0.82	0.63	7004
1	0.55	0.22	0.31	6996
accuracy			0.52	14000
macro avg	0.53	0.52	0.47	14000
weighted avg	0.53	0.52	0.47	14000

Figure 14. Bernoulli Naïve Bayes, Confusion matrix and classification report

Decision Tree. Decision tree can be used to visually and explicitly represent decisions and decision making. As the name goes, it uses a tree-like model of decisions. Though a commonly used tool in machine learning for deriving a strategy to reach a particular goal, it's also widely used in machine learning, which will be the main focus of this article. Growing a tree involves deciding on which features to choose and what conditions to use for splitting, along with knowing when to stop. As a tree generally grows arbitrarily, you will need to trim it down for it to look beautiful [7].

Implementation of Decision Tree algorithm in Python. To implement DT in Python we need to import it via this code:

`from sklearn.tree import DecisionTreeClassifier`

In Figure 15 Accuracy result is shown as 64% which better than KNN and SVM, Naïve Bayes as well.

```
from sklearn.tree import export_graphviz

clf = DecisionTreeClassifier()

clf = clf.fit(X_train,y_train)

y_pred = clf.predict(X_test)

y_pred
array([1, 0, 0, ..., 0, 0, 1], dtype=int64)

print("Accuracy:",metrics.accuracy_score(y_test, y_pred))
Accuracy: 0.641
```

Figure 15. DT accuracy

If we use DT entropy it is showed better result in Accuracy with 73% with following code:

`clf = DecisionTreeClassifier (criterion="entropy", max_depth=3)` in Figure 16.

```

clf = DecisionTreeClassifier(criterion="entropy", max_depth=3)

clf = clf.fit(X_train,y_train)

y_pred = clf.predict(X_test)

print("Accuracy:",metrics.accuracy_score(y_test, y_pred))

Accuracy: 0.7288571428571429

```

Figure 16. DT(entropy) accuracy

In Figure 17 below it is represented Decision Tree of our dataset.

Result: Accuracy of Decision Tree is 73%.

Random Forest. Random forest, like its name implies, consists of a large number of individual decision trees that operate as an ensemble. Each individual tree in the random

forest spits out a class prediction and the class with the most votes becomes our model's prediction [7].

Implementation in Python Random Forest algorithm. As we know it starts from importing classifier with following code:

```

from sklearn.ensemble import RandomForestRegressor
regressor = RandomForestRegressor(n_estimators=20,
random_state=0)

```

```

regressor.fit(X_train, y_train)

```

```

y_pred = regressor.predict(X_test)

```

In the Random Forest Mean Absolute Error: 0.3650785714285715

Mean Squared Error: 0.19626714285714283

Root Mean Squared Error: 0.4430204767921488.

```

graph.write_png('cardio.png')
Image(graph.create_png())

```

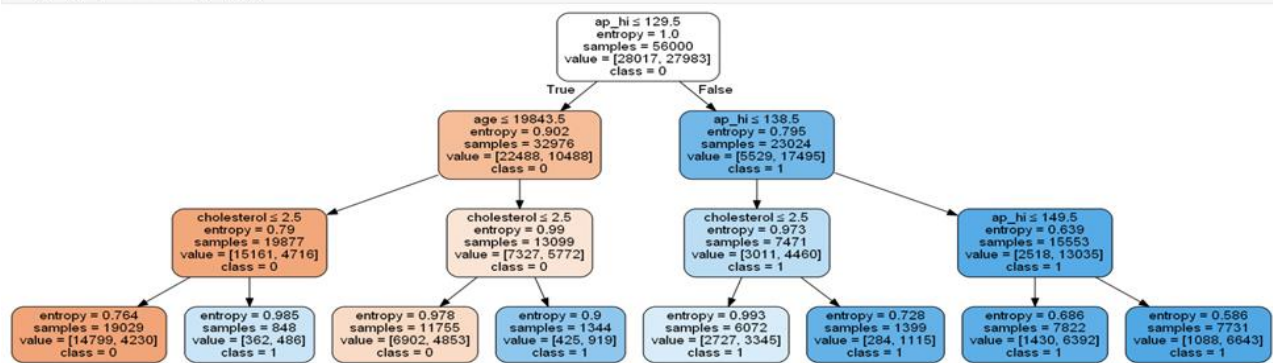


Figure 17. Visualization of DT

Random Forest gives the same accuracy as Decision Tree 73% in Figure 18 it is shown.

```

from sklearn import metrics

print("Accuracy:",metrics.accuracy_score(y_test, y_pred))

Accuracy: 0.728

```

Figure 18. RF accuracy

Result: Accuracy of Random Forest is 73%.

3. Results and discussion

In the previous section, we implemented the main algorithms of machine learning and tested them with a medical dataset. As a result of testing, the accuracy indicators of the algorithms turned out to be different. Their accuracy indicators are shown in Table 1.

Table 1. Accuracy of ML algorithms

№	Machine learning algorithm	Accuracy
1	KNeighborsClassifier	59%
2	Logistic Regression	69%
3	Support Vector Machine kernel=sigmoid	51%
4	Support Vector Machine kernel = rbf	60%
5	Gaussian Naïve Bayes	57%
6	Bernoulli Naïve Bayes	52%
7	Decision Tree	73%
8	Random Forest	73%

As shown in Table 1, the best accuracy algorithms are Decision Tree and Random Forest, accuracy=73%. During the test, two types of the Naïve Bayes method were considered and two different parameters of the Support Vector Machine algorithm were tested. But their accuracy is not higher than Decision Tree algorithm. Random Forest is an ensemble implementation of Decision Tree algorithm.

4. Conclusions

This paper consists of all main methods and algorithms of Machine learning in order to correctly use a medical dataset and get some essential information from useless information. In this article were given following algorithms of Machine learning such as KNN, SVM, DT, RF, Naïve Bayes and Logistic Regression. The results were quite surprising. Accuracy of all algorithms were not less than 50% and not higher than 73%. The best result was demonstrated by Decision Tree and Random Forest they have shown the same result 73% and Logistic Regression had slightly fewer percentages as about 71%. The worst result has represented by Naïve Bayes only 57% whereas KNN had shown 59%.

To sum up, I would probably say that Decision Tree algorithm and Random Forest made perfect job for dataset cardiovascular_train.csv. I suppose, it happened because of datatypes and our target which was Boolean 1 and 0.

References

- [1] Embi, P.J. & Payne, P.R. (2009). Clinical research informatics: challenges, opportunities and definition for an emerging do-

- main. *Journal of the American Medical Informatics Association*, 16(3), 316–327. <https://doi.org/10.1197/jamia.M3005>
- [2] Prokosch, H.U. & Ganslandt, T. (2009). Perspectives for medical informatics. Reusing the electronic medical record for clinical research. *Methods of information in medicine*, 48(1), 38–44
- [3] Wasserman, R.C. (2011). Electronic medical records (EMRs), epidemiology, and epistemology: reflections on EMRs and future pediatric clinical research. *Academic pediatrics*, 11(4), 280–287. <https://doi.org/10.1016/j.acap.2011.02.007>
- [4] Dean, B.B., Lam, J., Natoli, J.L., Butler, Q., Aguilar, D. & Nordyke, R.J. (2009). Review: use of electronic medical records for health outcomes research: a literature review. *Medical care research and review*, 66(6), 611–638. <https://doi.org/10.1177/1077558709332440>
- [5] Tannen, R.L., Weiner, M.G., & Marcus, S.M. (2006). Simulation of the Syst-Eur randomized control trial using a primary care electronic medical record was feasible. *Journal of clinical epidemiology*, 59(3), 254–264. <https://doi.org/10.1016/j.jclinepi.2005.08.008>
- [6] Williams, J.G., Cheung, W.Y., Cohen, D.R., Hutchings, H.A., Longo, M.F. & Russell, I.T. (2003). Can randomised trials rely on existing electronic data? A feasibility study to explore the value of routine data in health technology assessment. *Health technology assessment*, 7(26), iii–117. <https://doi.org/10.3310/hta7260>
- [7] Yamamoto, K., Matsumoto, S., Tada, H., Yanagihara, K., Teramukai, S., Takemura, T. & Fukushima, M. (2008). A data capture system for outcomes studies that integrates with electronic health records: development and potential uses. *Journal of medical systems*, 32(5), 423–427. <https://doi.org/10.1007/s10916-008-9147-7>

Медициналық деректерді өңдеу үшін машиналық оқыту алгоритмдерін пайдалану

Г. Мукажанова¹, Ж. Алибиева^{2*}, А. Касенхан², Н. Мукажанов²

¹Инновациялық Еуразия Университеті, Павлодар, Қазақстан

²Satbayev University, Алматы, Қазақстан

*Корреспонденция үшін автор: zh.alibiyeva@satbayev.university

Андатпа. Бұл мақалада machine learning алгоритмдерінің салыстырмалы талдауы қарастырылады. Алгоритмдерді салыстру келесі деректер жиыны бойынша жүргізіледі: cardio_train.csv from kaggle.com (сілтеме: <https://www.kaggle.com/sulianova/cardiovascular-disease-dataset>). Сонымен қатар, деректерді іздеу алгоритмдері cardio_train.csv дәлдігі бойынша ең жақсы алгоритмдерді анықтайды. Python 3.0 бағдарламалау тілінде орындалған процедураларды тексеру, матрица және жіктеу есебі, дәлдік көрсеткішін, еске түсіруді, f1 ұпайын және қолдауды көруге мүмкіндік береді. Сонымен қатар, осы мақалада келесі классификациялық модельдерді көруге болады: KNN алгоритмі, логистикалық регрессия, шешім ағашы, Random Forest, Naive Bayes және SVM. Нәтижесінде медициналық деректерді өңдеудің ең жоғары дәлдігі анықталады.

Негізгі сөздер: медициналық мәліметтер базасы, деректер жинау, алгоритмдер, KNN алгоритмі, логистикалық регрессия, шешім ағашы, Random Forest, SVM, Naive Bayes.

Использование алгоритмов машинного обучения для обработки медицинских данных

Г. Мукажанова¹, Ж. Алибиева^{2*}, А. Касенхан², Н. Мукажанов²

¹Инновационный Евразийский Университет, Павлодар, Казахстан

²Satbayev University, Алматы, Казахстан

*Автор для корреспонденции: zh.alibiyeva@satbayev.university

Аннотация. В статье рассматривается сравнительный анализ алгоритмов Machine learning для набора данных: cardio_train.csv от kaggle.com. (ссылка: <https://www.kaggle.com/sulianova/cardiovascular-disease-dataset>). Более того, с помощью алгоритмов machine learning будут обнаружены алгоритмы наилучшей точности для cardio_train.csv. Рассматривается процедура, разработанная на языке программирования Python 3.0, которая представляет собой confusion matrix и отчет о классификации, позволяет увидеть оценку точности, отзыв, оценку f1 и поддержку. Кроме того, в этой статье вы можете увидеть следующие модели классификации: алгоритм KNN, логистическая регрессия, дерево решений, Random Forest, наивный байесовский метод и SVM. В результате будет определена высочайшая точность обработки медицинских данных.

Ключевые слова: набор медицинских данных, сбор данных, алгоритмы, KNN алгоритм, логистическая регрессия, древо решений, Random Forest, SVM, наив байесовский метод.

Received: 13 December 2022

Accepted: 16 March 2023

Available online: 31 March 2023

A review of common practices and challenges in autonomous driving

K. Saidov*, A. Moldagulova

Satbayev University, Almaty, Kazakhstan

*Corresponding author: kirill.saidov.d@gmail.com

Abstract. Autonomous driving technologies have garnered significant attention in recent years, promising transformative impacts on transportation systems. The landscape of transportation is undergoing a profound transformation with a focus on achieving autonomy in vehicles, ranging from advanced driver assistance systems (ADAS) to the ambitious goal of fully autonomous vehicles. This article delves into the complexity of autonomous driving, exploring both the advancements driving this paradigm shift and the intricate challenges impeding its seamless integration into daily life. The journey toward autonomy involves breakthroughs in sensor technology, artificial intelligence, and connectivity, with a crucial emphasis on sensor fusion for precise navigation. The review highlights key advancements in machine learning, computer vision, and sensor technologies that underpin autonomous driving systems, offering insights into their current capabilities and limitations. The synthesis of this review aims to provide a holistic understanding of the current state of autonomous driving, facilitating informed discussions among researchers, practitioners, and the broader public. By shedding light on both achievements and challenges, this paper contributes to the ongoing discourse on the future of autonomous driving and informs the development of strategies to address the complexities inherent in achieving widespread adoption of this transformative technology.

Keywords: *autonomous driving, self-driving vehicles, artificial intelligence, sensor fusion, deep learning.*

1. Introduction

The landscape of transportation is undergoing a transformative shift, marked by the relentless pursuit of autonomy in vehicles. From cutting-edge advanced driver assistance systems (ADAS) to the ambitious vision of fully autonomous vehicles, delivery robots, the automotive industry is at the forefront of technological innovation. In this article, we delve into the intricate tapestry of autonomous driving, examining both the common practices that propel this paradigm shift and the multifaceted challenges that cast a shadow on its seamless integration into our daily lives.

The journey towards autonomous driving has been characterized by a convergence of breakthroughs in sensor technology, artificial intelligence and deep learning algorithms, connectivity, and mapping. Vehicles equipped with an array of sensors, including cameras, LiDAR, radar, and ultrasonic sensors, engage in the intricate dance of sensor fusion. This practice involves the meticulous integration of data from diverse sensors, creating a composite and nuanced understanding of the vehicle's surroundings. Sensor fusion is the linchpin of perception for autonomous driving, allowing to navigate the complex and dynamic environment with high precision and error tolerance.

However, according to Muhammad et al. significant improvements in sensor fusion technologies and hardware manufacturing still demand further attention in research and academia before full industry deployment as it serves the key role in reducing road accidents and saving human lives [1]. For instance, only in Kazakhstan approximately 15 thousand people were injured because of over 10 thousand road incidents occurring in the preceding year [2]. There has been a notable

21% increase in deaths counts. Further analysis suggests that over 86% of accidents are caused by drivers.

Machine learning and artificial intelligence technologies empower deep neural networks to learn from vast datasets, adapt to diverse driving conditions, and make split-second decisions. This can hopefully dramatically reduce the number of road accidents and subsequent deaths. According to the National High-way Traffic Safety Administration (NHTSA), over 94% of incidents are usually caused by human errors [3]. The advent of deep learning, the widespread deployment of Automated Driving Systems (ADSs) is projected to reduce not only the number of accidents, but also atmospheric emissions, stress, increase traffic efficiency and overall social wellbeing by nearly \$800 billion in monetary value by 2050 [4].

Despite these notable strides, the road to fully autonomous driving is fraught with challenges that demand meticulous consideration. Safety concerns loom large, requiring the industry to address unpredictable variables such as adverse weather conditions, erratic human drivers, and unexpected obstacles. Striking a balance between achieving technological reliability that matches or surpasses human drivers in all situations and ensuring the utmost safety is a formidable challenge.

The regulatory and legal landscape presents another intricate puzzle. The accelerated pace of technological development has outpaced the establishment of comprehensive frameworks governing autonomous driving. Questions of liability in the event of accidents, data privacy, and the standardization of testing procedures demand urgent attention to facilitate the widespread adoption of autonomous vehicles. Harmonizing these regulations on a global scale emerges as a pressing necessity.

Ethical dilemmas cast a philosophical shadow on the path to autonomy. Autonomous vehicles are confronted with situations where moral decisions must be made, such as choosing between minimizing harm to the vehicle occupants and avoiding harm to pedestrians. Resolving these ethical conundrums requires a delicate balance between societal values, legal considerations, and technological capabilities.

In an era where vehicles are increasingly connected and reliant on software, cybersecurity risks add an additional layer of complexity. Autonomous vehicles, with their extensive network connectivity, become susceptible to hacking, potentially leading to severe consequences. Establishing robust cybersecurity measures to protect vehicle systems from unauthorized access and manipulation is an ongoing challenge that requires constant vigilance.

As we navigate through the intricate landscape of autonomous driving, it becomes evident that this technological frontier is not merely a convergence of hardware and software but a holistic reimagining of transportation. The promise of safer, more efficient, and accessible mobility beckons, but not without overcoming the intricate web of challenges that intertwine with progress. Researchers, engineers, policymakers, and society at large are integral players in this transformative narrative, shaping the future of autonomous driving.

1.1. Autonomous driving prospects and social challenges

The widespread adoption of ADSs is imminent. According to Moreno et al. the availability of vehicles results in increased financial and biodiversity burden to cities [5]. Moreover, recent COVID-19 events have sped up Artificial Intelligence (AI) adoption and subsequently greater enhancements of Deep Learning (DL) algorithms. This has led to the re-emergence of an old concept of smart cities or rebranded as the «15-Minute City», closely discussed by Moreno et al. in his work. He mentions that this concept has been depicted as the Sustainable Development Goal 11 of United Nations. And one of the sub-goals revolves around removing or, at least, replacing non-green vehicles from cities with their electric counterparts. Consequently, this leads to autonomous driving as new opportunities arise.

Yurtsever and his team foresees the following potential impacts on society [6]:

- 1) ADSs will help mitigate traffic accidents, improve traffic efficiency, and reduce emissions by stabilizing the city ecosystem.

- 2) A new opportunity shall arise revolving around Mobility as a Service (MaaS), which already has a noticeable impact on logistics.

However, Maas can also play a major role in other areas of human life, not limited to logistics. Potentially, with an increased growth of elderly people, ADSs technology can help them improve their quality of life and productivity. Not to mention a steady shift towards MaaS consumption by masses as opposed to vehicle-ownership. According to Yurtsever's et al. research, vehicle ownership is projected to become 50:50 by 2030.

As ADSs become more advanced and intricate, they gain the ability to operate in indeterministic environments. With its fast-paced evolution, there is a need to monitor and classify the level of automation. According to the Society of Automotive Engineers (SAE) there are 5 levels of driving automation. The taxonomy mentions level zero as no automation at all. It's up to the driver to handle every aspect of driving. Level one de-

picts primitive driver assistance, whereas level two includes partial automation. These systems usually have emergency braking and collision avoidance mechanisms integrated into vehicles to support the drivers in emergency situations. SAE note that the difficulties start arising from level three and upward.

The challenge of level three automation lies in conditional automation: in cases when the driver needs to take over the control during an emergency. In addition, level three automation is limited to certain operational domains. For example, highways. During an investigation, it has been proven that the control takeover from automated mode to manual mode usually results in traffic accident risks. Thus, this is yet to be solved.

Level four automation adds on a whole new complexity layer on top of existing technology. It includes automatic departure, parking, and routing. Level five automation steps up the game by making the vehicle operate seamlessly on any road network, any weather condition or indeterministic situations. Nonetheless, both levels of automation require special domain infrastructure to operate well. At the current state of urban roads, the environmental variables are still highly indeterminate, which are difficult to predict accurately.

For example, Tesla's ADS failed to differentiate a white truck colliding with the vehicle and killing the driver [8]. Therefrom arise the ethics dilemma: who's responsible, and how the system should normally behave? Should it prioritize the driver's wellbeing or the pedestrians. These questions need careful consideration.

2. System components and architectures

This chapter describes the software and hardware used by the ADS researchers, developers and engineers and their intrinsic details. We explore the intricacies of deep learning-based decision-making architectures and their components. ADS are designed to operate independently by processing streams of incoming data from different on-board sensors. These can include cameras, radars, global positioning systems (GPS), light detection and rangings (LiDARs), ultrasonic sensors and many more.

Various components of ADS architectures are usually based on AI and Deep Learning technologies but are not limited to these. Sometimes a classical approach is taken that involves non-learning-based components. Nonetheless, these systems still have a common architecture at its core: perception, localization, high-level planning, low-level planning (behavior arbitration) and motion controllers. The system may consist of an end-to-end learning approach where the sensory input data is mapped to motion controllers or of an action pipeline-based approach where decisions are computed in a pipeline-based fashion [9].

2.1. Key technologies

The ADS systems involve multiple components such as computer systems, robotics, mechanical engineering, machine and deep learning, communication, systems engineering and many more resulting in a complex autonomous device of its own. The actual computing system usually includes a set of similar technologies to benefit from its advantages and overcome the disadvantages of the devices on-board. For example, the ADS system usually includes a variety of sensors: cameras, radar, LiDAR, ultrasonic sensors, and GPS.

Cameras play a pivotal role in the sensor suite of autonomous vehicles, contributing to their ability to perceive and

interpret the surrounding environment. Cameras are primarily employed for environmental perception, capturing visual information from the vehicle's surroundings. They act as the «eyes» of the autonomous system, providing a real-time feed of the road, traffic, pedestrians, and other relevant objects. The cameras give a straightforward 2D view of the surroundings, making it useful for object classification and lane detection. Lane markings, road edges, and other lane-related information are extracted from the camera feed. Additionally, cameras contribute to tracking the vehicle's position within the lane and adjusting its trajectory accordingly. However, cameras also face challenges such as adverse weather conditions, low-light situations, and potential occlusions. To mitigate these challenges, sensor fusion and redundancy strategies are often employed, combining camera data with information from other sensors to enhance overall reliability and safety in autonomous driving systems.

Whereas cameras capture visual information in the form of images or video frames using visible or infrared light, radars use radio waves to detect objects and measure their distance, speed, and angle. They provide long-range detection and are less affected by environmental conditions but typically have lower resolution. The generated data size of radars is small: around 10-100 KB per second [10].

Like radars, LiDARs emit laser beams to measure the time it takes for the light to reflect off objects, providing precise distance and 3D mapping. They have long-range detection and high resolution in three dimensions, providing detailed spatial information. The performance is also notably good. LiDARs are widely used in object detection, distance estimation and edge detection of still objects. The sensor is less effected by weather conditions than camera, but the competitive cost is high, which restricts its wide adoption in ADS.

2.2. Critical ADS tasks

The utilization of machine learning, particularly deep neural networks, is a cornerstone in autonomous driving. The entire task of navigating through a city can be subdivided into six major components: road detection, lane detection, vehicle detection, pedestrian detection, collision avoidance and traffic sign detection.

Road detection aims at recognizing road boundaries and other areas where autonomous vehicles are allowed to drive. A common practice is to use convolutional neural networks (CNNs) for such tasks. There are also other works presenting an end-to-end model called RNet for road detection in a single network [1]. Lane detection, like road detection, is responsible for keeping within the vehicle lane on roads, thus, ensuring vehicle safety and minimizing risk of collision.

Vehicle and pedestrian detection are vital part of ADS system. It must recognize other vehicles and objects, and estimate their sizes, shapes, and relative speed to navigate around the city. Pedestrian-vehicle accidents are a common issue. The ADS need to learn to differentiate humans from other objects, track all possible pedestrians to avoid collision. Wang et al. proposed a new system with pedestrian body parts semantic detection using DNNs and contextual information to build accurate location [11].

2.3. ADS architectures

A robust architecture is directly responsible for ADS system performance. It defines how the entire system is controlled and managed. A good system architecture can help auto-

nomous vehicles computer and analyze voluminous amounts of data more efficiently and produce better predictions. There are many approaches to how to design these systems.

One popular approach is an ego-only system. This system carries all the necessary technology on board and is independent of other ADS vehicles, always making driving decisions in a single self-sufficient manner. Whereas connected ADS may or may not depend on each other, which is decided by the infrastructure and situation on the road, and the need to exchange the information when such arises.

Another approach is a modular system [12]. It is structured as a pipeline linking together different components of sensory input. The typical pipeline looks like feeding data streams from sensors into object detection and locations modules. The produced information is then used for scenery prediction and navigation; decision-making is generated and fed to the control module. The advantage of such a system is in its modularity, but so is its disadvantage: the error is propagated along the entire pipeline with small errors resulting in major system failures.

The third approach was mentioned earlier: end-to-end driving. This approach revolves around the model trying to imitate an expert human driver. This approach is not fully end-to-end though, as it needs an additional step to generate the driving actions. But the question here is should the ADS system drive like a human or not? The end-to-end driving approach is an emerging promising technology. It learns to interact with the environment through repeated failures, but lacks safety-measures and interpretability, making it unpopular.

3. Challenges and corner cases

The pursuit of autonomous driving technology leads to groundbreaking advancements and rapid adoption, equally, a spectrum of intricate challenges and corner cases. As engineers and researchers push the boundaries of innovation, they grapple with scenarios that transcend the ordinary, demanding solutions that can navigate the unpredictable intricacies of real-world environments.

In recent years, the number of news or road accidents involving ADS systems that led to fatalities has increased. Early adoption of self-driving vehicles had led to five cases of ADS failures: four of which are attributed to Tesla and one to Uber [13]. The first two cases happened in 2016. The autopilot failed to recognize the truck in both cases, taking it for open space the second time. On the third incident the ADS system failed to recognize the highway divider in 2018. And in 2019 the autopilot crashed unable to recognize the semitrailer. With regards to Uber, the ADS system failed to recognize pedestrians walking.

There are many challenges that researchers and engineers must overcome to make automated driving safe. One such challenge is how to handle irrational or unpredictable human behavior. Human drivers remain a formidable challenge for autonomous systems. From unpredictable decision-making at intersections to sudden lane changes, interpreting and predicting human actions present complex challenges that require nuanced solutions. Handling the idiosyncrasies of human behavior in diverse cultural and driving contexts adds an additional layer of complexity.

Another challenge is adverse weather conditions. More particularly, poor illumination and changing appearance. The main drawback of using cameras has to do with lighting

conditions. It is inherently difficult to deal with low-light conditions. For example, snow may drastically change the appearance of city streets or roads, hiding the key features of scenes such as road lanes. To solve this issue, a sensor fusion strategy is employed as described above. Although, these strategies are not robust. Driving in direct sunlight may cause problems to ADS vision systems anyway as it is also susceptible to direct sun glare. Paul et al. proposed a combination of HDR algorithms to improve autopilot's performance [14].

There is also an ongoing debate on how to handle failure detection and diagnostics. How to define sensor failure? What constitutes failure? There is no pre-defined standard. Moreover, there is no reliable study or standard on how to detect sensor failures. Even if the sensors are working properly, how should the sensor data failure be detected in real time scenarios? The sensors may be working correctly, but the generated data may not reflect the actual scenario. An example may be sensor blocking or occlusion. The last type of failure is algorithmic. Hazardous weather conditions may directly affect algorithm's performance. Sometimes utilizing priorly collected information is important. Therefore, developing robust algorithms still proves to be a challenge.

Autonomous driving systems also pose a new challenge to their developers: cyberattack protection. With a wide adoption of autonomous cars cybersecurity becomes an important part of ADS. There is no absolute security, but basic protection from spoof attacks, denial of service makes it vital for human safety. The data streams must be checked before proceeding further to sensor fusion. For instance, an occluded roadblock detected by radars may be corrected by the camera data and vice versa.

Lastly, the final issue is to manage ADS energy consumption, effectiveness, and costs. Finding the balance between the three proves to be a real challenge.

4. Conclusions

The journey toward autonomy is defined by breakthroughs in sensor technology, artificial intelligence, deep learning algorithms, connectivity, and mapping. Sensor fusion, a linchpin of perception in autonomous driving, intricately weaves data from an array of sensors to create a nuanced understanding of the vehicle's surroundings. However, the journey is not without its tribulations.

Safety concerns loom large, necessitating meticulous attention to unpredictable variables such as adverse weather conditions, erratic human drivers, and unexpected obstacles. The regulatory and legal landscape poses a puzzle, demanding swift attention to establish comprehensive frameworks governing autonomous driving. Ethical dilemmas cast philosophical shadows, requiring a delicate balance between societal values, legal considerations, and technological capabilities.

The imminent widespread adoption of Automated Driving Systems (ADSs) holds promises and social challenges. The advent of smart cities, propelled by AI adoption and enhanced Deep Learning algorithms, envisions a shift towards sustainable, efficient urban living. The rise of Mobility as a Service (MaaS) emerges as a transformative force, potentially improving the quality of life for the entire population and reshaping the landscape of vehicle ownership.

The diverse architectures of ADSs, from ego-only systems to modular pipelines and end-to-end driving approach-

es, offer insights into the myriad ways researchers and engineers approach the design and functionality of autonomous systems. Each architecture comes with its advantages and challenges, emphasizing the need for ongoing innovation and refinement.

References

- [1] Muhammad, K., Ullah, A., Lloret, J., Del Ser, J., & de Albuquerque, V.H.C. (2020). Deep learning for safe autonomous driving: Current challenges and future directions. *IEEE Transactions on Intelligent Transportation Systems*, 22(7), 4316-4336. <https://doi.org/10.1109/TITS.2020.3032227>
- [2] Governments Report. (2022). The number of road accidents in Kazakhstan increased by 8.9%. Retrieved from: <https://www.gov.kz/memleket/entities/pravstat/press/news/detail/s/444039?lang=ru>
- [3] Singh, S. (2015). Critical reasons for crashes investigated in the national motor vehicle crash causation survey. Washington, DC, USA, Tech. Rep. DOT HS 812 115. Retrieved from: <https://crashstats.nhtsa.dot.gov/Api/Public/ViewPublication/812115>
- [4] Montgomery, W.D., Mudge, R., Groshen, E.L., Helper, S., MacDuffie, J.P. & Carson, C. (2018). America's workforce and the self-driving future: Realizing productivity gains and spurring economic growth. *Securing America's Future Energy*, Washington, DC, USA. Retrieved from: <https://avworkforce.secureenergy.org/wp-content/uploads/2018/06/Americas-Workforce-and-the-Self-Driving-Future-Realizing-Productivity-Gains-and-Spurring-Economic-Growth.pdf>
- [5] Moreno, C., Allam, Z., Chabaud, D., Gall, C. & Pratlong, F. (2021). Introducing the «15-Minute City»: Sustainability, resilience and place identity in future post-pandemic cities. *Smart Cities*, 4(1), 93-111. <https://doi.org/10.3390/smartcities4010006>
- [6] Yurtsever, E., Lambert, J., Carballo, A. & Takeda, K. (2020). A survey of autonomous driving: Common practices and emerging technologies. *IEEE access*, (8), 58443-58469. <https://doi.org/10.1109/ACCESS.2020.2983149>
- [7] Standatrs. (2016). Taxonomy and definitions for terms related to driving automation systems for on-road motor vehicles. Retrieved from: https://www.sae.org/standards/content/j3016_201401/
- [8] McFarland, M. (2019). Who's Responsible When an Autonomous Car Crashes? Retrieved from: <https://money.cnn.com/2016/07/07/technology/tesla-liability-risk/index.html>
- [9] Grigorescu, S., Trasnea, B., Cocias, T. & Macesanu, G. (2020). A survey of deep learning techniques for autonomous driving. *Journal of Field Robotics*, 37(3), 362-386. <https://doi.org/10.1002/rob.21918>
- [10] Liu, L., Lu, S., Zhong, R., Wu, B., Yao, Y., Zhang, Q. & Shi, W. (2020). Computing systems for autonomous driving: State of the art and challenges. *IEEE Internet of Things Journal*, 8(8), 6469-6486. <https://doi.org/10.1109/JIOT.2020.3043716>
- [11] Wang, S., Cheng, J., Liu, H., Wang, F. & Zhou, H. (2018). Pedestrian detection via body part semantic and contextual information with DNN. *IEEE Transactions on Multimedia*, 20(11), 3148-3159. <https://doi.org/10.1109/TMM.2018.2829602>
- [12] Chen, C., Seff, A., Kornhauser, A. & Xiao, J. (2015). DeepDriving: Learning affordance for direct perception in autonomous driving. *IEEE International Conference on Computer Vision (ICCV)*, 2722-2730. <https://doi.org/10.1109/ICCV.2015.312>
- [13] Paul, N. & Chung, C. (2018). Application of HDR algorithms to solve direct sunlight problems when autonomous vehicles using machine vision systems are driving into sun. *Computers in Industry*, 98, 192-196.

Автономды жүргізудің жалпы тәжірибелері мен міндеттеріне шолу

К. Саидов*, А. Молдагулова

Satbayev University, Алматы, Қазақстан

*Корреспонденция үшін автор: kirill.saidov.d@gmail.com

Андатпа. Автономды жүргізу технологиялары соңғы жылдары айтарлықтай назар аударып, көлік жүйелеріне трансформациялық әсер етеді. Көлік индустриясы жүргізушіге көмек көрсетудің озық жүйелерінен (ADAS) толық автономды көліктердің өршіл мақсатына дейін көліктердің автономиясына қол жеткізуге бағытталған терең өзгерістерді бастан кешіруде. Бұл мақалада автономды жүргізудің күрделілігіне тереңірек үңіліп, осы парадигманың ауысуына әкелетін жетістіктерді де, оның күнделікті өмірге біркелкі енуіне кедергі болатын күрделі қиындықтарды да зерттейміз. Автономияға апаратын жол дәл навигация үшін сенсорларды біріктіруге маңызды назар аудара отырып, сенсорлық технологиялар, жасанды интеллект және коммуникациялардағы жетістіктерді қамтиды. Шолу машинаны оқытудағы, компьютерлік көру және сенсорлық технологиялардағы негізгі жетістіктерді көрсетеді, олар автономды жүргізу жүйелерін негіздейді және олардың қазіргі мүмкіндіктері мен шектеулері туралы түсінік береді. Бұл шолудың синтезі зерттеушілер, практиктер және жалпы жұртшылық арасында ақпараттандырылған пікірталасқа жәрдемдесіп, автономды жүргізудің қазіргі жай-күйін тұтас түсінуді қамтамасыз етуге бағытталған. Жетістіктерге де, қиындықтарға да жарық түсіре отырып, бұл құжат автономды көлік жүргізудің болашағы туралы жалғасып жатқан дискурсқа үлес қосады және осы трансформациялық технологияны кеңінен енгізуге тән қиындықтарды шешуге арналған стратегияларды әзірлеу туралы хабарлайды.

Негізгі сөздер: автономды көлік, өзін-өзі басқаратын көлік құралдары, жасанды интеллект, сенсорлар, терең оқыту.

Обзор распространенных практик и задач автономного вождения

К. Саидов*, А. Молдагулова

Satbayev University, Алматы, Казахстан

*Автор для корреспонденции: kirill.saidov.d@gmail.com

Аннотация. В последние годы технологии автономного вождения привлекли к себе значительное внимание, обещая преобразующее воздействие на транспортные системы. Транспортная сфера претерпевает глубокую трансформацию с акцентом на достижение автономности транспортных средств, начиная от передовых систем помощи водителю (ADAS) и заканчивая амбициозной целью создания полностью автономных транспортных средств. В этой статье мы углубимся в сложность автономного вождения, исследуя как достижения, способствующие этому сдвигу парадигмы, так и сложные проблемы, препятствующие его плавной интеграции в повседневную жизнь. Путь к автономности предполагает прорывы в сенсорных технологиях, искусственном интеллекте и средствах связи, при этом решающий упор делается на объединение датчиков для точной навигации. В обзоре освещаются ключевые достижения в области машинного обучения, компьютерного зрения и сенсорных технологий, лежащих в основе систем автономного вождения, а также предлагается понимание их текущих возможностей и ограничений. Обобщение этого обзора направлено на то, чтобы обеспечить целостное понимание текущего состояния автономного вождения, способствуя информированным дискуссиям среди исследователей, практиков и широкой общественности. Проливая свет как на достижения, так и на проблемы, этот документ вносит свой вклад в продолжающийся дискурс о будущем автономного вождения и дает информацию для разработки стратегий для решения сложностей, присущих достижению широкого внедрения этой преобразующей технологии.

Ключевые слова: автономный транспорт, самоуправляемые транспортные средства, искусственный интеллект, сенсоры, глубокое обучение.

Received: 23 November 2022

Accepted: 16 March 2023

Available online: 31 March 2023

Unveiling the Future: Machine Learning Algorithms in Educational Predictive Modeling

O. Adilbek*, R. Satybaldiyeva

Satbayev University, Almaty, Kazakhstan

*Corresponding author: olzha.adilbek@gmail.com

Abstract. Education is an inherent entitlement of every individual. It refers to the process of acquiring knowledge about the items in our surroundings. It facilitates comprehension and resolution of any issue, while promoting equilibrium in all aspects of life. The field of education has seen significant transformations at all levels in the previous two decades. Several changes have occurred as a result of the growing number of technological breakthroughs in both teaching methods and student learning approaches. The emergence of artificial intelligence and machine learning has been a significant milestone in the progression of technology. These two advanced technologies have impacted every aspect of human existence, including commerce, finance, communication, travel, health, and education. Indeed, it is certain that teachers and educators are indispensable. However, technology will inevitably bring about many alterations to the role of a teacher and to the most effective methods of instruction. This review focuses on the overarching topic of machine learning in the field of education.

Keywords: machine learning, educational predictive modeling, student outcomes, personalized learning, algorithmic applications, educational technology, learning analytics, predictive analytics.

1. Introduction

There has been a significant shift in the area of education brought about by the use of machine learning algorithms into educational predictive modeling. Traditional statistical techniques, while their significance, have limits when it comes to capturing the intricate dynamics of student learning. The spread of machine learning has resulted in the development of flexible approaches that are capable of identifying detailed patterns and customizing educational experiences to meet the specific needs of individual students.

Because education is inherently diverse and is always growing, it requires technology that is adaptive in order to properly handle the complexities of student performance, engagement, and learning outcomes. The processing capacity and unique algorithms that machine learning algorithms possess make it possible for corporations and educational institutions to take advantage of a wide variety of possibilities that were previously unprecedented. The purpose of this research is to investigate the historical development of machine learning in educational contexts, as well as investigate its current applications and potential future applications [1].

Within the realm of education, the relevance of machine learning extends well beyond the realm of predictive modeling. The use of data-driven decision-making, early intervention tools, and adaptive learning environments are all included in its scope. The purpose of investigating certain machine learning algorithms, such as the interpretability of Decision Trees and the ensemble strength of Random Forests, is to get a comprehensive knowledge of the uses and limitations of these algorithms in educational environments.

In spite of this, there are ethical implications that may be drawn from this investigation. When it comes to the use of

student data, the reduction of bias, and the comprehension of complex models, key focal areas emerge. It will be necessary for the ethical framework that governs the use of machine learning in educational settings to develop in tandem with the progression of technology in the years to come [2].

It is possible to create opportunities for ongoing innovation by anticipating impending advances, such as reinforcement learning and natural language processing, among other technological advancements. There is the possibility that the use of machine learning in the field of education may lead to a future in which education will be individualized, adaptable, and comprehensive. The purpose of this study is not only to get an understanding of the current state of affairs, but also to encourage introspection on the ethical responsibilities and transformative powers that are inherent in the mutually beneficial link between education and machine learning.

The integration of machine learning algorithms and pedagogical predictive modeling signifies a significant paradigm shift within the field of education. Although conventional statistical methods are crucial, they are incapable of depicting the intricate dynamics of student learning. The expansion of machine learning yields versatile methods capable of deciphering intricate patterns and customizing educational experiences to meet the specific needs of each learner.

Education, being inherently diverse and ever-changing, necessitates flexible technologies that can effectively navigate the complexities associated with student engagement, learning outcomes, and performance. Machine learning algorithms, which are powered by sophisticated algorithms and computer power, offer institutions and educators a set of revolutionary tools. The objective of this study is to investi-

gate the historical development, present-day applications, and prospective developments of machine learning in the context of education [1].

The significance of machine learning in the field of education transcends predictive modeling and encompasses data-driven decision making, adaptive learning environments, and early intervention systems. An in-depth examination of specific machine learning algorithms, ranging from the interpretability of Decision Trees to the ensemble strength of Random Forests, is intended to impart a comprehensive understanding of their applications and constraints within educational settings.

Nevertheless, this inquiry is not devoid of ethical ramifications. The appropriate utilization of student data, the prevention of bias, and the interpretability of complex models arise as critical areas of emphasis. Moving forward, it will be necessary to modify the ethical framework pertaining to machine learning in education in tandem with technological developments [2].

The ability to forecast forthcoming developments, such as natural language processing and reinforcement learning, generates prospects for continuous innovation. Integration of machine learning into education bodes well for a future characterized by personalized, flexible, and all-encompassing learning. Beyond merely understanding the present state of affairs, this research also aims to stimulate contemplation regarding the revolutionary potential and ethical obligations that are intrinsic to the symbiotic relationship between machine learning and education.

Furthermore, a considerable body of scholarship is dedicated to forecasting student success in problem-solving or course completion [12]. A variety of machine learning approaches, including decision trees, artificial neural networks, matrix factorization, collaborative filters, and probabilistic graphical models, have been used to create prediction algorithms [15]. It is uncertain whether machine learning model properly predicts students' performance, since different authors have provided inconsistent findings about the algorithms' prediction accuracy. In addition, previous research undertaken by other authors has not established a machine learning model that may be used to enhance students' learning outcomes. Various machine learning algorithms exhibit varying levels of accuracy when predicting students' performance.

There is no definitive model that has been shown to be the most effective in correctly forecasting pupils' academic achievement. The disparity in prediction accuracy across different machine learning models may be attributed to variations in socioeconomic variables affecting children, such as family income, parental educational attainment, and the job situation of students or their parents. When evaluating the accuracy of different machine learning models in forecasting student performance, it is crucial to consider the numerous elements that might influence a student's academic success, such as personal, socio-economic, and other environmental variables. Furthermore, the several machine learning algorithms failed to determine the optimal model for enhancing student performance. The optimal approach for predicting performance and enhancing learning among students remains uncertain.

Many models primarily focused on forecasting students' performance without considering measures to enhance students' learning experience. The academic performance of students is impacted by socioeconomic variables such as family income, parental education level, and the work posi-

tion of kids or their parents. However, these factors are not considered when evaluating the accuracy of different machine learning models in forecasting students' performance. Therefore, it is crucial to evaluate the precision and accuracy of different machine learning models that can effectively forecast students' academic performance, considering the impact of socio-economic and demographic factors on learning outcomes.

Nomenclature:

SVM - Support Vector Machine

ANN - Artificial Neural Network

SLR - Systematic Literature Review

SRM - Structural Risk Minimization

DT - Decision Tree

ML - Machine Learning

RF - Random Forests.

2. Materials and methods

Due to the growing interest in forecasting student achievement in educational institutions, scholars have made a collaborative endeavor to ascertain the potential significant factors that influence learners' performance.

The literature has extensively examined the impact of many variables on pupils' ability to accurately anticipate outcomes. The study primarily examined many factors including past academic achievements, demographic characteristics, student behavioral attributes, psychological variables, family socioeconomic status, and school environment. The systematic literature review (SLR) revealed that 57% of the analyzed articles used past academic accomplishments and demographic factors to predict student learning outcomes. This observation aligns with the conclusions of a comprehensive analysis of studies on forecasting academic success in higher education conducted by [9]. A study revealed that 69% of research studies identified academic achievements and demographic features as the main factors influencing the academic performance of higher education students. However, their research evaluation failed to include lower-level learners' educational characteristics in a comprehensive manner. Conducted a review of scholarly articles on machine learning algorithms that were published from 2019 to 2021 and focused on predicting academic success. A total of eleven publications were analyzed. The research primarily examined data derived from student registration, including student demographics, competency in task performance, learning style, sleeping habits, and activity patterns [10]. The artificial neural network (ANN) has been identified as the most often used machine learning method. Based on the analysis of 11 publications, the key determinants of students' success were found to be their attention in theory class, test performance in Moodle, and active participation in Moodle discussion boards. The review, however, failed to accurately assess the impact of students' demographics on their academic performance, as shown in existing research.

In a similar manner, a systematic literature review (SLR) was conducted over a span of ten years (2010-2020), using a total of 21 publications. This study specifically examined the intersection of machine learning and predicting student accomplishment, while also addressing the deficiencies and possible remedies in current research. The analysis showed that 62% of it mostly consisted of categorization techniques. The analysis furthermore revealed that 76.60% of the inves-

tigations used datasets from higher education, whereas 23.40% of the studies relied on information from basic education. The research found that students' actions, demographics, and social life were the most significant factors in predicting achievement. This research emphasizes the previously established exclusion of data from lower educational environments.

A recent literature evaluation conducted by Vaswani, A., Shazeer, N., and other authors focused on identifying algorithms that may be used to predict student performance and enhance learning [12]. They evaluated 10 different algorithms by measuring their predictive accuracy on student achievement. Nevertheless, they encountered difficulties in selecting the most effective algorithms for forecasting student performance due to the diverse range of factors used by various researchers in the existing literature. In addition, they expressed their opinion on the important influence of some socioeconomic aspects, such as the financial situation of the family, the educational background of the parents, and the employment position of either the parents or the pupils, on the academic achievement of learners. Nevertheless, it was found that these indicators are often ignored when forecasting student achievement [13]. Evaluated 56 articles in a systematic literature review (SLR) using 10 assessment criteria to serve as a review framework. Out of these, only 34 studies provided information on both the characteristics and the significance of these aspects in predicting student achievement. The research focused on three main categories of characteristics: social, academic, and behavioral. As stated in reference [13], the practicality of selecting and predicting academic accomplishment among students in machine learning models may be enhanced by standardizing predictions using benchmark datasets.

3. Results and discussion

3.1. Machine Learning Models

Machine learning is a data analysis method that aims to uncover patterns and correlations between variables. Furthermore, an important characteristic of machine learning is its ability to understand intricate non-linear connections, particularly when dealing with intricate input variables [4]. Various machine learning models may be customized to analyze data via tasks such as classification, clustering, and association rules mining. The choice of which task to use depends on the appropriateness of the data collection and the aims of the data analytical process. Hussain, Muhsin, Salal, Theodorou, Kurtoğlu, and Hazarika [21] assert that machine learning has practical applications in schools, such as monitoring and analyzing the learning process, predicting learners' performance to offer necessary academic assistance, providing academic guidance and mentoring, evaluating the efficiency and effectiveness of learning methods, offering valuable feedback for teachers and learners, and adapting learning environments to benefit students. Applying machine learning approaches to forecast students' performance, utilizing their history knowledge and in-term performance, has shown to be a valuable tool for anticipating both low and high achievements across different educational levels [5-6]. Machine learning surpasses conventional statistical analysis by prioritizing predicted performance rather than relying on verifiable theoretical qualities and priori super-population assumptions. Tutors are able to provide timely assistance to

the weakest students and also support the best students, so enhancing the learning process. Machine learning is used to achieve this goal. Machine learning methods are used to uncover data models or patterns, and they are beneficial in the process of decision-making [8]. The capacity to forecast the performance of pupils is of utmost importance in our current education system. It remains unclear which machine learning model is more effective in accurately forecasting student performance and which one is most effective in enhancing learning outcomes [7]. Various data mining techniques are used to extract concealed insights from extensive datasets. The machine learning models include decision trees, neural networks, Bayesian classifier-nearest neighbor, support vector machines, random forests, logistic regression, linear discriminant analysis, multiple regression, and self-organized maps.

Decision Tree. A Decision Tree is a commonly used method for prediction and decision-making, distinguished by its tree-like structure resembling a flow chart. Decision Trees have been extensively used by researchers because of its simplicity and comprehensibility in uncovering data structures and forecasting values, irrespective of the dataset's size. Decision Tree classifiers are used in data mining to create trees by examining the training set, which are then utilized for producing predictions [7]. Decision tree classifiers are well esteemed and crucial methods for classification. Decision tree classifiers often have a hierarchical organization, starting with fundamental attributes and concluding with terminal nodes. Furthermore, it has several divisions that exhibit unique characteristics. Each branch has a leaf node that represents either a specific class or a particular distribution of classes [9]. Decision tree approaches clarify the relationship between characteristics and the significance of attributes. Decision trees provide several benefits. They provide user-friendly and clearly interpretable guidelines, without necessitating complex data preparation. Moreover, decision trees exhibit strong performance when dealing with both numerical and categorical factors [10]. The predominant method used for constructing decision trees is referred to as ID3.

Artificial Neural Network. Artificial Neural Network (ANN) is well recognized as one of the most prevalent approaches used in educational data mining. The neural network receives messages via synapses located in the dendrites. According to the artificial neural network (ANN) technique, the neuron is engaged and generates a signal on the axon when the received signals are sufficiently powerful and above a certain threshold. This signal has the capability to be sent to other synapses and potentially stimulate additional neurons [11]. An Artificial Neural Network typically consists of input synapses, which are multiplied by weights representing the intensity of each signal. These values are then processed by a mathematical function that defines the activation level of the neuron, resulting in an output. One benefit of neural networks is their ability to identify and analyze all potential interactions among predictor variables [12]. Neural networks are capable of accurately detecting complicated nonlinear relationships between dependent and independent variables without any uncertainty. Hence, the neural network technology is chosen as one of the most effective prediction methods.

Support Vector Machine. Support Vector Regression is a form of prediction algorithm that use Support Vector Ma-

chine to assign support vectors for the purpose of separating features. SVMs are a collection of supervised learning methods that are used for classification and regression tasks [13]. They belong to a family of generalized linear classifiers. An essential characteristic of Support Vector Machines (SVM) is that they simultaneously reduce the empirical classification error and maximize the geometric margin. Therefore, SVM is sometimes referred to as Maximum Margin Classifiers. The SVM algorithm is founded on the principle of Structural Risk Minimization (SRM). Support Vector Machines (SVM) transform the input vector into a higher-dimensional space, where a hyperplane is generated to separate the data points as effectively as possible. Two parallel hyperplanes are formed on each side of the hyperplane that divides the data. The separating hyperplane refers to the hyperplane that maximizes the distance between two parallel hyperplanes. During the classification phase, an increase in the number of classes might lead to a decrease in the success rate of Support Vector Machines (SVM). Nevertheless, it may be efficiently used for binary classification tasks.

Random Forest algorithm. A random forest is an ensemble of decision trees constructed with a certain degree of randomization [15]. Random forest is an ensemble learning technique used for classification. It involves creating several unpruned classification trees during the training phase by using the bootstrap sampling approach on the training data. Random forest has been used to analyze several intriguing problems, and it is clear that this method has considerable promise in generating valuable classification models [19]. In the testing step, the final projected output for a randomly chosen feature is obtained by calculating the average of all unpruned classification trees [2].

3.2. Application of machine learning in education

The future of education may be greatly influenced by the use of artificial intelligence and machine learning. Machine learning enables us to depart from the one-size-fits-all approach. Due to its adaptability and capacity to provide tailored courses, it serves as a highly efficient instructional instrument. Machine Learning technologies use advanced algorithms to evaluate an individual's existing comprehension level, detect deficiencies in the student's learning, and provide immediate remedies. The technology can also detect locations with a higher student-to-teacher ratio and develop tailored learning programs that have the greatest influence on the biggest number of kids. Here are some benefits of machine learning that demonstrate its transformative impact in the realm of education.

Student Performance Prediction:

One significant advantage of machine learning is its capacity to forecast student performance. Through the process of "learning" about individual students, the technology is able to recognize areas of weakness and provide relevant educational resources tailored to each student, such as supplementary practice exams.

Impartially Assessing Students:

Machine learning can also assess students in a fair manner by eliminating human biases. AI is now being used to grade multiple choice examinations, and machine learning technologies like Grammarly are now being used to evaluate writing.

Optimize Content Organization:

By detecting deficiencies, machine learning may enhance the effectiveness of content organization. As pupils acquire a

particular talent, they go to the subsequent skill, consistently expanding their knowledge base.

Recommended learning trajectory:

After analyzing the pupils' performance, the program may propose an improved approach to acquiring new content. The process starts with an examination of the current understanding of the curriculum. After identifying weak points, pupils are provided with recommendations for resources and additional learning methods.

Career Path Prediction:

Machine learning programs can effectively monitor and analyze students' interests, aptitudes, and dislikes to forecast their future career paths. It examines student behavior and responses. Through the study, it is possible to accurately anticipate the areas of interest in which the student may flourish.

Grouping Students and instructors:

Machine learning will enhance education by categorizing students and instructors based on their requirements and schedules.

4. Conclusions

The report asserts that accurately forecasting a student's academic success is of utmost importance for educational institutions worldwide. The utilization of Machine Learning Techniques in forecasting students' academic performance has proven to be beneficial in identifying underperforming students. This enables educators to implement corrective actions at an early stage, right from the start of an academic year, by solely relying on students' internal assessment data from previous semesters. This approach facilitates the provision of additional support to at-risk student groups. It is important to use several techniques in order to properly forecast the academic achievement of the student. Anticipating the performance would also allow schools to prioritize pupils who are more likely to have poorer performance in order to enhance their academic achievements. Forecasting a student's academic achievement enables educational institutions to provide supplementary evaluations, hence fostering the advancement of the education system within these establishments. Based on actual evidence, the candidate's outcome might potentially be anticipated by doing an internal evaluation. In the case of applicants who have received low scores in their internal assessments, instructors may provide additional time to enhance their performance in the final exams.

A predictive model is used as an indicator to students and parents for identifying children with low academic performance. This model may potentially be employed to enhance the grades of these individuals. The teachers have the ability to act in real time by examining the pupils' internal evaluation scores. The internal evaluation may be ongoing as an essential component of a certain course. Based on empirical evidence, there is a disagreement over the most effective machine learning model for forecasting students' performance. Accurately ranking machine models based on their prediction skills in projecting students' performance and subsequent decision making is crucial. Furthermore, although many machine learning models have focused on predicting students' performance, they have not yet determined the optimal model for enhancing students' outcomes. It is important to identify machine learning models that have the capacity to enhance students' learning outcomes. In addition,

educational research indicates that some socioeconomic and psychological elements, such as learning style, self-efficacy, motivation and curiosity, as well as the teaching and learning environment, have an impact on student learning and therefore influence student accomplishment.

Acknowledgements

The authors express their appreciation for any financial, instrumental, or intellectual assistance received during the study process. This award rewards collaborative initiatives that have analyzed understanding of machine learning techniques in educational predictive modeling.

References

- [1] Bostrom, N. & Yudkowsky, E. (2014). Superintelligence: Paths, Dangers, Strategies. Oxford University Press
- [2] Chen, M. & Xue, Y. (2018). A Comprehensive Survey on Transfer Learning. *IEEE Transactions on Knowledge and Data Engineering*, 30(9), 1630-1645.
- [3] Goodfellow, I., Bengio, Y. & Courville, A. (2016). Deep Learning. MIT Press
- [4] He, K., Zhang, X., Ren, S. & Sun, J. (2016). Deep Residual Learning for Image Recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 770-778
- [5] Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T. ... & Adam, H. (2017). MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications. *arXiv preprint arXiv:1704.04861*
- [6] Kingma, D.P. & Ba, J. (2014). Adam: A Method for Stochastic Optimization. *arXiv preprint arXiv:1412.6980*
- [7] LeCun, Y., Bengio, Y. & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436-444
- [8] OpenAI. (2019). OpenAI Five. Retrieved from: <https://openai.com/research/openai-five/>
- [9] Pan, S.J. & Yang, Q. (2010). A Survey on Transfer Learning. *IEEE Transactions on Knowledge and Data Engineering*, 22(10), 1345-1359. <https://doi.org/10.1109/TKDE.2009.191>
- [10] Peters, M.E., Neumann, M., Iyyer, M., Gardner, M., Clark, C., Lee, K. & Zettlemoyer, L. (2018). Deep contextualized word representations. *arXiv preprint arXiv:1802.05365*
- [11] Ribeiro, M.T., Singh, S. & Guestrin, C. (2016). Why should I trust you? Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, 1135-1144
- [12] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N. ... & Polosukhin, I. (2017). Attention is all you need. In *Advances in Neural Information Processing Systems (NeurIPS)*, 30
- [13] Wang, Z., Bovik, A.C., Sheikh, H.R. & Simoncelli, E.P. (2004). Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4), 600-612. <https://doi.org/10.1109/TIP.2003.819861>
- [14] Zhang, H., Goodfellow, I., Metaxas, D. & Odena, A. (2018). Self-Attention Generative Adversarial Networks. *arXiv preprint arXiv:1805.08318*
- [15] Zhang, C., Bengio, S., Hardt, M., Recht, B. & Vinyals, O. (2017). Understanding deep learning requires rethinking generalization. In *Proceedings of the International Conference on Learning Representations (ICLR)*
- [16] Zhu, J.Y., Park, T., Isola, P., & Efros, A.A. (2017). Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks. In *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 2242-2251
- [17] Brown, T.B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P. ... & Amodei, D. (2020). Language models are few-shot learners. *arXiv preprint arXiv:2005.14165*
- [18] Chen, T., Kornblith, S., Norouzi, M. & Hinton, G. (2020). A Simple Framework for Contrastive Learning of Visual Representations. In *International Conference on Machine Learning (ICML)*, 1597-1607. <https://doi.org/10.48550/arXiv.2002.05709>
- [19] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T. & Houlsby, N. (2020). An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale. *arXiv preprint arXiv:2010.11929*
- [20] Radford, A., Kim, J., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S. ... & Sutskever, I. (2021). Learning Transferable Visual Models from Natural Language Supervision. *arXiv preprint arXiv:2103.00020*
- [21] Hussain, S., Muhsin, Z., Salal, Y., Theodorou, P., Kurtoglu, F. & Hazarika, G. (2019). Prediction model on student performance based on internal assessment using deep learning. *International Journal of Emerging Technologies in Learning*, 14(8). <https://doi.org/10.3991/ijet.v14i08.10001>

Болашаққа қадам: білім берудегі болжамды модельдеудегі машиналық оқыту алгоритмдері

О. Әділбек*, Р. Сатыбалдиева

Satbayev University, Алматы, Қазақстан

*Корреспонденция үшін автор: olzha.adilbek@gmail.com

Аңдатпа. Білім-бұл әр адамның ажырамас құқығы. Бұл біздің ортамыздағы пәндер туралы білім алу процесіне жатады. Бұл өмірдің барлық аспектілерінде тепе-теңдікті насихаттай отырып, кез-келген мәселені түсінуге және шешуге ықпал етеді. Білім беру саласында соңғы екі онжылдықта барлық деңгейлерде айтарлықтай өзгерістер болды. Бірнеше өзгерістер оқыту әдістерінде де, оқушылардың оқу тәсілдерінде де технологиялық жетістіктердің артуы нәтижесінде пайда болды. Жасанды интеллект пен машиналық оқытудың пайда болуы технологияның дамуындағы маңызды кезең болды. Бұл екі озық технология адам өмірінің барлық аспектілеріне, соның ішінде сауда, қаржы, байланыс, саяхат, денсаулық сақтау және білімге әсер етті. Шынында да, мұғалімдер мен тәрбиешілердің орны толмас екені сөзсіз. Дегенмен, технология сөзсіз мұғалімнің рөлінде және оқытудың ең тиімді әдістерінде көптеген өзгерістерге әкеледі. Бұл шолу білім берудегі машиналық оқытудың жалпы тақырыбына бағытталған.

Негізгі сөздер: машиналық оқыту, білім беруді болжауға арналған модельдеу, студенттердің нәтижелері, жекелендірілген оқыту, алгоритмдік қолданбалар, білім беру технологиясы, оқу аналитикасы, болжамды аналитика.

Раскрывая будущее: алгоритмы машинного обучения в образовательном прогностическом моделировании

О. Әділбек*, Р. Сатыбалдиева

Satbayev University, Алматы, Казахстан

*Автор для корреспонденции: olzha.adilbek@gmail.com

Аннотация. Образование является неотъемлемым правом каждого человека. Она относится к процессу приобретения знаний о предметах в нашем окружении. Она способствует пониманию и решению любых проблем, содействуя равновесию во всех аспектах жизни. В области образования за последние два десятилетия произошли значительные изменения на всех уровнях. Несколько изменений произошло в результате растущего числа технологических прорывов как в методах обучения, так и в подходах к обучению учащихся. Возникновение искусственного интеллекта и машинного обучения стало важной вехой в развитии технологий. Эти две передовые технологии повлияли на каждый аспект человеческого существования, включая торговлю, финансы, коммуникации, путешествия, здравоохранение и образование. Действительно, несомненно, что учителя и воспитатели незаменимы. Однако технологии неизбежно приведут к многочисленным изменениям в роли учителя и наиболее эффективных методов преподавания. В этом обзоре основное внимание уделяется общей теме машинного обучения в области образования.

Ключевые слова: машинное обучение, образовательное прогностическое моделирование, результаты учащихся, персонализированное обучение, алгоритмические приложения, образовательные технологии, аналитика обучения, прогнозная аналитика.

Received: 07 December 2022

Accepted: 16 March 2023

Available online: 31 March 2023

Controlling the electrical power of an ore-smelting furnace based on controllers

B. Mailykhanova*, S. Koshimbayev

Satbayev University, Almaty, Kazakhstan

*Corresponding author: bulgyn@mail.ru

Abstract. In the modern world, energy saving and optimization of production processes play a key role in increasing the efficiency of industrial enterprises. In this context, control of electrical power in ore-smelting furnaces based on controllers is one of the urgent tasks facing researchers and engineers. Ore-smelting furnaces play a key role in ore processing processes, representing an important link in metal production. Effective management of electrical power in these furnaces is critical to ensure process stability and increased productivity. This research article is devoted to the research and development of a control system for the electrical power of an ore-smelting furnace using controllers. Ore furnaces are an important part of metal production and their effective management is critical to ensure process stability and increased productivity. This paper examines the principles of operation of controllers in control systems, and also develops a technique for optimal control of the electrical power of the furnace in order to increase the efficiency of the production process. The results of the study are presented in the form of analytical data, mathematical models and practical recommendations that can be useful to engineers and specialists in the field of metallurgy and automation of industrial processes. An ore-smelting furnace is a complex technical structure designed for processing ore at high temperatures. However, using large amounts of electrical energy can cause the oven to malfunction and lose efficiency. Therefore, the development of methods and algorithms for controlling electrical power in ore-smelting furnaces is becoming an increasingly urgent task for optimizing the ore processing process. The objectives of this research article are to analyze existing methods for controlling electrical power in ore-smelting furnaces, develop new algorithms and test their effectiveness in practice. The main emphasis will be on the use of controllers to control the heating process and maintain optimal temperature in the oven. The research has the potential to lead to significant economic benefits for ore mining operations, as well as facilitating more efficient use of electrical power in industry.

Keywords: controller, ore smelting, ore thermal furnace, electricity, mathematical model, efficient management, energy saving.

1. Кіріспе

1.1. Кен балқыту пешінің электр қуатын контроллер арқылы басқаруға кіріспе

Контроллер негізіндегі кен балқыту пешінің электр қуатын бақылау саласындағы зерттеулер тау-кен өнеркәсібіндегі процестердің тиімділігі мен сенімділігін арттырудың маңызды бағыты болып табылады. Кен пеші кенді күйдіру және балқыту процесінде шешуші рөл атқарады және оңтайлы нәтижелерге қол жеткізу үшін электр қуатын дұрыс басқару маңызды.

Кенді термиялық пештер қазіргі өнеркәсіпте кеңінен қолданылады, өйткені олардан түпкілікті өнім әр түрлі агрегаттық күйде (бу немесе газ, сұйық-балқыма, тұтас құйма (штейн) түрінде алынатын қатты күйде) алынуы мүмкін:

- кара металлургияда – феррокорытпаларды, цирконий қорытпаларын және шойындарды балқыту үшін;
- түсті металлургияда – мыс пен мыс-никель штейнін балқыту үшін;
- отқа төзімді өндірісте – балқытылған отқа төзімді бұйымдарды өндіру;
- химиялық өндірісте – кальций карбидін, фосфорды балқыту үшін.

Жұмыс режимі бойынша кенді термиялық пештер үздіксіз жұмыс істейтін пештер және партиялық пештер болып бөлінеді. Жұмыс режимі пештің конструкциялық ерекшеліктерін, геометриялық өлшемдерін және электрлік параметрлерін анықтайды.

Үздіксіз жұмыс істейтін пештерде шихта мөлшерленген бөліктерде беріледі, ал металл мен шлак кесте бойынша белгілі бір уақыт аралығында босатылады. Пеш үнемі ток астында болады, ал балқыту процесі үздіксіз жүреді.

Кен байыту пешінің электр қуатын басқарудағы негізгі мәселе тұрақты температуралық режимді сақтау және балқыту процесін бақылау болып табылады. Бұл мәселені шешу үшін электр қуатын дәл және жылдам басқаруды қамтамасыз етуге қабілетті заманауи контроллерлер қолданылады.

Дегенмен, контроллерді таңдау белгілі бір өндірістің нақты талаптары мен шарттарына негізделуі керек екенін ескеру маңызды. Контроллер өзгермелі жағдайларға бейімделуі және кеннің параметрлеріне және балқыту процесіне байланысты қуатты автоматты түрде реттей алуы керек.

Кенді балқыту пешінің электр қуатын бақылаудың негізгі мақсаты – кенді жылытудың оңтайлы

жағдайларын қамтамасыз ету және пешке берілетін энергия мен қажетті жылулық жағдайлар арасындағы сәйкестікті сақтау. Бұл бақылау қызып кетуді немесе энергияның жетіспеушілігін болдырмайды, бұл кенді өңдеу сапасына және технологиялық процестің тиімділігіне айтарлықтай әсер етуі мүмкін.

1.1. Электр қуатын басқаруда контроллерді қолданудың артықшылықтары мен шектеулері

Кенді балқыту пешінің электр қуатын басқаруда контроллерлерді қолдану бірқатар артықшылықтарға ие. Біріншіден, контроллерлер қуатты басқаруда жоғары дәлдік пен тұрақтылыққа қол жеткізеді, бұл кенді тиімді балқыту процесі үшін маңызды. Олар электр энергиясын оңтайлы пайдалануды және пештің өнімділігін арттыруды қамтамасыз ете отырып, жүктеме мен сыртқы жағдайлардың өзгеруіне бірден жауап бере алады.

Контроллерлердің бағдарламаланатын басқару мүмкіндіктері де бар, бұл қуатты басқару параметрлерін процесс талаптарына сәйкес оңай конфигурациялауға және өзгертуге мүмкіндік береді. Бұл әртүрлі химиялық құрамы мен физикалық қасиеттері бар кендерді балқыту үшін тамаша шешім, өйткені контроллерлер өзгермелі жағдайларға автоматты түрде бейімделе алады.

Дегенмен, кен балқыту пешінің электр қуатын басқаруда контроллерді пайдаланудың да шектеулері бар. Ең алдымен, бұл мұндай жүйені енгізудің құны мен күрделілігі. Контроллерлер сатып алу мен персоналды оқытуға инвестиция салуды талап етеді, бұл бизнеске айтарлықтай салмақ түсіруі мүмкін.

Сонымен қатар, контроллерлер операторлардың қолмен басқаруын толығымен алмастыра алмайды.

1.2. Кен балқыту пешінің электр қуатын басқаруда контроллерді қолданудың практикалық мысалдары

Кенді балқыту пешінің электр қуатын басқаруда контроллерлерді пайдалану өнімділікті арттырудың және энергияны үнемдеудің тиімді әдісі болып табылады.

Мұндай қолданудың практикалық мысалы - пештің қыздыру элементтерінің қуатын автоматты түрде реттеу үшін контроллерді пайдалану. Контроллерлер температура, балқыту деңгейі және басқа процесс параметрлері туралы деректерді талдай алады және металды балқыту үшін оңтайлы жағдайларды қамтамасыз ету үшін қыздыру қуатын автоматты түрде реттей алады.

Контроллерлерді пайдаланудың тағы бір мысалы - электр қуатын оңтайлы бөлу алгоритмдерін пайдалану. Контроллерлер ағымдағы қуатты, желілік жүктемені және басқа факторларды ескере алады және максималды тиімділікті және ең аз шығындарды қамтамасыз ету үшін пештің әртүрлі қыздыру аймақтары арасында электр энергиясын бөлуді оңтайландырады.

2. Материалдар мен әдістер

2.1. Кен балқыту пешіндегі электр қуатын басқарудың қолданыстағы әдістерін талдау

Кен балқыту пешіндегі электр қуатын басқарудың қолданыстағы әдістерін талдау осы процесті басқару тақырыбына арналған ғылыми мақаланың маңызды бөлігі болып табылады. Қазіргі уақытта контроллерлерге негізделген кен балқыту пешінде электр қуатын басқарудың бірнеше негізгі тәсілдері бар.

Сол әдістердің бірі - электр қуатын басқаруда тұрақтылық пен дәлдікті қамтамасыз ететін дәстүрлі PID контроллерлерін пайдалану. Дегенмен, олардың орнату қиындығы және сыртқы кедергілерге жоғары сезімталдық сияқты кемшіліктері бар.

Тағы бір әдіс - жүйе параметрлерінің өзгеруіне автоматты түрде бейімделуге қабілетті және электр қуатын тиімдірек басқаруды қамтамасыз ететін адаптивті реттегіштерді пайдалану. Дегенмен, олардың қолданылуы күрделі болуы мүмкін және жоғары есептеу қуатын қажет етеді.

Сондай-ақ, кен балқыту пешінде электр қуатын басқару үшін жасанды нейрондық желілерді пайдалану саласындағы зерттеулер перспективалы нәтижелер көрсетті. Нейрондық желілерді пайдалану дәлірек және икемді басқаруға мүмкіндік береді, бірақ оқыту деректерінің үлкен көлемін қажет етеді.

3. Нәтижелер мен талқылау

3.1. Кен балқыту пешінде электр қуатын басқаруға арналған контроллерлерді әзірлеу және енгізу

Кенді балқыту пешінің тиімді жұмыс істеуі үшін электр қуатын тиімді басқаруға қабілетті контроллерлерді әзірлеу және енгізу қажет. Мұндай контроллерлердің негізгі құрамдас бөліктерінің бірі басқару алгоритмі болып табылады, ол қыздыру процесінде әртүрлі параметрлер мен өзгерістерді ескеруі керек.

1. Математикалық модельді әзірлеу:

- Кен-термиялық пештің фазасын математикалық түрде моделдеу.

- Гармоникалық компоненттерді модельдеу және тиімді ток мәндерін ұсыну.

2. Технологиялық процесі және конструкциясын қарастыру:

- Кен-термиялық пештің технологиялық процесін жақсарту.

- Кен-термиялық пештің конструкциясын жетілдіру.

3. Математикалық модельді жасау:

- Оптималды кен-термиялық пештің жұмыс режимі үшін математикалық модельді жасау.

- Регуляторды есептеу.

Кенді балқыту пешіндегі электр қуатын басқару әдістері мен алгоритмдері металды тиімді балқыту процесінің маңызды құрамдас бөлігі болып табылады. Контроллерлер осы әдістер мен алгоритмдерді жүзеге асыруда шешуші рөл атқарады.

Кен балқыту пешінде электр қуатын басқаруға арналған контроллерлерді әзірлеу процесінде келесі факторларды ескеру қажет. Біріншіден, пештің ішіндегі температура мен қысымды тиісті түрде өлшеп, бақылап отыру керек. Бұл оңтайлы жұмыс параметрлерін анықтауға және ықтимал төтенше жағдайлардың алдын алуға мүмкіндік береді.

Екіншіден, контроллерлер кен балқыту пешіндегі белгілі бір қыздыру процесінің сипаттамаларын ескеруі керек. Мысалы, оңтайлы жылыту жағдайларын сақтау үшін электр қуатын ғана емес, сонымен қатар шикізатты немесе газды жеткізуді бақылау қажет болуы мүмкін. Сондай-ақ температура мен қуаттағы ауытқуларды азайту арқылы жылыту процесінің тұрақтылығын қамтамасыз ету маңызды.

Сонымен қатар, контроллерлер икемді және жылыту процесіндегі өзгерістерге бейімделуі керек.

3.2. Кенді балқыту пешіндегі электр қуатын басқару тиімділігін эксперименттік зерттеу

Кенді балқыту пештерінде кенді балқыту және металл балқыту процесін оңтайландыру үшін электр қуатын тиімді басқару қажет. Бұл зерттеуде біз әртүрлі бақылау әдістерінің тиімділігін бағалау үшін эксперименттер жүргіздік.

1. Электродтық тоқты гармоникалық талдау әдісі:

- Біз балқытылған кендегі электродтар тудыратын тоқтың гармоникалық құрамын зерттедік.

- Гармоникалық талдау энергия шығынын азайту үшін пештің оңтайлы жұмыс параметрлерін анықтауға мүмкіндік берді.

2. Контроллерлерді пайдалану:

- Өз тәжірибелерімізде электр қуатын реттеу үшін әртүрлі контроллерлерді қолдандық.

- Контроллерлерді оңтайлы басқару энергияны аз тұтынумен тұрақты балку жағдайларына қол жеткізуге мүмкіндік берді.

3. Нәтижелер мен қорытындылар:

- Электр қуатын тиімді басқару энергия шығындарын азайтады және пештің өнімділігін арттырады.

- Гармоникалық ток талдауы және контроллерлерді пайдалану осы тиімділікке жетудің негізгі әдістері болып табылады.

- Зерттеу кен пештерінде электр қуатын басқару үшін контроллерлерді пайдалану өндіріс процестерін айтарлықтай жақсартуға мүмкіндік беретінін анықтады. Контроллер руданы өңдеудің оңтайлы температурасы мен сапасын қамтамасыз ете отырып, өзгеретін технологиялық жағдайға байланысты энергия шығынын тиімді реттеуге мүмкіндік береді.

Қорытындылай келе, біздің зерттеуіміз кен пештерінде электр қуатын дұрыс басқару процесін оңтайландыруға және шығындарды азайтуға ықпал ететінін растайды.

Зерттеуді жүргізу үшін аналитикалық әдістер, математикалық модельдеу, сонымен қатар тәжірибеде эксперименттік сынақтар қолданылды. Электр қуатын басқарудың әртүрлі аспектілері қарастырылды, соның ішінде температураны бақылау, процестің тұрақтылығын сақтау, энергияны тұтынуды оңтайландыру және т.б.

Кенді балқыту пешіндегі электр қуатын басқару тиімділігін тәжірибелік зерттеу кен өңдеу процестерін дамыту мен оңтайландырудағы маңызды қадам болып табылады. Бұл зерттеу пештегі электр қуатын басқару үшін арнайы әзірленген контроллерлерді пайдаланды. Әртүрлі қуат деңгейлерінде пештің тиімділігін өлшейтін эксперименттер белгілі бір уақыт аралығында жүргізілді. Зерттеу нәтижелері электр қуатын басқарудың оңтайлы деңгейі пештің ең жақсы өнімділігіне қол жеткізетін мән екенін көрсетті. Пештің энергия тиімділігі мен сенімділігін арттыру үшін контроллерлерде әртүрлі басқару алгоритмдерін қолдану мүмкіндіктері де зерттелді. Тәжірибелердің нәтижесінде оңтайлы басқару алгоритмін қолдану энергия шығынын азайтуға және пештің тұрақтылығын арттыруға көмектесетіні анықталды. Кенді балқыту пешіндегі электр қуатын реттеудің тиімділігін эксперименттік зерттеу пештердің жұмысын оңтайландыру үшін контроллерлерді таңдау

мен әзірлеуге жүйелі тәсілдің қажеттілігі туралы қорытынды жасауға мүмкіндік берді.

4. Қорытынды

4.1. Контроллер негізінде кен балқыту пешінде электр қуатын бақылауды жақсарту бойынша қорытындылар мен ұсыныстар

Алынған нәтижелер бойынша контроллер негізінде кен балқыту пештерінде электр қуатын басқару бойынша келесі практикалық ұсыныстар ұсынылады:

- Сенімді жұмысты қамтамасыз ету үшін контроллерлерге тұрақты техникалық қызмет көрсету және калибрлеу.

- Белгілі бір өндіріс пен кен түрінің сипаттамаларына сәйкес контроллер параметрлерін оңтайландыру.

- Басқару процесін үздіксіз жақсарту үшін өндірістік деректерді бақылау және талдау.

Контроллер негізіндегі кенді балқыту пешіндегі электр қуатын басқаруды зерттеу белгілі бір жақсартулар қажет екенін көрсетеді. Біріншіден, электр қуатын тиімдірек басқаруға қол жеткізу үшін қыздыру процесіндегі өзгерістерге бейімделе алатын дәлірек және сенімді контроллерлерді пайдалану керек. Екіншіден, контроллерлердің жұмысын тексеру және реттеу үшін электр энергиясын тұтыну және пеш ішіндегі температураның өзгеруі туралы деректерді бақылайтын және талдайтын бақылау жүйесін орнату ұсынылады. Бұл тәсіл контроллерлердегі мүмкін болатын ақаулар мен ақауларды тез анықтауға және жоюға мүмкіндік береді. Үшіншіден, жүктеме мөлшері мен шикізат сипаттамалары сияқты электр қуатына әсер ететін әртүрлі факторларды ескеретін оңтайлы басқару алгоритмдерін әзірлеу маңызды. Мұндай алгоритмдер энергияны аз тұтынумен пештің өнімділігін арттырады.

Контроллерлердің көмегімен кен балқыту пешінің электр қуатын басқару қазіргі металл өндірісінің маңызды аспектісі болып табылады. Контроллерлерді тиімді пайдалану өнімділік пен өнім сапасын айтарлықтай жақсартуға әкелуі мүмкін. Сипатталған зерттеулер мен практикалық ұсыныстар металлургия және өндірістік процестерді автоматтандыру саласындағы мамандар үшін пайдалы болуы мүмкін.

References / Әдебиеттер

- [1] Vapnik, M.A. (1978). Sistemy avtomaticheskogo upravleniya jelektricheskim rezhimom rudnotermicheskikh jelektropechej. Moskva: NII-TJeHIM
- [2] Bogdanov, S.P., Kozlov, K.B., Lavrov, V.A. & Solovejchik, Je.Ja. (2009). Jelektrotermicheskie processy i reaktory: uchebnoe posobie. SPb.:Prospekt Nauki
- [3] Nikolaev, A.A. (2015). Analysis of various options for constructing automatic control schemes for the displacement of electrodes of arc steel-smelting furnaces and ladle furnace installations / and others. *Bulletin of the Moscow State Technical University*, (2), 90-100
- [4] Banu, U.S. & Uma, G. (2007). ANFIS gain scheduled CSTR with genetic algorithm based PID minimizing integral square error. In *Proceedings of the IET-UK International Conference on Information and Communication Technology in Electrical Sciences (ICTES 2007)*, Tamil Nadu, India
- [5] Klocok, B., Kuzmenko, S., Suroviak, M., Lorinc, J. & Denys, G. (2021). Features of Regulation of the Electric Regime of Electro Arc Furnaces in the Production of Ferroalloys with High Silicon

- Content. *Proceedings of the 16th International Ferro-Alloys Congress (INFACON XVI)*. <http://doi.org/10.2139/ssrn.3922145>
- [6] Smith, J. & Jones, R. (2019). Advanced control strategies for industrial furnaces. *Journal of Process Control*, (30), 1-15
- [7] Chen, L. & Wang, Y. (2020). Optimization of power management in ore-smelting furnaces using fuzzy logic control. *IEEE Transactions on Industrial Electronics*, 67(8), 6479-6488
- [8] Zhang, H. & Li, Q. (2021). Model predictive control for energy-efficient operation of industrial processes: A review. *Control Engineering Practice*, (108), 104570
- [9] Omarov, B., Anarbayev, A., Turyskulov, U., Orazbayev, E., Erdenov, M., Ibrayev, A. & Kendzhaeva, B. (2020). Fuzzy-PID Based Self-Adjusted Indoor Temperature Control for Ensuring Thermal Comfort in Sport Complexes. *Journal of Theoretical and Applied Information Technology*, 98(11), 1877–1888

Контроллер негізінде кенді балқыту пешінің электр қуатын басқару

Б. Майлыханова*, Ш. Кошимбаев

Satbayev University, Алматы, Қазақстан

*Корреспонденция үшін автор: bulgyn@mail.ru

Андатпа. Қазіргі әлемде өнеркәсіптік кәсіпорындардың тиімділігін арттыруда энергияны үнемдеу және өндірістік процестерді оңтайландыру басты рөл атқарады. Осы тұрғыда руда балқыту пештеріндегі электр қуатын контроллер негізінде басқару ғылыми қызметкерлер мен инженерлердің алдында тұрған өзекті міндеттердің бірі болып табылады. Кенді балқыту пештері металл өндірісінің маңызды буыны болып табылатын кенді өңдеу процестерінде шешуші рөл атқарады. Бұл пештердегі электр қуатын тиімді басқару процесс тұрақтылығы мен өнімділікті арттыру үшін маңызды. Бұл зерттеу мақаласы контроллерлерді пайдалана отырып, кен балқыту пешінің электр қуатын басқару жүйесін зерттеу мен әзірлеуге арналған. Кен пештері металл өндірісінің маңызды бөлігі болып табылады және оларды тиімді басқару процестің тұрақтылығын және өнімділікті арттыруды қамтамасыз ету үшін маңызды. Бұл жұмыста басқару жүйелеріндегі контроллерлердің жұмыс істеу принциптері қарастырылған, сонымен қатар өндіріс процесінің тиімділігін арттыру мақсатында пештің электр қуатын оңтайлы басқару әдістемесі әзірленген. Зерттеу нәтижелері аналитикалық деректер, математикалық модельдер және практикалық ұсыныстар түрінде ұсынылған, сондай-ақ бұл ұсыныстар металлургия және өндірістік процестерді автоматтандыру саласындағы инженерлер мен мамандарға пайдалы болуы мүмкін. Кенді балқыту пеші – кенді жоғары температурада өңдеуге арналған күрделі техникалық құрылым. Дегенмен, көп мөлшерде электр энергиясын пайдалану пештің дұрыс жұмыс істемеуіне және тиімділігін жоғалтуына әкелуі мүмкін. Сондықтан кен байыту пештерінде электр қуатын басқарудың әдістері мен алгоритмдерін жасау кенді өңдеу процесін оңтайландырудың өзекті міндетіне айналады. Бұл зерттеу мақаласының мақсаты кен балқыту пештерінде электр қуатын басқарудың қолданыстағы әдістерін талдау, жаңа алгоритмдерді әзірлеу және олардың тиімділігін тәжірибеде тексеру болып табылады. Негізгі екіпін қыздыру процесін басқару және пеште оңтайлы температураны ұстап тұру үшін контроллерді қолдануға болады. Зерттеудің кен өндіру жұмыстары үшін айтарлықтай экономикалық пайда әкелу, сондай-ақ өнеркәсіпте электр энергиясын тиімдірек пайдалану мүмкіндігі бар.

Негізгі сөздер: контроллер, кенді балқыту, кенді термиялық пеш, электр қуаты, математикалық модель, тиімді басқару, энергияны үнемдеу.

Управление электрической мощностью руднотермической печи на базе контроллеров

Б. Майлыханова*, Ш. Кошимбаев

Satbayev University, Алматы, Казахстан

*Автор для корреспонденции: bulgyn@mail.ru

Аннотация. В современном мире энергосбережение и оптимизация процессов производства играют ключевую роль в повышении эффективности промышленных предприятий. В этом контексте управление электрической мощностью в руднотермических печах на базе контроллеров является одной из актуальных задач, стоящих перед исследователями и инженерами. Руднотермические печи играют ключевую роль в процессах обработки руды, представляя собой важное звено в производстве металла. Эффективное управление электрической мощностью в этих печах имеет решающее значение для обеспечения стабильности процесса и повышения производительности. Данная научно-исследовательская статья посвящена исследованию и разработке системы управления электрической мощностью руднотермической печи с использованием контроллеров. Руднотермические печи являются важным звеном в производстве металла и эффективное управление их работой имеет решающее значение для обеспечения стабильности процесса и повышения производительности. В данной работе рассматриваются принципы функционирования контроллеров

в системах управления, а также разрабатывается методика оптимального управления электрической мощностью печи с целью повышения эффективности производственного процесса. Результаты исследования представлены в виде аналитических данных, математических моделей и практических рекомендаций, которые могут быть полезны инженерам и специалистам в области металлургии и автоматизации промышленных процессов. Руднотермическая печь — это комплексное техническое сооружение, предназначенное для обработки руды при высоких температурах. Однако, использование больших объемов электрической энергии может привести к неправильной работе печи и потерям эффективности. Поэтому разработка методов и алгоритмов управления электрической мощностью в руднотермических печах становится все более актуальной задачей для оптимизации процесса обработки руды. Задачи данной исследовательской статьи заключаются в анализе существующих методик управления электрической мощностью в руднотермических печах, разработке новых алгоритмов и проверке их эффективности на практике. Основной акцент будет сделан на использовании контроллеров для управления процессом нагрева и поддержания оптимальной температуры в печи. Исследование имеет потенциал привести к значительным экономическим выгодам для предприятий, работающих с рудой, а также способствовать более эффективному использованию электрической мощности в промышленности.

Ключевые слова: контроллер, рудоплавка, руднотермическая печь, электроэнергия, математическая модель, эффективное управление, энергосбережение.

Received: 27 November 2022

Accepted: 16 March 2023

Available online: 31 March 2023

Applying machine learning methods for analysis socio-economic survey data

G.S. Rysmendeyeva*

Satbayev University, Almaty, Kazakhstan

*Corresponding author: g.rysmendeyeva@satbayev.university

Abstract. To ensure the content of decision-making information systems in the individuals' assets management process requires the development of mathematical models of complex social systems. Studying the expectations of young people on socio-economic issues is of great importance for understanding the future development of the state for developing social policy strategies. A priority throughout the life cycle of an individual is happy and stable marriage, for the stability of which material and moral well-being is important. The next important factor of growing up is closely related to solving the housing problem. The strategic goal of most universities is to train highly paid specialists who are capable to develop the country and support the well-being of own family. Planning expected income is one of the steps of the family welfare planning algorithm. The purpose of this work is to study factors that are important for maturation and well-being. Using machine learning methods, the work explores socio-economic problems from the point of view of first-year university students. The influence of various factors for making decisions regarding the expected age of marriage, solving the housing problem, and expected job income is considered in the research work. Pre-processing of survey data applies data mining techniques. A comparative analysis of the forecast accuracy of classification methods is carried out: logistic regression, neural networks, support vector machines. Students are clustered using the K-means method.

Keywords: machine learning methods, mathematical model, forecasting, behavior patterns, youth problems.

1. Introduction

The long-term priority in the field of social policy in the direction of ensuring social guarantees and increasing the personal responsibility of the country's citizens is the development strategy of our state. Having a conscious management trajectory, a person can plan personal finances at a young age without resorting to the help of a credit institution. Information decision-making systems in the field of asset management for individuals require the development of mathematical models of complex functional social systems and algorithms for their solutions.

Research into issues of investment and human capital during the life cycle of an individual is relevant. Thus, in [1] it is noted that transferring subsidies to an earlier age increases aggregate welfare and human capital. Determining the optimal portfolio for an investor with increased risk aversion in the stock market suggests that older investors should reduce their allocations to risky assets, which is consistent with the empirical relationships between age, wealth and portfolios [2]. Work [3] notes the importance of conducting research on the life cycle of individuals in families in urban conditions. Optimal strategies for financing and investing a defined contribution pension plan when changing consumption, attitude to risk and the level of human capital. Assessing intergenerational social mobility is important for understanding the role of family in explaining income inequality. Issues of modeling optimal investment behavior considering personality factors such as abilities, human capital, strength, etc. are studied within the framework of the dynamic conflict model.

This paper aims to describe issues of family financial well-being using mathematical modeling methods for further use in the development of an information system. To achieve this goal, it is first necessary to describe the processes of creating family financial well-being and determine the limitations of the mathematical model. A study of the literature allows for the application of existing institutional investment management solutions to family finances. By family we mean three generations: children, parents and the older generation of grandparents. The welfare of the family is considered as family savings, the income of family members is considered as private cash flows. Earnings can be attributed to individuals, while family savings are distributed among family members. Savings also serve as reserves and insurance for rainy days. Since the well-being of an individual is interdependent on the well-being of the family, the totality of family assets can be considered as a long-term fund, which implies the financing of certain long-term goals. Examples of goals could be financing the education of children, financial assistance in creating a young family, purchasing real estate, and financial assistance for elderly parents. Planning expected income is one of the steps of the family welfare planning algorithm [4-6]. The purpose of this work is to study factors that are important for maturation and well-being.

1.1. Data collection and preparation

This work uses supervised learning models for labeled data for classification problem and an unsupervised learning model

for student clustering. The methods help to identify factors that divide students into clusters and classes. Required data set for learning models is collected by anonymously surveying students using online surveys via Google forms. One of the main tasks is systematization and processing of survey data. It requires preliminary data preparation, data analysis, selection of features for training models, and assessment of forecasting efficiency. Pre-processing of survey data Excel spreadsheets is carried out using intelligent and exploratory data analysis methods using the Python programming language. The volume of data is 90 questionnaires of first year students of various specialties who voluntarily filled out questionnaires. A variety of data were obtained: structured, semi-structured, unstructured regarding values and attitudes of students on different aspects of future life. Traditionally, forecasting is based on numbers, indicators, coefficients based on statistics and mathematical modeling. However currently one uses machine learning methods (MLM), since the use of unstructured data allows making predictions deeper and information technology - doing calculations faster [7-10].

Data preparation and ranking involves cleaning and organizing the source data into a consolidated format so that the resulting data set is suitable for further analysis and training of the model. The quality of the data influences the training of the selected machine learning models. Since there is relatively small data set, deleting incorrectly completed questionnaires is not suitable. All missing, blank responses were replaced with minimum, maximum, or average values. To make the data suitable for analysis, text and unstructured data were digitized. Aggregation of similar data into one variable was not produced, although some of the questions were for assessment of value factors of family well-being, for example, attitude towards future work, spouse, and income. To aggregate or isolate factors that do not have an impact and are not associated with output data, additional expert knowledge is required. No external expertise was involved, so all collected answers left without aggregation or removal. Some data relates to material and time factors like the age of marriage, the age of purchasing own home, income. In general, resulting variables refer to different types: ordered, continuous, categorical. For example, income is continuous, the choice to live with or not with parents is categorical, the assessment of the degree of importance of real estate is discrete and ordered. After converting text data, it was received up to 90 observations and 56 features of digital data set. The data obtained are not only of different types, but also lie in different ranges of values. Normalization is sensitive to outliers and is used when the data distribution unknown. Standardization is less sensitive to outliers, since it depends on the average value and standard deviation. In this work, the data were standardized according to formula:

$$x_{norm} = ((x - \mu)) / \sigma$$

2. Materials and methods

2.1. Preliminary data analysis

Visualization of data in the form of histograms allows you to make a preliminary analysis of students' attitudes to issues of interest. Data analysis shows that the majority of modern young people prefer to join into marriage after they buy a house, a car and settle down. They prefer to conclude marriage at 23 or 28 years. A minority would like to get married before

the age of 21 or after 33. It also shows that most students want to earn no less than 400 thousand tenge per month. Few people expect to maintain an income of 150-300 thousand tenge. There are students who want to earn more than 1 million tenge. The overwhelming majority believe that an apartment is the most important property that can be acquired by spouses during marriage. The survey of students about what does a family spend most of its income on gives the results that most students think that this is buying a real estate.

2.2. Nearest neighbors' method

Scatterplot and nearest neighbor method are used to understand the relationships between two features. So, there is no clear connection between the following characteristics: how much do parents spend on average in month for student support and at what age does the student plan to decide his own family's housing issue. In most cases the boundaries of separation between students take complex forms. However, there is decision boundary between students whose expected age of marriage before and after 27 years old. Some division occurs between students regarding opinion, that saving on a deposit is ineffective or the costs of purchasing real estate is not the largest family expenses. One can interpret the division as different attitude towards a saving and real estate among different groups of students. There is a division of opinions regarding the point that the most important thing is to pass on healthy genes to the heirs.

2.3. Clustering

The clustering method of k-means is used to determine k-groups of students with similar answers to questionnaires. In this case, the structure of the division is unknown in advance, in contrast to the classification based on a predetermined feature. The elbow method was used, which helps to select the optimal number of clusters for the clustering task. Based on the distance metric between the clusters it follows that the students are optimally divided into three clusters considering all 56 characteristics. Fig. 2 shows a two-dimensional visualization of three clusters based on the first two features, since it is impossible to display 56 features on the graph. Fig. 2 shows the centroid of three clusters according to coordinates x = feature "real estate purchase" and y = "feature financing education of children". It is possible to conclude that students who consider the financing of children's education as insignificant expenses fall into the first cluster. The rest students, who do not agree with this, are divided into two groups according to attitude to the cost of buying real estate. Thus, the second cluster estimates the cost of buying real estate on the scale of 3 and below (middle priority), the third cluster - on the scale 4 – 5 (high priority).

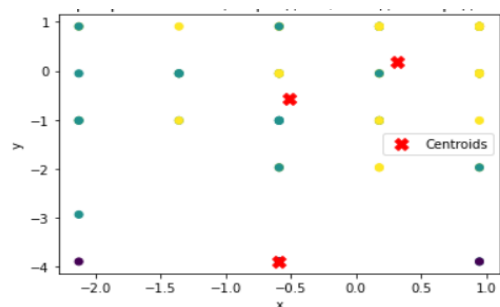


Figure 1. Scatter diagram of three clusters on x = "real estate purchase" and y = "financing education of children"

2.4. Classification

The choice of training method depends on the classification task, type of data and data set size. Most machine learning methods involve training on structured data, so the unstructured data was initially processed and organized into a structured format. In our case there were 90 observations and 56 features in the data set, so it was decided to use several methods and compare the results to study the advantages of each method. In this research work, the most important factor in family well-being is the expected income of a student. Therefore, the forecast of an expected student income was taken as the output variable of the classification model. The model for predicting student income can be used in the process of education and training of highly paid specialists. Collected from the questionnaires, the expected income data set was marked and divided into two classes: label 1, if the expected income is at least 500 thousand tenge and label 0 if it below 500 thousand tenge. As seen in Figure 2 such a division divides student into two almost equal classes.

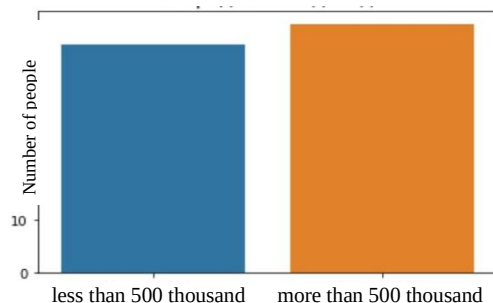


Figure 2. Output data of income

3. Results and discussion

Three classification methods were used in the research work: logistic regression, neural networks (NN) and the SVM (support vector machine method). To apply classification methods, 80% of the data was used as training data, 20% as test data. The neural network architecture consists of three layers. There are 64 neurons in the hidden layers and one neuron in the output layer. The activation function was the relu function in the hidden layers and the sigmoid function at the output layer. Forecast results for test data are presented in the error matrix (Figures 3-5). For test data, error analysis showed that logistic regression works better at reducing errors Type I, and the SVM method is better in reducing the Type II error. The NN method is optimal.

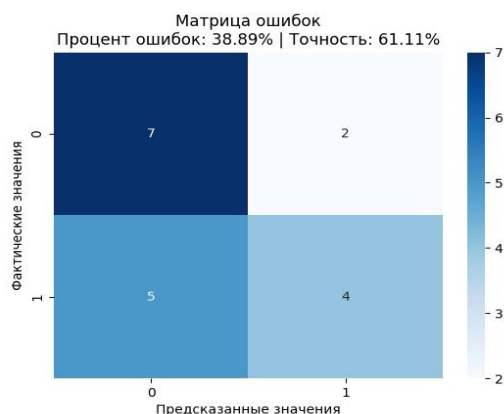


Figure 3. Logistic Regression Error Matrix

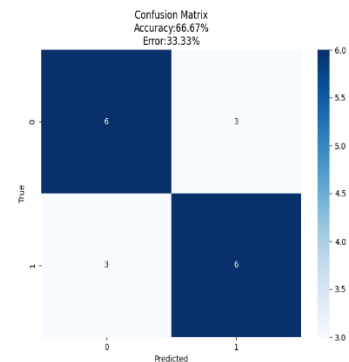


Figure 4. Error matrix of the neural network method

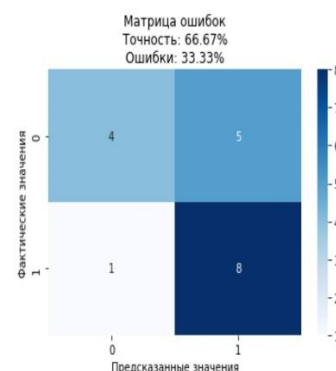


Figure 5. Error matrix, SVM method

Table 2 shows the classification performance indicators of each method, of which one can see that considering the sample size, heterogeneous data types, the ratio of the number of observations and features, the NN and SVM methods give better results than logistic regression. Moreover, the NN method gives balanced performance measures in the sense of type I and type II errors, which can be explained by the fact that in the data set the number of observations exceeded the number of features. In addition, most of the features had low variability, since it took values from a limited set {1; 2; 3; 4; 5}. Thus, the forecast accuracy of the learning model for forecasting expected income according to the survey data was 67%. This is higher than just guessing 50%, but the result is not efficient enough.

4. Conclusions

Thus, the research work collected and analyzed data from a sociological survey of university students. Structured and unstructured survey data were digitized and preprocessed by exploratory data analysis methods for further processing by machine learning methods. The clustering method of k-means was used to determine three groups of students with similar answers to questionnaires. Several machine learning methods were applied for classification students by expected future income. The research work provides a comparative analysis of the effectiveness of classification methods: logistic regression, neural networks, support vector machines – for specific task of classifying students, where the output attribute is the expected income, and the input features are answers to the socio-economic questions of the questionnaire. Received models allow studying behavior patterns and classifying and clustering students by expected income depending on the expected age of entry into adulthood, the

financial capabilities of students' parents and other factors. One can conclude that the students are optimally divided into three clusters considering 56 characteristics. However further detailed analysis is needed to interpret the data and apply for long-term forecasting and planning of family wealth.

References

- [1] Caucutt, E.M. & Lochner, L. (2020). Early and Late Human Capital Investments, Borrowing Constraints, and the Family. *Journal of Political Economy*, 128(3), 1065-1147
- [2] Back, K., Liu, R. & Tegui, A. (2019). Increasing risk aversion and life-cycle investing. *Mathematics and Financial Economics*, (13), 287–302. <https://doi.org/10.1007/s11579-018-0228-1>
- [3] Caplin, A. (2018). Introduction to symposium on «Engineering Data on Individual and Family Decisions Over the Life Cycle». *Economic inquiry*, 56(1), 9-12. <https://doi.org/10.1111/ecin.12468>
- [4] Rysmendeyeva, G.S. (2023). Modeling the age of adulthood depending on economic factors using machine learning method. *International Satbayev Conference*, (4), 148-160
- [5] Rysmendeyeva, G.S. (2021). Development of mathematical models for the information system of decision-making in the process of asset management. *Vestnik KazNITU*, (3), 65-75
- [6] Rysmendeyeva, G.S. (2020). Development of visual models of the information system of decision-making in asset management process of physical persons. *Vestnik KazNITU*, (5), 434-439
- [7] Zakharova, I.G. (2018). Machine learning methods of providing informational management support for students' professional development. *Obrazovanie I Nauka-Education and Science*, 20(9), 91-114
- [8] Dhruvil, S., Devarsh, P., Jainish, A., Pruthvi, H. & Manan, S. (2021). Integrating machine learning and blockchain to develop a system to veto the forgeries and provide efficient results in education sector. *Visual Computing for Industry Biomedicine and Art*, 4(1), 18-28
- [9] Shuo-Chang Tsai, Cheng-Huan Chen, Yi-Tzone Shiao, Jin-Shuei Ciou, Trong-Neng Wu. (2020). Precision education with statistical learning and deep learning: a case study in Taiwan. *International Journal of Educational Technology in Higher Education*, 17(12), 12-25
- [10] Zubarev, A., Bekirova, O. (2020). Analysis of Bank Default Factors in 2013-2019. *Ekonomicheskaya politika*, 15(3), 106-133

Әлеуметтік-экономикалық сауалнамалардың мәліметтерін талдау үшін машиналық оқыту әдістерін қолдану

Г.С. Рысмендеева*

Satbayev University, Алматы, Қазақстан

*Корреспонденция үшін автор: g.rysmendeyeva@satbayev.university

Аңдатпа. Жеке активтерді басқару процесінде шешім қабылдаудың ақпараттық жүйелерінің мазмұнын қамтамасыз ету үшін күрделі әлеуметтік жүйелердің математикалық үлгілерін жасау қажет. Жастардың әлеуметтік-экономикалық мәселелер бойынша үміттерін зерделеу мемлекеттің даму болашағын түсіну және әлеуметтік саясаттың стратегияларын әзірлеу үшін үлкен маңызға ие. Адамның бүкіл өмірлік цикліндегі басымдық - бұл бақытты және тұрақты неке, оның тұрақтылығы үшін материалдық және моральдық әл-ауқат маңызды. Бұл өсудің маңызды факторы тұрғын үй мәселесін шешумен тығыз байланысты. Көптеген университеттердің стратегиялық мақсаты – елді дамытуға және отбасының әл-ауқатына қолдау көрсетуге қабілетті, жалақысы жоғары мамандарды дайындау. Күтілетін табысты жоспарлау отбасының байлығын жоспарлау алгоритмінің кезеңдерінің бірі болып табылады. Бұл жұмыстың мақсаты - жетілу және әл-ауқат үшін маңызды факторларды зерттеу. Машиналық оқыту әдістерін қолдана отырып, жұмыс университеттің бірінші курс студенттері тұрғысынан әлеуметтік-экономикалық проблемаларды қарастырады. Зерттеу жұмысы күтілетін неке жасына, тұрғын үй шешімдеріне және жұмыстан күтілетін табысқа қатысты шешім қабылдауға әртүрлі факторлардың әсерін зерттейді. Сауалнама деректерін алдын ала өңдеу деректерді іздеу әдістерін қолдану арқылы жүзеге асырылады. Жіктеу әдісінің болжам дәлдігіне салыстырмалы талдау жүргізіледі: логистикалық регрессия, нейрондық желілер, тірек векторлық машиналар. Оқушылар «К-орталары» әдісі арқылы топтастырылады.

Негізгі сөздер: машиналық оқыту, математикалық модель, болжау, мінез-құлық үлгілері, жастар мәселелері.

Применение методов машинного обучения для анализа данных социально-экономических опросов

Г.С. Рысмендеева*

Satbayev University, Алматы, Казахстан

*Автор для корреспонденции: g.rysmendeyeva@satbayev.university

Аннотация. Для обеспечения содержания информационных систем принятия решений в процессе управления активами личности необходима разработка математических моделей сложных социальных систем. Изучение ожиданий молодежи по социально-экономическим вопросам имеет большое значение для понимания перспектив

развития государства и разработки стратегий социальной политики. Приоритетом на протяжении всего жизненного цикла личности является счастливый и стабильный брак, для стабильности которого важно материальное и моральное благополучие. Данный важный фактор взросления тесно связан с решением жилищного вопроса. Стратегической целью большинства университетов является подготовка высокооплачиваемых специалистов, способных развивать страну и поддерживать благополучие своей семьи. Планирование ожидаемого дохода является одним из этапов алгоритма планирования благосостояния семьи. Целью данной работы является изучение факторов, важных для взросления и благополучия. С помощью методов машинного обучения в работе исследуются социально-экономические проблемы с точки зрения студентов-первокурсников вуза. В исследовательской работе рассматривается влияние различных факторов на принятие решений относительно предполагаемого возраста вступления в брак, решения жилищного вопроса, ожидаемого дохода от работы. Предварительная обработка данных опроса осуществляется методами интеллектуального анализа данных. Проводится сравнительный анализ точности прогноза методов классификации: логистической регрессии, нейронных сетей, опорных векторов. Проводится кластеризация студентов методом К-средних.

Ключевые слова: машинное обучение, математическая модель, прогнозирование, паттерны поведения, проблемы молодежи.

Received: 28 November 2022

Accepted: 16 March 2023

Available online: 31 March 2023

Development of an optimal control system for the drying tower of the sodium tripolyphosphate production process

A.O. Ulaskan*, M.M. Orynbet

Satbayev University, Almaty, Kazakhstan

*Corresponding author: aishuak2000@mail.ru

Abstract. Sodium tripolyphosphate ($\text{Na}_5\text{P}_3\text{O}_{10}$) is an important chemical compound widely used in various industries, including food, chemical and pharmaceutical. Used as a stabilizer, leavening agent and water softener. The development of an effective control system for the extraction process of sodium tripolyphosphate is an urgent task to improve production efficiency and ensure product quality. Today's industry is increasingly focusing on developing and optimizing chemical extraction processes to improve efficiency and reduce costs. One such item is food, pharmaceuticals, glass, etc. Sodium tripolyphosphate, which is widely used in various industries. The development of a control system for the sodium tripolyphosphate extraction process is of great importance for optimizing and increasing the efficiency of this process. The relevance of this task is due to the fact that the process of producing sodium tripolyphosphate has certain difficulties and depends on temperature, pressure, ratio of reagents, etc. This is due to the fact that it is necessary to control and adjust various parameters. As a result, an automated model of drying towers was developed to improve efficiency and reduce costs in the sodium tripolyphosphate industry.

Keywords: automatic control system, extraction process, drying towers, mathematical model.

1. Introduction

The drying tower is a vertical cylindrical device with a conical bottom. The top of the tower is equipped with an explosion valve and two mixing baffles that mix two heat flows: the first from the turbocalciner and the second from the combustion of natural gas in the upper burner.

First of all, it should be noted that none of the installations of the drying and heating shop has an optimal process control system, which is closely related to the lack of an adequate mathematical model of these processes.

The quality of SPTP increases by 10-15% with the introduction of an optimal control system in the drying tower of the sodium tripolyphosphate production process. Moreover, the process is carried out with minimal fuel consumption and low consumption of finished products [4].

2. Materials and methods

In modern industry, more and more attention is paid to the development and optimization of chemical extraction processes in order to increase efficiency and reduce costs. One of these items is food, pharmaceuticals, glass, etc. sodium tripolyphosphate, which is widely used in various industries.

The technological scheme of production of sodium tripolyphosphate using a drying-heating furnace, as well as the device and working principle of this process were studied. The calculation of the material balance has been completed. Based on the results of the calculations, a promising and effective integrated method scheme was selected, which allows to achieve high efficiency.

To achieve this goal, it is necessary to solve the following tasks:

- analysis of existing methods and technologies for obtaining sodium tripolyphosphate;
- research and identification of the main factors affecting the process of obtaining sodium tripolyphosphate;
- creating a mathematical model of the sodium tripolyphosphate production process;
- development of a management algorithm based on a mathematical model and considering the main factors;
- development of software for implementation of management system;
- conducting experiments to check the effectiveness of the developed management system;

It is expected that the results of this work can be used in various fields where it is necessary to obtain sodium tripolyphosphate. Development of management system allows to reduce costs

Sodium tripolyphosphate (sodium tripolyphosphate) production technology is an important process in the chemical industry. Sodium tripolyphosphate is an inorganic compound used in many applications such as the food processing, laundry and metallurgical industries.

The production process of sodium tripolyphosphate begins with the main raw materials - phosphates, for example, rock and bone phosphates. Phosphates usually contain 16-18% tribasic phosphate, which is the main single component for the production of sodium tripolyphosphate.

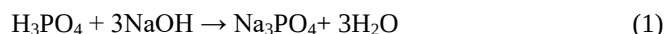
One of the first steps in the process is to treat the phosphates with a sulfuric acid solution to remove unwanted impurities such as iron and aluminum. The resulting phosphate solution is then filtered to remove phosphate deposits.

After filtration, the phosphate solution is mixed with sodium hydroxide solution, resulting in a white precipitate of sodium tripolyphosphate. The resulting sediments are separated from the liquid through the sedimentation process.

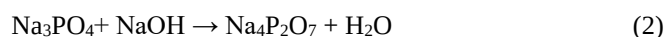
Sodium tripolyphosphate precipitates undergo a drying process that results in a stable material that is easily stored and transported.

The production process of sodium tripolyphosphate usually involves several steps:

Preparation of sodium phosphate: Primary components are phosphoric acid (H_3PO_4) and sodium hydroxide (NaOH). These reagents react in the reactor to form sodium phosphate (Na_3PO_3) and water (H_2O).



Neutralization process: After sodium phosphate is produced, its solution is neutralized by adding an alkali such as sodium hydroxide (NaOH). This leads to the formation of basic sodium phosphate.



Precipitation of tripolyphosphate: The resulting basic sodium phosphate ($\text{Na}_4\text{P}_2\text{O}_7$) is further processed to produce sodium tripolyphosphate. The basic phosphate solution is heated and excess water is removed by evaporation. As a result, a solid precipitate of sodium tripolyphosphate is formed.



Purification and drying: The resulting sodium tripolyphosphate is purified to remove impurities and undesirable compounds. It is then processed to the proper particle size and dried to remove residual moisture.

Packaging and Storage: Purified and dried sodium tripolyphosphate is packaged according to customer requirements and stored in a safe and dry place until use.

The obtained sodium tripolyphosphate has a wide range of applications. In the food industry, it is used as an emulsifier and stabilizer that improves the structure and shelf life of food. In the metallurgical industry, sodium tripolyphosphate is used as an additive to remove rust and increase process productivity. It is also used in the production of detergents to remove grease and prevent scale formation.

The production technology of sodium tripolyphosphate is complex and many factors such as temperature, concentration of reagents and reaction time can affect the final quality of the product. Effective manufacturing technology requires careful control of these factors and continuous process improvement to achieve optimal results.

The melting point is 622°C , in its pure state it is quite stable up to the melting point, when the temperature is further increased, it decomposes into meta and pyrophosphate.

Sodium tripolyphosphate is also soluble in water. Dissolves up to 50 g in 100 ml of water at a temperature of 20°C .

As for the production technology at the sodium tripolyphosphate plant:

The production of sodium tripolyphosphate is a complex chemical process carried out in a factory. These include:

Raw materials: The main raw materials for the production of sodium tripolyphosphate are phosphoric acid (H_3PO_4) and sodium hydroxide (NaOH).

Dosing and mixing: Phosphoric acid and sodium hydroxide are dosed and mixed in special reactors. Control additives and catalysts are added to ensure optimal reaction conditions.

Reaction: In reactors, a reaction takes place between phosphoric acid and sodium hydroxide, resulting in the formation of sodium tripolyphosphate ($\text{Na}_5\text{P}_3\text{O}_{10}$) and water (H_2O). This is an exothermic reaction accompanied by heat release.

Separation and purification: after the completion of the reaction, the resulting mixture passes through special separation units that separate the liquid phase (water) from the solid phase (sodium tripolyphosphate). Purification of sodium tripolyphosphate may involve the use of filtration, drying, and other methods.

Forming and Packaging: After purification, sodium tripolyphosphate can be crushed, graded and packaged according to customer requirements. It usually comes in powder or pellet form.

Quality control: Quality control is carried out at every stage of production to ensure that the product conforms to established standards. This may include analysis of active ingredients, purity tests and physicochemical measurements.

Waste Disposal: The production of sodium tripolyphosphate can generate waste and wastewater. The plant must provide a system for the treatment and disposal of these wastes in accordance with environmental and safety standards.

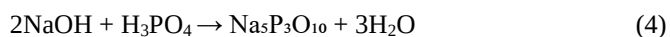
On an industrial scale, sodium tripolyphosphate is produced from solutions of sodium orthophosphate in a one- or two-step process. In the first version, drying and heating are carried out in the same device, often in a rotary kiln that recycles the product.

The entire sodium tripolyphosphate manufacturing process requires compliance with strict safety and regulatory requirements to ensure high quality products and minimize potential adverse environmental impacts [1].

Sodium tripolyphosphate ($\text{Na}_5\text{P}_3\text{O}_{10}$) is a chemical widely used in various industries such as food processing, laundry, detergents, dishwashing detergents, glass, ceramics and other materials.

Raw materials and reactions

The production of sodium tripolyphosphate is based on the use of phosphoric acid (H_3PO_4) and sodium hydroxide (NaOH) as starting materials. The reaction between phosphoric acid and sodium hydroxide leads to the formation of sodium tripolyphosphate:



Production process

The production process of sodium tripolyphosphate can be divided into several main stages:

1. **Preparation of the reaction mixture:** Phosphoric acid and sodium hydroxide are mixed in a reactor in known proportions. Usually, an excess of sodium hydroxide is used to ensure complete conversion of phosphoric acid.

2. **Reaction:** The mixture is heated to a certain temperature and kept liquid by mechanical stirring. A chemical reaction occurs between phosphoric acid and sodium hydroxide, resulting in the formation of sodium tripolyphosphate.

3. **Cooling and crystallization:** After the reaction is completed by cooling the mixture, the obtained sodium

tripolyphosphate comes out of the solution in the form of crystals.

4. Separation and drying: The resulting sodium tripolyphosphate crystals are separated from the solution and subjected to a drying process to remove excess moisture.

Quality and control

Product quality control is an important aspect in the production process of sodium tripolyphosphate. This includes the analysis of the composition of the main substance, the identification of impurities and contaminants, as well as physico-chemical and microbiological tests. Quality control ensures product compliance with established standards and safety requirements [2].

The production of sodium tripolyphosphate is based on the reaction between phosphoric acid and sodium hydroxide. This process includes preparation of the reaction mixture, chemical reaction, cooling and crystallization, separation and drying of the obtained product. Quality control is an important part of production to ensure product compliance with requirements and standards.

Below is detailed information on some of the physicochemical properties of sodium tripolyphosphate:

1. Solubility

Sodium tripolyphosphate is highly soluble in water. Easily dissolves up to 10 g/ml at room temperature. This makes it effective for use in processes that require rapid and complete dissolution of water-soluble components.

2. Chemical stability

Sodium tripolyphosphate is chemically stable under most conditions. It is stable in alkaline solutions and has high thermal stability. But under acidic conditions or at high temperatures, it can decompose into phosphates with a lower oxidation state.

3. pH

Sodium tripolyphosphate is an alkaline compound. Therefore, its solutions have a high pH, typically around 11 in a 1% solution. This property makes it useful in processes where the pH needs to be controlled or maintained, such as detergent production or food stabilization.

4. Formation of complexes

Sodium tripolyphosphate forms complexes with various metal ions, including calcium, magnesium and iron ions. This allows it to be used in processes such as degreasing and defoaming in industrial systems to remove solid deposits and precipitated salts. Hardening properties can also be used to stabilize foods and prevent ingredients from settling or mixing.

5. Source of phosphorus

Sodium tripolyphosphate is a good source of phosphate, which is important in biological processes. Phosphorus is an important element for the growth and development of plants, as well as for the normal functioning of the body. In the food industry, sodium tripolyphosphate can be used as an additive to improve the structure and texture of food products.

6. Form and appearance

Sodium tripolyphosphate appears as colorless or white crystals, usually in powder or granular form. It has a very high density of about 2.5 g/cm³.

The drying tower is a vertical cylindrical device with a conical bottom. The top of the tower is equipped with an explosion valve and two mixing baffles that mix two heat flows: the first from the turbocalciner and the second from the combustion of natural gas in the upper burner.

First of all, it should be noted that none of the installations of the drying and heating shop has an optimal process control system, which is closely related to the lack of an adequate mathematical model of these processes.

The quality of SPTP increases by 10-15% with the introduction of an optimal control system in the drying tower of the sodium tripolyphosphate production process. Moreover, the process is carried out with minimal fuel consumption and low consumption of finished products [3].

The scheme of automatic adjustment of the orthophosphate drying process is as follows (Figure 1).

Various devices are used in raw material drying, especially spread drying towers are considered widespread.

Diffusion drying towers are designed for drying sodium orthophosphate and are designed to reduce moisture content to 1%. The drying tower (9) is a vertical cylindrical apparatus, conical downwards. The upper part of the tower is equipped with an explosion (explosion) valve and two mixing baffles that mix two coolant flows: the first from the turbo calciner and the second from the upper natural gas combustion burner. The upper pressure manifold is located outside the tower under the mixing chamber. Decomposes orthophosphate solution consists of 32 nozzles (10) with a hole diameter of 0.8 mm.

In the nozzle, orthophosphate is distributed and sprayed under a pressure of 9-15 MPa through a pipe drive. Injector burners (8) are located in the tower. Dispersed orthophosphate is dried by burning gas in a burner. Dried orthophosphate is dehydrated and crushed in a drying tower goes down. From here it is sent directly to the turbo calciner (12).

The scheme of automatic adjustment of the orthophosphate drying process is as follows (Figure 1).

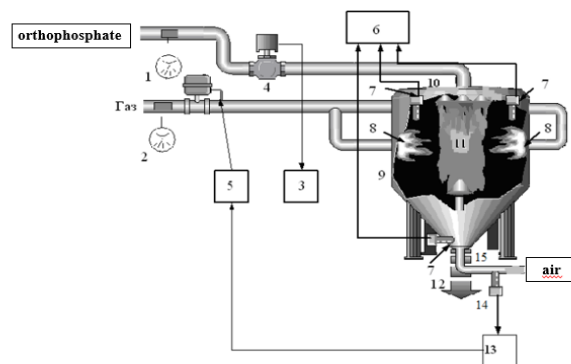


Figure 1. Diagram of automatic adjustment of the orthophosphate drying process

Various devices are used in raw material drying, especially spread drying towers are considered widespread.

Diffusion drying towers are designed for drying sodium orthophosphate and are designed to reduce moisture content to 1%. The drying tower (9) is a vertical cylindrical apparatus, conical downwards. The upper part of the tower is equipped with an explosion (explosion) valve and two mixing baffles that mix two coolant flows: the first from the turbo calciner and the second from the upper natural gas combustion burner. The upper pressure manifold is located outside the tower under the mixing chamber. Decomposes orthophosphate solution consists of 32 nozzles (10) with a hole diameter of 0.8 mm [2].

In the nozzle, orthophosphate is distributed and sprayed under a pressure of 9-15 MPa through a pipe drive. Injector burners (8) are located in the tower. Dispersed orthophosphate

is dried by burning gas in a burner. Dried orthophosphate is dehydrated and crushed in a drying tower goes down. From here it is sent directly to the turbo calciner (12).

Control of gas and orthophosphate pressure is carried out by manometers 1 and 2. Temperature control is carried out by means of thermocouples (7), whose readings are recorded in a multi-point automatic potentiometer (6) located at different points of the tower. Loss of orthophosphate is measured by an induction flow meter consisting of a sensor (4) and a measuring block (3).

The automatic control system of the process is implemented as follows: the moisture meter measures the moisture content of the dried orthophosphate. Humidity 1%-is lower than, the moisture meter sends a signal to the controller (13). It itself sends a signal to the executive mechanism [1].

3. Results and discussion

A drying tower was considered as an object. The transfer function of the object was given by the second-order aperiodic cut-off equation. This section is expressed by a second-order differential equation:

$$W_0(s) = \frac{1}{6s^2 + 5s + 1} \quad (5)$$

The task of the designer when choosing the type of regulator should be to provide a quality job of the regulator at minimum cost and maximum reliability. The designer may choose a relay, continuous, or discrete (digital) controller type [3].

To choose the type of regulator and determine its flexibility, you need to know the following:

1. Static and dynamic description of the control object
2. Regulatory quality process requirement
3. Regulator quality indicator for serial regulators
4. Exciting character that affects the regulatory process.

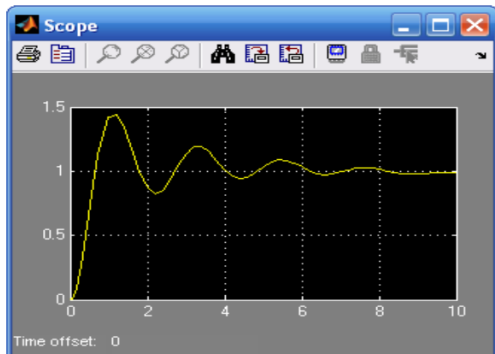
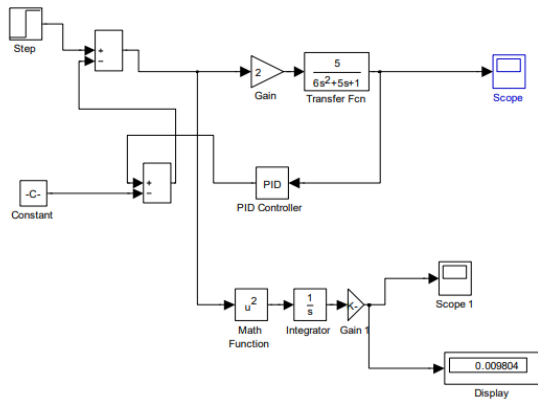


Figure 2. Graph of the automatic regulation system

The dynamic equation of the heat balance in the drying tower:

$$\rho_{ca} \times V_{ca} \times c_{ca} \times \frac{d\theta}{dt} = G_{ca} \times c_{pca} \times \theta_{ca} + G_{bm} \times c_{pca} \times \theta_{ca} - G_{ca} \times c_{pem} \times \theta_{ca} - G_{cm} \times c_{pem} \times \theta_{cm} - W_m \times r \quad (6)$$

Where: V_{ca} - is object volume; c - concentrations; G - expenses; θ - temperatures.

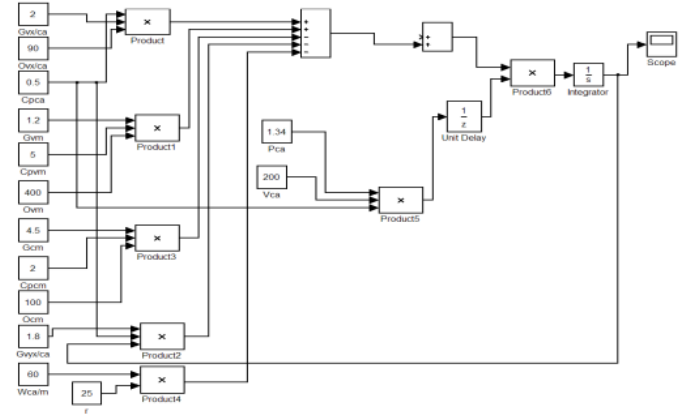


Figure 3. Compilation of the dynamic equation of the heat balance in Simulink

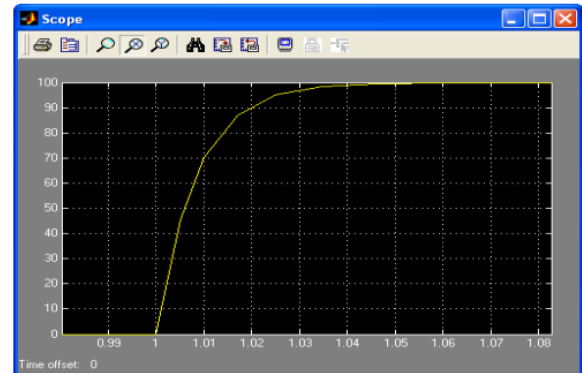


Figure 4. Graph of the dynamic equation of heat balance

The dynamic equation of the material balance depending on the amount of moisture in the product:

$$\rho_{ca} \times V_{cm} \times \frac{d\omega}{dt} = G_{bm} \times \omega_{bm} - G_{cm} \times \omega_{cm} - W_m \quad (7)$$

Where ω_{cm} , ω_{bm} - material humidity.

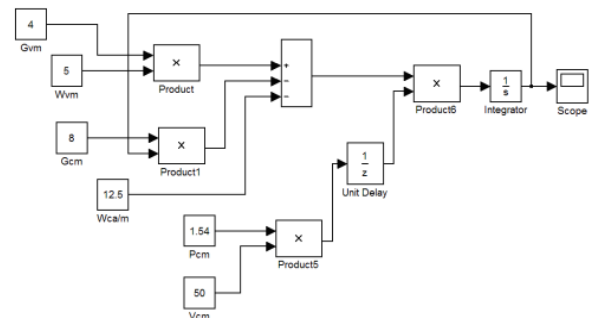


Figure 5. Compilation of the dynamic equation of the material balance in Simulink

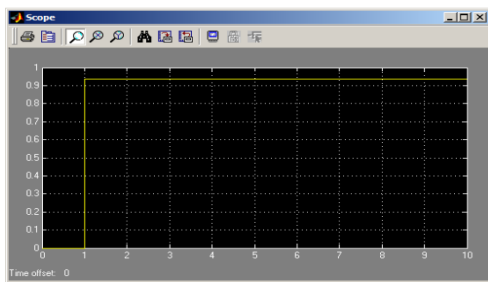


Figure 6. The graph of the dynamic equation of the material balance

Dynamic equation of the material balance in the drying process:

$$\rho_{ca} \times V_{cm} \times \frac{d\omega}{dt} = G_{bm} \times \omega_{bm} + G_{bm} \times \varphi_{ca} - G_{cm} \times \omega_{cm} - G_{ca} \times \varphi_{ca} \quad (8)$$

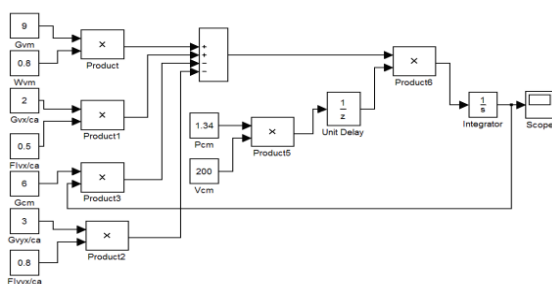


Figure 7. Compilation of the dynamic equation of the material balance in Simulink

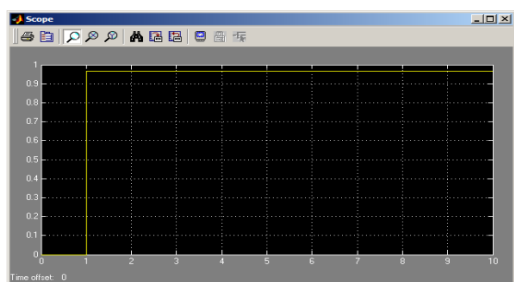


Figure 8. The graph of the dynamic equation of the material balance

It is expected that the results of the work can be used in various fields where it is necessary to obtain sodium tripolyphosphate. Development of a control system reduces costs.

4. Conclusions

Global production of sodium tripolyphosphate will continue to increase until algal blooms become a problem in water bodies. Therefore, saving the world from this problem is the main goal of this project. Russia, China, Kazakhstan: the main producing countries of sodium tripolyphosphate.

In order to optimize the production of sodium tripolyphosphate and improve product quality, a control system is being created that allows for effective monitoring and regulation of the sodium tripolyphosphate production process.

It is expected that the results of this work can be used in various fields where it is necessary to obtain sodium tripolyphosphate. Development of a control system reduces costs.

References

- [1] Nikolaev, A.A. (2015). Analysis of various options for constructing automatic control schemes for the displacement of electrodes of arc steel-smelting furnaces and ladle furnace installations. *Bulletin of the Moscow State Technical University after named G.I. Nosov*, (2), 90-100
- [2] Banach, M., Kowalski, Z., Wzorek, Z. & Gorazda, K. (2009). A chemical method of the production of "heavy" sodium tripolyphosphate with the high content of Form I or Form II. *Polish Journal of Chemical Technology*, 11(2), 3-20.
- [3] Tarajko, M. (2017). Ecological and economic assessment methods of technological processes modernization on the example of chromium and phosphorus compounds production. *Kraków, Poland*
- [4] Farrell, A.E, Plevin, R.J, Turner, B.T., Jones, A.D. & Kammen, M.D. (2006). Ethanol can contribute to energy and environmental goals. *Science*, (311), 506-508. <https://doi.org/10.1126/science.1121416>

Триполифосфат натрий өндіру үрдісіне кептіргіш мұнараға оптималды басқару жүйесін әзірлеу

А.Ө. Ұласқан*, М.М. Орынбет

Satbayev University, Алматы, Қазақстан

*Корреспонденция үшін автор: aishuak2000@mail.ru

Андатпа. Бұл мақалада натрий триполифосфатын алудың автоматты басқару және басқару жүйесін әзірлеу процесі қарастырылған. Натрий триполифосфаты ($\text{Na}_3\text{P}_3\text{O}_{10}$) әртүрлі салаларда, соның ішінде тамақ, химия және фармацевтика салаларында кеңінен қолданылатын маңызды химиялық қосылыс. Тұрақтандырғыш, қопсытқыш және су жұмсартқыш ретінде қолданылады. Натрий триполифосфатының экстракция процесін бақылаудың тиімді жүйесін жасау өндіріс тиімділігін арттыру және өнім сапасын қамтамасыз етудің кезек күттірмейтін міндеті болып табылады. Осы процесі оңтайландыру және тиімділігін арттыру үшін натрий триполифосфатының экстракция процесін басқару жүйесін жасаудың үлкен маңызы бар. Бұл мәселенің өзектілігі натрий триполифосфатын алу процесінің белгілі бір қиындықтарға ие болуымен және температураға, қысымға, реагенттер қатынасына және т.б. тәуелді болуына

байланысты. Бұл әртүрлі параметрлерді бақылау және реттеу қажеттігіне байланысты. Нәтижесінде натрий триполифосфат өнеркәсібінің тиімділікті арттыру және шығындарды азайту үшін кептіру мұнараларының автоматтандырылған моделі әзірленді.

Негізгі сөздер: автоматты басқару жүйесі, экстракция процесі, кептіру мұнаралары, математикалық модель.

Разработка оптимальной системы управления сушильной башней процесса производства триполифосфата натрия

А.О. Уласкан*, М.М. Орынбет

Satbayev University, Алматы, Казахстан

*Автор для корреспонденции: aishuak2000@mail.ru

Аннотация. В данной статье рассмотрен процесс разработки автоматической системы управления и контроля производством триполифосфата натрия. Триполифосфат натрия ($\text{Na}_5\text{P}_3\text{O}_{10}$) — важное химическое соединение, широко используемое в различных отраслях промышленности, включая пищевую, химическую и фармацевтическую. Используется в качестве стабилизатора, загустителя и умягчителя воды. Создание эффективной системы управления процессом экстракции триполифосфата натрия является актуальной задачей для повышения эффективности производства и обеспечения качества продукции. Для оптимизации этого процесса и повышения его эффективности большое значение имеет создание системы управления процессом экстракции триполифосфата натрия. Актуальность данной проблемы обусловлена тем, что процесс получения триполифосфата натрия имеет определенные трудности и зависит от температуры, давления, соотношения реагентов и др. в зависимости от зависимости. Это связано с необходимостью контроля и корректировки различных параметров. В результате была разработана автоматизированная модель сушильных башен для повышения эффективности и снижения затрат в промышленности триполифосфата натрия.

Ключевые слова: система автоматического управления, экстракционный процесс, сушильные башни, математическая модель.

Received: 17 November 2022

Accepted: 16 March 2023

Available online: 31 March 2023

CONTENTS

<i>Omar A., Kalpeyeva Zh.</i> BLOCKCHAIN-BASED VOTING SYSTEM: A SYSTEMATIC LITERATURE REVIEW.....	1
<i>Yerimbetova A., Daiyrbayeva E., Nechta I., Lukpanova L.</i> MATRIX COMBINATION IN STRIP CONVERSION FOR IMPLEMENTING HIDDEN MESSAGES IN THE IMAGE.....	6
<i>Mukazhanova G., Alibiyeva Zh., Kassenkhan A., Mukazhanov N.</i> USING MACHINE LEARNING ALGORITHMS FOR PROCESSING MEDICAL DATA.....	13
<i>Saidov K., Moldagulova A.</i> A REVIEW OF COMMON PRACTICES AND CHALLENGES IN AUTONOMOUS DRIVING.....	20
<i>Adilbek O., Satybaldiyeva R.</i> UNVEILING THE FUTURE: MACHINE LEARNING ALGORITHMS IN EDUCATIONAL PREDICTIVE MODELING.....	25
<i>Mailykhanova B., Koshimbayev S.</i> CONTROLLING THE ELECTRICAL POWER OF AN ORE-SMELTING FURNACE BASED ON CONTROLLERS.....	31
<i>Rysmendeyeva G.S.</i> APPLYING MACHINE LEARNING METHODS FOR ANALYSIS SOCIO-ECONOMIC SURVEY DATA.....	36
<i>Ulaskan A.O., Orynbet M.M.</i> DEVELOPMENT OF AN OPTIMAL CONTROL SYSTEM FOR THE DRYING TOWER OF THE SODIUM TRIPOLYPHOSPHATE PRODUCTION PROCESS.....	41

МАЗМҰНЫ

<i>Омар А., Қалпеева Ж.</i> БЛОКЧЕЙНГЕ НЕГІЗДЕЛГЕН ДАУЫС БЕРУ ЖҮЙЕСІ: ӘДЕБИЕТТЕРГЕ ЖҮЙЕЛІ ШОЛУ.....	1
<i>Еримбетова А., Дайырбаева Э., Нечта И., Лукпанова Л.</i> СТРИП-ТҮРЛЕНДІРУДЕ МАТРИЦАЛАРДЫ КЕСКІНГЕ ҚҰПИЯ ХАБАРЛАМАЛАРДЫ ЕНГІЗУ ҮШІН БІРІКТІРУ.....	6
<i>Мукажанова Г., Алибиева Ж., Касенхан А., Мукажанов Н.</i> МЕДИЦИНАЛЫҚ ДЕРЕКТЕРДІ ӨНДЕУ ҮШІН МАШИНАЛЫҚ ОҚЫТУ АЛГОРИТМДЕРІН ПАЙДАЛАНУ.....	13
<i>Саидов К., Молдагулова А.</i> АВТОНОМДЫ ЖҮРГІЗУДІҢ ЖАЛПЫ ТӘЖІРИБЕЛЕРІ МЕН МІНДЕТТЕРІНЕ ШОЛУ... <i>Әділбек О., Сатыбалдиева Р.</i>	20
БОЛАШАҚҚА ҚАДАМ: БІЛІМ БЕРУДЕГІ БОЛЖАМДЫ МОДЕЛЬДЕУДЕГІ МАШИНАЛЫҚ ОҚЫТУ АЛГОРИТМДЕРІ.....	25
<i>Майлыханова Б., Кошимбаев Ш.</i> КОНТРОЛЛЕР НЕГІЗІНДЕ КЕНДІ БАЛҚЫТУ ПЕШІНІҢ ЭЛЕКТР ҚУАТЫН БАСҚАРУ.. <i>Рысмендеева Г.С.</i>	31
ӘЛЕУМЕТТІК-ЭКОНОМИКАЛЫҚ САУАЛНАМАЛАРДЫҢ МӘЛІМЕТТЕРІН ТАЛДАУ ҮШІН МАШИНАЛЫҚ ОҚЫТУ ӘДІСТЕРІН ҚОЛДАНУ.....	36
<i>Ұласқан А.Ө., Орынбет М.М.</i> ТРИПОЛИФОСФАТ НАТРИЙ ӨНДІРУ ҮРДІСІНЕ КЕПТІРГІШ МҰНАРАҒА ОПТИМАЛДЫ БАСҚАРУ ЖҮЙЕСІН ӨЗІРЛЕУ.....	41

СОДЕРЖАНИЕ

<i>Омар А., Қалтеева Ж.</i> СИСТЕМА ГОЛОСОВАНИЯ НА ОСНОВЕ БЛОКЧЕЙНА: СИСТЕМАТИЧЕСКИЙ ОБЗОР ЛИТЕРАТУРЫ.....	1
<i>Еримбетова А., Дайырбаева Э., Нечта И., Лукпанова Л.</i> КОМБИНИРОВАНИЕ МАТРИЦ В СТРИП-ПРЕОБРАЗОВАНИИ ДЛЯ ВНЕДРЕНИЯ СКРЫТЫХ СООБЩЕНИЙ В ИЗОБРАЖЕНИЕ.....	6
<i>Мукажанова Г., Алибиева Ж., Касенхан А., Мукажанов Н.</i> ИСПОЛЬЗОВАНИЕ АЛГОРИТМОВ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ОБРАБОТКИ МЕДИЦИНСКИХ ДАННЫХ.....	13
<i>Саидов К., Молдагулова А.</i> ОБЗОР РАСПРОСТРАНЕННЫХ ПРАКТИК И ЗАДАЧ АВТОНОМНОГО ВОЖДЕНИЯ.....	20
<i>Әділбек О., Сатыбалдиева Р.</i> РАСКРЫВАЯ БУДУЩЕЕ: АЛГОРИТМЫ МАШИННОГО ОБУЧЕНИЯ В ОБРАЗОВАТЕЛЬНОМ ПРОГНОСТИЧЕСКОМ МОДЕЛИРОВАНИИ.....	25
<i>Майлыханова Б., Кошимбаев Ш.</i> УПРАВЛЕНИЕ ЭЛЕКТРИЧЕСКОЙ МОЩНОСТЬЮ РУДНОТЕРМИЧЕСКОЙ ПЕЧИ НА БАЗЕ КОНТРОЛЛЕРОВ.....	31
<i>Рысмендеева Г.С.</i> ПРИМЕНЕНИЕ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ ДЛЯ АНАЛИЗА ДАННЫХ СОЦИАЛЬНО-ЭКОНОМИЧЕСКИХ ОПРОСОВ.....	36
<i>Уласкан А.О., Орынбет М.М.</i> РАЗРАБОТКА ОПТИМАЛЬНОЙ СИСТЕМЫ УПРАВЛЕНИЯ СУШИЛЬНОЙ БАШНЕЙ ПРОЦЕССА ПРОИЗВОДСТВА ТРИПОЛИФОСФАТА НАТРИЯ.....	41

Учредитель: Satbayev University

Регистрация:

Министерство информации и общественного развития Республики Казахстан
№ KZ85VPY00060870 от 09.12.2022

Официальный сайт: <https://ce.journal.satbayev.university/index.php/journal>

Основан в 2023 г. Выходит 4 раз в год

Адрес редакции:

г. Алматы, ул. Сатпаева, 22
тел.: 292-63-46