ANNOTATION
of the Venera Ismailovna Nakhipova dissertation for the Philosophy Doctor degree (PhD) in the educational program 8D01503 - "Training of computer science teachers"

Research topic: Scientific and methodological foundations for the use of machine learning methods in predicting students' academic performance.

Relevance of the Research: In the modern world, the education system plays a crucial role in enhancing the quality of human capital and ensuring the socio-economic development of a country. The application of artificial intelligence (AI) and machine learning (ML) in education represents a significant step towards developing the national education system in line with current requirements. From presidential addresses to government decrees, scientific research, and international cooperation, all measures support the integration of AI into education. This approach not only aims to improve the quality of education but also to enhance the country's competitiveness in the global educational space.

The President of the Republic of Kazakhstan, Kassym-Jomart Tokayev, has repeatedly emphasized the importance of education in his addresses, stating that the future of the country lies in the hands of educated youth. The President noted that "Access to quality education is the most effective way to ensure equal opportunities for citizens". This highlights the need for research-based methods to improve students' academic performance. The relevance of the topic is emphasized by the fact that high-quality education in the President's address determines the direction of development of the country and is the main prerequisite for achieving socio-economic progress. President Tokayev has emphasized that "Digital technologies should help raise education and science to a new level".

The use of machine learning methods is becoming an essential tool for predicting and improving academic performance. Its relevance is determined by the need to enhance the quality of education in the modern education system, making the assessment process more objective and personalized. Machine learning methods enable the analysis of large datasets, providing a clear assessment of each student's performance and identifying their strengths and weaknesses. Additionally, this technology facilitates the personalization of learning programs, adapting them to the needs of each student, making the learning process more efficient and engaging. This further confirms the importance and relevance of the research topic.

This field continues to develop in Kazakhstan, where several universities and research centers, building on studies in this area, are developing new methods and techniques for applying artificial intelligence in the educational process. Among the educators and researchers of Kazakhstan conducting such studies are B. Berikkozha, A. Mukhiyadin, M. Serik, L. A. Suleimenova, Z. Abdiakhmetova, M. Orynbassar, M. Zhumadilova, Z. Zulpykhar, A. Tlebaldiyeva, N.O. Izimbetov and others.

Internationally, the use of artificial intelligence in education is widespread.

For example, in countries such as the United States, China and Germany, AI-based platforms and systems are widely used in educational institutions. These platforms play an essential role in developing adaptive learning systems tailored to the individual needs of students, predicting academic performance, and automating the learning process.

The issues of applying machine learning (ML) and artificial intelligence (AI) are extensively covered in the works of foreign scholars. Among the prominent researchers in this field are Mahmoud Goodayer, Diego Buenaño-Fernández, Jiang H., Fu W, Cristobal Romero, Sebastian Ventura, George Siemens, Annisa Uswatun Khasanah, Harwati and others.

Despite the numerous studies conducted by the above-mentioned national and international researchers, there are several problematic situations and unresolved issues in Kazakhstan regarding the application of machine learning for predicting students' academic performance. For instance, one of the primary issues is the lack of data and their quality. Effective ML models require large volumes of high-quality data containing comprehensive information. Unfortunately, in the field of education in Kazakhstan, the accumulation and availability of such data are limited. In particular, in some higher education institutions, student performance data are not stored in electronic format or are incomplete. This reduces the accuracy of models and affects the reliability of prediction results.

Another challenge is the lack of infrastructure and technological support. The complete absence of modern computer equipment and the necessary software in many universities makes it difficult to implement ML methods. Low or unstable internet speed in higher education institutions hinders data processing and analysis.

Additionally, there is a shortage of qualified specialists. The lack of skilled professionals in the field of machine learning has become one of the pressing problems in education in Kazakhstan. The effective integration of ML technology into the education system requires specialists and researchers with specialized training.

Although Kazakhstan actively promotes initiatives to integrate artificial intelligence into the education system, several urgent issues hinder its full implementation:

1. Insufficient Digital Competence of Educators: Teachers lack sufficient digital skills related to AI technologies, which prevents their effective integration into the learning process.
2. Limited Adaptation of Educational Materials: The inadequate adaptation of learning materials, platforms, and tools creates regional inequality.
3. Incomplete Legal Framework: The lack of clear regulations and standards for the ethical and safe use of AI raises concerns about data privacy and security.

Another critical issue is funding. Successful implementation of ML projects requires significant financial resources. Funding should comprehensively cover all needs related to data collection, infrastructure updates, specialist training, and software acquisition. However, the level of funding in this area in the country remains relatively low, which constrains the implementation of specialized

projects.

One more aspect that deserves special attention is the legal and ethical aspects. When collecting and processing students' personal data, it is crucial to comply with legal and ethical standards. The lack of clear regulations and standards in this area leads to issues related to data confidentiality and security.

Although the potential for applying ML to predict students' academic performance in Kazakhstan is vast, the above-mentioned issues hinder this process. Therefore, the objective contradictions between the need to use ML for predicting students' academic performance and the lack of a scientific and methodological basis for its implementation define the problem of this study.

The need to resolve these contradictions has formed the basis for selecting the topic of the study: **"Scientific and methodological foundations for the use of machine learning methods in predicting students' academic performance".**

Purpose of the research study is to determine the scientific and methodological foundations for applying machine learning in predicting students' academic performance, to identify, enhance, and implement effective predictive models within the "Informatics" educational program.

The object of the research is the learning process at the university, the forecast and the factors contributing to the academic performance of students.

The subject of the research is the scientific and methodological foundations of machine learning and the improvement of its application in predicting the academic performance of students in the educational program "Informatics".

Scientific hypothesis of the study: If scientific and methodological foundations for the use of machine learning in predicting academic performance are developed, and effective machine learning models are introduced into the university's educational process, this will lead to increased accuracy in predicting academic performance, improved use of artificial intelligence tools based on machine learning in predicting students' academic performance, expanded opportunities for improving the educational process and improving the quality of education. Since the use of an artificial intelligence tool is based on the previous achievements of students, the analysis of teaching methods and other factors makes it possible to accurately predict their future results.

Main objectives of the research:

- to determine the current state of forecasting the academic performance of university students and the scientific and methodological foundations of the use of machine learning methods;

- substantiation of the purpose and necessity of collecting and pre-processing data related to the academic performance of university students;

- application and improvement of machine learning models for predicting academic performance in the Computer Science educational program and determining criteria for evaluating their effectiveness;

- to conduct an experimental study of the forecast of academic performance and propose ways to introduce the research results into the education system.

Research methods:

Theoretical and empirical methods were employed in this study, including the analysis of scientific, methodological, and computer science-mathematical literature related to the research problem, as well as the examination of informatics educational programs at higher education institutions. The study also involved a comprehensive review of the experience of teaching mandatory subjects and elective courses in higher education institutions. In addition, psychological and pedagogical methods were applied, specifically aimed at using machine learning to predict and enhance students' academic performance. These methods included modeling, pedagogical experiments, surveys, interviews, and methods of mathematical statistics. The modeling method was specifically employed to predict students' academic performance.

Scientific novelty of the research:

1. The scientific and methodological foundations of the use of machine learning methods in predicting the academic performance of students at universities have been determined.

2. The university has defined the purpose of collecting and pre-processing data on students' academic performance, and justified its necessity.

3. To predict the academic performance of students in the Computer Science educational program, the use of machine learning has been improved by creating a hybrid model combining collaborative filtering and Nive Bayesian methods, and criteria for evaluating the effectiveness of the model have been defined.

4. In order to integrate the hybrid model and the learning management system (LMS) developed as part of the research, a learning platform was developed and experimentally implemented, the effectiveness of the model was experimentally confirmed, and specific recommendations and recommendations were developed for the implementation of research results in the education system.

Theoretical significance of the research:

The study identified the factors influencing students' academic performance and highlighted the importance of machine learning in the educational system. The methods for collecting and preprocessing educational data, as well as the theoretical foundations of machine learning, were clearly defined. Furthermore, new approaches to predicting student learning outcomes were proposed based on a hybrid model combining collaborative filtering and the Naive Bayes method. The study also provided a theoretical analysis of how to personalize and optimize the learning process through the integration of machine learning models into a Learning Management System (LMS). The practical implementation of this approach was demonstrated, experimentally validated, and supported by specific recommendations for integrating the research findings into the educational system. The study also contributed to enhancing the application of machine learning models in preparing future computer science teachers. It promoted the

development of students' applied knowledge, practical skills, and competencies in line with the advancements in science, technology, and engineering.

Practical significance of the research:

According to the results of the study, students of the Computer Science University educational program (OP) were offered new approaches to predicting student learning outcomes based on a hybrid model combining collaborative filtering and Naive Bayes methods, a machine learning model was introduced into the LMS system, as a result of which educational organizations are invited to optimize the learning process, create learning trajectories based on individual learners' needs. According to the needs of students, the possibility of improving the effectiveness of education based on forecasting academic performance has been identified. As a result, the introduction of artificial intelligence and machine learning into the education system contributes to improving the quality of the educational process.

Furthermore, the LEARNING platform was developed within the framework of students' research activities (project work, diploma theses, scientific-practical conferences, etc.). An educational and methodological manual was published based on the dissertation topic. Two certificates of authorship were obtained: "Methodology for Determining Students' Academic Performance Using Machine Learning," Certificate No. 37586, dated June 29, 2023. "Integration of Collaborative Filtering into the Naive Bayes Method to Improve Academic Performance Prediction," Certificate No. 57665, dated May 6, 2025.

The research results can be usefully applied in the process of training future teachers at universities, improving their professional qualifications, distance learning and professional development institutions.

It should be noted that the author has made a significant contribution to the practical implementation of the model for predicting students' academic performance. The author works as a junior employee of the project for the grant competition for 2023-2025 years AP19680169 "Integration of machine learning for predicting learning strategies in LMS in the formation of professional competencies of students". The results of this project have confirmed the relevance of the proposed model and its effectiveness in the educational environment.

Key Provisions Submitted for Defense:

1. The scientific and methodological foundations for applying machine learning methods to predict students' academic performance in the "Informatics" educational program are of significant importance. The integration of new methodological approaches of artificial intelligence into the educational process, aimed at enhancing the effectiveness of predicting student achievement in higher education institutions, contributes to improving the quality of education.

2. A high-quality analysis of data related to students' academic performance in higher education institutions requires systematic data collection and preprocessing, which are essential prerequisites for accurate outcome prediction and personalized learning.

3. The application of a hybrid model that combines collaborative filtering and

the Naive Bayes method for predicting students' academic performance within the "Informatics" educational program is a modern approach to enhancing the effectiveness of machine learning and creating personalized learning trajectories. The establishment of clear criteria for evaluating the model's quality ensures the reliability and practical significance of the results.

4. Integrating the hybrid machine learning model designed for predicting learning outcomes into a Learning Management System (LMS) facilitates the digitalization and personalization of the learning process. The practical implementation of this model, its effectiveness confirmed through experimentation, and the development of specific methodological recommendations for its integration into the educational process all contribute to improving the quality of education through the use of innovative technologies.

Validity and reliability of the research results are ensured by the following factors: a comprehensive review of scientific, theoretical, and methodological literature; the application of a set of research methods appropriate to the subject and objectives of the study; an in-depth analysis of various perspectives on the use of machine learning for predicting students' academic performance in higher education institutions; the use of statistical methods to process the obtained data; verification of the consistency between the experimental research results and the initial hypotheses.

Publications based on the research results:

A total of 17 scientific publications have been issued on the dissertation topic, including: 3 articles in international peer-reviewed journals indexed in Scopus and Web of Science; 1 article in a scientific publication recommended by the Committee for Quality Assurance in the Sphere of Science and Higher Education of the Ministry of Science and Higher Education of the Republic of Kazakhstan; 3 articles in proceedings of international scientific-practical conferences held abroad; 7 articles in international conferences held in the Republic of Kazakhstan; 2 copyright certificate; educational and methodical manual-1.

Articles published in Scopus/Web of Science-indexed journals:

1. Use of the Naive Bayes Classifier Algorithm in Machine Learning for Student Performance Prediction. International Journal of Information and Education Technology, Vol. 14, No. 1, 2024, pp. 92–98. [CiteScore 2023: 2.8; 61 percentile] https://orcid.org/0000-0002-4248-8943

2. Integration of Collaborative Filtering into Naive Bayes Method to Enhance Student Performance Prediction. International Journal of Information and Communication Technology Education (IJICTE), Vol. 20, No. 1, 2024, pp. 1–18. [CiteScore 2023: 4.2; 76 percentile] https://doi.org/10.4018/IJICTE.352512

3. Applying Computer Vision and Machine Learning Techniques in STEM-Education Self-Study. International Journal of Advanced Computer Science &

Applications, Vol. 15, No. 1, 2024, pp. 819–827. [CiteScore 2023: 2.3; 43 percentile] https://doi.org/10.14569/IJACSA.2024.0150182

Articles in works suggested by the Ministry of Education and Science of the Republic of Kazakhstan's Committee for Quality Assurance in the Field of Science and Higher Education:

4. Машиналық оқыту әдістерін қолдану арқылы білім алушылардың үлгерімін анықтау. Bulletin of the "National Academy of Sciences of the Republic of Kazakhstan," Pedagogical Series. Vol. 412, No. 6, 2024, pp. 171–186 https://doi.org/10.32014/2024.2518-1467.862

Articles published in international conferences held abroad:

5. EDM Methods for Predicting the Academic Performance of University Students Based on Regression Analysis. Proceedings of the 4th International Scientific Conference "Progress in Science," Brussels, Belgium, Nov. 9–10, 2023, pp. 70–79.

6. The Use of Machine Learning Methods for the Formation of Professional Competencies of Future Computer Science Teachers. Norwegian Journal of Development of the International Science, No. 123/2023, pp. 62–66.

7. The Integration of Machine Learning into Learning Strategies within the Learning Management System for Improving Professional Competence in Students. 6th International Conference on Educational Research and Practice (ICERP), Universiti Putra Malaysia (UPM), July 3–4, 2024, pp. 768–778.

Articles published in international conferences held in Kazakhstan:

8. Determination of Students' Academic Performance Using Machine Learning Methods. ILIM, Vol. 41, No. 3, 2024, pp. 5–20.

9. Methods for Using Machine Learning Tools to Develop a Learning Assessment Platform: A Comprehensive Study. TASHENEV BULLETIN, No. 1 (1), 2024, pp. 11–21.

10. Application of Machine Learning in Education. Proceedings of the International Scientific-Theoretical Conference "National Science: Current Approaches and Interdisciplinary Research Issues." – Shymkent: Alem Publishing, 2024, pp. 97–106.

11. Overview of Machine Learning Based on Systems for Predicting Student Performance. Proceedings of the International Scientific-Practical NEO Conference "Trends and Prospects for the Development of Education and Science." – Shymkent: Alem Publishing House, Vol. 1, 2023, pp. 96–101.

12. Application of Machine Learning in Education. Proceedings of the 10th Republican Scientific-Practical Conference of Students "Baytanayev Readings–X," dedicated to the Day of Science – April 12. – South Kazakhstan State Pedagogical University, Vol. 1, Shymkent, 2022, pp. 262–265.

13. Applying Machine Learning Methods to Web Systems. Proceedings of the International Scientific-Practical Conference Dedicated to the 80th Anniversary of Academician A. Kuatbekov: "Education and Science: Challenges of the Fourth Industrial Revolution." – Shymkent, Vol. 1, 2022, pp. 86–92.

14. Application of Machine Learning Methods. Same conference as above – Vol. 2, 2022, pp. 62–65.

Educational and methodological manual:

15. Educational and methodological manual "Foundations of Machine Learning and Intelligent Data Analysis in Education" was reviewed by the Academic Council of the O. Zhanibekov South Kazakhstan Pedagogical University and approved by the decision of the protocol meeting No. 5 dated April 15, 2025.

Copyright Certificate:

16. "Methodology for Determining Students' Academic Performance Using Machine Learning." State Certificate of Registration of Rights to Copyright-Protected Objects No. 37586, dated June 29, 2023.

17. Integration of Collaborative Filtering into the Naive Bayes method to improve student performance forecasting, No. 57665, May 6, 2025.

Structure of the dissertation:

The dissertation consists of an introduction, two main chapters, a conclusion, a list of references, and appendices. The content of the dissertation and its implementation in the educational process, including the model developed during the research, were carried out on the basis of the LEARNING platform designed for integration with a Learning Management System (LMS).