International Journal of Electrical and Computer Engineering (IJECE)
Vol. 15, No. 2, April 2025, pp. 2172~2180
ISSN: 2088-8708, DOI: 10.11591/ijece.v15i2.pp2172-2180

Application of machine learning methods to analysis and evaluation of distance education

Ainur Mukhiyadin[1], Manargul Mukasheva[2], Ulzhan Makhazhanova[1], Aislu Kassekeyeva[1], Gulmira Azieva[1], Zhanat Kenzhebayeva[3], Alfiya Abdrakhmanova[1]
[1]Department of Information systems, Eurasian National University named L.N.Gumilyov, Astana, Republic of Kazakhstan
[2]Digital Learning Research Laboratory, National Academy of Education named after I. Altynsarin, Astana, Republic of Kazakhstan
[3]Department of Computer Science at the Caspian University of Technology and Engineering, named after Sh. Yessenov, Aktau, Republic of Kazakhstan

Article Info

Article history:

Received Aug 13, 2024
Revised Oct 26, 2024
Accepted Nov 20, 2024

Keywords:

Distance learning
Machine learning
Quality assessment
Term frequency-inverse document frequency
Text data analysis
Word2Vec

ABSTRACT

In recent decades, distance learning has become an essential component of the modern educational system, providing students with flexibility and access to knowledge regardless of location. This paper discusses creating a hybrid machine-learning model for assessing the quality of distance learning based on survey data. The model combines two feature extraction methods: Term frequency-inverse document frequency (TF-IDF) and Word2Vec. Combining these methods allows for a more complete and accurate representation of text data, improving the quality of machine learning models. The study aims to develop and evaluate the effectiveness of the proposed hybrid model for analyzing survey data and assessing the quality of distance learning. The paper considers the tasks of collecting and preprocessing text data, experimentally comparing various feature extraction methods and their combinations, training and evaluating a machine learning model based on a combination of TF-IDF and Word2Vec features, as well as analyzing the results and assessing the effectiveness of the proposed model using various metrics. In conclusion, the prospects for further development and application of the proposed model in educational institutions to improve the quality of distance learning are discussed.

Corresponding Author:

Manargul Mukasheva
Digital Learning Research Laboratory, National Academy of Education named after I. Altynsarin
010000 Astana, Republic of Kazakhstan
Email: manargul.mukasheva@mail.ru

1. INTRODUCTION

In recent decades, distance learning [1]–[3] has become an essential component of the modern educational system [4], providing students with flexibility and access to knowledge regardless of location. The development of technology and the internet has significantly expanded the possibilities of distance learning [5]–[7], making it available to millions of students worldwide. At the same time, the growth in distance learning programs has led to the need to develop methods for assessing their quality, which is essential for ensuring effective learning and student satisfaction. Traditional methods of evaluating the quality of education [8], such as questionnaires and surveys, often do not provide a complete and objective picture. Modern approaches to data analysis based on machine learning and natural language processing methods [9]–[11] offer new opportunities for more accurate and detailed analysis of text data obtained from student and teacher surveys. One promising area is the use of hybrid models that combine various feature extraction methods to improve the quality of analysis. This article discusses creating a hybrid machine-

learning model for assessing the quality of distance learning based on survey data. The model combines two feature extraction methods: term frequency-inverse document frequency (TF-IDF) and Word2Vec. TF-IDF allows you to highlight essential terms in the text, while Word2Vec represents words as dense vectors that reflect their semantic meaning. Combining these methods allows for a more complete and accurate representation of text data, improving the quality of machine learning models.

This study aims to develop and evaluate the effectiveness of the proposed hybrid model for analyzing survey data and assessing the quality of distance learning. The paper considers the following tasks: i) Collection and preprocessing of text data from student and teacher surveys; ii) Experimental comparison of various feature extraction methods and their combinations; iii) Training and evaluation of a machine learning model based on a combination of TF-IDF and Word2Vec features; and iv) Analysis of the results and evaluation of the effectiveness of the proposed model using various metrics.

To solve these problems, the studies propose to use modern analytical and information technologies at the stages of collecting, accumulating, storing, and processing empirical information. One of the essential components of modern global changes is forming a single digital space, which can become the key to deepening integration processes in society. The Internet has become an essential channel for disseminating information trusted by the population, which makes it necessary to reliably verify the information received and the possibility of its critical understanding. To obtain objective, prompt, and reliable information about the state of society, it is advisable to organize timely and objective monitoring of the development of its structures. Within the framework of complex studies, it is proposed to use a research complex that includes an automated information system and a sociological research methodology associated with it [12]–[14]. This complex is intended not only for studying social capital and other structures of society but also for forming the necessary culture of thinking and scientific culture of management.

Firdaus *et al.* [15] uses the information system (IS) success model to analyze the data on the use of information technology in distance learning in schools during the pandemic in order to identify the factors that influence the use of e-learning in distance learning systems. The study uses a quantitative approach to identify the elements that affect the effectiveness of using technology in distance learning activities. The study used methods, analytical techniques, and tools consistent with the quantitative approach to obtain accurate results. The study successfully examined the use of information technology in distance learning activities using the Delaunay and McLean models. As a result, the following conclusions can be drawn: Google Classroom is the second most popular program for use in distance learning after Zoom. The quality of information, quality of service, user characteristics, and use of the system determine user satisfaction with the use of applications in this study.

Gurcan and Cagiltay [16] examines the main topics and trends in distance learning by analyzing 27,735 articles published over the past decade. The study, based on semantic content analysis using the n-gram text categorization method, identified ten main topics such as “systems creation,” “media,” “evaluation,” “methods,” “content,” “levels of education,” “learner,” “research methods,” “interaction-communication,” and “resources-materials-tools.” The obtained results are intended to provide essential insights for further research and practice in distance learning, as well as for the development of standards and continuous improvement in this field. Dogan *et al.* [17] examines using artificial intelligence (AI) technologies in online distance learning by analyzing 276 publications. The study found three dominant thematic trends: the use of AI in online learning processes, algorithms to recognize, identify, and predict student behavior, and adaptive and personalized learning with AI. The leading countries in this area are China, India, and the United States, and the main research areas are computer science, engineering, and social sciences. Cui *et al.* [18] discusses the challenges faced by the education industry due to the spread of coronavirus disease 2019 (COVID-19) and the use of online education and big data to overcome them. The study analyzes the development of online education, the impact of the combination of online education and big data technology, and innovative methods and platforms such as MOOC and DingTalk, which have been widely adopted. Based on the current epidemic situation analysis, the article predicts the prospects and development of online education and big data technology, emphasizing their significance and expected impact on the education field and other industries.

The critical task of improving the measurement system in sociology is to develop a particular automated information system that will increase the accuracy of economic and sociological research results by improving the quality of measuring empirical information. Based on the accumulated experience, Kazakhstani sociology can significantly contribute to the development of data analysis methods. To obtain objective fundamental knowledge about the structure of society, it is necessary to ensure the quality of the data obtained that meets the requirements of modern science [19]. This requires developing and verifying research tools that allow obtaining accurate and objective information. The study results demonstrate that the proposed hybrid model is superior to traditional methods using only one of the approaches and provides a more accurate and detailed assessment of the quality of distance learning. In conclusion, the possibilities of further development and application of the proposed model in educational institutions to improve the quality

of distance learning are discussed. Pinto *et al.* [20] focuses on applying artificial intelligence and machine learning in higher education, where most universities are just beginning to implement them. A systematic literature review was conducted, which revealed that the main areas of research are predicting student academic performance and employability. These findings highlight the importance of using artificial intelligence (AI) and machine learning (ML) in universities without compromising academic integrity. Okagbue *et al.* [21] explores the use of AI and ML in pedagogy, where they contribute to the transformation of traditional educational processes into digital and practical forms. A bibliometric analysis of publications on this topic was conducted using Scopus data, revealing global trends and the impact of AI and ML in education from 2000 to 2021. The authors call on school administrators and policymakers to support the implementation of AI and ML to improve the quality of pedagogical services. Bozkurt and Sharma [22] examines the role of generative artificial intelligence in distance and online education, highlighting its potential and challenges. The technology can improve learning through personalization, automation, and content creation but raises concerns about bias, data security, and overreliance on AI. The author calls for rethinking educational roles such as tutors and learning resource developers and emphasizes the importance of human intervention to reduce the "transactional distance" in the educational process. This paper aims to develop and implement a hybrid model capable of more accurately analyzing survey data and assessing the quality of distance learning. The use of modern machine learning methods, such as TF-IDF and Word2Vec, will not only enhance the accuracy of the assessment but also create more adaptive and practical tools for managing educational processes in the face of global challenges.

2. METHOD

In conducting complex sociological research, one of the critical tasks is to ensure the accuracy and correctness of measurements [23]–[25]. Inaccuracies may arise at various stages of the study: when forming a system of observed variables, converting these variables into numerical data, and during mathematical and statistical processing and data analysis. These problems can lead to incorrect scientific conclusions that do not reflect reality, which is especially critical in complex studies, where errors can have a cumulative effect. They are also mainly due to the peculiarities of socio-economic data, which are often difficult to formalize, and their quantitative indicators are predominantly qualitative, which is not always clear to the researcher at the task planning stage. Analysis of data collection and processing in complex studies requires a comprehensive approach that considers the specifics of socio-economic indicators. At the research planning stage, special attention should be paid to the questionnaire's development and the questions' wording [26]–[28] to minimize the likelihood of systematic errors and bias in results. It is essential to ensure that all variables are measurable and relevant to the study's objectives, and that their interpretation is unambiguous for all survey participants. This stage also includes pilot testing of questionnaires to identify potential problems and their adjustments before mass data collection. At the data processing stage, it is necessary to apply adequate statistical methods that consider the specifics of socio-economic indicators [29], [30]. Using factor analysis, cluster analysis, regression models, and other statistical techniques allows us to identify hidden patterns and relationships between variables. An important aspect is checking the reliability and validity of the methods used, which requires conducting appropriate tests and applying adjustments, if necessary. Modern software and big data processing technologies open up new possibilities for analysis, but their use requires highly qualified researchers and careful control over all stages of the study.

3. RESULTS AND DISCUSSION

To study the level of educational activity of schoolchildren during the pandemic, factors influencing it were selected. Empirical methods of observation, interviewing respondents, and analysis of the results were used. The survey was conducted as an online questionnaire of 32 questions. The survey data were considered in four contexts: contextual block, content block, ergonomic block, technical support, and psycho-emotional block. Since the characteristics of the survey participants are heterogeneous, with students of different ages, social classes, and family support, many factors can influence the adoption of a new type of education, such as emergency distance learning. The survey involved 35,950 students of comprehensive schools in 16 regions and cities of republican significance in the Republic of Kazakhstan. Of these, 17,170 students are in urban schools, 18,780 are in rural schools, and 700 are in small schools. Survey participants were divided by school status, student status, and language of instruction, which made it possible to obtain a representative sample and consider the diversity of learning conditions. The sample included participants of different ages, social statuses, and levels of family support, which may have affected the generalizability of the findings. The diversity of the sample allowed for a comprehensive analysis of the impact of distance learning in different

demographic contexts, which increased the reliability of the results while recognizing that certain regional or socioeconomic conditions may have affected their broader applicability.

Developing a data analysis model for studying the behavior of social institutions requires a well-structured approach that begins with defining a question base. The question base is the foundation of the study, as it is at this stage that the main areas to be studied are formulated. It is necessary to determine which aspects of the behavior of social institutions are of most significant interest, be it their effectiveness, interaction with other institutions, or their impact on society. This requires a deep understanding of the subject area and a thorough literature analysis. For example, one could use questions on the quality of education, the availability of educational resources, and student satisfaction to study educational institutions. For political institutions, essential aspects could be citizens' trust in the government, transparency, and accountability of management processes. In this case, it is necessary to consider modern research and theoretical concepts, such as the theory of functionalism, which views social institutions as systems that maintain the stability of society. Defining the question base may also involve consulting with experts in the field of social sciences and conducting pilot studies to clarify the wording and structure of the questions.

The next step involves defining key questions. Once the question base has been created, it is necessary to narrow the focus of the research to a few key aspects. This allows you to focus on the most significant issues and provides a deeper understanding of the problem under study. Key questions should be specific, measurable, and relevant to help obtain meaningful and valuable data. For example, when studying home-based distance learning, key questions may include: "Are you comfortable participating in online lessons at home?" (answers: yes, no, not really), "How many hours a day do you sit at the computer?" (answers: 1 hour, 2-3 hours, more than 4 hours), "How much time does it take to complete assignments for one subject during distance learning?" (answers: less than 30 minutes, up to 1 hour, more than 1 hour), "How many times a day do you do gymnastic warm-ups during distance learning?" (answers: once in the morning, after each class, every hour of classes, every two hours of classes, do not do). Other key questions may be aimed at identifying barriers, such as: "What prevents you from studying remotely?" (answers: poor internet connection, no personal computer/laptop, unfriendly teachers, excessive control by parents/guardians, other) or assessing the impact of distance learning on health and well-being: "I feel that with distance learning ..." (answers: increased strain on my eyesight, increased strain on my hearing, increased strain on my posture, disrupted daily routine, disrupted sleep, increased anxiety, other). Such questions not only identify specific problems but also allow them to be measured using quantitative data. The specificity and measurability of questions are essential to ensure the accuracy and replicability of the study. The key questions must reflect current issues and trends in the area of study, which requires considering previous studies' results and analyzing current data. For example, a study of the interaction of social institutions may focus on issues related to the coordination between different institutions and their impact on social processes. Specific and measurable key questions help structure data collection and provide a basis for subsequent analysis, facilitating a more in-depth and comprehensive interpretation of the results.

Applying principal component analysis (PCA) factor analysis methods using machine learning or SPSS is the next important step. PCA reduces the dimensionality of data by identifying the main components that explain most of the variation in the data. This helps to simplify the analysis and focus on the most significant variables. For example, in a distance learning study, PCA can analyze responses to questionnaire questions such as "How many hours a day do you sit at the computer?" or "What prevents you from studying remotely?" PCA transforms the original variables into a new set of variables (principal components), which are linear combinations of the original variables, and explains the variation in the data as much as possible. A hybrid machine learning model combining TF-IDF and Word2Vec methods was developed to analyze and assess the quality of distance learning using 32 questionnaire questions. The initial stage is collecting and preprocessing questionnaire data, including critical questions aimed at studying various aspects of distance learning. The question texts are broken down into individual tokens, then normalized by converting to lowercase, removing stop words, and converting words to their original form. The normalized tokens are used to create vector representations using the Word2Vec model, which allows us to reflect the semantic relationships between words.

PCA also helps reduce the problem of multivariate ness and multicollinearity, which is a common problem in high-dimensional data sets. For example, when analyzing the factors that influence the effectiveness of distance learning, PCA can reveal that variables such as "internet quality" and "owning a personal computer" explain a significant portion of the variation in the data and are vital components that influence learning outcomes. In addition, PCA helps improve the interpretability of machine learning models by reducing the number of input variables and eliminating noise, resulting in more accurate and reliable predictions. As a result, using PCA in combination with machine learning is a powerful tool for analyzing complex data sets and obtaining meaningful and valuable insights for studying social institutions. Figure 1 illustrates the process of processing text data to find similar questions. In the first step, a set of critical

questions is used, tokenized to separate them into individual words. Next, the data is normalized, after which two approaches are used to represent texts: the Word2Vec model, which transforms words into vectors considering semantic proximity, and the TF-IDF method to highlight important words based on their frequency. Then, dimensionality is reduced using the principal component analysis (PCA), simplifying further analysis. The resulting vectors are combined, and a calculation is made using the cosine similarity metric to find similar questions. As a result, questions most similar to the original ones are output, which can be used for text analysis and natural language processing tasks.

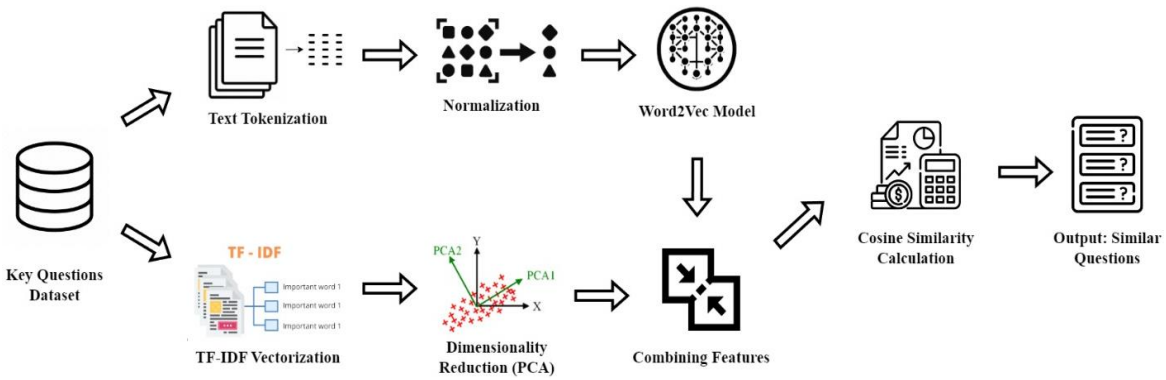

Figure 1. The process of developing a model of behavior of social institutions

In parallel with Word2Vec, the texts of questions are vectorized using the TF-IDF method, which determines the importance of each word in the text relative to the entire data corpus. Vectors obtained using TF-IDF undergo a dimensionality reduction procedure using the PCA, which reduces the number of features while preserving important information. The combined features obtained using TF-IDF and Word2Vec form a single set that provides a more complete representation of the texts of questions. Based on these features, the cosine similarity between the questions is calculated, allowing us to determine their similarity's degree. The analysis results present a list of the most similar questions, which can be used to improve the quality of the questionnaire and obtain insights into critical aspects of distance learning. Using a hybrid model combining various feature extraction methods allows us to increase the accuracy and informativeness of the analysis, which, in turn, contributes to a deeper understanding of the factors affecting the quality of distance learning. In the future, the proposed approach can be adapted and used to analyze other educational data and develop recommendations for improving academic programs and teaching methods.

The format of the responses also plays a vital role in ensuring the accuracy of the data. Using different types of questions, such as multiple choice, Likert scales, and open-ended questions, allows for more diverse and comprehensive information to be collected. For example, respondents can be asked to select multiple answer options for questions about barriers to distance learning, which will more accurately reflect their experiences and opinions. It is also essential to ensure that the wording is neutral to avoid bias against certain answers. In addition, it is necessary to consider the socio-demographic characteristics of respondents, such as age, gender, level of education, and socio-economic status. This allows for a more in-depth analysis of the data and the identification of differences in the responses of different groups. For example, differences in access to high-quality Internet and the availability of a personal computer can significantly affect the results of distance learning and require separate analysis. Thus, careful design of the questionnaire, based on critical questions and principal component analysis results, plays a crucial role in ensuring the accuracy and reliability of the data, which in turn contributes to obtaining meaningful and valuable research results.

Figure 2 shows the performance of three models, PCA (blue), Word2Vec (red), and Hybrid (green), on four metrics: Accuracy, Precision, Recall, and F1-Score. The Hybrid model shows the highest results on all metrics, with an accuracy of 0.92, a prediction accuracy of 0.87, a recall of 0.89, and an F1-Score of 0.88. The PCA model comes in second, achieving an accuracy of 0.91, a prediction accuracy of 0.85, a recall of 0.88, and an F1-Score of 0.87. The Word2Vec model shows the lowest values, with an accuracy of 0.89, a prediction accuracy of 0.83, a recall of 0.86, and an F1-Score of 0.85. Thus, the histogram illustrates the superiority of the hybrid model over the other two models on all metrics.

As a result of the conducted research, a methodology was developed, including automated information systems for collecting, verifying, and analyzing data. This ensures a high level of control over the quality of information and allows for the formation of reasonable and reliable conclusions, which is a critical factor for the successful conduct of comprehensive studies of social institutions and the development of recommendations on their basis for improving socio-economic policy. In the study, a hybrid machine learning model was chosen to assess the quality of distance learning, which combines the TF-IDF and Word2Vec methods. The choice of these methods is because TF-IDF allows you to highlight important terms based on their frequency, and Word2Vec represents words as dense vectors reflecting their semantic meaning. This combination provides a more complete and accurate representation of text data, improving machine learning models' quality. However, each of these methods has its limitations: TF-IDF does not consider the context of words, and Word2Vec requires a large amount of data for training. To solve these problems, a combination of these methods was used, which reduced their limitations and increased the accuracy of the model and the interpretability of the results through the use of the PCA dimensionality reduction method.

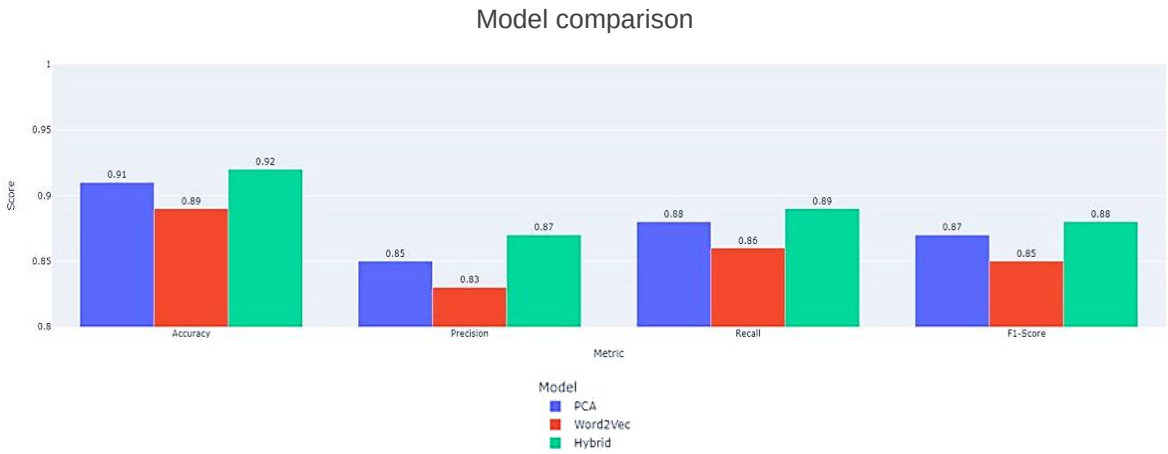

Figure 2. Comparative result of accuracy by models

4. CONCLUSION

As a result of the study, a hybrid machine learning model combining TF-IDF and Word2Vec methods was developed and evaluated for analyzing survey data and assessing the quality of distance learning. The results showed that the proposed model outperforms traditional methods using only one of the approaches and provides a more accurate and detailed assessment of the quality of distance learning. The hybrid model demonstrated high accuracy, F1-measure, prediction accuracy, and recall. Using PCA to reduce the dimensionality of the data improved the interpretability and reliability of the model. In conclusion, the possibilities of further development and application of the proposed model in educational institutions to improve the quality of distance learning are discussed. The importance of carefully forming the questionnaire and taking into account the respondents' socio-demographic characteristics to ensure the data's accuracy and reliability is also emphasized. Summarizing the results of the data analysis and assessment of the statistical significance of the criteria in studies of the behavior of social institutions, it is essential to emphasize that the correctness and accuracy of measurements are fundamental aspects of any sociological research. It is important to minimize errors that may lead to incorrect scientific conclusions at various stages, from variable generation to data analysis.

REFERENCES

[1] A. Mukhiyadin *et al.*, "Application of information technologies and methods for processing big data to the management of the educational process during the pandemic," *Journal of Theoretical and Applied Information Technology*, vol. 101, no. 2, pp. 458–470, 2023.

[2] M. Mukasheva, A. Mukhiyadin, U. Makhazhanova, and S. Serikbayeva, "The behaviour of the ensemble learning model in analysing educational data on COVID-19," *International Journal of Information and Education Technology*, vol. 13, no. 12, pp. 1868–1878, 2023, doi: 10.18178/ijiet.2023.13.12.2000.

[3] S. Serikbayeva *et al.*, "Development of queries using the Z39.50 protocol in distributed information systems to support scientific and educational activities," *Eastern-European Journal of Enterprise Technologies*, vol. 5, no. 2, pp. 66–79, Oct. 2021, doi: 10.15587/1729-4061.2021.239540.

[4] S. Sandugash, T. Jamalbek, S. Madina, Y. Akbota, and A. Ainur, "Building a standard model of an information system for working with documents on scientific and educational activities," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 9, 2021, doi: 10.14569/ijacsa.2021.0120951.
[5] C. A. Onggirawan, J. M. Kho, A. P. Kartiwa, Anderies, and A. A. S. Gunawan, "Systematic literature review: The adaptation of distance learning process during the COVID-19 pandemic using virtual educational spaces in metaverse," *Procedia Computer Science*, vol. 216, pp. 274–283, 2023, doi: 10.1016/j.procs.2022.12.137.
[6] F. Bashir and N. F. Warraich, "Systematic literature review of semantic web for distance learning," *Interactive Learning Environments*, vol. 31, no. 1, pp. 527–543, Aug. 2020, doi: 10.1080/10494820.2020.1799023.
[7] R. Guzal, "Prospects for distance education," *(in Russian) Academic Research in Educational Sciences*, vol. 4, no. 11, pp. 69–472, 2023.
[8] M. Alenezi, S. Wardat, and M. Akour, "The need of integrating digital education in higher education: Challenges and opportunities," *Sustainability*, vol. 15, no. 6, Mar. 2023, doi: 10.3390/su15064782.
[9] Z. Sadirmekova *et al.*, "Ontology engineering of automatic text processing methods," *International Journal of Electrical and Computer Engineering*, vol. 13, no. 6, pp. 6620–6628, 2023, doi: 10.11591/ijece.v13i6.pp6620-6628.
[10] D. Khurana, A. Koli, K. Khatter, and S. Singh, "Natural language processing: state of the art, current trends and challenges," *Multimedia Tools and Applications*, vol. 82, no. 3, pp. 3713–3744, Jul. 2022, doi: 10.1007/s11042-022-13428-4.
[11] S. C. Fanni, M. Febi, G. Aghakhanyan, and E. Neri, "Natural language processing," in *Introduction to Artificial Intelligence*, Springer International Publishing, 2023, pp. 87–99.
[12] G. Ritschard, T. F. Liao, and E. Struffolino, "Strategies for multidomain sequence analysis in social research," *Sociological Methodology*, vol. 53, no. 2, pp. 288–322, Apr. 2023, doi: 10.1177/00811750231163833.
[13] J. Daikeler *et al.*, "Assessing data quality in the age of digital social research: A systematic review," *Social Science Computer Review*, Apr. 2024, doi: 10.1177/08944393241245395.
[14] M. Inghilleri, "Sociological approaches," in *The Routledge Handbook of Translation Theory and Concepts*, Routledge, 2023, pp. 241–262.
[15] D. R. Firdaus, B. Harto, and H. Ferdiansyah, "The effectiveness analysis of distance learning by optimizing the use of information technology," *Journal of Information System, Technology and Engineering*, vol. 1, no. 1, pp. 1–5, May 2023, doi: 10.61487/jiste.v1i1.4.
[16] F. Gurcan and N. E. Cagiltay, "Research trends on distance learning: a text mining-based literature review from 2008 to 2018," *Interactive Learning Environments*, vol. 31, no. 2, pp. 1007–1028, Sep. 2020, doi: 10.1080/10494820.2020.1815795.
[17] M. E. Dogan, T. Goru Dogan, and A. Bozkurt, "The use of artificial intelligence (AI) in online learning and distance education processes: A systematic review of empirical studies," *Applied Sciences*, vol. 13, no. 5, Feb. 2023, doi: 10.3390/app13053056.
[18] Y. Cui *et al.*, "A survey on big data-enabled innovative online education systems during the COVID-19 pandemic," *Journal of Innovation & Knowledge*, vol. 8, no. 1, Jan. 2023, doi: 10.1016/j.jik.2022.100295.
[19] M. O. Turaqulovich, "Applied to the current training process requirements," *The Multidisciplinary Journal of Science snd Technology*, vol. 3, no. 1, pp. 54–63, 2024.
[20] A. S. Pinto, A. Abreu, E. Costa, and J. Paiva, "How machine learning (ML) is transforming higher education: A systematic literature review," *Journal of Information Systems Engineering and Management*, vol. 8, no. 2, Apr. 2023, doi: 10.55267/iadt.07.13227.
[21] E. F. Okagbue *et al.*, "A comprehensive overview of artificial intelligence and machine learning in education pedagogy: 21 Years (2000–2021) of research indexed in the Scopus database," *Social Sciences & Humanities Open*, vol. 8, no. 1, 2023, doi: 10.1016/j.ssaho.2023.100655.
[22] A. Bozkurt and R. C. Sharma, "Challenging the status quo and exploring the new boundaries in the age of algorithms: Reimagining the role of generative AI in distance education and online learning," *Asian Journal of Distance Education*, vol. 18, no. 1, 2023.
[23] M. Palmieri and R. Aprile, "Data quality in social survey research," in *Encyclopedia of Information Science and Technology, Sixth Edition*, IGI Global, 2024, pp. 1–13.
[24] D. Alonzo and S. Teng, "Trustworthiness of teacher assessment and decision-making: Reframing the consistency and accuracy measures," *International Journal of Instruction*, vol. 16, no. 3, pp. 1075–1094, Jul. 2023, doi: 10.29333/iji.2023.16357a.
[25] F. Kouam Arthur William, "Mastering validity and reliability in academic research: Meaning and significance," *International Journal of Research Publications*, vol. 144, no. 1, Feb. 2024, doi: 10.47119/ijrp1001441320246160.
[26] R. Fernández-García, E. Melguizo-Ibáñez, F. Zurita-Ortega, and J. L. Ubago-Jiménez, "Development and validation of a mental hyperactivity questionnaire for the evaluation of chronic stress in higher education," *BMC Psychology*, vol. 12, no. 1, Jul. 2024, doi: 10.1186/s40359-024-01889-1.
[27] F. Krieglstein, M. Beege, G. D. Rey, C. Sanchez-Stockhammer, and S. Schneider, "Development and validation of a mental hyperactivity questionnaire for the evaluation of chronic stress in higher education," *Educational Psychology Review*, vol. 35, no. 1, Jan. 2023, doi: 10.1007/s10648-023-09738-0.
[28] T. Zhang, X. Lu, X. Zhu, and J. Zhang, "The contributions of AI in the development of ideological and political perspectives in education," *Heliyon*, vol. 9, no. 3, Mar. 2023, doi: 10.1016/j.heliyon.2023.e13403.
[29] M. Mukasheva, O. Chorosova, Z. Zhilbayev, and Y. Payevskaya, "Integrated approach to the development and implementation of distance courses for school computer science teachers," in *2020 IEEE 14th International Conference on Application of Information and Communication Technologies (AICT)*, Oct. 2020, pp. 1–6, doi: 10.1109/aict50176.2020.9368817.
[30] M. Mukasheva, Z. Kalkabayeva, and A. Zhanasbayeva, "Teachers' perspectives on the use of virtual reality in collaborative learning," in *2023 13th International Conference on Advanced Computer Information Technologies (ACIT)*, Sep. 2023, pp. 670–674, doi: 10.1109/acit58437.2023.10275565.

BIOGRAPHIES OF AUTHORS

Ainur Mukhiyadin began her higher education in 2003 with a degree in applied mathematics in L. N. Gumilyov Eurasian National University. In 2009-2011, she studied at the master's degree in information systems and defended her scientific dissertation on the topic "Application of distance educational technologies in Kazakhstan" in Kazakh economy, finance and international trade university. Currently, She is a doctoral student at the Department of Information Systems of the L. N. Gumilyov Eurasian National University. Her research interests focus on the application of statistical methods in educational research, especially survey data analysis and factor extraction methods. Actively participates in publications and work on optimization of survey measurements and evaluation of their effectiveness, making significant contributions to the development of methodologies in this area. She can be contacted at email: amukhiyadin@gmail.com.

Manargul Mukasheva currently holds the position of chief researcher at the Digital Learning Research Laboratory of the Altynsarin National Academy of Education in Astana, Kazakhstan. She has the academic title of associate professor at the Eurasian National University named after L.N. Gumilyov, where she has been working since 2004 at the Department of Information Systems. She has a distinguished career that spans various positions, including head of the Information and Analytical Support Department of the Center for the Bologna Process and Academic Mobility in Astana, and associate professor and assistant professor at Aktobe Regional State University named after K. Zhubanov. She received a Candidate of Pedagogical Sciences (PhD) degree from the Kazakh National Pedagogical University named after Abai, specializing in teaching mathematics and computer science. Her research interests mainly focus on virtual and augmented reality technologies in education, computational thinking, educational programming and inclusive education methodologies. She is the author of numerous scientific works, including articles, conference reports and books, making significant contributions to the development of educational technologies and teaching practice in. She can be contacted at email: manargul.mukasheva@mail.ru.

Ulzhan Makhazhanova in 2008 she graduated from Eurasian National University with a degree in computer science. In 2011 he received a master's degree in computer science. In 2021, she graduated from Eurasian National University doctoral studies with a specialty 6D070300 – information systems. From 2020 to the present, she is a senior lecturer at the Department of Information Systems, L.N. Gumilyov Eurasian National University, Astana. She is the author of more than 22 works. Her research interests include data analysis, big data, machine learning, fuzzy logic. She can be contacted at email: makhazhan.ut@gmail.com.

Aislu Kassekeyeva received her PhD degree in 2021 in the specialty "Information systems" named after. L.N. Gumilyov, Kazakhstan. Currently he is a senior lecturer at the Department of Information Systems named after. L.N. Gumilyov. Her research interests include computational linguistics, data analytics, artificial intelligence, and data mining. She can be contacted at email: aibike_7474@mail.ru.

Gulmira Azieva completed her PhD in information systems at L. Gumilyov Eurasian National University in 2021. In her dissertation, she investigated the application of a simulation model of the crude oil and petroleum products supply chain aimed at optimizing logistics processes and improving the efficiency of resource management in this area. Gulmira's research interests cover a wide range of issues in the field of information systems, including the development and implementation of models for analyzing and predicting complex systems. She began her education at the Faculty of Physics and Mathematics of the same university, where she completed her bachelor's degree in 2004. During her studies and research at Aktobe University, Gulmira focused on the application of mathematical and simulation models to solve practical problems, which became the basis for her subsequent research in the field of information systems. She can be contacted at email: gulmira_azieva@mail.ru.

Zhanat Kenzhebayeva graduated with a degree in "Professional training in informatics and computing technology" in 1998 from Aktau State University named after Sh. Yessenov. In 2006, she completed her postgraduate studies in "Systems analysis, management, and information processing" at the Kazakh National Technical University named after K. Satpayev, and in 2010, she received the degree of Candidate of Technical Sciences. Currently, she works in the Department of Computer Science at the Caspian University of Technology and Engineering named after Sh. Yessenov. She is the author of over 40 scientific papers, including 5 articles in the Scopus database, 6 textbooks, and 2 monographs. Her research interests include machine learning, systems analysis, information processing, information management, and GIS. She can be contacted at email: zhanat.kenzhebaeva@yu.edu.kz.

Alfiya Abdrakhmanova received a bachelor's degree in technical sciences in specialty 6M070300 - "Information systems" at the Eurasian National University (ENU) named after L.N. Gumilyov, Astana, Kazakhstan, 2022. She is a co-author of 3 publications. Her research interests include knowledge bases, big data, artificial intelligence and machine learning. She received a master's degree in technical sciences in the field of information systems from the Caspian State University of Technology and Engineering named after Sh. Yesenov in 2009. She can be contacted at email: alfiya_zagievna@mail.ru.