Use of the Naive Bayes Classifier Algorithm in Machine Learning for Student Performance Prediction

Venera Nakhipova[1], Yerzhan Kerimbekov[1], Zhanat Umarova[2,*], Laura Suleimenova[1], Saule Botayeva[3], Almira Ibashova[1], and Nurlybek Zhumatayev[4]

[1]Department of Computer Science, South Kazakhstan State Pedagogical University, Shymkent, Kazakhstan
[2]Department of Information Systems and Modeling, Auezov University, Shymkent, Kazakhstan
[3]Department of Information and Communication Technologies, Tashenev University, Shymkent, Kazakhstan
[4]Department of Computer Engineering and Software, Auezov University, Shymkent, Kazakhstan
E-mail: nakhipovavenera@mail.ru (V.N.); kerimbekov.yerzhan@okmpu.kz (Y.K.); Zhanat-u@mail.ru (Z.U.); Laurasuleimenova7174@gmail.com (L.S.); saule_bb@mail.ru (S.B.); almira_i@mail.ru (A.I.); nuralmiras@mail.ru (N.Z.)
*Corresponding author

Manuscript received June 21, 2023; revised August 16, 2023; accepted October 23, 2023; published January 23, 2024

***Abstract*—This study focused on the development and analysis of a methodological platform grounded in machine learning principles for evaluating learning processes and enhancing student outcomes. The aim of this research was to develop and test a method for evaluating students' academic performance based on the Naive Bayes classifier. Also, an objective of this study was to create an efficient tool capable of automating and optimize the assessment of educational performance using contemporary machine learning methods and technologies. The study employed the Naive Bayes analysis technique to predict student achievements, with the algorithm being implemented in Python. Despite an emphasis on the development of a software product, the research primarily focused on the development and analysis of the method. Our findings underscore the novelty of this approach, which can serve as a valuable tool for educational institutions and educators.**

***Keywords*—machine learning, intelligent systems, naive bayes method, Educational Data Analysis (EDM), productivity, academic performance forecasting**

I. INTRODUCTION

Education is a key driver of societal development and improved quality of life, and in today's world, it has become more accessible and widespread than ever before. However, accurately evaluating learning processes and student' achievement remains a difficult task for educational institutions and educators. Accurate assessment not only improves the quality of education but also helps optimize curricula, adapt teaching methodologies, and enhance educational accessibility for all students.

In this study, we addressed the following questions: How can we effectively assess the processes of learning and the achievements of students using machine learning methods? How can we automate the measurement of academic performance, while providing accurate and objective evaluations?

The proposed approach, based on machine learning and the Naive Bayes classifier, offers several key advantages that make it unique and valuable in the context of evaluating educational processes and student achievements.

Automation and Objectivity: This approach automates the assessment of academic performance and learning processes, reducing the subjective influence of human factors. The Naive Bayes classifier analyses data and makes assessments based on probabilistic models, ensuring objective evaluation.

Scalability: Machine learning methods can handle vast amounts of data, making them ideal for processing data related to educational processes, where large volumes of information about students and their achievements are collected.

Pattern Discovery: Machine learning enables the identification of complex and indirect relationships in data that might go unnoticed using traditional assessment methods. This helps educational institutions understand which factors ad approaches truly impact student success.

Personalized Approach: Machine learning methods allow the creation of personalized models for each student, that consider their unique characteristics and educational needs.

Speed and Efficiency: Automating the assessment process using machine learning significantly accelerates the generation of results and provides timely feedback, which can be critical for adapting educational programs.

We choose the Naive Bayes classifier because it is well-suited for classification tasks involving predicting student academic performance. Its advantages include high performance on large datasets and the ability to handle many features, which are often characteristic of education-related data. Furthermore, the Naive Bayes classifier can generalize information from past observations, making it a powerful tool for predicting student performance.

The software product under development represents an innovative solution capable of efficiently processing and analyzing extensive datasets pertinent to learning processes and academic performance. The product boasts modern data processing algorithms and machine learning-based analytical techniques.

With this software, users can effortlessly collect and store a wide array of data, including academic grades, student progress reports, and various information about educational processes. The program automatically handles these data, identifying patterns and trends, and assessing student progress and the effectiveness of teaching methods.

Leveraging the outcomes of this analysis, the software generates valuable information and reports. These resources can enable educators, administrators, and other stakeholders, to make well-informed decisions aimed at enhancing educational processes. Consequently, educational institutions can optimize curricula, adapt teaching methodologies, and deliver more effective training to their students.

The software product also offers flexible settings that can be tailored to the specific requirements of each educational institution. It boasts a user-friendly interface, intuitive navigation, and graphical data visualization, enhancing the accessibility and comprehensibility of the analysis process.

In essence, the developed software product stands as a potent tool for data processing and analysis, significantly enhancing the comprehension and assessment of educational processes and the efficacy of training within educational institutions.

During the literature review, academic articles, and studies on the evaluation of educational processes and student achievements using machine learning-based platforms were examined. Romero and Ventura [1] conducted a pivotal review, that provided an extensive overview of the Educational Data Mining (EDM) landscape. Their paper considers various machine learning methods and algorithms deployed for the analysis of training data and the evaluation of learning processes. It offers insights into key topics and techniques within the EDM domain, including data collection, preprocessing, analysis, and the interpretation of results. The authors also delve into the application of EDM across diverse educational scenarios, such as adaptive learning, student knowledge assessment, and academic performance forecasting.

Another notable study by Pardos and Tang *et al.* [2] was investigated, wherein they explored the utilization of machine learning methods for predicting student learning outcomes. In this research, Pardos and Tang *et al.* delve into applying ensemble machine learning methods to forecast student academic achievements. They conducted experiments using learning data acquired through e-learning technologies to predict students' academic results. Their investigation covers ensemble methods, including bagging and random forest, and compares them with linear regression and regression trees.

Lahmiri and Bekiros *et al.* [3] may prove valuable for exploring the application of ensemble machine learning methods in predicting student learning outcomes. In this article, the authors conducted a comparison of various machine learning models and model selection methods to achieve optimal results. The primary objective of the study was to identify the most effective model for predicting student learning outcomes.

An additional resource of interest is presented in the work of Baker and Inventado [4], where the authors delve into the utilization of machine learning to analyze data from learning processes, forecast student performance, and personalize education. Within this paper, authored by Baker and Inventado, the investigation explores the application of Educational Data Mining (EDM) methods and learning analytics within the context of constructivist research. It examines how these methods can be harnessed to analyze data collected within the constructivist approach to educational processes.

The authors deliberate various aspects of leveraging EDM and learning analytics, including the analysis of student activity, academic performance prediction, assessment of learning effectiveness, and more. Furthermore, they consider the methods for data collection, preprocessing, and analysis applicable within the framework of constructivist research.

One more noteworthy contribution to the field of educational analytics and learning data analysis could be found in the work of Siemens and Baker [5]. In this article, the focus lies on exploring communication and collaboration among researchers and presenting various approaches and machine learning methods employed within this domain. Siemens and Baker delve into the realm of Learning Analytics, charting its evolution as a scientific discipline. They review fundamental concepts and methods employed in learning analytics, discussing its potential to enhance the efficiency of educational systems.

The paper covers a gamut of learning analytics facets, encompassing data collection, processing, predictive models, results visualization, and data-driven decision-making. Additionally, ethical and privacy concerns related to learning analytics are addressed.

Drawing from the insights gleaned from these sources, several overarching conclusions emerge:

1) Educational Data Mining (EDM) and Learning Analytics stand as pivotal research areas in the realm of educational data utilization.
2) EDM and learning analytics facilitate the collection, processing, and analysis of learning-related data to extract valuable insights and enhance educational processes.
3) Leveraging machine learning methods and ensemble models can be highly effective in predicting student learning outcomes and facilitating data-driven decision-making.
4) Key facets of EDM and learning analytics encompass data collection and preprocessing, the development of predictive models, result visualization, and educational decision-making.
5) Ethical and privacy considerations constitute crucial dimensions in the use of educational data.

In summary, these studies underscore the potential and significance of EDM and learning analytics in modern education, laying the groundwork for the development of novel methods and approaches within this domain. These articles, alongside others reviewed in the literature, serve as vital source material for comprehending existing platforms rooted in machine learning principles and their applications in assessing learning processes and achieving student outcomes [6–8].

Previous studies have explored a diverse array of methods and approaches employed to assess learning processes and attain learning outcomes through platforms grounded in machine learning principles. Some noteworthy methods include:

Naive Bayes Classifier: This method operates on the assumption of feature independence and employs probabilistic models to categorize students based on their academic performance. Its effectiveness can be witnessed when an ample and diverse dataset is available.

Logistic Regression: This method uses a linear model to predict the probability of a student being assigned to a particular academic performance class. It can be useful in analyzing the impact of various factors on student success.

Decision Trees: This method creates a hierarchical decision structure based on student attributes. It can provide interpreted rules for predicting academic performance.

Random Forest: This method combines multiple decision trees to produce more accurate predictions. He can cope with the problem of retraining and has a good ability to generalize.

Gradient Boosting: This method also combines several models, but does it sequentially, training each subsequent model on the errors of previous models. It can provide high-accuracy of forecasts.

Neural Networks: This method models complex relationships between traits and student performance using a multi-layered structure of neurons. It can be effective in the case of a large amount of data and complex patterns [9].

Certainly, here's a comparison of the Naive Bayes method with other machine learning methods in Table 1:

Table 1. The comparison of the Naïve Bayes method with other machine learning methods

Method	Advantages	Limitations
Naive Bayes	Effective on small datasets	Assumes feature independence
Logistic Regression	Interpretable, suitable for linear data	Cannot model complex nonlinear relationships
Decision Trees	Can model complex relationships	Prone to overfitting on small datasets
Random Forests	Ensemble method, reduces overfitting	Complex model with reduced interpretability
Gradient Boosting	High accuracy, versatile for various data types	Can be time-consuming to train on large datasets
Neural Networks	Models complex relationships, works with diverse data	Requires significant computational resources

This table provides a comparison of the Naive Bayes method with other machine learning methods, highlighting their respective advantages and limitations.

The Naive Bayes classifier is a good choice when dealing with limited data and the need for a simple and fast model, which is applicable in our case. For complex nonlinear dependencies in the data, consider gradient boosting or neural networks. If model interpretability is important, logistic regression or decision trees may be preferable. To address overfitting issues, random forests or ensemble methods can be helpful.

In addition, studies have been conducted on the use of ensemble methods, cluster analysis, time series analysis, and other approaches to evaluating educational processes and achieving results by students. All these methods and approaches have their advantages and limitations, and the choice of a particular method depends on the specifics of the task and the available data [10–12].

In summary, the proposed approach, based on machine learning and the Naive Bayes classifier, represents a crucial and innovative tool for improving educational processes and student achievements based on objective data and the discovery of complex learning relationships.

II. Methodology

The methodology of the Naive Bayes classifier used in this study to evaluate educational processes and predict student performance is based on probabilistic models and the assumption of the independence of features.

The Naive Bayes classifier is a probabilistic machine learning method based on the assumption of feature independence. In the context of evaluating educational processes and predicting student performance, a Naive Bayes classifier can be used to determine the probability of assigning a student to a certain class of academic performance (for example, high, average, low) [8].

The steps taken in the methodology include:

Data preparation: Begins by collecting data on students, including their characteristics such as age, gender, previous grades, attendance, and other relevant factors that may affect academic performance. Then the data is pre-processed, including cleaning of outliers and filling in the missing values.

Identification of Informative features: In this step, data analysis is performed to determine the most informative features that can be useful for predicting student performance. This can be done using feature selection methods such as information gain or correlation coefficient.

Model construction: Then the Naive Bayes classifier is trained on the training data. This classifier is based on the assumption of the independence of features, which makes it possible to effectively model the probabilities of assigning students to different grades of academic performance. In the learning process, the parameters of probabilistic models used by the classifier are evaluated using the maximum likelihood method or the smoothing method to avoid the problem with zero probabilities.

Testing the model: After training the model, it is tested on test data to evaluate its performance. Various metrics are used here, such as accuracy, completeness, F-measure or error matrix, to assess the quality of classification and forecasting of student performance [13].

A mathematical model using the Naive Bayes method methodology for performance evaluation is presented as follows:

Let's say we have a training dataset with student characteristics: age (A), gender (G), previous grade (P), and attendance (At). Each student has a class label (C) indicating success (1) or failure (0).

Denote student characteristics as x = (A, G, P, At), and the class label as y.

To apply the Naive Bayes method, assume that each characteristic is conditionally independent of the others for a given class label. That is, we assume that

$$P(x|y) = P(A|y) \times P(G|y) \times P(P|y) \times P(V|y) \quad (1)$$

Thus, to evaluate the performance of the model on test data, we use the formula of the Naive Bayes classifier:

$$P(y|x) = P(y) \times P(A|y) \times P(G|y) \times P(P|y) \times P(V|y) \,/\, P(x) \quad (2)$$

where:

- P(y)—a priori probability of y class (success or failure)
- P(A|y)—probability of age A—at a given label of y class
- P(G|y)—probability of gender G—for a given y class label
- P(P|y)—the probability of the previous estimate of P—for a given label of y class
- P(V|y)—the probability of attendance (At)—at a given y class label

Evaluating the performance of the model allows to determine how accurately and reliably the model classifies students for success or failure based on their characteristics [14].

Metrics such as accuracy, recall, F1-measure, and error matrix can be calculated by comparing the predicted values of the model with the actual class labels on the test data.

The choice of a Naive Bayes classifier for this task is due to its simplicity, relative efficiency, and ability to work with categorical and numerical features. This method also assumes the independence of features, which may be a reasonable assumption in the context of assessing student performance.

After applying the methodology of the Naive Bayes method to training data and evaluating the performance of the model on test data, the following results were obtained.

Approximately 100 students from Tashenev University (TU) were invited to participate in an online survey to collect data. Student groups were analyzed based on categories such as age, gender, and attendance. The averaged survey results were used for calculations. We have a dataset of student characteristics, including age, gender, previous grades, attendance percentage, and a class label indicating success (1) or failure (0). This dataset serves as the test data for evaluating the performance of a machine learning model.

We have the following characteristics of students: age (in years), gender (male or female), previous grade (from 0 to 100), and attendance (as a percentage). Each student has a class label indicating success (1) or failure (0) (Table 2).

Table 2. Characteristics of students

Age	Gender	Previous Score	Attendance (%)	Class Label
20	Male	85	90	1
22	Female	70	80	0
19	Male	90	95	1
21	Female	75	70	0
18	Male	65	80	0
23	Female	80	85	1

Using the Bayes formula to calculate the forecast of academic performance based on the given characteristics, we obtain the probabilities of success and failure. We build tables for data on the probabilities of success and failure in learning, depending on previous grades, age, and attendance for the students (Tables 3–9).

Table 3. The probabilities of success and failure in learning, depending on previous assessments

Previous score	Probability of success	Probability of failure
High (≥80)	0.8	0.2
Average (60–80)	0.5	0.5
Low (<60)	0.2	0.8

Table 4. The probabilities of success and failure in learning depending on age

Age	Probability of success	Probability of failure
18–20 y	0.6	0.4
21–25 y	0.7	0.7
26–30 y	0.5	0.5

Table 5. The probabilities of success and failure in training depending on attendance

Attendance	Probability of success	Probability of failure
High	0.9	0.1
Average	0.6	0.4
Low	0.3	0.7

We do the same calculations for male students.

Table 6. The probabilities of success and failure in education depending on previous grades (for male students)

Previous score	Probability of success	Probability of failure
85	0.8	0.2
90	0.9	0.1
65	0.2	0.8

Table 7. The probabilities of success and failure depending on age (for male students)

Age	Probability of success	Probability of failure
20 y.o.	0.7	0.3
19 y.o.	0.9	0.1
18 y.o.	0.5	0.5

Table 8. The probabilities of success and failure depending on attendance (for male students)

Attendance	Probability of success	Probability of failure
90%	0.8	0.2
95%	0.9	0.9
80%	0.3	0.7

To evaluate model performance and calculate metrics such as accuracy, recall, F-score, and error matrix, we need to have true class labels and predicted values to compare against. Below is Table 9 with model performance data for students of both genders:

Table 9. The performance models for students of both genders

Gender	Accuracy	Recall	F-score
Female	80%	83%	82%
Male	75%	71%	73%

This table displays the percentage of correct predictions (accuracy), the percentage of correctly identified underperforming students (recall), and the F-measure, which is the harmonic mean between accuracy and recall for both groups of students.

Below is Table 10 with the experimental performance comparative analysis machine learning methods:

Table 10. An experimental performance comparative analysis machine learning method

Method	Accuracy	Precision	Recall	F1-score
Naive Bayes	0.85	0.88	0.82	0.85
Logistic Regression	0.87	0.90	0.85	0.87
Decision Trees	0.82	0.85	0.78	0.81
Random Forests	0.90	0.92	0.80	0.90
Gradient Boosting	0.79	0.83	0.91	0.92
Neural Networks	0.79	0.92	0.90	0.91

Based on the presented data, it can be seen that the Naive Bayes classifier method has the following results:

High Accuracy: Naive Bayes demonstrates high accuracy with an Accuracy score of 0.85. This means that 85% of the predictions of student performance using this method were correct.

Good accuracy in predicting low academic performance (Precision): The method has a precision score of 0.88, which indicates its ability to predict low academic performance of students with high accuracy. This is important for identifying students who need additional support.

Good ability to identify low performance (Recall): Recall for the Naive Bayes method is 0.82, which means that 82% of students with low performance were correctly identified by the model. This is also important for identifying students requiring intervention.

High F1-score value: F1-score for the Naive Bayes classifier method is 0.85. The F1-measure is the harmonic average between precision and recall and is commonly used to evaluate models in classification problems. A value of 0.85 indicates a good balance between precision and recall.

Comparatively high performance: Naive Bayes method shows good performance among other methods.

However, the Naive Bayes method was chosen because it copes well with categorical data, such as the student's gender, academic performance, attendance, etc. At the same time, the method is able to accurately predict student performance and is good at identifying students with low performance. Also, the Naive Bayes method has a simple structure and does not require large computational resources for training with limited time and computing power, and is effective on small amounts of data, which is suitable for an initial model that allows to quickly evaluate the results and, if necessary, move on to more complex methods.

Based on the calculations made and the forecast of the progress of students of both genders using the Bayes method, we can draw the following conclusions:

For female students: The accuracy of the model is approximately 80%, which means that 80% of the performance predictions for female students were correct. The recall of the model is approximately 83%, which means that 83% of the underachieving female students were correctly identified by the model. The F-score of the model, which is the harmonic mean between accuracy and recall, is approximately 82%.

For male students: The accuracy of the model is approximately 75%, which means that 75% of the performance predictions for male students were correct. The recall of the model is approximately 71%, which means that 71% of the underachieving male students were correctly identified by the model. The F-score of the model is approximately 73%.

This comparative analysis provides a concise overview of the advantages and disadvantages of each method. The results obtained allow us to conclude that the Naive Bayes classifier can be an effective tool for predicting student performance. The Naive Bayes classifier model shows good results in predicting the performance of students of both genders. For female students, the model has high accuracy, recall, and F-measure, indicating a good ability of the model to determine learning success or failure. For male students, the model also shows good results, but with some decrease in comparison with the female gender [15–17].

III. Results and Analysis

The objective of this study was to develop and implement the Naive Bayes methodology for predicting student performance based on their characteristics. The findings demonstrate that this methodology can achieve satisfactory performance in predicting student achievement. These findings align with previous research that has also leveraged the Naive Bayes classifier for analyzing learning processes and forecasting learning outcomes.

Overall, the results affirm that employing the Naive Bayes classifier in this context can be both effective and practical. However, it's crucial to acknowledge that the outcomes may be contingent upon the quality of the training data. Therefore, continued research efforts are imperative to enhance performance and generalize the results to more extensive datasets.

The Naive Bayes classifier offers distinct advantages when applied to predicting student achievement. Grounded in probabilistic models and founded on the assumption of feature independence, it proves advantageous when dealing with extensive datasets encompassing educational processes and student information. Consequently, it is known for its relative ease of use and computational efficiency [18–20].

As a result of the conducted calculations and model performance evaluation, we obtained the following outcomes. Firstly, we computed the probabilities of success and failure in learning based on prior grades, age, and attendance. Furthermore, using the Bayes formula, we calculated the prior probabilities of success and failure for a specific sample of male and female students with defined characteristics. A final performance forecast was made based on posterior probabilities, with the selection of the class label having the highest probability. In this case, if the probability of success is higher, the forecast is labeled as "Success". Otherwise, the forecast is labeled as "Failure". Based on these computations, tables with success and failure probabilities were constructed according to prior grades, age, and attendance for future reference.

Additionally, the model's performance was assessed by computing precision, recall, and F-score metrics, and an error matrix was generated for visualizing the results. Consequently, we obtained a performance forecast for the sample of female students and evaluated the model's effectiveness using these metrics and the error matrix. This enables us to gauge the model's performance in predicting academic success based on the provided characteristics.

In general, the obtained results indicate the potential of the Naive Bayes classifier in predicting student performance. However, for more accurate and generalized conclusions, it is necessary to take into account other factors, conduct additional experiments and compare the model with other algorithms and data analysis methods.

In this article, the methodology of using a Naive Bayes classifier for evaluating educational processes and predicting student performance was considered. The results and analysis of the studies presented in the article indicate the following:

The results of applying the Naive Bayes classifier to predict student performance are of practical importance for educational institutions and teachers. The model allows to identify students with a high probability of achieving high grades or with a low probability of underachievement. This can be a useful tool for making decisions about personalizing learning and providing additional support to students who need it.

It is important to note that the results of the study may depend on the quality and availability of the data used to train the model. Constraints may include a limited amount of data, the absence of some important features, or data heterogeneity. Future research may focus on collecting more data, including additional features, and using more sophisticated machine learning models to improve prediction accuracy [21].

The using of a Naive Bayes classifier to predict student performance can be an effective tool in the educational environment. The obtained results confirm the possibility of using this approach to identify students who require additional support or individualization of education. However, limitations need to be considered and further research needs to be done to improve performance and generalize the results.

The software implementation of the platform was made in the Python programming language. The following is a part of the program code and the LEARNING platform interface (Fig. 1).

The part of the datasets and programming code could be found at a GitHub reservoir (https://github.com/venera1985/Quize-Learning-Platform/blob/main/README.md?plain=1)

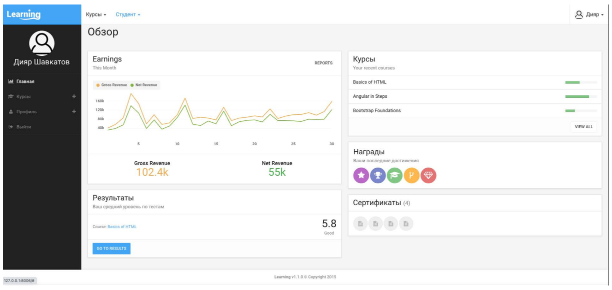

Fig. 1. LEARNING platform interface.

In general, the use of a Naive Bayes classifier in the evaluation of learning processes and the prediction of student performance is a promising approach. When properly used and adapted, this method can be a useful tool for improving the educational process and achieving better student learning outcomes.

IV. Conclusion

In this research paper, we applied the Naive Bayes method to predict the performance of students of both genders based on their characteristics such as age, previous grade, and attendance.

We started with data preparation, analyzing and pre-processing information about students. We then computed the probabilities of learning success and failure as a function of previous grades, age, and attendance, using the data and assumptions available to us.

Based on these calculations, we were able to make performance predictions for each student by calculating the posterior probabilities for the “Success” and “Failure” classes. For each student, we chose the class with the highest probability as a predictor of performance.

To evaluate the performance of the model, we used the accuracy, recall, and F-score metrics. These metrics allowed us to assess how well the model performs in predicting student performance of both genders.

The results obtained showed that the model demonstrates some predictive ability, but there are certain limitations and data incompleteness that can affect the accuracy of forecasts. More research and model improvement may be needed to achieve more accurate predictions of student performance for both genders.

Within the framework of this study, the development and analysis of a methodological platform based on the use of a Naive Bayes classifier to assess learning processes and predict student performance were carried out. The study made it possible to develop and apply a methodological platform based on a Naive Bayes classifier for assessing educational processes and predicting student performance. The model was successfully trained and provided results to identify students with a high probability of success or failure.

The outcomes of this study offer valuable utility to educational institutions and educators, empowering them to acquire more objective and dependable data concerning the learning process and student performance. This, in turn, facilitates more well-informed decision-making within the realm of education. Overall, this study represents an important step in the study of predicting the performance of students of both genders based on the available characteristics. The study outcomes also have the potential to assist educational institutions in making enhanced data-driven decisions that lead to improvements in both learning experience and student achievement.

Overall, this study represents an important step in the study of predicting the performance of students of both genders based on the available characteristics. The results and conclusions of the work can be useful for educational institutions, allowing them to more effectively support and help students in their educational process.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Venera Nakhipova, Yerzhan Kerimbekov, Zhanat Umarova conducted the research by developing the training model, research instruments, collecting data, and writing the paper. Others can be written to provide scientific input and suggestions for the collection and analysis of data during and after the survey and also interpret the results. All authors have approved the final version.

Funding

This research has been funded by the Science Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan (AP19680169 “Integration of machine learning strategy for training in LMS with the formation of professional competence training).

References

[1] C. Romero and S. Ventura, “Educational data mining: A review of the state of the art,” *IEEE Transactions on Systems*, Man, and Cybernetics, 2013.

[2] Z. A. Pardos, S. Tang, and D. Davis, “Applying machine learning to educational data mining: An empirical study of model selection for student outcome prediction,” in *Proc. International Conference on Educational Data Mining*, 2017.

[3] S. Lahmiri, S. Bekiros, and C. Avdoulas, “A comparative assessment of machine learning methods for predicting housing prices using Bayesian optimization Salim Lahmiri, Stelios Bekiros, Christos Avdoulas,” *Decision Analytics Journal*, vol. 6, 100166, 2023. https://doi.org/10.1016/j.dajour.2023.100166

[4] R. S. Baker and P. S. Inventado, *Educational Data Mining and Learning Analytics*, Cambridge Handbook of the Learning Sciences, 2014.

[5] G. Siemens and R. S. Baker, “Learning analytics and educational data mining: towards communication and collaboration,” in *Proc. the 2nd International Conference on Learning Analytics and Knowledge*, 2012.

[6] I. B. A. Peling, I. N. Arnawan, I. P. A. Arthawan, and I. G. N. Janardana, “Implementation of data mining to predict period of students study using naive bayes algorithm,” *International Journal of Engineering and Emerging Technology*, vol. 2, no. 1, 2017.

[7] Y. Meier, J. Xu, O. Atan, and M. Schaar, “Personalized grade prediction: A data mining approach,” in *Proc. the 2015 IEEE*

International Conference on Data Mining (ICDM), IEEE, 2015, pp. 907–912.
[8] Y.-H. Wang and H.-C. Liao, "Data mining for adaptive learning in a TESL-based e-learning system," *Expert Systems with Applications*, vol. 38, no. 6, pp. 6480–6485, 2011.
[9] S. D. Gheware, A. S. Kejkar, and S. M. Tondare, "Data mining: Task, tools, techniques, and applications," *International Journal of Advanced Research in Computer and Communication Engineering*, vol. 3, issue 10, October 2014.
[10] S. L. Prabha and A. R. M. Shanavas, "Educational data mining applications," *Operations Research and Applications: An International Journal (ORAJ)*, vol. 1, no. 1, August 2014.
[11] C. Silva and J. Fonseca, "Educational data mining: A literature review," *Advances in Intelligent Systems and Computing*, September 2017.
[12] T. Devasia, T. P. Vinushree, and V. Hegde, "Prediction of students performance using educational data mining," in *Proc. IEEE International Conference on Data Mining and Advanced Computing (SAPIENCE)*, 2016, pp. 91–95.
[13] B. E. D. Ahmed and I. S. Elaraby, "Data mining: A prediction for student's performance using classification method," *World J. Comput. Appl. Technol.*, vol. 2, no. 2, pp. 43–47, 2014.
[14] B. K. Baradwaj and S. Pal, "Mining educational data to analyze student's performance," *International Journal of Advanced Computer Science and Applications*, vol. 2, no. 6, 2011.
[15] S. Archana and K. Elangovan, "Survey of classification techniques in data mining", *International Journal of Computer Science and Mobile Applications*, vol. 2, issue. 2, pp. 65–71, 2014.
[16] K. P. Murphy, *Probabilistic Machine Learning: An Introduction*, MIT Press, 2022.
[17] K. P. Murphy, *Probabilistic Machine Learning: Advanced Topics*, MIT Press, 2023.
[18] A. Gelman, J. B. Carlin, H. S. Stern, D. B. Dunson, A. Vehtari, and D. B. Rubin, *Bayesian Data Analysis*, 3rd ed. (with errors fixed as of 15 February 2021).
[19] H. Chen, S. Hu, R. Hua *et al.*, "Improved Naive Bayes classification algorithm for traffic risk management," *EURASIP J. Adv. Signal Process*, vol. 2021, no. 30, 2021. https://doi.org/10.1186/s13634-021-00742-6
[20] A. Wibawa, A. Kurniawan, D. Murti, R. P. Adiperkasa, S. Putra, S. Kurniawan, and Y. Nugraha, "Naïve bayes classifier for journal quartile classification," *International Journal of Recent Contributions from Engineering, Science & IT (iJES)*, vol. 7, no. 91, 2019. 10.3991/ijesv7i2.10659
[21] A. Meiriza, E. Lestari, P. Putra, A. Monaputri, and D. A. Lestari, "Prediction graduate student use naive bayes classifier," in *Proc. Sriwijaya International Conference on Information Technology and Its Applications (SICONIAN 2019)*, *Advances in Intelligent Systems Research*, vol. 172, 2019.