International Journal of Applied Mathematics
Volume 38 No. 11s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

EVALUATION AND VALIDATION OF MACHINE LEARNING MODELS TO SUPPORT EDUCATIONAL DECISION-MAKING

Kainizhamal Iklassova[1], Anna Shaporeva[1], Aigul Shaikhanova[2]
Corresponding author: Anna Shaporeva (*Email: annvolkova88@gmail.com)*
[1]Manash Kozybayev North Kazakhstan University, Kazakhstan
kiklasova1205@gmail.com
ORCID: https://orcid.org/0000-0002-8330-4282

[2]Manash Kozybayev North Kazakhstan University, Kazakhstan
E-mail: annvolkova88@gmail.com
ORCID: http://orcid.org/0000-0002-6211-5634

[3]L.N. Gumilyov Eurasian National University, Kazakhstan
E-mail: aigul.shaikhanova@gmail.com
ORCID: https://orcid.org/0000-0001-6006-4813

Abstract

Artificial intelligence is one of the key factors in the development of modern information systems today. Its integration makes it possible to expand the functionality of traditional information systems, increase the efficiency of data processing and provide support for intelligent decision-making. One of the directions of artificial intelligence development is its integration into the information systems of universities for making managerial decisions. The article presents a comparative analysis of three machine learning models – decision tree, random forest, and gradient boosting – to solve the diagnostic problem of identifying the causes of students' academic failure. In the context of digitalization of education, predictive analytics is shifting the focus from simple forecasting to understanding the factors underlying learning difficulties. The research aims to evaluate models not only in terms of accuracy, but also in terms of interpretability of their results, which is a key factor for their practical application in the decision-making process.

Based on a synthetic dataset simulating the behavior of students in a digital environment, training and model testing were conducted. The results showed high predictive efficiency of all models (accuracy from 98% to 100%). The decision tree has demonstrated 99% accuracy with full transparency of logic, which makes it a valuable tool for generating effective recommendations. XGBoost has achieved 100% accuracy, confirming its status as a stateoftheart algorithm for tasks with clearly structured data.

The paper concludes that it is necessary to choose a model based on the balance between accuracy and interpretability, and outlines the prospects for integrating high-precision models with explicable AI (XAI) methods to create effective decision support systems in education.

Keywords: artificial intelligence, inverse problems, explicable artificial intelligence, data analysis

International Journal of Applied Mathematics
Volume 38 No. 11s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

1. Introduction

The modern educational environment is undergoing a fundamental transformation due to the widespread introduction of digital technologies. Learning management Systems (LMS), massive open online courses (MOOCs) and other educational platforms generate unprecedented amounts of data on students' interaction with educational content, their behavior and academic performance [1]. This phenomenon, often described as a "gold mine of educational data," opens up unique opportunities for in-depth analysis and improvement of educational processes. In response to this challenge, an interdisciplinary field has emerged – Educational Data Mining (EDM) [2], which applies machine learning, statistics, and data mining methods to solve pressing educational challenges.

The main goal of EDM is to move from reactive to proactive educational strategies. Instead of detecting academic failure at the end of a semester, predictive analytics allows students at risk to be identified at the earliest stages of their studies. This provides an opportunity for timely and personalized intervention (justintime, JIT) aimed at preventing academic difficulties and improving educational outcomes [3].

However, as the field evolves, the focus of research shifts from simple forecasting to a more complex diagnostic task. While early work in the field of EDM focused on the question of whether a student would be able to successfully complete a course, today the question of why they might encounter difficulties is becoming more relevant. This concept, which can be described as an "inverse problem" in educational analytics, involves not just predicting the fact of academic failure, but identifying its root causes [4]. Such a transition from prediction to diagnosis requires not only accurate, but also interpretable models, whose internal logic can be analyzed to understand the key factors influencing the outcome. Providing the teacher with information that the student is at risk has limited value. It is much more important to provide him with actionable insights explaining exactly which aspects, whether it's skipping classes, low activity on the platform, or difficulties with certain topics, are causing the predicted risk [5].

Thus, modern educational analytics faces a number of fundamental challenges. It is necessary to find a balance between the accuracy of predictive models and their interpretability, ensure the ethical validity of algorithmic solutions, eliminating bias and injustice, and, most importantly, create models that not only predict, but provide teachers and administrators with practical, explicable and effective recommendations [6].

The purpose of this study is to compare and justify the use of machine learning models – decision trees, random forests, and gradient boosting - to solve the inverse problem of identifying the causes of student academic failure, as well as evaluating their effectiveness in terms of accuracy and interpretability.

The scientific novelty of the work lies in shifting the focus from a purely technical competition of models based on accuracy metrics to a more holistic assessment of their suitability as decision support tools in the educational process. The study directly addresses a gap in the literature related to the lack of systematic comparisons of models from different points of the accuracy-interpretability spectrum to solve a diagnostic rather than a predictive task in educational analytics.

Received: August 07, 2025 Published: November 29, 2025

International Journal of Applied Mathematics
Volume 38 No. 11s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

2. Literature review

Choosing a machine learning model for predicting student academic performance is a key decision that determines not only the accuracy of forecasts, but also their practical applicability in the educational process. In the EDM literature, a continuum of approaches can be distinguished, with simple and transparent models at one end and complex, high–precision, but opaque "black boxes" at the other. Choosing a particular model often represents a compromise between accuracy and interpretability, two fundamental but often conflicting goals.

The work [7] emphasizes that there is no universally best algorithm; the optimal choice depends on the specific task, the characteristics of the data and, what is especially important in the educational context, on the needs of end users – teachers, students and administrators.

The authors of [8] argue that the strategic choice of models for analysis covering various points of this spectrum allows us to systematically explore the trade-off between accuracy and interpretability, which is a deeper task than a simple comparison of performance metrics.

The category of transparent or "whiteboxes" includes models whose internal structure and decision-making logic are intuitive to humans. Their main advantage lies in their simulability and decomposability – the ability of a person to trace and reproduce the process of obtaining a result, understanding the contribution of each component of the model [9].

According to the authors of [10], Decision Trees (DT) occupy a central place among such models in EDM. This algorithm builds a hierarchical structure consisting of decision rules of the form "IF THERE is". Each rule checks the value of a certain feature and, depending on the result, directs the object to the next node until the final "sheet" containing the forecast is reached.

Considering the Decision Trees model, the authors [11] note that visualization of the decision tree makes it possible to visually demonstrate the logic of classification, which contributes to the confidence of non-technical specialists in the model.

In [12], the strengths and weaknesses of the Decision Trees model are considered. The strengths of decision trees are their high interpretability, ease of visualization, and minimal data preparation requirements. They are often used as a baseline for comparison with more complex algorithms, both in terms of accuracy and clarity of results. As the authors note, this approach has significant weaknesses. Decision trees are prone to overfitting, especially on data with a large number of features, which means that they can perfectly fit the training sample, but generalize poorly to new data. In addition, they may have difficulty modeling complex nonlinear relationships in the data, which often leads to lower accuracy compared to ensemble or deep models. Small changes in the training data can lead to the construction of a completely different tree, which indicates their instability.

In an effort to overcome the limitations of simple models and achieve maximum predictive accuracy, EDM researchers are increasingly turning to ensemble methods and deep learning models. These approaches, as a rule, demonstrate significantly higher performance, but at the cost of loss of transparency, which is why they are often called "black boxes" [13, 14].

Ensemble methods, such as random forest and gradient boosting, build their forecast based on an aggregated solution of a set of basic models, most often decision trees.

International Journal of Applied Mathematics
Volume 38 No. 11s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

According to the authors of [15,16], Random Forest (RF) is one of the most popular and efficient algorithms in EDM. It creates a set of decision trees, each of which is trained on a random subsample of data and features, and the final forecast is determined by "voting" (in classification tasks) or averaging (in regression tasks). This approach can significantly reduce the tendency to overfitting, which is typical for single trees, and increase the stability and accuracy of the model. However, combining hundreds of trees into one ensemble makes it impossible to simply follow the logic of decision-making, as in the case of a single tree, which reduces direct interpretability.

Studies [17, 18] have shown that Gradient Boosting and its high-performance implementation XGBoost represent an even more powerful class of ensemble methods. Unlike a random forest, where trees are built independently, in gradient boosting they are created sequentially: each subsequent tree is trained on the mistakes of the previous one. This iterative process allows the model to fine-tune to the data and achieve the highest accuracy, making XGBoost one of the most effective tools for solving forecasting problems in education. However, this consistent error correction makes the model even more complex and opaque than a random forest.

Deep Learning models, in particular, artificial neural networks (ANN) and their more complex variants, such as recurrent neural networks (LSTM), represent the pinnacle of predictive accuracy, especially when working with complex sequential data, for example, with students' clickstreams in online courses [19]. Their ability to automatically extract features from raw data and model high-order nonlinear dependencies allows them to achieve results that are not available for most other methods. Nevertheless, they are the quintessence of a "black box": the decision-making process is hidden in a multi-layered architecture of thousands of interconnected neurons with weights, which makes it almost impossible for human understanding [20].

Thus, there is a fundamental compromise between accuracy and interpretability in the field of EDM. According to the authors of [21], as the range of models moves from a simple decision tree to a random forest, gradient boosting, and finally to deep neural networks, predictive accuracy usually increases, but the internal transparency of the model steadily decreases. This trade-off is a central problem that the rapidly developing field of explicable artificial intelligence (XAI) is striving to solve.

In the study [22], the desire for ever higher accuracy of forecasts led to the dominance of complex models, but in the educational context, accuracy itself is insufficient. A decision made by a "black box", even if it is statistically correct, remains useless or even harmful if the end users teachers, administrators, and students themselves) cannot trust it and act on it. Thus, explainability becomes not just a desirable technical characteristic, but an ethical and pedagogical imperative.

As the authors of [23] note, the success of any predictive model in education directly depends on two key components: the quality and relevance of the data (features) used, as well as the methodological rigor of their processing and selection. Modern educational datasets are characterized by high dimensionality, which requires the use of advanced approaches to engineering and feature selection to build accurate and robust models.

As described in the study [24], modern educational datasets can include hundreds or even thousands of potential features, especially when analyzing detailed interaction logs (clickstreams). Such a high data dimension creates a number of problems: the risk of overfitting the model increases, computational complexity increases, and "noise" is introduced into the model due to irrelevant or

International Journal of Applied Mathematics
Volume 38 No. 11s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

redundant features. To solve these problems, feature selection methods are used, which make it possible to identify the most informative subset of predictors.

Despite significant progress, several significant gaps can be identified in existing research that limit the practical application of predictive analytics in the real educational process. Firstly, there is a clear bias towards forecasting rather than diagnosis. Most studies compare different models based on their ability to predict the final outcome – whether a student will pass the exam or not, whether they will receive a certificate, or whether they will be expelled. However, as noted earlier, it is much more important for a teacher to understand the reasons for potential failure. There is a clear lack of work that systematically evaluates and compares models in terms of their suitability for solving the "inverse problem" of diagnosing the factors underlying academic difficulties. This task requires models not only to be accurate, but also to be able to provide causally meaningful and interpretable explanations.

Secondly, there is no systematic comparison of models located in different parts of the accuracy-interpretability spectrum to solve this particular diagnostic problem. Many papers either focus on models of the same class (for example, comparing several high-precision "black boxes"), or use simple models as a base level, without going into depth analysis of the quality and nature of their explanations. It is rare to find studies that, like this work, conduct a controlled comparison of models strategically selected from different points of this spectrum – from the fully transparent Decision Tree Classifier to the balanced Random Forest Classifier and the highly accurate but opaque XGBoost – in order to analyze not only their predictive power, but also the quality of their logical conclusions in relation to the diagnosis of causes. academic failure.

3.Materials and methods

They represent different points on the "accuracy-interpretability" spectrum and provide a balance between stability and performance.:

The Decision Tree Classifier was chosen as the basic, easily interpretable model. Its structure in the form of a set of rules "IFLITO" allows you to visualize the logic of decision-making, which is critically important in the context of explicable AI (XAI) and for practical application by educators.

The Random Forest Classifier, as an ensemble model, was included to increase resistance to overfitting and evaluate the impact of the ensemble approach on accuracy. It aggregates the solutions of multiple trees, which reduces variance and increases the generalizing ability of the model.

XGBoost (Extreme Gradient Boosting) was selected as one of the most powerful and effective tools of modern machine learning. It presents advanced boosting techniques and is particularly effective when dealing with unbalanced classes, often demonstrating stateoftheart accuracy results.

The following is a mathematical description of the models.

Decision Tree. The model builds a classification by recursively dividing the feature space based on a criterion that minimizes uncertainty, for example, the Gini index.:

$$Gini(D) = 1 - \sum_{k=1}^{k} p_k^2 \qquad (1)$$

Received: August 07, 2025 Published: November 29, 2025

International Journal of Applied Mathematics
Volume 38 No. 11s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

pk – the proportion of objects of class k in the subset D

Random Forest. The model is an ensemble of N independent trees, each of which is trained on a random subsample. The final decision is made by voting:

$$\hat{y} = mode\{h_1(x), h_2(x), \dots, h_N(x)\} \quad (2)$$

$h_i(x)$ – the result of the ith tree

Random Forest reduces variance compared to a single tree and increases resistance to overfitting.

XGBoost. The model consistently builds an ensemble of trees, where each subsequent tree corrects the errors of the previous one. The regularized loss function is minimized:

$$\tau = \sum_{i=1}^{n} l(y_i, \hat{y}_i) + \sum_{t=1}^{T} \Omega(f_t) \quad (3)$$

$\Omega(f_t) = \gamma T + \frac{1}{2}\lambda\|\varpi\|^2$ – regulation of the number of leaves and weights;

$l(y_i, \hat{y}_i)$ – is an error function, for example, logistic or quadratic.

To conduct the experiment and avoid working with personal data, a synthetic dataset was generated, consisting of 1,000 records simulating student activity. The set included the following attributes:

– percentage of task completion,

– average grade,

– time on the platform,

– chat activity;

the number of missed classes.

The target variable (class) was the reason for the lag, which was assigned based on logical rules. For example, if a student had more than 5 passes, he was assigned the "Passes" class. This made it possible to create a controlled environment for evaluating the diagnostic ability of models.

The data was divided into training (70%) and test (30%) samples. All three models were trained on the same dataset. The Accuracy, Precision, Recall, F1score and error matrix metrics were used for the assessment.

The following hyper-parameters were used in the experiment, selected to ensure a balance between accuracy and protection against overfitting, shown in Table 1.

International Journal of Applied Mathematics
Volume 38 No. 11s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

Table 1– Hyperparameters

Model	Main parameters	Interpretability	Stability	Justification of choice
Decision Tree	max_depth=4, random_state=42	High	Average	Simplicity of structure, explainability of logic
Random Forest	n_estimators=100, max_depth=6	Average	High	The ensemble approach reduces the variance
XGBoost	n_estimators=100, learning_rate=0.1	Average	Very high	Better accuracy and resistance to rare classes

The use of these three models provides a balance between interpretability, stability and accuracy, which allows not only to build an effective classification, but also to make a meaningful comparison of artificial intelligence strategies in an educational environment.

4. Results and discussion

The training and testing of the models was carried out using hyperparameters selected to ensure a balance between accuracy and protection against overfitting. The results of training the DecisionTreeClassifier model are presented in Table 2.

Table 2. – Learning outcomes of the DecisionTreeClassifier model

Class	precision	recall	f1score	support
0: Ок	1.00	0.95	0.97	78
1: I don't understand	1.00	1.00	1.00	52
2: Inactive	0.80	1.00	0.89	4
3: Motivation	0.92	1.00	0.96	36
4: Omissions	1.00	1.00	1.00	130

This model showed high accuracy: accuracy 99%, macroF1score 0.96, for most classes precision and recall reached 1.00. A graph of the distribution of classes has also been constructed, demonstrating an adequate correlation of causes in synthetic data (the predominance of omissions, the balance of the remaining classes), presented in accordance with Figure 1.

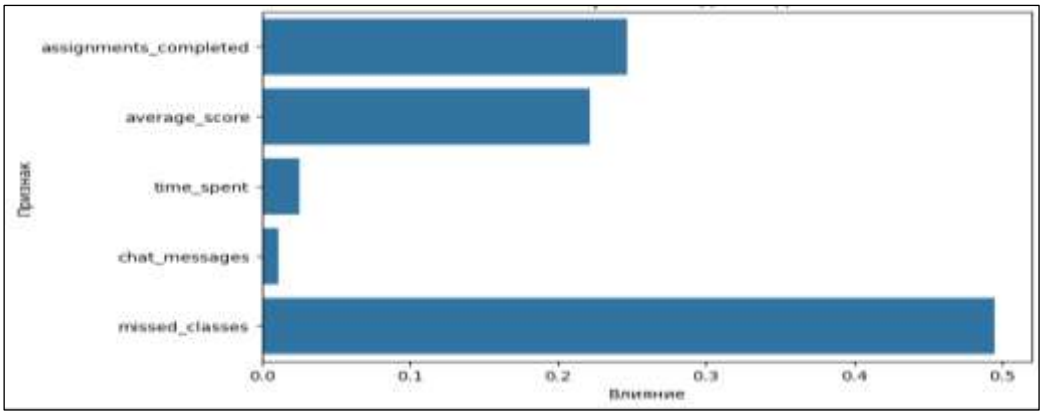

Figure 1– Importance of the features of the DecisionTreeClassifier model

International Journal of Applied Mathematics
Volume 38 No. 11s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

An error matrix was constructed that showed minor errors in classifying the "Ok" class as "motivation" or "inactive". The remaining classes were recognized without errors, presented in accordance with Figure 2.

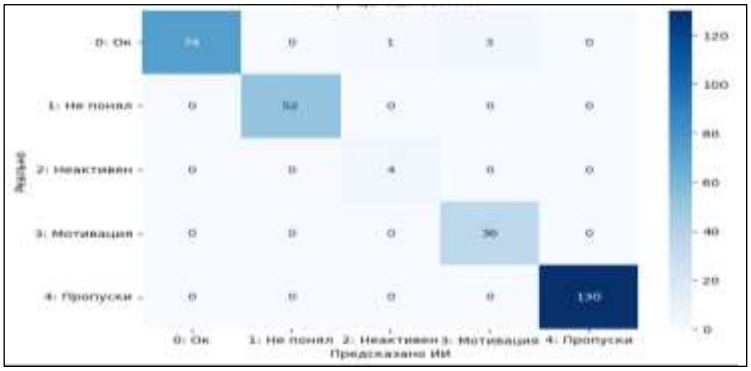

Figure 2 – Error matrix of the DecisionTreeClassifier model

The analysis of the importance of the signs was carried out – skipping classes had the greatest impact, followed by completing tasks and an average score. A decision tree has been built that visualizes the AI logic: an initial check for the number of passes, followed by an analysis of completed tasks and activities, presented in accordance with Figure 3.

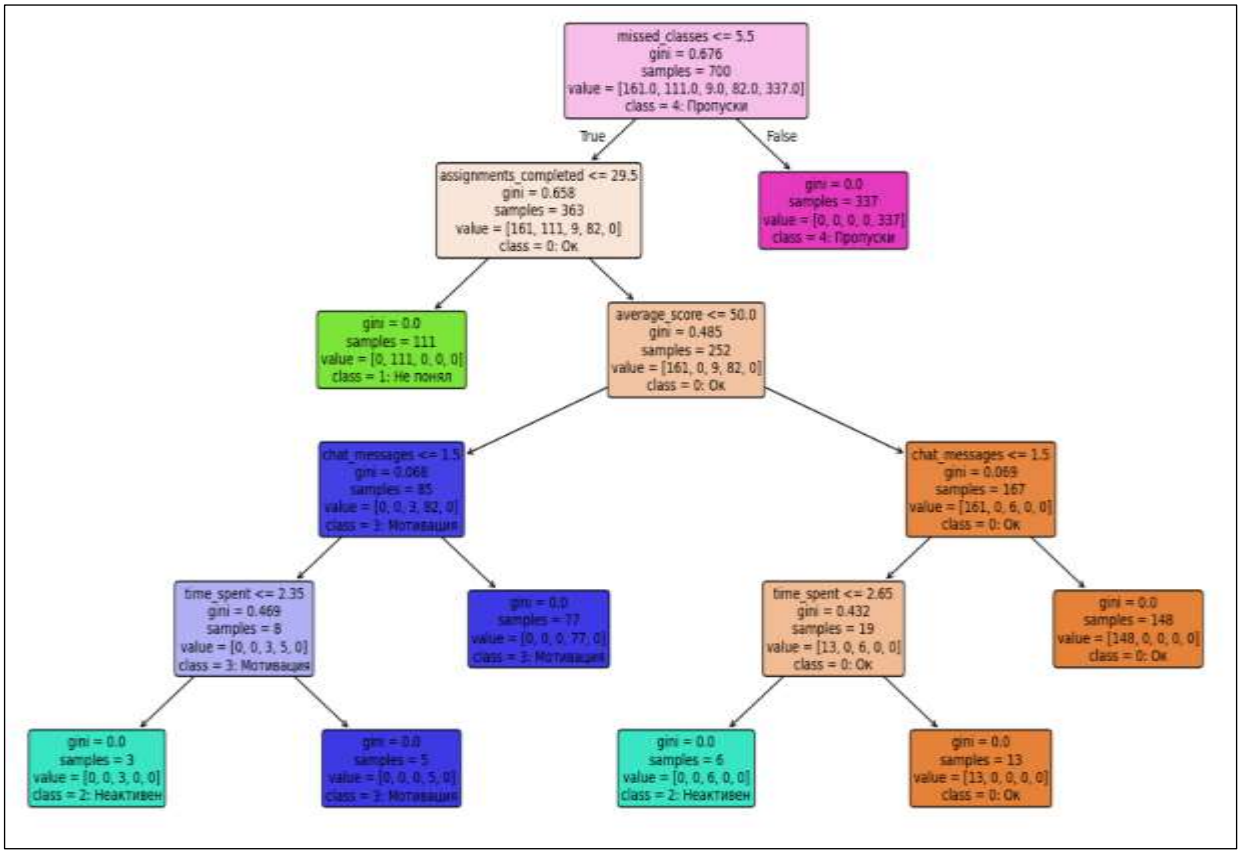

Figure 3– Decision tree for determining the backlog

Visualization of the decision tree (for the Decision Tree model) clearly demonstrated the logic of classification. The primary node of separation is checking the number of passes. If their number

International Journal of Applied Mathematics
Volume 38 No. 11s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

exceeds the threshold, the student is immediately classified as lagging due to "Missing Out". Otherwise, the model analyzes academic performance and activity to identify other causes.

The experiment confirmed the possibility of effectively using AI models to solve inverse problems in education. The model successfully identifies the causes of lag, is characterized by high accuracy and transparency of decision-making, which is important for practical application in educational systems. At the next stage, it is planned to expand the experimental base and comparative analysis of models in order to select the best solution according to the criteria of accuracy, stability and interpretability.

The next model to be trained was the Random Forest model. The learning outcomes are presented in table 3.

Table 3 – Results of Random Forest model training

Class	Класс	precision	recall	f1score	support
0: Ок	0: Ок	0,99	0,96	0,97	78
1: I don't understand	1: Не понял	1,00	1,00	1,00	52
2: Inactive	2: Неактивен	1,00	0,25	0,40	4
3: Motivation	3: Мотивация	0,88	1,00	0,94	36
4: Omissions	4: Пропуски	1,00	1,00	1,00	130

This model showed high accuracy: accuracy 98%, macroF1score 0.86, for most classes precision and recall reached 1.00. A graph of class distribution was constructed for the model, which demonstrated an adequate correlation of causes in synthetic data (the predominance of omissions, the balance of the remaining classes), presented in accordance with Figure 4.

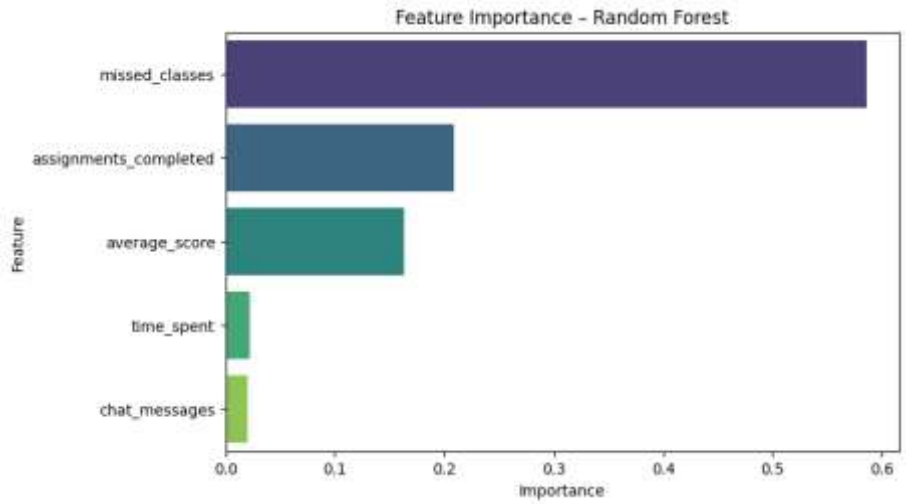

Figure 4 – The importance of the Random Forest model features

An error matrix was constructed that showed minor errors in classifying the "Ok" class as "motivation" or "inactive". The remaining classes were recognized without errors, presented in accordance with Figure 5.

International Journal of Applied Mathematics
Volume 38 No. 11s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

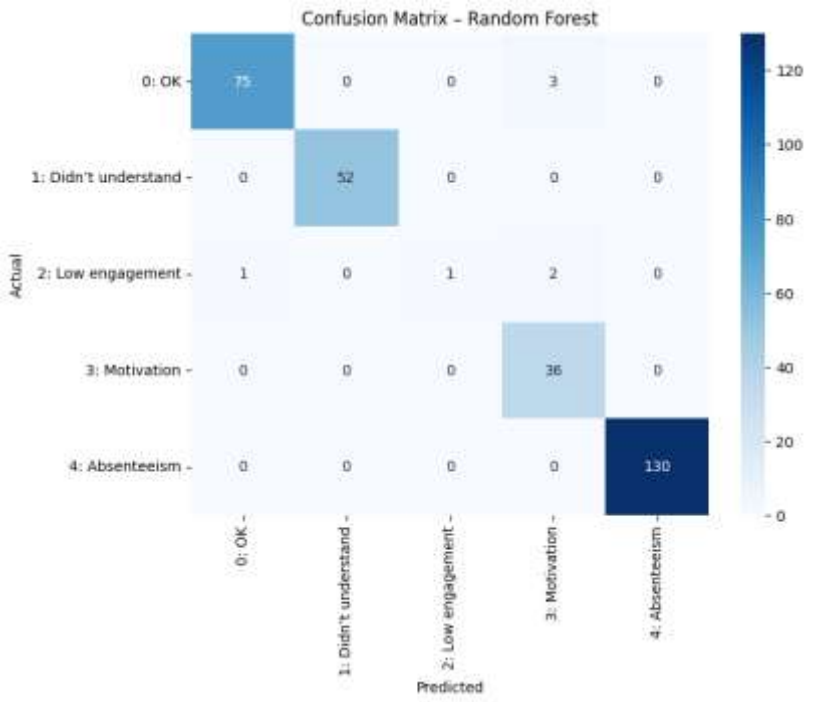

Figure 5 – Error matrix of the Random Forest model

These Random Forest models are almost identical to the DecisionTreeClassifier model. The last trained model was the XGBoost model. The results of the model training are presented in Table 4.

Table 4 – Results of XGBoost model training

Class	precision	recall	f1score	support
0: Ок	1.00	1.00	1.00	78
1: I don't understand	1.00	1.00	1.00	52
2: Inactive	1.00	1.00	1.00	4
3: Motivation	1.00	1.00	1.00	36
4: Omissions	1.00	1.00	1.00	130

This model showed the highest accuracy: accuracy 100%, macro F1score 1.00, all precision and recall classes reached 1.00. A graph of class distribution was constructed for the model, which demonstrated an adequate correlation of causes in synthetic data (the predominance of omissions, the balance of the remaining classes), presented in accordance with Figure 6.

Received: August 07, 2025 Published: November 29, 2025

International Journal of Applied Mathematics
Volume 38 No. 11s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

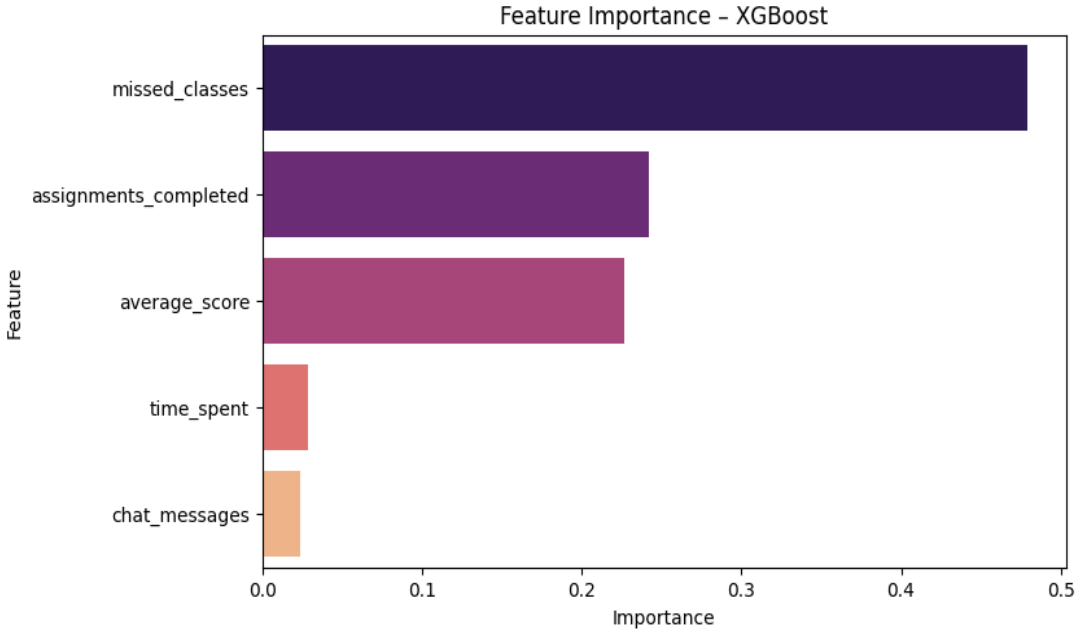

Figure 6 – The importance of the features of the XGBoost model

An error matrix was constructed that showed minor errors in classifying the "Ok" class as "motivation" or "inactive". The remaining classes were recognized without errors, presented in accordance with Figure 7.

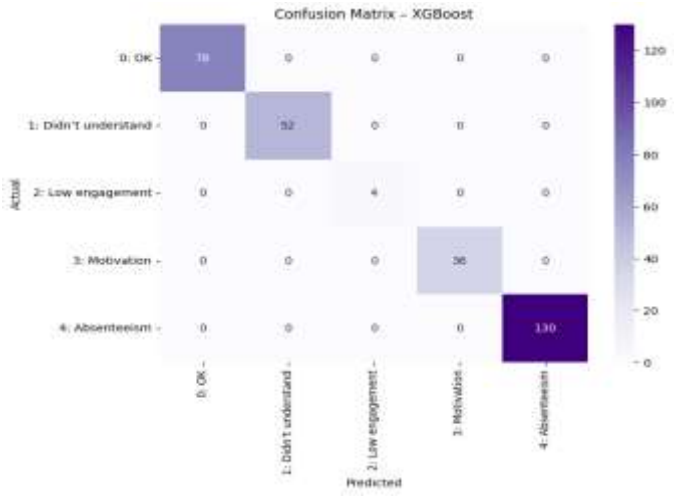

Figure 7– Error matrix of the XGBoost model

Figure 8 shows a comparative analysis of the accuracy of machine learning models. Table 5 shows the comparative performance metrics of the models in the test sample.

Received: August 07, 2025 Published: November 29, 2025

International Journal of Applied Mathematics
Volume 38 No. 11s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

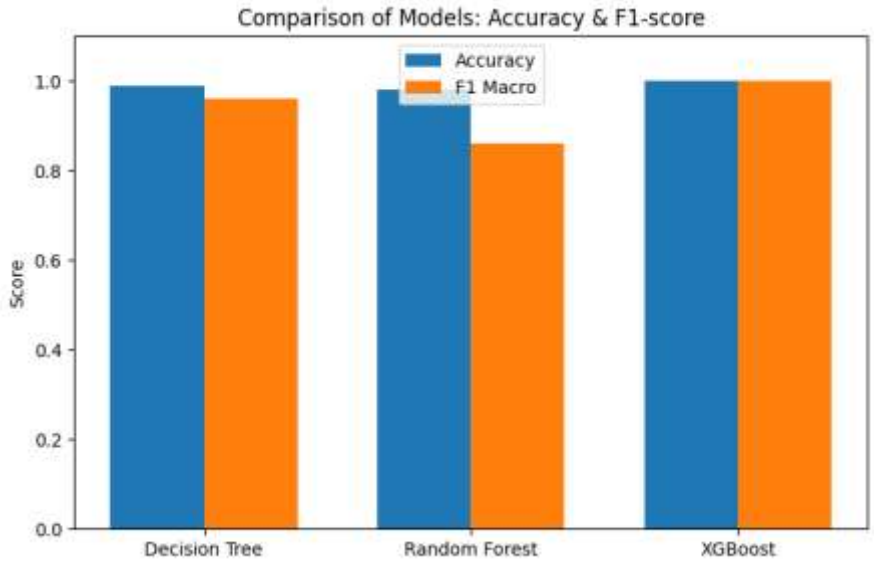

Figure 8– Comparative analysis of the accuracy of machine learning models

Table 5–Comparative metrics of model performance in the test sample.

Model	Accuracy	Macro Avg F1score	Key observations
Decision Tree	99.0%	0.96	High accuracy with full interpretability. Minor errors are only for the "Inactive" class (low support).
Random Forest	98.0%	0.86	High overall accuracy, but significantly lower F1score for the rare "Inactive" class (0.40), indicating difficulties with unbalanced data.
XGBoost	100%	1.00	Perfect accuracy on all metrics, error-free classification of all classes.

The analysis of the importance of signs conducted for all three models showed similar results: the most dominant factor for determining the cause of academic failure is the "number of missed classes." It is followed by the "completion percentage" and "average score". Signs such as "chat activity" and "time on the platform" turned out to be the least significant.

The experimental results confirm the high efficiency of machine learning models for solving the "inverse problem" in educational analytics – diagnosing the causes of academic failure. The comparative analysis revealed the key differences between the approaches, which allows us to formulate recommendations for their application.

Comparison of performance and interpretability. The XGBoost model has demonstrated perfect accuracy (100%), which confirms its reputation as one of the most powerful algorithms for working with structured data. However, such flawless results obtained on synthetic data require careful interpretation. In real conditions, educational data contains significantly more "noise", omissions, and nonlinear dependencies, and achieving 100% accuracy is unlikely. The main disadvantage of

XGBoost in the educational context remains its "black box" nature, which makes it difficult for educators to directly use its findings without using additional XAI tools such as SHAP or LIME.

The Decision Tree, showing only slightly lower accuracy (99%), is an excellent alternative when transparency is a priority. Its main advantage is its complete interpretability. The logical rules generated by the tree can be directly translated into recommendations for teachers, which corresponds to the request for effective and understandable insights. This makes it an ideal tool for systems where not only a forecast is required, but also an explanation for the end user.

Random Forest (98% accuracy) took an intermediate position. Although it is an ensemble, which complicates its interpretation compared to a single tree, it has shown weakness in dealing with unbalanced classes (F1score 0.40 for the "Inactive" class). This is an important limitation, because in real educational data, groups of students with certain problems are often small.

The main limitation of this study is the use of synthetic data. This made it possible to conduct a controlled experiment, but for practical implementation, validation of models on real data from LMS systems is necessary. Real-world data will allow us to evaluate the models' resistance to noise and identify more complex, implicit dependencies. Nevertheless, the study demonstrates the high potential of models as diagnostic tools capable of providing teachers with valuable information for timely intervention.

5.Conclusion

The conducted research has demonstrated that modern machine learning algorithms can be successfully applied not only for forecasting, but also for automatically identifying the causes of student lag, thereby solving an important diagnostic task in educational analytics.

All three models considered (Decision Tree, Random Forest, XGBoost) showed high predictive accuracy on structured synthetic data, confirming their applicability for educational tasks. XGBoost is the most accurate model, while Decision Tree remains a valuable tool due to its complete transparency.

The analysis of the importance of signs confirmed that behavioral indicators such as class absences and academic performance are key predictors, which allows us to create models based on pedagogically relevant data.

For high-precision "black box" models such as XGBoost, it is necessary to introduce modern methods of explicable AI (for example, SHAP, LIME). This will combine the high accuracy of forecasts with the provision of understandable and effective explanations, which is a key factor for the successful implementation of AI in the university's information system.

Funding: This research was funded by the Science Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan (Grant No. AP23488869).

Transparency: The authors confirm that the manuscript is an honest, accurate, and transparent account of the study; that no vital features of the study have been omitted; and that any discrepancies from the study as planned have been explained. This study followed all ethical practices during writing.

Competing Interests: The authors declare that they have no competing interests.

Authors' Contributions: All authors contributed equally to the c on ce pt i on and design of the study. All authors have read and agreed to the published version of the manuscript

REFERENCES

1. Shaporeva, A ., Kopnova, O., Shmigirilova, I ., Kukharenko, Y., Aitymova, A . (2022). Development of comprehensive decision support tools in distance learning quality management processes. EasternEuropean Journal of Enterprise Technologies, 4 (3 (118)), 43–50. doi: https://doi.org/10.15587/17294061.2022.263285
2. LópezMeneses, E., MelladoMoreno, P. C., Gallardo Herrerías, C., & PelícanoPiris, N. (2025). Educational Data Mining and Predictive Modeling in the Age of Artificial Intelligence: An InDepth Analysis of Research Dynamics. *Computers*, *14*(2), 68. https://doi.org/10.3390/computers14020068
3. Z. Chen, G. Cen, Y. Wei and Z. Li, Student Performance Prediction Approach Based on Educational Data Mining, in *IEEE Access*, vol. 11, pp. 131260131272, 2023, doi: 10.1109/ACCESS.2023.3335985 https://ieeexplore.ieee.org/document/10327720
4. Haleem, A., Javaid, M., Qadri, M. A., & Suman, R. (2022). Understanding the Role of Digital Technologies in Education: A Review. *Sustainable Operations and Computers, 3,* 275285. https://doi.org/10.1016/j.susoc.2022.05.004
5. Mastour H, Dehghani T, Moradi E, Eslami S. Explainable artificial intelligence for predicting medical students' performance in comprehensive assessments. Sci Rep. 2025 Jul 3;15(1):23752. doi: 10.1038/s41598025074601. PMID: 40610576; PMCID: PMC12229488. https://pmc.ncbi.nlm.nih.gov/articles/PMC12229488/
6. Ab Rahman, N. F., Wang, S. L. . ., Ng, T. F. . ., & Ghoneim, A. S. . . (2024). Artificial Intelligence in Education: A Systematic Review of Machine Learning for Predicting Student Performance. *Journal of Advanced Research in Applied Sciences and Engineering Technology*, *54*(1), 198–221. https://doi.org/10.37934/araset.54.1.198221
7.Tang, Z., Jain, A., & Colina, F. E. (2024). A Comparative Study of Machine Learning Techniques for College Student Success Prediction. *Journal of Higher Education Theory and Practice*, *24*(1). https://doi.org/10.33423/jhetp.v24i1.6764
8. Bhushan, M., Vyas, S., Mall, S. *et al.* A comparative study of machine learning and deep learning algorithms for predicting student's academic performance. *Int J Syst Assur Eng Manag* 14, 2674–2683 (2023). https://doi.org/10.1007/s13198023021603
9. Puthanveettil Madathil, A., Luo, X., Liu, Q., Walker, C., Madarkar, R., & Qin, Y. (2025). A review of explainable artificial intelligence in smart manufacturing. International Journal of Production Research, 1–44. https://doi.org/10.1080/00207543.2025.2513574
10. Bulut, O., Tan, B., Mazzullo, E., & Syed, A. (2025). Benchmarking Variants of Recursive Feature Elimination: Insights from Predictive Tasks in Education and Healthcare. *Information*, *16*(6), 476. https://doi.org/10.3390/info16060476
11. Gomez, M. J., Armada Sánchez, Á., AlbaladejoGonzález, M., Garcia Clemente, F. J., & RuipérezValiente, J. A. (2025). Utilising Explainable AI to Enhance RealTime Student

Performance Prediction in Educational Serious Games. Expert Systems, 42, e70008. https://onlinelibrary.wiley.com/doi/10.1111/exsy.70008
12. Gunasekara, S., & Saarela, M. (2025). Explainable AI in Education: Techniques and Qualitative Assessment. *Applied Sciences*, *15*(3), 1239. https://doi.org/10.3390/app15031239
13. Ngwu, C., Liu, Y. & Wu, R. Reinforcement learning in dynamic job shop scheduling: a comprehensive review of AIdriven approaches in modern manufacturing. J Intell Manuf (2025). https://doi.org/10.1007/s10845025025856
14. Romero, C., & Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 10(3), e1355. https://arxiv.org/abs/2402.07956
15. Granitto, P. M., Furlanello, C., Biasioli, F., & Gasperi, F. (2006). Recursive feature elimination with random forest for PTRMS analysis of agroindustrial products. Chemometrics and Intelligent Laboratory Systems, 83(2), 8390. https://www.scihub.ru/10.1016/j.chemolab.2006.01.007?ysclid=mfvtsdzped426137211
16. Breiman, L. Random Forests. Machine Learning 45, 5–32 (2001). https://doi.org/10.1023/A:1010933404324
17. Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. In Advances in neural information processing systems (pp. 47654774) https://arxiv.org/abs/1705.07874
18. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining (pp. 785794). https://arxiv.org/pdf/1603.02754
19. Salem, M., Shaalan, K. Unlocking the power of machine learning in Elearning: A comprehensive review of predictive models for student performance and engagement. Educ Inf Technol 30, 19027–19050 (2025). https://doi.org/10.1007/s10639025135264
20. Lee, J. E., Jindal, A., Patki, S. N., Gurung, A., Norum, R., & Ottmar, E. (2023). A comparison of machine learning algorithms for predicting student performance in an online mathematics game. Interactive Learning Environments, 32(9), 5302–5316. https://doi.org/10.1080/10494820.2023.2212726
21 VillegasCh, W., Gutierrez, R., GarcíaOrtiz, J. et al. Explainable educational assistant integrated in Moodle: automated semantic assessment and adaptive tutoring based on NLP and XAI. Discov Artif Intell 5, 191 (2025). https://doi.org/10.1007/
22. M. Adnan *et al*., "Predicting atRisk Students at Different Percentages of Course Length for Early Intervention Using Machine Learning Models," in *IEEE Access*, vol. 9, pp. 75197539, 2021, doi: 10.1109/ACCESS.2021.3049446 https://ieeexplore.ieee.org/document/9314000
23. Misinem, M., Kurniawan, T., Dewi, D., Zakaria, M., & Nazmi, C. (2024). Leveraging Data Analytics for Student Grade Prediction: A Comparative Study of Data Features. *Journal of Applied Data Sciences, 5*(4), 20252038. doi:https://doi.org/10.47738/jads.v5i4.442
24. Nnadi, L. C., Watanobe, Y., Rahman, M. M., & JohnOtumu, A. M. (2024). Prediction of Students' Adaptability Using Explainable AI in Educational Machine Learning Models. *Applied Sciences*, *14*(12), 5141. https://doi.org/10.3390/app14125141