PREDICTIVE MODELING OF STUDENT SUCCESS USING MACHINE LEARNING

Arbër H. Hoti, Xhemal Zenuni, Jaumin Ajdari, Florije Ismaili*

South East European University, Tetovo;
Republic of North Macedonia

* Corresponding Author, e-mail: ah29750@seeu.edu.mk

Abstract: This paper presents a comprehensive study of predictive models for student success, leveraging machine learning techniques to analyze diverse factors affecting academic performance in higher education. Using a dataset of students from various institutions in Kosovo, encompassing demographic, socioeconomic, and academic attributes, the research employs and evaluates several machine learning algorithms to predict the dropout rate. To address the issue of unbalanced data, SMOTE and ADASYN techniques were implemented, enhancing the predictive accuracy. The study identifies significant predictors of student success, and the results indicate that machine learning can accurately predict student outcomes, providing educational institutions with actionable insights to mitigate dropout rates and enhance student support.

Keywords: Predictive modeling, student performance, machine learning, educational data mining, dropout prevention.

1. INTRODUCTION

One of the primary goals of educational institutions is to deliver high-quality education experiences and knowledge to learners. Identifying students who need additional support to enhance their performance is essential to achieving this objective [1]. With technological advancements and the widespread availability of internet access, knowledge acquisition has become virtually limitless through online and in-person education, courses, and training [2]. Predicting dropout rates is vital for universities and governmental institutions, to improve student retention and success [3, 4]. By identifying at-risk students early, institutions can implement targeted interventions, such as academic support or counselling, to address challenges before they lead to dropout. This proactive approach enhances student outcomes and boosts graduation rates, contributing to the overall quality of education. Reducing dropout rates also minimizes financial losses for both institutions and governments. Fewer dropouts mean more consistent tuition revenue for universities and a more educated workforce for society [5]. And machine learning techniques can have an important impact on finding, analysing, and predicting any student's future performance, making decisions [6], and giving guidance and support for decision-making [7]. One of the challenges identified is the limited availability of data, particularly in Kosovo, where there is a significant lack of open data on demographic, socio-economic, and academic factors influencing student performance. To address this gap, data was collected through a questionnaire, which served as the foundation for various analyses aimed at predicting student performance and understanding the factors that impact academic decisions [8].

The current study introduces a novel approach by exploring various machine learning techniques to establish a baseline and assess the complexity of predicting student performance in Kosovo—an area not previously addressed in similar studies. The training and testing processes were conducted using real data from multiple universities in Kosovo. To address the issue of unbalanced data, the study implemented SMOTE and ADASYN techniques, improving predictive accuracy.

The Structure of the paper is organized as follows: Section II provides a review of the relevant literature. Section III outlines the methods and techniques employed to achieve the paper's objectives. Section IV presents the results from evaluating the algorithms, followed by a discussion of the findings. Finally, Section V concludes with the key findings, limitations, and suggestions for future research.

2. RELATED WORK

The field of education has witnessed a significant transformation in recent years, with the advent of artificial intelligence playing a crucial role in enhancing the understanding and analysis of student performance. Researchers have explored various machine learning techniques to predict student academic success, address the issue of student dropouts, and provide targeted interventions to improve educational outcomes [9, 10, 11]. Nevertheless, researchers have made significant progress in this area, utilizing a range of machine-learning algorithms to analyse student data and generate accurate predictions. A study [9] presents a neural network approach testing linear and logistic functions to predict school dropouts in a developing country context. The researchers highlight the importance of addressing the growing need for AI-based solutions in developing countries, where data limitations can pose a significant challenge in identifying dropout students. The study [12] offers a comprehensive analysis of the various machine learning techniques employed in predicting student performance, including Collaborative Filtering, Recommender Systems, and Artificial Neural Networks. The application of AI-powered predictive models can provide schools and educators with valuable knowledge about student behaviour, enabling them to proactively identify students at risk of dropping out and implement targeted interventions to support their academic success [9]. Authors in [13] highlight the importance of predicting student performance as well. In their evaluations, XGBoost outperformed other algorithms such as Random Forrest, OLS and logistic regression. The study [14] emphasized that monitoring factors like exams, attendance, and class tests helps identify students at risk of failure, enabling timely intervention. The Levenberg Marquardt Algorithm (MLA) achieved 88.6% accuracy in predicting student performance [15]. A study [16] found high prediction accuracy across algorithms, with Extra Trees achieving the highest at 98.15%. Alongside [17], demonstrated the impact of Signed Graph Neural Networks (SGNNs) and Large Language Models (LLMs), like GPT-3, that show predictive accuracy and significance on personalized learning and adaptive educational systems. Study [18] identify key factors influencing student success, including the instructor, socioeconomic status, institution, parents, and physical health. Asia Mat et al. [19] suggest that teachers could benefit from performance predictions, allowing them to adjust their teaching methods. Research [20, 21] highlight additional factors such as GDP, inflation rate, unemployment, and special educational needs in evaluating student performance. Using unbalanced data, the authors applied SMOTE and ADASYN balancing techniques, achieving algorithm accuracies of 61% for Logistic Regression, 60% for Support Vector Machine, 65% for Decision Tree, and 72% for Random Forest. Findings [20] show that gender and nationality have minimal impact on student success, with a recommendation for institutions to enhance learning opportunities. In a dataset of 1000 records, Linear Regression and Support Vector Regressor (SVR) performed well, with MSE values of 0.0028 and R^2 of 0.8866.

Despite the advancements in using machine learning for dropout prediction, several knowledge gaps persist. From the literature review, it can be concluded that while many studies focus on specific algorithms, there is a lack of robust comparisons across different educational contexts and diverse datasets. The reliance on imbalanced datasets poses significant challenges for predictive accuracy and model generalization. Future research should prioritize the development of multiclass prediction models that can effectively address these issues while considering the unique characteristics of different learning environments. Additionally, the integration of user-generated content and qualitative data, as suggested in the research on MOOCs, represents an underexplored area. By including student feedback and experiences within predictive models, researchers can increase a more nuanced knowledge of the factors influencing student retention and performance.

In this paper, we aim to address these gaps by employing and evaluating machine learning techniques on data collected from students in Kosovo. The dataset includes 39 attributes, covering a range of domains such as socioeconomic factors, academic performance, and demographic information, among others. In addition, we evaluate the performance of the most used machine learning algorithms using standard metrics such as precision and recall. Finally, to enhance classification accuracy, SMOTE and ADASYN techniques were applied.

3. METHODOLOGY

To employ and evaluate student performance prediction using machine learning, a well-defined methodology is essential. The paper follows a standard and systematic approach: data collection, data pre-processing, feature selection, model selection, model training and evaluation, and finally interpretability and discussion on the results.

Data is crucially important in machine learning. It forms the foundation of any machine learning model's success. A mixed methodology was employed to collect data about students in Kosovo [8]. Initially, 39 socio-economic, educational and demographic attributes mainly were considered as important and having an impact on student performance. However, after the correlation analysis between the values of these attributes with the target class, it was noticed that not all of them have strong correlation with the output class.

Data was collected through a questionnaire distributed to institutions across Kosovo, including public universities and private colleges in cities like Pristina, Prizren, Gjakova, and Gjilan. These institutions share similar educational programs but serve students from diverse regions, making the research more challenging. The final dataset consists of 2555 records and has undergone the process of transformation to a suitable format (csv format) and has been pre-processed in terms of data cleaning and data transformation.

For the graduation class, the dataset shows differing outcomes: 1,608 students graduated, accounting for 62.90% of the total; 616 students, or 24.10%, were still enrolled; and 331 students, or 13%, dropped out. This dataset is unbalanced, necessitating the use of data balancing techniques. To address this, we applied SMOTE and ADASYN, observing notable improvements in the results after their application.
After data collection and pre-processing phase, the algorithms were executed, and results were evaluated using Accuracy, F1-Score, Precision, and Recall. This study does not assume that any specific algorithm will outperform the others. Instead, multiple algorithms were implemented, applied, and evaluated against the described dataset to ensure a comprehensive analysis. The algorithms included in the evaluation are Support Vector Machine (SVM), Decision Tree, Random Forest, Logistic Regression, GaussianNB, MultinomialNB, ComplementNB, BernoulliNB, MLP (Multi-Layer Perceptron), and KAN (Kolmogorov-Arnold Networks). This systematic approach allowed for the assessment of each algorithm's performance in managing the dataset effectively.

This methodology integrates a systematic approach for predicting student performance using machine learning. The process ensures that data is adequately prepared, suitable models are selected, and performance is rigorously evaluated, all while addressing issues like class imbalance and feature selection. By doing so, institutions can not only predict student success and failure but also derive actionable insights to improve educational outcomes.

4. RESULTS AND DISCUSSION

This section presents the results and discussion on applying the selected algorithms on the collected dataset. The analysis covers the standard model, boosting models, and the top 10 attributes correlated with the target. For this study Python, Jupyter Notebook, and several ready-made libraries such as ScikitLearn and Numpy were used.

4.1. Results for the standard model

Table 1 presents the evaluation results for each algorithm. Overall, all algorithms performed well, with accuracy exceeding 0.91 for unbalanced data. However, the Naïve Bayes algorithms showed lower accuracy, ranging from 0.75 to 0.76. The highest accuracy for unbalanced data was achieved by KAN (0.95), Logistic Regression (0.95), Random Forest (0.93), and MLP (0.93). For F1-score, Precision and Recall, the highest values were observed in Logistic Regression (0.95) and KAN (0.94).

When the SMOTE technique was applied, Logistic Regression achieved the highest accuracy (0.98), followed by MLP and KAN (0.97), with the Support Vector Machine and Decision Tree both achieving 0.95. At the same time, BernoulliNB had the lowest accuracy (0.66). For the F1-score on balanced data with SMOTE, Random Forest had the highest value (0.98), and MLP (0.97), and Precision and Recall, had the highest value, had LR and KAN with 0.98.

With the ADASYN technique, Logistic Regression again led in accuracy (0.98), while the highest F1-score, Precision and Recall was shared by LR and KAN (0.98). The lowest F1-score and Precision was recorded by BernoulliNB and for Recall lowest value was MultinomialNB and ComplementiNB with 0.74 values. Overall, the Naïve Bayes algorithms had the lowest performance, with an average accuracy of 0.75 and an average F1-score of 0.77.

Table 1. Accuracy results for standard algorithms

Standard Algorithms												
	Without Imbalanced Data				*With SMOTE*				*With ADASYN*			
Algorithms	Accuracy	F1-Score	Precision	Recall	Accuracy	F1-Score	Precision	Recall	Accuracy	F1-Score	Precision	Recall
Support Vector Machine	0.92	0.91	0.93	0.92	0.95	0.95	0.95	0.95	0.89	0.90	0.92	0.89
Decision Tree	0.91	0.91	0.92	0.92	0.95	0.92	0.92	0.92	0.97	0.96	0.96	0.96
Random Forest	0.93	0.92	0.93	0.91	0.86	0.87	0.91	0.86	0.86	0.87	0.91	0.86
Logistic Regression	**0.95**	**0.95**	**0.95**	**0.95**	**0.98**	**0.98**	**0.98**	**0.98**	**0.98**	**0.98**	**0.98**	**0.98**
GaussianNB	0.75	0.78	0.83	0.75	0.76	0.79	0.85	0.76	0.76	0.79	0.85	0.76
MultinomialNB	0.75	0.78	0.84	0.75	0.76	0.78	0.85	0.76	0.74	0.77	0.85	0.74
ComplementiNB	0.76	0.79	0.85	0.76	0.76	0.78	0.85	0.76	0.74	0.77	0.85	0.74
BernoulliNB	0.75	0.77	0.83	0.75	0.75	0.77	0.83	0.75	0.75	0.77	0.83	0.75
MLP (Multi-layer Perceptron)	0.93	0.93	0.93	0.93	**0.97**	**0.97**	0.97	0.97	0.96	0.96	0.96	0.96
KAN (Kolmogorov-Arnold Networks)	**0.95**	**0.94**	**0.94**	**0.94**	**0.97**	0.94	**0.98**	**0.98**	0.97	**0.98**	**0.98**	**0.98**

To avoid overfitting the decision tree, hyperparameter tuning was performed using RandomizedSearchCV, with the best combination controlling the depth and complexity of the tree (ccp alpha = 0.0, max depth = 10, min samples leaf = 5, min samples split = 10). The model was evaluated with 5-fold cross-validation, achieving a mean accuracy of 0.88.

Figure 1 illustrates the decision tree classifier, which uses entropy as the criterion to measure impurity at each node. The tree first classifies students based on whether they complete fewer than or equal to 5.5 courses, predicting graduation or dropout. The entropy value at this stage is 0.65, indicating the uncertainty about whether students will finish 5.5 courses.

The decision tree continues with further nodes, evaluating conditions such as whether students can graduate if they complete at least 7.5 courses, with an entropy of

0.41. The tree further analyzes students progressing to the fourth year, requiring at least 6.5 courses. The analysis focuses on 884 students who are likely to graduate.

In the penultimate node, students need a GPA of 6.55 to graduate, with the condition of no more than 3.5 years of study. If students exceed this time, they are predicted to drop out. The entropy at this node is 0.19, with a sample of 553 students 13 predicted to drop out (entropy = 0.39) and 540 expected to graduate (entropy = 0.076).

Important factors influencing graduation include the number of courses completed each year, GPA, and parental education, which highlight the socio-economic influence on educational outcomes. Parental qualifications are identified as a key factor in student success.

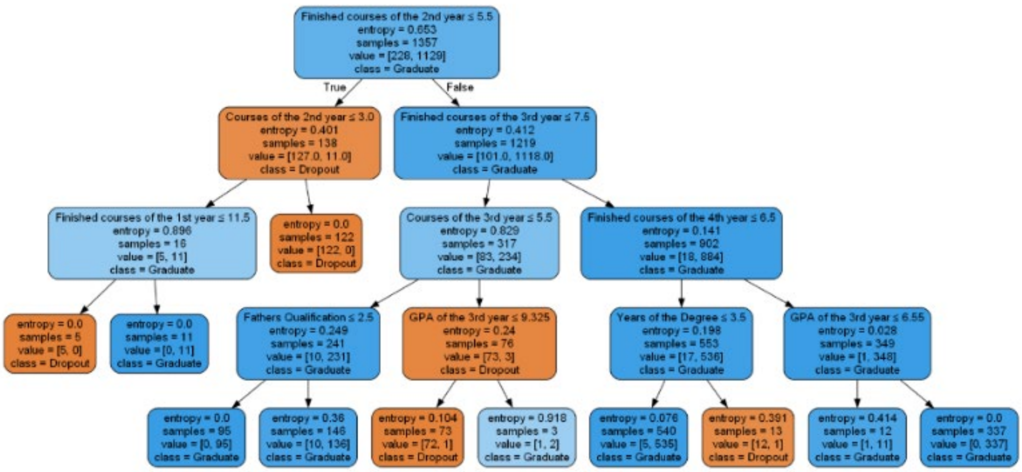

Figure 1. Decision tree classifier

4.2. Results for boosting models

After applying the standard algorithms, we evaluated boosting algorithms for Accuracy, F1-score, Precision and Recall across all three cases, including unbalanced data. A 70% training and 30% testing split was used, focusing on the graduate and dropout classes.

The results in Table 2 show that all boosting algorithms performed well, with Accuracy, F1-score, Precision, and Recall above 0.91. For unbalanced data, Gradient Boosting, CAT Boost, and Light GBM achieved the highest performance, with an accuracy of 0.98, and F1-score for Gradient Boosting of 0.98 and CAT Boost, Light GBM 0.96. The lowest performance came from Extreme Gradient Boosting, with an accuracy of 0.91 and an F1-score of 0.91, also for Precision and Recall the highest value had Gradient Boosting and Light GBM with 0.98.

With the SMOTE technique, Light GBM achieved the best performance with 0.99 accuracies, while Gradient Boosting, ADA Boost, and CAT Boost achieved an accuracy of around 0.98. In terms of F1-score, Gradient Boosting was 0.98, CAT Boost, and Light GBM scored 0.97, ADA Boost was at 0.95, and Extreme Gradient Boosting was at 0.91. For Precision and Recall, the highest value was Light GBM.

When using the ADASYN technique, Light GBM again had the highest accuracy (0.99), while Extreme Gradient Boosting showed the lowest (0.91). For the F1-score, Gradient Boosting was 0.98 and Light GBM with CAT Boost both scored 0.97, while Extreme Gradient Boosting had the lowest at 0.91. And for Precision and Recall Gradient Boosting and Light GBM was 0.98.

Table 2. Results of boosting algorithms

Boosting Algorithms												
	Without Imbalanced Data				*With SMOTE*				*With ADASYN*			
Algorithms	Accuracy	F1-Score	Precision	Recall	Accuracy	F1-Score	Precision	Recall	Accuracy	F1-Score	Precision	Recall
Gradient Boosting	**0.98**	**0.98**	**0.98**	**0.98**	0.98	**0.98**	0.98	0.98	0.98	**0.98**	**0.98**	**0.98**
Extreme Gradient Boosting	0.91	0.91	0.91	0.91	0.91	0.91	0.91	0.91	0.91	0.91	0.91	0.91
ADA Boost	0.97	0.94	0.95	0.92	0.98	0.95	0.95	0.92	0.97	0.94	0.95	0.93
CAT Boost	**0.98**	**0.96**	0.96	0.96	0.98	0.97	0.97	0.97	0.98	0.97	0.97	0.97
Light GBM	**0.98**	**0.96**	**0.98**	**0.98**	**0.99**	0.97	**0.99**	**0.99**	**0.99**	0.97	**0.98**	**0.98**

4.3. Results only for TOP10 correlation attributes with the target.

After analyzing the accuracy of the algorithms with respect to general attributes, the top 10 most relevant attributes were identified, as shown in Table 3. Additionally, a more specific analysis of these attributes was conducted by evaluating their accuracy and F1-score. For the top 10 correlated attributes, 70% training data and 30% test data split was used, focusing on the graduate and dropout classes.

Table 3. Top correlation attribute between features

Top correlations in **Demographic** variables:	Age - Marital Status: 0.58
Top correlations in **Academic** data variables:	Level of Studies - Previous Qualification: 0.98 Level of Studies - GPA of Previous Qualification: 0.89 GPA of Previous Qualification - Previous Qualification: 0.88 Target - Enrolled Semester: 0.77 Department - Level of Studies: 0.76 Department - Previous Qualification: 0.75 Years of the Degree - GPA of Previous Qualification: 0.67 Department - GPA of Previous Qualification: 0.65
Top correlations in **Academic performance** variables:	GPA of the 4th year - Finished courses of the 4th year: 0.96 GPA of the 3rd year - Finished courses of the 3rd year: 0.89 Finished courses of the 4th year - Courses of the 4th year: 0.83 GPA of the 4th year - Courses of the 4th year: 0.83 Courses of the 3rd year - Courses of the 2nd year: 0.78 GPA of the 3rd year - Finished courses of the 2nd year: 0.76 Finished courses of the 3rd year - Finished courses of the 2nd year: 0.75 Finished courses of the 3rd year - Courses of the 3rd year: 0.71 GPA of the 3rd year - Courses of the 3rd year: 0.7 Finished courses of the 2nd year - Finished courses of the 1st year: 0.57

Table 4 presents the accuracy and F1-score for algorithms such as Support Vector Machine, Decision Tree, Random Forest, Logistic Regression, and ADA Boost, using the SMOTE and ADASYN techniques for data balancing. When applying the SMOTE technique, the Decision Tree algorithm achieved the highest accuracy and F1-score at 0.98, while ADA Boost had the lowest values at 0.86. With the ADASYN technique, Support Vector Machine and Random Forest both achieved the highest accuracy and F1-score at 0.94, while ADA Boost again had the lowest values.

These results indicate high accuracy across all algorithms. However, the accuracy was influenced by data preprocessing, including the removal of irrelevant attributes that had minor impact on the outcome. Removing these attributes helped improve the performance of the algorithms.

Table 4. Results from top 10 correlations

Standard Algorithms - TOP10 correlation attributes				
	With SMOTE		*With ADASYN*	
Algorithms	*Accuracy*	*F1- Score*	*Accuracy*	*F1- Score*
Support Vector Machine	0.97	0.97	**0.94**	**0.94**
Decision Tree	**0.98**	**0.98**	0.91	0.91
Random Forest	0.97	0.97	**0.94**	**0.94**
Logistic Regression	0.96	0.96	0.90	0.90
ADA Boost	0.86	0.86	0.80	0.80

5. CONCLUSION

The study demonstrates that machine learning algorithms, particularly when combined with data balancing techniques like SMOTE and ADASYN, can significantly improve the prediction of student performance and dropout rates. This approach can aid educational institutions in identifying at-risk students early, enabling interventions that can improve student retention and graduation rates. Compared to the benchmarks in [21] and [22], this study demonstrates better results for each of the algorithms used.

However, the study also highlights the need for improved data collection and quality, especially in regions with limited open data. Future research could focus on expanding datasets, refining models, and integrating qualitative student feedback for more accurate predictions.

This research lays a solid foundation for further exploration into machine learning applications in education, particularly in regions like Kosovo where data availability remains a challenge.

REFERENCES

[1] Hussein Altabrawee, Osama Abdul Jaleel Ali, and Samir Qaisar Ajmi. Predicting students' performance using machine learning techniques. *Journal pf univeristy of Babylon for pure and applied sciences,* vol. 27, no. 1, 2019, pp. 194-205.

[2] Fidelia A. Orji and Julita Vassileva. Machine learning approach for predicting students academic performance and study strategies based on their motivation. *arXiv preprint arXiv:2210.08186*, 2022, pp. 1-7.

[3] José Orlando Quintana Quispe, Osmar Cuentas Toledo, Maryluz Cuentas Toledo, Elmer Elio Calizaya Llatasi, and Elizabeth Marina Ramos Saira. Early prediction of university student dropout using machine learning models. *Nanotechnology Perceptions*, 2024, pp.659-669.

[4] Holmes Yesid Ayala-Yaguara, Gina Maribel Valenzuela-Sabogal, and Alexander Espinosa-García. Obtaining a data mining model to be applied to university desertion from the Systems Engineering program of the University of Cundinamarca. *Revista Ontare,* vol. 7, 2019, pp. 134-150.

[5] T. Susnjak. Beyond predictive learning analytics modelling and onto explainable artificial intelligence with prescriptive analytics and ChatGPT. *International Journal of Artificial Intelligence in Education,* vol. 34, no. 2, 2024, pp. 452-482.

[6] Opeyemi Ojajuni, Foluso Ayeni, Olagunju Akodu, Femi Ekanoye, Samson Adewole, Timothy Ayo, Sanjay Misra, and Victor Mbarika. Predicting student academic performance using machine learning. in *Proceedings of the Computational Science and Its Applications—ICCSA 2021, 21st International Conference*, Cagliari, Italy, 2021, p. 481-491.

[7] Egle Gedrimiene, Ismail Celik, Antti Kaasila, Kati Mäkitalo, and Hanni Muukkonen. Artificial intelligence (AI)-enhanced learning analytics (LA) for supporting career decisions: Advantages and challenges from user perspective. *Education and Information Technologies,* vol. 29, no. 1, 2024, pp. 297-322.

[8] Arbër H. Hoti, and Xhemal Zenuni. Factors influencing student academic performance and career choices. in *2024 8th International Artificial Intelligence and Data Processing Symposium (IDAP)*, Malatya, Türkiye, 2024.

[9] K. Kalegele. Enabling proactive management of school dropouts using neural networ. *Journal of Software Engineering and Applications,* vol. 13, no. 10, 2020, pp. 245-257.

[10] Khalfan Al Mayahi, and Mahmood Al-Bahri. Machine learning based predicting student academic success. in *2020 12th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT)*, IEEE, 2020, pp 264-268.

[11] Olaf Zawacki-Richter, Victoria I. Marín, Melissa Bond, and Franziska Gouverneur. Systematic review of research on artificial intelligence applications in higher education--where are the educators?. *International Journal of Educational Technology in Higher Education,* vol. 16, no. 1, 2019, pp. 1-27.

[12] Juan L. Rastrollo-Guerrero, Juan A. Gómez-Pulido, and Arturo Durán-Domínguez. Analyzing and predicting students' performance by means of machine learning: A review. *Applied sciences,* vol. 10, no. 3, 2020, p. 1042.

[13] Kelli A. Bird, Zachary Mabel, and Yifeng Son. Bringing transparency to predictive analytics: A systematic comparison of predictive modeling methods in higher education. *AERA Open,* vol. 7, no. 1, 2021, p. 1–19.

[14] Sazol Sarker, Mahit Kumar Paul, Sheikh Tasnimul Hasan Thasin and Md. Al Mehedi Hasan. Analyzing students' academic performance using educational data mining. *Computers and Education: Artificial Intelligence*, 2024, p. 100263.

[15] Muhammad Mubashar Hussain, Shahzad Akbar, Syed Ale Hassan, Muhammad Waqas Aziz, and Farwa Urooj. Prediction of student's academic performance through data mining approach. *Journal of Informatics and Web Engineering,* vol. 3, no. 1, 2024, pp.241--251.

[16] Mustafa S. Ibrahim Alsumaidaie, Ahmed Adil Nafea, Abdulrahman Abbas Mukhlif, Ruqaiya D. Jalal, and Mohammed M AL-Ani. Intelligent system for student performance prediction using machine. *Baghdad Science Journal,* vol. 21, no. 12, 2024, pp. 1-15.

[17] Lin Ni, Sijie Wang, Zeyu Zhang, Xiaoxuan Li, Xianda Zheng, Paul Denny, and Jiamou Liu. Enhancing student performance prediction on learnersourced questions with. *Proceedings of the AAAI Conference on Artificial Intelligence,* vol. 38, no. 21, 2024, pp.23232-23240.

[18] Mariel F. Musso, Eduardo C. Cascallar, Neda Bostani, and Michael Crawford. Identifying reliable predictors of educational outcomes through machine-learning predictive modeling. *Frontiers in Education,* vol. 5, 2020, p. 104.

[19] Asiah Mat, Nik Zulkarnaen Khidzir, Safaai Deris, Nik Nurul Hafzan Mat Yaacob, Mohd Saberi Mohamad, and Siti Syuhaida Safaai. A review on predictive modeling technique for student academic performance monitoring. *MATEC Web of Conferences,* vol. 255, 2019, p. 03004.

[20] E. Deniz. Evaluating the effectiveness of machine learning models in predicting student academic achievement. *ADBA Computer Science,* vol. 1, no. 1, 2024, pp. 8-13.

[21] Mónica V. Martins, Daniel Tolledo, Jorge Machado, Luís M. T. Baptista, and Valentim Realinho. Early prediction of student's performance in higher education: A case study. *Trends and Applications in Information Systems and Technologies,* vol. 1, no. 9, 2021, pp. 166-175.

[22] Valentim Realinho, Jorge Machado, Luís Baptista, and Mónica V. Martins. Predicting student dropout and academic success. *Data,* vol. 7, no. 11, 2022, p. 146.

Information about the authors:

Arbër H. Hoti – is a PhD candidate at South East European University within the Faculty of Contemporary Sciences and Technologies. Also, he is a teaching assistant at the Faculty of Education, University of Prishtina. His interests in the research field are Artificial Intelligence, Machine Learning, E-learning, Software Development, and ICT in Education.

Xhemal Zenuni – is a full-time professor at the South East European University, within the Faculty of Contemporary Sciences and Technologies. His interests in the research fields are Artificial Intelligence, Machine Learning, Semantic Web, Data Analytics, and Web Services.

Jaumin Ajdari – is a full-time professor at the South East European University, within the Faculty of Contemporary Sciences and Technologies. His interests in the research field are Database, Data Processing, Parallel Processing, High Performance Computing, and Algorithms.

Florije Ismaili - is a full-time professor at the South East European University, within the Faculty of Contemporary Sciences and Technologies. Her research interests include Data Science, Distributed Systems, Semantic Web, and Web Programming.

Manuscript received on 09 January 2025