AGU INTERNATIONAL JOURNAL OF SCIENCES
p-ISSN 3093 - 3536

A COMPARATIVE ANALYSIS OF MACHINE LEARNING MODELS FOR PREDICTING ACADEMIC PERFORMANCE OF INFORMATION TECHNOLOGY AND SOFTWARE ENGINEERING STUDENTS AT AN GIANG UNIVERSITY

Nguyen Minh Vi[1], Thieu Thanh Quang Phu[1], Nguyen Van Vu[1]
[1]*An Giang University, VNU-HCM*

Information:
Received:17/10/2025
Accepted:13/11/2025
Published:12/2025

Keywords:
Machine learning, academic performance prediction, educational data mining, Information Technology, Software Engineering

ABSTRACT

Predicting student academic performance is a critical application of artificial intelligence in education, enabling personalized learning pathways and early intervention for students at risk of academic difficulties. This study evaluates and compares the performance of nine machine learning models, including basic algorithms (KNN, SVM, DCT, MLP) and ensemble models (RF, Bagging, AdaBoost, XGBoost, CatBoost), in forecasting the graduation rankings of students in the Information Technology (IT) and Software Engineering (SE) programs at An Giang University. The dataset, comprising 788 student records from graduates between 2019 and 2024, was preprocessed, balanced, and standardized. Results indicate that SVM achieved the highest performance on the IT dataset, while CatBoost outperformed the others models on the SE dataset. The study confirms that selecting an appropriate machine learning model should be based on the specific characteristics of each program's data, highlighting the potential of these models for integration into academic advising and early warning systems in higher education institutions.

1. INTRODUCTION

In recent years, the application of machine learning (ML) in education has garnered increasing attention. One of the prominent applications is predicting student academic performance, which aims to enhance training quality and support students in achieving optimal learning outcomes. For disciplines within the field of information technology, achieving strong academic results requires students to possess logical thinking, programming skills, and academic discipline. However, many students face challenges in certain courses, particularly foundational ones, leading to overall academic performance that falls short of expectations. Consequently, predicting academic performance plays a pivotal role in providing timely support to students and improving training quality.

Although numerous studies worldwide have extensively explored machine learning techniques for predicting academic performance, achieving high accuracy across diverse disciplines and addressing imbalanced data remain significant challenges. In Vietnam, research on this topic is still developing, with

few studies focusing on comparing model performance across different disciplines within the same educational institution. Motivated by this context, this study evaluates and compares the effectiveness of multiple machine learning algorithms in two undergraduate programs at the Faculty of Information Technology, An Giang University: Information Technology (IT) and Software Engineering (SE). Utilizing a dataset comprising academic transcripts of students who graduated over the past five years, this study aims to identify the most accurate machine learning model for predicting academic performance and to determine the key features influencing students' academic outcomes.

The primary contribution of this study is to provide a comparative analysis of the performance of machine learning algorithms in the IT and SE programs, clarifying the suitability of each model based on the specific characteristics of each discipline. This work underscores the potential of artificial intelligence applications in Vietnamese higher education.

2. BACKGROUND

The application of machine learning in predicting student academic performance has emerged as a focal research area, contributing to enhanced educational quality through timely interventions and personalized learning pathways (Jagwani, 2019). Recent studies, both globally and in Vietnam, have demonstrated the efficacy of ML algorithms in analyzing academic data, forecasting performance, and identifying students at risk of underachievement (Albreiki et al., 2021; Bellaj et al., 2024). This literature review synthesizes key studies on the use of ML in predicting academic outcomes, focusing on methodologies, achieved results, and persistent challenges, while identifying the research gap addressed by this study.

Numerous international studies have elucidated the potential of ML in predicting academic performance across diverse educational contexts. Jagwani (2019) reviewed ML applications in education, emphasizing its their capacity to personalize learning pathways and optimize instructional content. Ahmed (2024) employed algorithms such as Support Vector Machine (SVM), Decision Tree (DCT), and K-Nearest Neighbors (KNN) on a dataset of 32,582 students in Ethiopia, achieving 96% accuracy with SVM. Similarly, Dervenis et al. (2022) applied Random Forest (RF) and XGBoost at a Greek university, demonstrating superior performance by these models in classifying academic outcomes. Pallathadka et al. (2023) compared multiple ML models, including RF, SVM, and Neural Networks, confirming that RF frequently achieves high accuracy in academic performance prediction tasks.

Other studies have explored specific aspects of ML in education. Luo et al. (2024) analyzed online learning behavior data in blended courses, achieving 74.6% accuracy with RF and Neural Networks. Musso et al. (2020) utilized ML to predict key educational outcomes, highlighting its capacity to identify risk factors early. Zeineddine et al. (2021) developed an automated ML system, achieving high performance with Gradient Boosting. Rastrollo-Guerrero et al. (2020) provided a comprehensive review, noting that models like XGBoost and Deep Neural Networks often outperform others in complex prediction tasks. However, these studies also identified challenges such as heterogeneous data, high computational costs, and the need for hyperparameter optimization (Albreiki et al., 2021; Bellaj et al., 2024).

In Vietnam, ML research in education is gaining momentum amid the digital transformation of the sector. Sang et al. (2020) applied multilayer neural networks to a large dataset (3,828,879 samples) at Can Tho University, achieving low error rates (measured by RMSE and MAE), and demonstrating potential for supporting underperforming students and nurturing high

achievers. Thuy (2023) utilized models such as Logistic Regression, KNN, RF, SVM, and XGBoost to predict on-time graduation for 6,696 students at the Banking Academy, achieving 92% accuracy with RF. Uyen and Tam (2019) employed Logistic Regression and Naïve Bayes to predict the risk of academic dismissal for Information Technology students at Vinh University, providing a basis for early academic interventions.

Minh et al. (2024) at Dong Thap University applied Naïve Bayes, DCT, and Neural Networks to a dataset of 233,510 records, achieving an average accuracy of 80.98% with Naïve Bayes. Mai and Duong (2023) underscored the potential of ML in Vietnamese higher education but highlighted challenges such as heterogeneous data and frequent curriculum changes. Dien and Nguyen (2020) combined factor analysis and deep learning to predict academic performance, showing improvements through data transformation techniques. These studies affirm the feasibility of ML in Vietnamese education, but its application in local universities remains limited due to the lack of high-quality data (Mai and Duong, 2023).

Despite significant progress, current research exhibits several limitations. First, most studies focus on general disciplines or large universities, while specialized fields such as IT and SE at local institutions like An Giang University remain underexplored. Second, class imbalance in academic datasets, where categories like "Excellent" are underrepresented, is rarely addressed comprehensively in Vietnamese studies, leading to potential model bias (Mai and Duong, 2023). Third, feature selection and the evaluation of specific factors, such as course grades, impacting academic performance are insufficiently explored in the Vietnamese educational context.

This study addresses these gaps by comparing the performance of multiple ML models (KNN, DCT, SVM, RF, BagDCT, MLP, AdaBoost, CatBoost, XGBoost) on a dataset of academic transcripts from IT and SE students at An Giang University. It employs the Synthetic Minority Over-sampling Technique (SMOTE) to address class imbalance and Principal Component Analysis (PCA) for feature selection. By focusing on the local context and leveraging advanced techniques, this research contributes to improving the accuracy of academic performance prediction and advancing innovation in Vietnamese higher education.

3. METHODOLOGY

The experiments in this study include the main steps: data collection, preprocessing, machine learning model training, and performance evaluation.

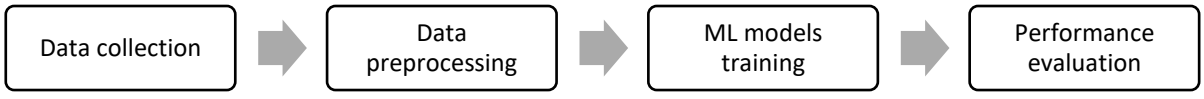

Figure 1. Main steps in the experimental procedure

3.1. Data

The dataset was collected from the academic management system of An Giang University. It includes academic records of full-time students from the Faculty of Information Technology who graduated between 2019 and 2024. After filtering, a total of 788 graduate records were compiled, divided into two groups: 490 students from the Information Technology (IT) program and 298 from the Software Engineering (SE) program. Each record corresponds to an individual student and includes the following main data fields after processing:

- Student ID (anonymized);
- Course grades: final grades for approximately 70 courses defined in the training program, spanning general

education, foundational, and specialized courses;

- Graduation classification: grouped into three levels -Average (TB), Good (K), and Excellent (G) - with the Outstanding (XS) category merged into Excellent due to its limited sample size.

Table 1. Dataset description

Student ID	BUS528	…	SEE910	Classification
DTH206101	8.8	…	9.2	G
DTH206102	8.3	…	7.4	K
…	…	…	…	…

The initial dataset underwent quality checks to eliminate records with incomplete information, such as missing course grades or final graduation rankings. Only students with complete and accurate data were retained to ensure the reliability of the models. Following data cleaning, the dataset was standardized and encoded to facilitate the training of machine learning models.

3.2. Data preprocessing

To ensure the quality of input data for machine learning models, a meticulous preprocessing pipeline was implemented, encompassing data cleaning, encoding, standardization, and balancing. This process aimed to eliminate noise, standardize formats, and optimize the dataset to enhance the performance in predicting graduation rankings.

Data Cleaning: Invalid records, including those with missing course grades or unrecorded graduation rankings, were removed to ensure dataset integrity and minimize the impact of erroneous data on model performance.

Encoding Categorical Data: Graduation ranking values were converted into integer representations to meet the requirements of machine learning algorithms, enabling efficient processing of categorical variables.

Addressing Data Imbalance: The distribution of graduation ranking classes revealed significant imbalance, with the "Good" category substantially outnumbering others. To address this, the study conducted experiments on a reprocessed dataset using SMOTE, which generates synthetic samples for minority classes. This approach not only balances class distribution but also improves classification accuracy, particularly for rare cases.

3.3. Machine Learning Models

This study implemented and compares nine machine learning algorithms, encompassing both simple and ensemble methods, to predict the graduation rankings of students in the IT and SE programs. The selected algorithms were chosen based on their suitability for classification tasks, compatibility with the dataset, and proven performance in similar studies. The rationale for selecting each algorithm and its relevance to the problem context are outlined below:

Simple Algorithms:

- K-Nearest Neighbors: KNN classifies data points based on their proximity to the nearest neighbors, using the number of closest neighbors to make predictions. This algorithm is suitable for datasets with clear clustering structures, such as course grades that may cluster according to graduation rankings. However, KNN may underperform

with high-dimensional or noisy data, necessitating careful preprocessing.

- Decision Tree: DCT construct branching rules based on feature values to generate predictions. This algorithm is interpretable and effective for identifying courses with significant influence on graduation rankings. However, it is prone to overfitting without proper tuning, particularly with student grade data.
- Support Vector Machine: SVM identifies an optimal hyperplane to separate classes, performing well with data exhibiting linear or near-linear boundaries. In this study, SVM is well-suited for handling IT program grade data, which may exhibit clear separation between ranking classes, especially after feature selection.
- Multi-Layer Perceptron: MLP is an artificial neural network with multiple layers and excels at modeling complex non-linear relationships. This algorithm is appropriate for uncovering hidden patterns in grade data, particularly when courses have intricate interactions affecting graduation outcomes.

Ensemble Algorithms:

- Random Forest: RF aggregates multiple decision trees to enhance stability and reduce overfitting. It is well-suited for datasets with numerous features and noise, such as diverse course grades, while also providing insights into the importance of individual courses.
- Bootstrap Aggregating: Bagging reduces variance by training multiple sub-models on different sampled datasets. This algorithm is effective for improving accuracy on data with high variability, such as SE program grades with diverse elective courses.
- AdaBoost: AdaBoost focuses on difficult-to-classify samples by adjusting weights during training. It is suitable for handling mildly imbalanced data after SMOTE application, enhancing prediction accuracy for minority classes.
- XGBoost: XGBoost employs gradient boosting to optimize a loss function, delivering high performance on complex structured data. It is well-suited for predicting graduation rankings due to its ability to handle non-linear relationships between grades and rankings.
- CatBoost: CatBoost is a gradient boosting algorithm optimized for categorical data, performs effectively with highly heterogeneous features. In this context, CatBoost is particularly suitable for SE program data, where course grades exhibit significant diversity and complex feature interactions.

The combination of simple and ensemble algorithms enables a comprehensive evaluation of predictive performance while leveraging the distinct characteristics of IT and SE datasets. The algorithms were implemented with hyperparameter tuning to optimize performance, as detailed in subsequent sections.

3.4. Model Training and Optimization

The training and tuning process was conducted meticulously, integrating cross-validation and hyperparameter optimization techniques to ensure that the machine learning models achieve optimal performance and high generalization capability on student grade data.

Stratified 10-Fold Cross Validation was employed to evaluate the models' generalization ability. By dividing the dataset into 10 folds while maintaining the proportion of graduation ranking classes in each fold, this method ensures that the models are tested on representative data subsets, minimizing bias due to class imbalance.

To fully exploit the dataset's structure, GridSearchCV was utilized to identify the optimal hyperparameter set for each algorithm. Parameters such as the number of neighbors (k)

in KNN, the regularization parameter C and kernel in SVM, or the number of trees and depth in RF were tested across a predefined grid of values. This process enables the models to achieve a balance between accuracy and generalization, particularly on educational data with significant variability.

The performance of the models was assessed using the following key metrics:

- Accuracy: The proportion of correct predictions across the entire test set, reflecting the model's overall classification capability.
- F1-Score: The harmonic mean of precision and recall, prioritized in this study due to class imbalance in the graduation rankings. This metric is particularly critical for evaluating prediction performance on minority classes, ensuring the model does not favor dominant classes.

4. EXPERIMENTAL RESULTS

This section presents the training and evaluation results of the machine learning models on two separate datasets corresponding to the two academic programs.

4.1. Results on the IT Dataset

Experimental results on the IT dataset demonstrate that the Support Vector Machine model outperformed others, achieving an accuracy of 91.19% and the highest F1-score compared to the remaining algorithms. This impressive performance reflects the favorable feature space structure of the IT dataset, where the graduation ranking classes (Excellent, Good, Average) exhibit clear linear or near-linear separation boundaries. SVM, with its capability to optimize the separating hyperplane, effectively leverages this characteristic to deliver high accuracy and stable classification performance.

Ensemble models, including Random Forest, AdaBoost, and Bagging, also recorded notable performance, with accuracies ranging from 86% to 89%. These models demonstrated high stability due to their ability to combine multiple sub-models, mitigating bias and adapting well to variability in the grade data. In contrast, simpler algorithms such as K-Nearest Neighbors and Decision Tree showed slight signs of overfitting, particularly when the data dimensionality was not optimized through feature selection. This limitation underscores the necessity of thorough data preprocessing to enhance the effectiveness of simpler models in academic performance prediction tasks.

Comparison of model performance on the IT dataset:

Table 2. Experimental results on IT dataset

Model	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)	Training Time (s)
KNN	87.12	87.80	87.12	84.83	0.80416
SVM	91.19	91.79	91.19	90.64	1.61645
MLP	86.30	86.81	86.30	85.09	35.33526
DCT	74.43	76.05	74.43	74.09	0.35725
BagDCT	85.28	86.21	85.28	83.77	39.94685
AdaBoost	75.86	77.49	75.86	75.35	0.63547
RF	87.33	88.17	87.33	85.47	6.82053

Model	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)	Training Time (s)
CatBoost	89.77	90.58	89.77	88.48	146.1459
XGBoost	84.04	85.23	84.04	82.76	13.98467

4.2. Results on the SE Dataset

On the SE dataset, the CatBoost model stood out with an accuracy of 88.50%, demonstrating superior capability in handling highly heterogeneous data and complex relationships among features. The diversity of course grades, particularly from elective courses, resulted in a non-linear data structure, where CatBoost, a gradient boosting algorithm optimized for categorical data, excels in modeling hidden patterns, ensuring accurate and stable predictions.

Other ensemble models, including XGBoost and Random Forest, also achieved impressive results, with accuracies closely trailing CatBoost, ranging from 85% to 86%. These algorithms leverage the strength of combining multiple sub-models to minimize errors and adapt to the variability in the SE dataset. In contrast, simpler algorithms such as K-Nearest Neighbors and Decision Tree recorded significantly lower accuracies, primarily due to the diversity in course grades, which reduces clustering tendencies and hinders these models' ability to identify clear separation boundaries. These results underscore the critical role of ensemble models in addressing complex educational data.

Comparison of model performance on the SE dataset:

Table 3. Experimental results on SE dataset

Model	Accuracy (%)	Precision (%)	Recall (%)	F1 Score (%)	Training Time (s)
KNN	85.42	82.90	85.42	83.56	0.27621
SVM	82.71	83.67	82.71	81.95	0.61692
MLP	79.58	80.62	79.58	78.96	8.09886
DCT	71.17	72.30	71.17	70.65	0.18553
BagDCT	85.29	84.54	85.29	83.90	21.30604
AdaBoost	72.42	73.94	72.42	72.25	0.21572
RF	87.25	86.37	87.25	85.75	8.21889
CatBoost	88.50	87.34	88.50	87.44	49.30386
XGBoost	85.29	84.92	85.29	84.45	3.36326

5. DISCUSSION

The experimental results on the datasets from the IT and SE programs at An Giang University highlight the distinct characteristics of each program while underscoring the critical role of data preprocessing techniques and machine learning models in predicting graduation

rankings. The following key points are derived from the study:

First, differences between the two programs: Despite both belonging to the Faculty of Information Technology, the grade datasets for the IT and SE programs exhibit distinct distribution and separability characteristics, leading to variations in machine learning model performance. For the IT dataset, the Support Vector Machine achieved a superior accuracy of 91.19%, leveraging the dataset's tendency for clear linear or near-linear class separation among graduation rankings (Excellent, Good, Average). SVM's ability to optimize separating hyperplanes effectively capitalizes on this structure. In contrast, the SE dataset displays greater diversity and non-linearity, driven by the variety of elective courses. In this context, CatBoost, a gradient boosting algorithm, excelled with an accuracy of 88.5%, owing to its capability to model complex feature relationships. This disparity emphasizes the importance of selecting models tailored to the specific data characteristics of each program.

Second, the impact of handling data imbalance: The original dataset exhibited significant class imbalance, with the "Good" ranking class overwhelmingly dominant. The Synthetic Minority Over-sampling Technique proved instrumental in enhancing model performance, particularly for minority classes. By generating synthetic samples for underrepresented classes, SMOTE not only improved overall accuracy but also significantly boosted the F1-score, ensuring more effective predictions for rare cases. This finding confirms that addressing class imbalance is an essential step in academic classification tasks, especially given the uneven class distribution often observed in educational data.

Third, the efficacy of ensemble models: Ensemble models, including Random Forest, AdaBoost, XGBoost, and CatBoost, demonstrated consistent and superior performance across both programs. These models leverage the strength of combining multiple sub-models to minimize errors and effectively capture non-linear relationships among features, such as complex interactions between course grades. Notably, XGBoost and CatBoost stood out due to their gradient optimization capabilities, making them particularly suitable for heterogeneous data like that of the SE program. However, the computational cost of ensemble models is significantly higher than that of simpler algorithms, necessitating careful consideration when deploying them on resource-constrained systems.

Finally, practical applications: The study's findings unlock significant potential for applications in higher education, particularly in academic advising and management. The developed machine learning models can assist instructors and advisors in early identification of students at risk of underperformance, enabling timely interventions such as tailored study strategies or adjusted learning pathways. These insights not only support educational management but also lay the foundation for developing integrated early warning systems, contributing to the optimization of students' learning experiences.

6. CONCLUSION

This study has demonstrated the potential of machine learning in predicting the graduation rankings of students in the Information Technology and Software Engineering programs at An Giang University through a performance comparison of nine algorithms on a dataset of 788 student records. The results indicate that the Support Vector Machine model achieved the highest accuracy on the IT dataset, benefiting from the clear linear separability of the data, while CatBoost excelled on the SE dataset due to its ability to handle complex non-linear feature relationships. The study provides a comparative analysis of machine learning algorithm performance across the two programs,

highlighting the suitability of each model for the specific data characteristics of each discipline. The analysis reveals significant performance differences among the models, underscoring that no single machine learning model is universally optimal across all contexts, and that model selection should be based on the specific data characteristics.

However, the study has limitations, as the data primarily relied on course grades and does not incorporate other factors such as demographic characteristics or student learning behaviors. Future research could expand by integrating diverse data sources to enhance model accuracy and comprehensiveness. Additionally, integrating these models into academic management systems as intuitive dashboards represents a promising direction for implementation, enabling instructors and educational administrators to make more informed decisions.

Acknowledgments

This research is funded by An Giang University (AGU), Vietnam National University HoChiMinh City (VNU-HCM) under grant number 24.02.CT.

REFERENCES

A. Jagwani. (2019). A Review of Machine Learning in Education. *Journal of Emerging Technologies and Innovative Research*, Volume 6, Issue 5.

Balqis Albreiki, Nazar Zaki, and Hany Alashwal. (2021). *A Systematic Literature Review of Student' Performance Prediction Using Machine Learning Techniques*. Education Science, 11(9), 552.

Bellaj, M., Ben Dahmane, A., Boudra , S. ., & Lamarti Sefian, M. . (2024). Educational Data Mining: Employing Machine Learning Techniques and Hyperparameter Optimization to Improve Students' Academic Performance. *International Journal of Online and Biomedical Engineering* (iJOE), 20(03), pp. 55–74.

Charalampos Dervenis, Vasileios Kyriatzis, Spyros Stoufis, Panos Fitsilis. (2022). *Predicting Students' Performance Using Machine Learning Algorithms*. ICACS '22: Proceedings of the 6th International Conference on Algorithms, Computing and Systems, Article No.: 6, Pages 1-7.

Dien Tran Thanh, Nguyen Thai-Nghe, (2020). Deep Learning with Data Transformation and Factor Analysis for Student Performance Prediction. *International Journal of Advanced Computer Science and Applications* 11(8):711-721.

Esmael Ahmed. (2024). *Student performance prediction using machine learning algorithms*. Applied Computational Intelligence and Soft Computing, Volume 2024, Issue 1.

H. L. U. Minh, P. T. Trinh, and N. V. Nhut. (2024). Predicting on-time graduation of students: A case study at Dong Thap University. *Journal of Education*, Volume 24, no. 1.

Harikumar Pallathadka, Alex Wenda, Edwin Ramirez-Asís, Maximiliano Asís-López, Judith Flores-Albornoz, Khongdet Phasinam. (2023). *Classification and prediction of student performance data using various machine learning algorithms*. Materials Today: Proceedings, Volume 80, Part 3.

Hassan Zeineddine, Udo Braendle, Assaad Farah. (2021). *Enhancing prediction of student success: Automated machine learning approach*. Computers & Electrical Engineering, Volume 89.

Iatrellis, O., Savvas, I.K., Fitsilis, P. et al. (2021). *A two-phase machine learning approach for predicting student outcomes*. Educ Inf Technol 26, 69–88.

Juan L. Rastrollo-Guerrero, Juan A. Gómez-Pulido, and Arturo Durán-Domínguez. (2020). *Analyzing and Predicting Students' Performance by Means of Machine Learning: A Review*. Appl. Sci., 10(3), 1042.

L. H. Sang, N. T. Hai, T. T. Dien, and N. T. Nghe. (2020). Predicting academic performance using deep learning with multilayer neural networks. *Can Tho University Journal of Science*, vol. 56, no. 3.

Lee, C.-A., Tzeng, J.-W., Huang, N.-F., & Su, Y.-S. (2021). *Prediction of Student Performance in Massive Open Online Courses Using Deep Learning System Based on Learning Behaviors*. Educational Technology & Society, 24(3), 130–146.

Luo, Y., Han, X. & Zhang, C. (2024). *Prediction of learning outcomes with a machine learning algorithm based on online learning behavior data in blended courses*. Asia Pacific Educ. Rev. 25, 267–285.

Musso, M.F., Hernández, C.F.R. & Cascallar, E.C. (2020). *Predicting key educational outcomes in academic trajectories: a machine-learning approach*. High Educ 80, 875–894.

N. T. Uyen and N. M. Tam. (2019). Predicting student academic performance using data mining techniques. *Journal of Science, Vinh University*, vol. 48, no. 3A.

N. V. Thuy. (2023). Using machine learning models to predict students' on-time graduation status. *Journal of Banking Science & Training*, no. 255.

Ofori, F., Maina, E. & Gitonga, R. (2020). Using Machine Learning Algorithms to Predict Students' Performance and Improve Learning Outcome: A Literature Based Review. *Journal of Information and Technology*, Vol. 4(1), 33-55.

P. T. T. Mai and P. M. Duong. (2023). Machine learning and its potential applications in higher education institutions. *Journal of Educational Equipment: Education Management*, vol. 1, no. 288.

Psyridou, M., Prezja, F., Torppa, M. et al. (2024). *Machine learning predicts upper secondary education dropout as early as the end of primary school*. Scientìic Reports 14, 12956.

Richard Lamb, Knut Neumann, Kayleigh A. Linder. (2022). *Real-time prediction of science student learning outcomes using machine learning classification of hemodynamics during virtual reality and online learning sessions*. Computers and Education: Artificial Intelligence, Volume 3, 2022, 100078.

Tai Tan Mai, Marija Bezbradica, Martin Crane. (2022). *Learning behaviours data in programming education: Community analysis and outcome prediction with cleaned data*. Future Generation Computer Systems, Volume 127, Pages 42-55.

Tiwari, M., and Jain, N. (2024). Student Performance Prediction Using Machine Learning Algorithms. ShodhKosh: *Journal of Visual and Performing Arts*, 5(6), 1112–1122.

W. Chen, C. G. Brinton, D. Cao, A. Mason-Singh, C. Lu and M. Chiang. (2019). *Early Detection Prediction of Learning Outcomes in Online Short-Courses via Learning Behaviors*. IEEE Transactions on Learning Technologies, vol. 12, no. 1, pp. 44-58.

Yakubu, M.N. and Abubakar, A.M. (2022). *Applying machine learning approach to predict students' performance in higher educational institutions*. Kybernetes, Vol. 51 No. 2, pp. 916-934.