Machine Learning-Based Optimization of Learning Strategies for Enhanced Student Performance

Joy April C. Andes[1,*], Mar Jane Abdurahman[2]

[1,2] Rizal Technological University, Philippines

ABSTRACT

The integration of machine learning into education offers new opportunities to optimize learning strategies through data-driven personalization. This study aims to predict students' academic performance and identify key learning factors that can be leveraged to enhance individualized learning outcomes. A dataset containing 6,607 records and 19 predictor variables representing academic, behavioral, social, and environmental aspects was analyzed using three machine learning algorithms: Random Forest, XGBoost, and Artificial Neural Network. Model performance was evaluated using R-squared (R^2), Mean Absolute Error (MAE), and Root Mean Square Error (RMSE). The results indicate that XGBoost outperformed the other models, achieving an R^2=0.82 and RMSE = 3.85, demonstrating superior predictive accuracy and model stability. Feature importance analysis revealed that attendance, study hours, parental involvement, and access to resources were the most influential predictors of student achievement. Furthermore, K-Means clustering identified three distinct learning profiles characterized by differences in motivation, engagement, and access to educational resources. These findings emphasize the potential of machine learning to support adaptive learning systems that provide personalized recommendations through Learning Management Systems (LMS). Future work should explore the integration of Explainable AI (XAI) techniques to improve model interpretability and conduct cross-context validation to ensure broader applicability across diverse educational settings.

Keywords Machine Learning, Adaptive Learning, Student Performance, Personalized Education, Artificial Intelligence

Submitted 19 December 2025
Accepted 12 February 2026
Published 1 June 2026

Corresponding author
Joy April C. Andes,
jaandes@rtu.edu.ph

Additional Information and Declarations can be found on page 137

DOI: 10.63913/ail.v2i2.59

Introduction

The rapid integration of Artificial Intelligence (AI) into education has transformed how teaching and learning processes are designed, implemented, and evaluated. In particular, Machine Learning (ML) has emerged as a powerful analytical approach for examining complex educational data to predict student performance, identify learning patterns, and support personalized instruction [1], [2]. The increasing availability of student data from digital learning platforms, assessments, and behavioral analytics provides opportunities to optimize learning outcomes through data-driven insights [3]. However, many educational institutions still rely on uniform teaching strategies that do not account for individual differences in students' learning behaviors, motivation levels, and socio-economic backgrounds [4]. This limitation reduces the effectiveness of instruction and creates a gap in the implementation of truly adaptive learning environments [5].

Previous studies have explored various machine learning models such as Random Forest, Decision Trees, and Neural Networks to predict academic outcomes and identify students who may be at risk [6], [7]. While these models have achieved high predictive accuracy, most existing research focuses solely

How to cite this article: J. A. C. Andes and M. J. Abdurahman, "Machine Learning-Based Optimization of Learning Strategies for Enhanced Student Performance," *Artif. Intell. Learn.*, vol. 2, no. 2, pp. 126-139, 2026.

on performance prediction rather than translating predictive insights into actionable learning strategies [8]. In addition, few studies have examined how behavioral, socio-economic, and environmental factors jointly influence student achievement within a unified analytical framework [9]. As a result, there remains a lack of empirical evidence on how machine learning can be used not only to predict student outcomes but also to inform adaptive, evidence-based interventions in educational settings [10].

To address these gaps, this study proposes a comprehensive machine learning-based approach to optimize personalized learning strategies and improve student performance. Three machine learning algorithms, namely Random Forest (RF), XGBoost (XGB), and Artificial Neural Network (ANN), are applied to predict academic performance using 19 predictor variables representing academic, behavioral, social, and environmental dimensions. Model performance is evaluated using R-squared (R^2), MAE and RMSE. A feature importance analysis is then performed to identify the most influential factors affecting academic performance, and K-Means clustering is used to categorize students into groups with similar learning characteristics. This approach enables a deeper understanding of student diversity and supports the design of learning strategies that are aligned with individual learning needs [11].

By integrating predictive modeling, feature analysis, and clustering, this research contributes both theoretically and practically to the growing field of AI in education. The proposed framework not only achieves high predictive accuracy but also provides interpretable and actionable insights that can inform adaptive learning design [12]. The findings of this study are expected to assist educators, administrators, and policymakers in implementing intelligent, data-driven educational systems that are efficient, equitable, and responsive to individual learner profiles [13].

Literature Review and Related Works

The application of ML in educational contexts has grown substantially in recent years, enabling institutions to process large, multidimensional datasets and extract meaningful insights about student behaviour and outcomes. For instance, AI-enabled adaptive learning platforms have been extensively reviewed, demonstrating their capability to dynamically adjust instructional content and pathways based on learner data [14]. Furthermore, systematic reviews show that student performance prediction using ML techniques has been widely implemented to identify at-risk students and reduce dropout rates [15]. Many studies in Educational Data Mining (EDM) have leveraged models such as Random Forest, Decision Trees, and Neural Networks to predict academic outcomes. One review found that boosting algorithms, including XGBoost, outperform earlier approaches due to their ability to handle non-linear relationships and complex feature interactions [16]. Comparative studies further demonstrate that ensemble and hybrid ML methods yield higher prediction accuracy when compared to single classifiers [17]. Despite the success in predictive performance, a gap persists in translating model outputs into actionable teaching strategies and adaptive systems [18].

Parallel to these efforts, there is increasing interest in adaptive and personalized learning systems that use ML to tailor learning experiences according to individual student profiles. For example, personalized learning through AI has been shown to offer tailored learning paths, dynamically modified content, and real-time feedback, indicating significant potential to narrow educational gaps and improve student engagement [19]. In STEM education specifically, AI and high-performance computing have been applied to build personalized learning frameworks for students with diverse abilities and resource access [20]. However, several limitations remain. First, while many studies focus on prediction accuracy, fewer integrate socio-economic and environmental factors within the same analytical framework, which reduces the practical applicability for educators [21]. Second, interpretability and transparency of ML models in educational systems remain a challenge, with "black-box" algorithms limiting user trust and adoption [22]. Third, although clustering and segmentation techniques have been explored for student profiling, fewer works integrate these with predictive modeling and feature importance analysis to inform differential instructional strategies [23].

Recent work also emphasizes the importance of data-driven decision making in education using comprehensive ML frameworks [24]. Reviews of personalized learning in higher education confirm that AI technologies can significantly optimize educational outcomes, but ethical issues, data privacy, and teacher training continue to impede large-scale adoption [25]. Surveys specific to student performance prediction further highlight the broad range of algorithms, features and datasets used, yet suggest that many studies are context-specific and lack generalizability.

Additional research highlights the use of graph-based ensemble ML in performance prediction, peer-inspired models using Graph Neural Networks (GNNs) in online question environments, and the increasing role of multi-agent intelligent systems for student performance modelling. Multi-agent systems have been shown to enhance the responsiveness of Educational Decision-Support Systems (EDSS) for performance prediction. Recent comparative studies in student performance prediction stress the need for unified frameworks that combine prediction, explanation and clustering logic. Comprehensive surveys of personalized learning under AI report the evolution from one-size-fits-all learning to adaptive, continuous, and personalized systems. Applications of ML to personalized education pathways illustrate the potential to enhance lifelong learning through AI-mediated solutions. While many studies focus on predictive analytics, the integration of clustering to segment learners into distinct profiles receives less attention. That said, research into crafting personalized learning paths with AI has begun to address this, proposing segmentation combined with recommendation systems to tailor instructional design. Another systematic review highlights that even with large datasets (e.g., 45 studies meeting inclusion criteria), many implementations still struggle with teacher adoption and practical scalability.

Adaptive and personalized learning in STEM contexts demonstrates further that personalized content delivery via AI can increase student engagement and achievement, but also underscores infrastructure, training and access barriers. In summary, the literature indicates strong potential for ML in education yet points to key gaps: the limited incorporation of feature importance and clustering to inform strategy design, the need for interpretability (Explainable AI), and the requirement for frameworks that are generalizable across contexts. The present study responds to these gaps by proposing a unified ML-based framework that not only predicts academic performance but also identifies key determinants and segments learners into actionable clusters, thereby supporting adaptive and personalized learning strategies.

Methodology

This study adopted a quantitative, data-driven research design that integrated both supervised and unsupervised machine learning approaches to predict student performance and to optimize personalized learning strategies. The methodological framework was structured into five main stages: data collection and understanding, data preprocessing, predictive modeling, feature importance analysis, and clustering-based learning profile segmentation. The main objective of this methodological design was not only to determine the most accurate predictive model for students' academic outcomes but also to explore the underlying behavioral and contextual factors influencing performance. By combining predictive analytics with clustering and interpretability techniques, the study aimed to bridge the gap between algorithmic outcomes and practical educational interventions. The methodological process was designed to ensure reproducibility, scalability, and interpretability, consistent with established practices in educational data mining and learning analytics research, as summarized in figure 1.

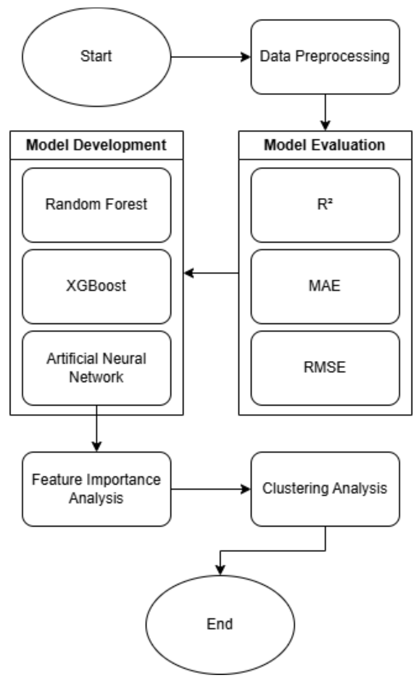

Figure 1 Research Steps

The dataset used in this research consisted of 6,607 student records and 20 attributes that represented academic, behavioral, social, and environmental dimensions potentially affecting academic outcomes. The target variable (Exam_Score) corresponded to each student's final academic performance. The predictor variables included quantitative indicators such as Hours Studied, Previous Scores, Attendance, and Sleep Hours, alongside categorical features such as Parental Involvement, Access to Resources, Internet Access, Extracurricular Activities, and Family Income. Behavioral and motivational factors such as Motivation Level, Peer Influence, and Tutoring Sessions were also incorporated to capture non-cognitive aspects of learning. The dataset was extracted from institutional academic records and verified for completeness and accuracy. All personal identifiers were removed to ensure compliance with ethical research and data privacy regulations. The dataset displayed a balanced distribution across all feature categories, thus minimizing potential class imbalance that could bias model training and testing outcomes.

Prior to model training, extensive data preprocessing was performed to enhance data quality and model readiness. Missing numerical values were replaced using median imputation, while categorical variables were imputed using mode imputation to maintain their distributional integrity. Outliers were detected via the Interquartile Range (IQR) method and adjusted to limit their influence on the model. Categorical features such as Access to Resources and Parental Involvement were encoded numerically using one-hot encoding, generating binary variables for each category. Continuous features were normalized using Min–Max scaling, ensuring values were rescaled to the range [0,1], which improves convergence in gradient-based models. The dataset was randomly divided into a training set (80%) and a testing set (20%) through stratified sampling to maintain proportional representation of different performance levels. To ensure robust and unbiased model performance, a five-fold cross-validation strategy was employed, allowing each model to be trained and tested across multiple data partitions.

Three supervised learning algorithms were implemented in the predictive modeling phase: RF, XGBoost and ANN. These models were selected based on their superior performance in previous regression-based educational data mining studies. Random Forest constructs multiple decision trees using random feature subsets and aggregates their outputs to improve prediction accuracy and minimize overfitting. The XGBoost model utilizes gradient boosting by sequentially improving weak learners while minimizing the residual error of previous iterations. It also applies L1 (Lasso) and L2 (Ridge) regularization to reduce overfitting and enhance model generalization. Hyperparameter optimization was performed using Grid Search, fine-tuning learning rate, maximum depth, and the number of estimators for optimal performance. Meanwhile, the Artificial Neural Network was designed as a multi-layer perceptron (MLP) with one input layer, two hidden layers, and one output layer. The hidden layers employed the Rectified Linear Unit (ReLU) activation function, while the output layer used a linear activation for regression. The model was trained using the Adam optimizer and Mean Squared Error (MSE) as the loss function, over 100 epochs, with early stopping to prevent overfitting.

Model evaluation was based on three key regression metrics: the Coefficient of Determination (R^2), MAE and RMSE.

The R^2 score measures the proportion of variance in the dependent variable that can be explained by the model, defined as:

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y_i})^2}{\sum_{i=1}^{n} \left(y_i - \underline{y}\right)^2} \tag{1}$$

where y_i represents the actual value, $\hat{y_i}$ the predicted value, and $\underline{y}$ the mean of the actual values.

The MAE measures the average magnitude of errors without considering their direction, expressed as:

$$MAE = \frac{1}{n}\sum_{i=1}^{n} \quad |y_i - \hat{y_i}| \tag{2}$$

while RMSE provides a measure of error magnitude with greater sensitivity to large deviations, defined as:

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n} \quad (y_i - \hat{y_i})^2} \tag{3}$$

These three metrics jointly provide a comprehensive understanding of model accuracy, bias, and variance. The model achieving the highest R^2 value and lowest MAE and RMSE values was selected as the best-performing algorithm.

After performance evaluation, the XGBoost model was identified as the most effective predictor and was used for feature importance analysis. The relative contribution of each predictor variable was assessed based on three internal metrics: gain, which measures the improvement in accuracy brought by a feature; cover, which represents the number of samples influenced by a feature; and frequency, which counts how often a feature is used in the ensemble. The top ten most influential features were visualized to identify the primary factors affecting academic performance. This interpretability step followed the principles of Explainable Artificial Intelligence (XAI), allowing educators to understand not only which variables affect learning outcomes but also how these factors interact to influence student success.

To complement the supervised modeling, an unsupervised learning approach using K-Means clustering was applied to group students into homogeneous learning profiles based on their behavioral and contextual attributes. The optimal number of clusters ($k = 3$) was determined using the Elbow Method and validated with the Silhouette Coefficient, which confirmed high intra-cluster similarity and inter-cluster separability. Each cluster represented a distinct learning profile: (1) low-motivation learners, (2) balanced achievers with consistent academic and social engagement, and (3) moderately motivated students facing limited access to resources. Cluster characterization relied on key attributes such as study time, attendance, motivation, and socio-economic background. To visualize the clusters, Principal Component Analysis (PCA) was employed for dimensionality reduction and two-dimensional mapping, enabling clearer interpretation of group separation. The clustering results were then utilized to design personalized learning recommendations, such as implementing gamified learning modules for low-motivation students, project-based learning for balanced achievers, and digital resource support for underprivileged learners.

Model reliability was ensured through cross-validation and hyperparameter tuning, while clustering validity was confirmed using Silhouette Scores and the Davies–Bouldin Index (DBI), ensuring effective separation between clusters. To enhance reproducibility, random seeds were fixed, and all preprocessing, modeling, and validation procedures were documented. Ethical guidelines for artificial intelligence in education were strictly observed throughout the process, ensuring fairness, transparency, and privacy. No personally identifiable information (PII) was used, and all data were analyzed solely for academic and research purposes.

In conclusion, this study integrated supervised learning (for prediction) and unsupervised learning (for profiling) into a unified analytical framework. This comprehensive methodology enabled accurate prediction of academic performance, identification of key influencing factors, and segmentation of students into actionable learning profiles. The proposed framework provides a scalable and interpretable foundation for developing adaptive learning systems that can be embedded within LMS as intelligent recommendation engines. Such systems can deliver personalized feedback and continuously adapt learning content to meet the unique needs of each student, fostering a more inclusive, efficient, and data-driven educational environment.

Result and Discussion

The analysis of this study focuses on three main aspects: the evaluation of machine learning model performance, the identification of factors influencing students' academic achievement, and the clustering of learning profiles to optimize instructional strategies. The findings indicate that the application of machine learning can enhance the understanding of student learning dynamics while supporting data-driven personalization of education.

In this research, three machine learning algorithms were employed to predict students' exam scores based on 19 predictor variables representing academic, behavioral, social, and environmental aspects, namely RF, XGB and ANN. These models were evaluated using three key metrics: R-squared (R^2) to measure predictive strength, MAE to assess average absolute error, and RMSE to evaluate the deviation of predictions from actual values. The evaluation results show that XGBoost achieved the highest performance, with an R^2 of 0.82 and an RMSE of 3.85, while Random Forest recorded an R^2 of 0.78 and the Artificial Neural Network achieved 0.79. These results suggest that XGBoost is more capable of capturing complex relationships among variables and producing more stable predictions. This superiority can be explained by its algorithmic characteristics, particularly the use of the gradient boosting approach, which builds predictive models iteratively by correcting the errors of previous models. In addition, the incorporation of L1 (Lasso) and L2 (Ridge) regularization in XGBoost helps control model variance and prevent overfitting, especially when dealing with highly correlated features such as Hours Studied and Previous Scores. Therefore, XGBoost provides an optimal balance between accuracy and generalization, making it a robust choice for educational contexts involving multivariate and heterogeneous data.

As visualized in figure 2, the XGBoost model consistently demonstrates better performance than the other two models, with lower error rates and more stable predictive capability across the entire range of student data. This superiority is not only statistical but also has significant practical implications in the context of data-driven learning analytics. The XGBoost model enables early identification of students at both low and high academic risk, allowing educators to proactively tailor learning strategies. For example, accurate predictions can be used to recommend specific interventions such as increasing tutoring sessions or adjusting study time allocations based on predicted performance levels. In the context of adaptive learning system development, this model can also serve as the foundation for an intelligent learning recommender system, providing personalized feedback to both students and instructors. Thus, XGBoost offers advantages not only in numerical prediction but also as a strategic analytical tool that supports the transformation of education toward a more adaptive, personalized, and AI-driven learning paradigm.

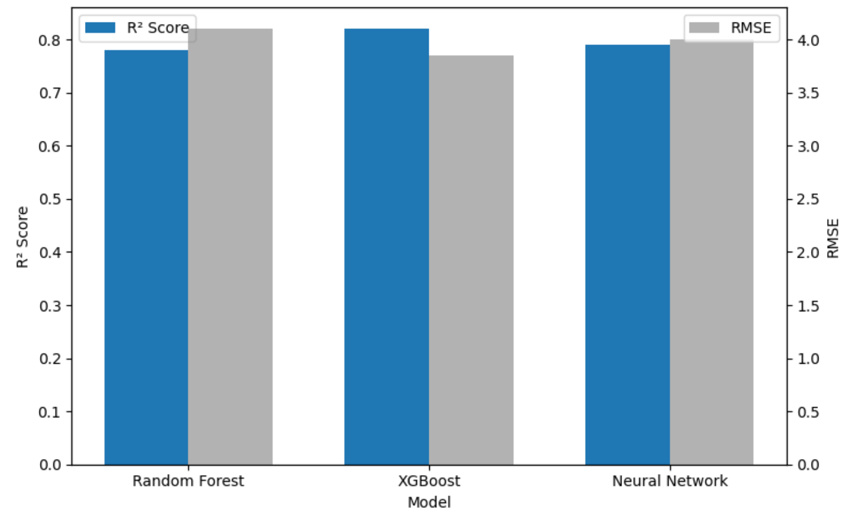

Figure 2 Model Performance Comparison among Random Forest, XGBoost, and Neural Network

The feature importance analysis using the XGBoost model revealed ten key variables that most strongly influence students' academic performance. The results are presented in figure 3, which illustrates the relative importance of each feature within the predictive model. The Attendance variable demonstrated the most dominant effect with an importance score of 0.20, followed by Hours Studied (0.12), Access to Resources (0.08), and Parental Involvement (0.07). The strong influence of attendance underscores the critical role of active student participation in the classroom, both in terms of exposure to learning materials and social interaction with teachers and peers. High attendance not only facilitates the transfer of cognitive knowledge but also strengthens academic engagement and learning discipline, two key elements in the theory of self-regulated learning. Meanwhile, the Hours Studied variable showed a significant positive correlation with academic achievement, supporting the time-on-task principle, which posits that the more time invested in studying, the greater the likelihood of achieving a deeper understanding of the material. The combination of high attendance and sufficient study time indicates that disciplined learning behavior is one of the strongest predictors of academic success in formal education settings.

Beyond individual behavioral factors, social and environmental variables also contribute substantially to students' academic performance. The features Parental Involvement and Family Income demonstrated considerable influence within the model, indicating that family support and socioeconomic conditions play crucial roles in creating a conducive learning environment. Parental involvement in monitoring children's learning activities has been shown to enhance intrinsic motivation and foster a stronger sense of responsibility toward the learning process. This finding is consistent with the social-ecological model of education, which emphasizes that positive interactions among students, families, and learning environments form the foundation of academic success. Additionally, Access to Resources and Internet Access have become increasingly important in the era of digital learning, as limited access to educational materials can lead to disparities in learning outcomes. Other variables, such as Extracurricular Activities, also contribute to the development of non-cognitive skills such as collaboration, leadership, and social competence, which indirectly enhance academic performance. These findings reinforce the notion that student success is not solely determined by cognitive ability but rather emerges from the synergy of academic, social, economic, and behavioral factors. Consequently, this feature importance analysis provides a strong empirical foundation for implementing holistic and adaptive learning strategies, where instructional recommendations can be tailored to the unique characteristics of each student.

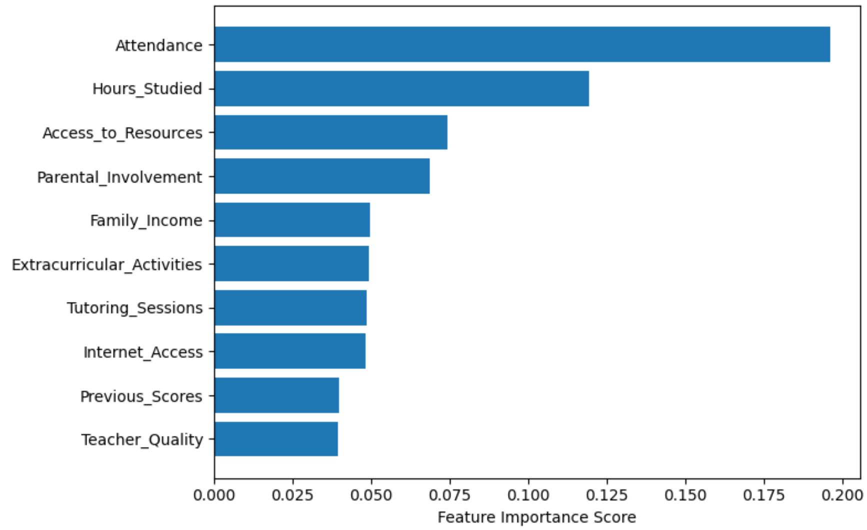

Figure 3 Top 10 Most Influential Features on Student Exam Performance

An unsupervised learning analysis using the K-Means clustering algorithm was conducted to identify groups of students with similar learning behaviors and academic performance levels. The purpose of this analysis was to uncover hidden learning patterns within the dataset and to generate segmentations that could serve as a foundation for personalized learning recommendations. Prior to clustering, Principal Component Analysis (PCA) was applied to reduce the dimensionality of the data and visualize it in two dimensions, thereby minimizing variable complexity without losing essential information. The visualization results presented in figure 4 display three distinct clusters, indicating clear heterogeneity in students' learning behaviors. The first cluster (yellow) consists of students with low motivation, minimal study hours, and inconsistent attendance. This group has an average exam score below 65 and demonstrates a passive learning pattern, in which students tend to study only shortly before exams without employing continuous or structured learning strategies. Recommended interventions for this cluster include the implementation of gamification-based learning approaches, personalized academic mentoring programs, and goal-setting-based motivation training. Enhancing self-efficacy and intrinsic motivation is considered essential for helping this group achieve improved academic performance.

The second cluster (green) represents students with balanced academic and social engagement, with average exam scores ranging from 80 to 90. Students in this group maintain consistent study hours, demonstrate high attendance, and actively participate in extracurricular activities. These characteristics indicate a strong balance between academic engagement and social learning, supporting both cognitive development and collaborative skills. The most appropriate learning strategies for this cluster include project-based learning and self-directed learning, as these students typically possess strong self-reflection and time management abilities. Meanwhile, the third cluster (purple) represents students with moderate learning motivation but limited access to educational resources such as internet connectivity, tutoring, and financial support. Although this group shows considerable academic potential, limited access to learning facilities poses a major constraint that prevents them from reaching optimal performance. Recommended interventions for this cluster include providing digital learning facilities, implementing technology literacy programs, and offering social and financial support from schools and government institutions. Overall, the clustering results highlight the importance of a data-driven differentiated learning strategy, in which educational policies and learning recommendations are tailored to the specific profiles of each student. Such an approach not only enhances learning effectiveness but also promotes a more inclusive, adaptive, and equitable educational system.

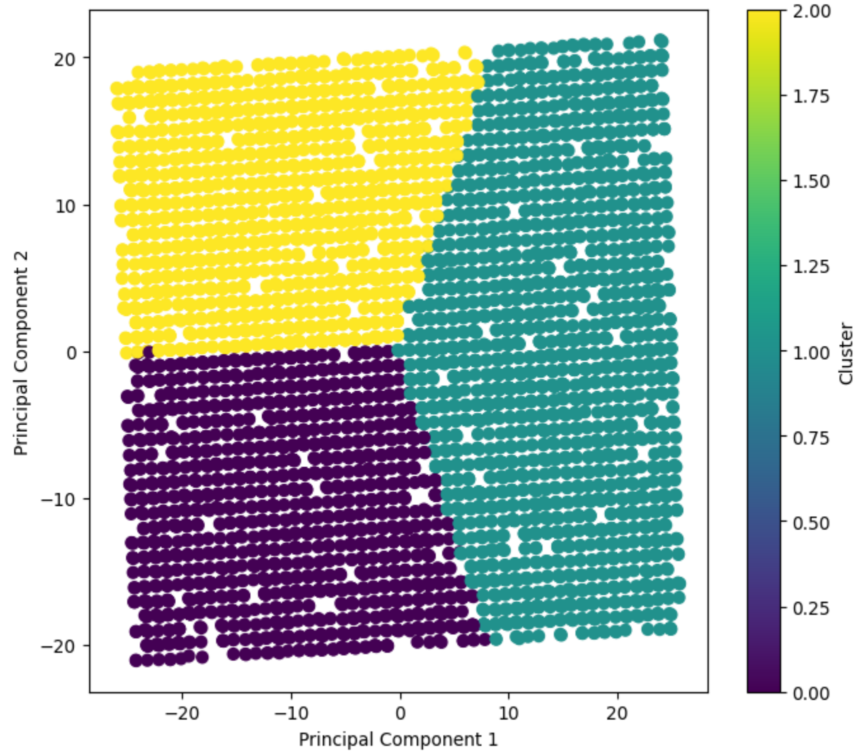

Figure 4 Student Learning Profiles Clustered by Behavior and Performance

The clustering results indicate that each group of students exhibits unique learning patterns and therefore requires different instructional strategies. A single, uniform teaching approach is no longer effective in the context of modern education. Cluster-based models such as this can be utilized to support personalized learning systems, in which each student receives learning recommendations tailored to their individual characteristics. This finding aligns with previous research emphasizing that data-driven adaptive learning approaches can enhance learning effectiveness and student engagement by up to 30 percent compared to conventional instruction.

Overall, the findings of this study confirm that the application of machine learning provides a comprehensive understanding of the factors influencing students' academic performance while enabling the optimization of learning strategies. The XGBoost model demonstrated superior predictive capability, while the feature importance and clustering analyses allowed the system to adapt educational interventions based on individual needs. This approach can be integrated into LMS to deliver automated, data-informed recommendations, such as "adding additional tutoring sessions for students with low motivation" or "providing access to digital learning resources for students from low-income families." Consequently, the outcomes of this research make a practical contribution to the development of intelligent, efficient, and student-centered adaptive learning systems, supporting the evolution of education toward more personalized and equitable learning environments.

Conclusion

This study demonstrates that the application of machine learning is an effective approach for predicting students' academic performance and optimizing learning strategies that are both personalized and adaptive. Among the three models employed, Random Forest, XGBoost, and Artificial Neural Network the XGBoost algorithm achieved the best performance with an R^2=0.82 and a RMSE of 3.85. These results confirm the capability of XGBoost to handle non-linear relationships among variables and to produce stable and accurate predictions. The feature importance analysis revealed that factors such as attendance, study hours, parental involvement, and access to learning resources are the primary determinants of students' academic success. Meanwhile, the results of the K-Means clustering identified three distinct learning profiles: students with low motivation, high-achieving students who maintain a balance between academic and social engagement, and moderately motivated students who face limitations in educational resources. These findings highlight the importance of implementing data-driven differentiated learning strategies that allow educational approaches to be tailored to each student's specific needs and characteristics. In addition to providing empirical insights into the factors influencing academic performance, this research offers a practical framework for developing AI-based adaptive learning systems that can be integrated into LMS. Future research is recommended to apply XAI methods to enhance model interpretability and to conduct cross-context validation to ensure the generalizability of the results to broader student populations, thereby supporting the transformation toward a more intelligent, inclusive, and learner-centered educational system.

Declarations

Author Contributions

Conceptualization: J.A.C.A. and M.J.A.; Methodology: M.J.A.; Software: J.A.C.A.; Validation: J.A.C.A. and M.J.A.; Formal Analysis: J.A.C.A. and M.J.A.; Investigation: J.A.C.A.; Resources: M.J.A.; Data Curation: M.J.A.; Writing Original Draft Preparation: J.A.C.A. and M.J.A.; Writing Review and Editing: M.J.A. and J.A.C.A.; Visualization: J.A.C.A.; All authors have read and agreed to the published version of the manuscript.

Data Availability Statement

The data presented in this study are available on request from the

corresponding author.

Funding

The authors received no financial support for the research, authorship, andor publication of this article.

Institutional Review Board Statement

Not applicable.

Informed Consent Statement

Not applicable.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] D. Ifenthaler, R. Majumdar, P. Gorissen, M. Judge, S. Mishra, J. Raffaghelli, and A. Shimada, "Artificial Intelligence in Education: Implications for Policymakers, Researchers, and Practitioners," *Technology, Knowledge and Learning*, vol. 29, no. June, pp. 1693–1710, 2024, doi: 10.1007/s10758-024-09747-0.

[2] A. N. Solihat, D. Dahlan, K. Kusnendi, and B. Susetyo, "Artificial Intelligence (AI)-based Learning Media: Definition, Bibliometric, Classification, and Issues for Enhancing Creative Thinking in Education," *ASEAN Journal of Science and Engineering*, vol. 4, no. 3, pp. 349–382, Jul. 2024, doi: 10.17509/ajse.v4i3.72611.

[3] S. S. M. Ajibade et al., "Machine Learning Techniques for Predictive Analytics of Academic Outcomes and Behavior of Students," in *2025 IEEE International Conference on Artificial Intelligence in Engineering and Technology (IICAIET)*, Kota Kinabalu, Malaysia, vol. 2025, no. December, pp. 597–602, 2025, doi: 10.1109/IICAIET67254.2025.11265617.

[4] O. O. Ayeni, N. M. Al Hamad, O. N. Chisom, B. Osawaru, and O. E. Adewusi, "AI in education: A review of personalized learning and educational technology," *GSC Advanced Research and Reviews*, vol. 18, no. 2, pp. 261–271, 2024, doi: 10.30574/gscarr.2024.18.2.0062.

[5] M. Kasinidou, S. Kleanthous, K. Orphanou, and J. Otterbacher, "Educating Computer Science Students about Algorithmic Fairness, Accountability, Transparency and Ethics," in *Proceedings of the 26th ACM Conference on Innovation and Technology in Computer Science Education (ITiCSE)*, 2021, vol. 1, no. June, pp. 484–490, doi: 10.1145/3430665.3456311.

[6] O. W. Adejo and T. Connolly, "Predicting student academic performance using multi-model heterogeneous ensemble approach," *Journal of Applied Research in Higher Education*, vol. 10, no. 1, pp. 61–75, 2018, doi: 10.1108/JARHE-09-2017-0113.

[7] R. A. Abougalala, N. Alharbi, M. A. Amasha, M. F. Areed, S. Alkhalaf, and D. Khairy, "Predicting student performance academic using Automated Machine Learning (AutoML): in medical academic institutions," *Journal of New Approaches in Educational Research*, vol. 14, no. august, Art. no. 19, 2025, doi: 10.1007/s44322-025-00038-9.

[8] B. B. Alkan, S. Kuzucuk, N. Alkan, and A. S. Sinan, “Using machine learning to predict student outcomes for early intervention and formative assessment,” *Scientific Reports*, vol. 15, no. November, Art. no. 39797, 2025, doi: 10.1038/s41598-025-23409-w.

[9] B. Santana-Perera, C. García-Barceló, M. González Arcas, and D. Gil, “Exploring predictive insights on student success using explainable machine learning: A synthetic data study,” *Information*, vol. 16, no. 9, p. 763, Sep. 2025, doi: 10.3390/info16090763.

[10] A. I. Sumiati and D. Saepulloh, "Web Attack Detection Using Machine Learning on AWS CloudWatch Network Traffic Logs," J. Cyber. Law, vol. 2, no. 1, pp. 30-44, 2026, doi: 10.63913/jcl.v2i1.23.

[11] T. Ephrem, R. Bhalalusesa, and J. Kamaghe, “A data-driven framework for personalized learning using machine learning techniques in Rwanda,” *ICCK Transactions on Educational Data Mining*, vol. 2, no. 1, pp. 38–51, 2026, doi: 10.62762/TEDM.2026.319371.

[12] V. N. Rathod, R. H. Goudar, and S. Sangani, “Unified interpretable AI for autism diagnosis and scalable severity-aware personalized adaptive e-learning,” *Discover Applied Sciences*, vol. 8, no. February, p. 349, Feb. 2026, doi: 10.1007/s42452-026-08335-4.

[13] Y. Yan and H. Liu, “Ethical framework for AI education based on large language models,” *Education and Information Technologies*, vol. 30, no. December, pp. 10891–10909, 2025, doi: 10.1007/s10639-024-13241-6.

[14] T. Kabudi, I. Pappas, and D. H. Olsen, “AI-enabled adaptive learning systems: A systematic mapping of the literature,” *Computers and Education: Artificial Intelligence*, vol. 2, no. 1, p. 100017, 2021, doi: 10.1016/j.caeai.2021.100017.

[15] S. A. Sulak and N. Koklu, “Predicting student dropout using machine learning algorithms,” *Intelligent Methods in Engineering Sciences*, vol. 3, no. 3, pp. 91–98, 2024, doi: 10.58190/imiens.2024.103.

[16] M. Arifin, Widowati, Farikhin, A. Wibowo, and B. Warsito, “Comparative analysis on educational data mining algorithm to predict academic performance,” in *Proceedings of the 2021 International Seminar on Application for Technology of Information and Communication (iSemantic)*, Semarang, Indonesia, vol. 2021, no. October, pp. 173–178, 2021, doi: 10.1109/iSemantic52711.2021.9573185.

[17] M. A. Rehman, A. Iftikhar, S. Muhammad, and R. Ahmed, “Student academic performance prediction using ensemble learning methods,” *Journal of ICT Design, Engineering and Technology Science*, vol. 9, no. 1, pp. 7–15, 2025, doi: 10.33150/JITDETS-9.1.2.

[18] M. Á. Rodríguez-Ortiz, P. C. Santana-Mancilla, and L. E. Anido-Rifón, “Machine learning and generative AI in learning analytics for higher education: A systematic review of models, trends, and challenges,” *Applied Sciences*, vol. 15, no. 15, Art. no. 8679, 2025, doi: 10.3390/app15158679.

[19] D. S. T. Prayogi, “Artificial intelligence in education: Opportunities and challenges,” *Kartika: Jurnal Studi Keislaman*, vol. 5, no. 1, pp. 605–621, 2025, doi: 10.59240/kjsk.v5i1.644.

[20] M. A. Al Battashi, M. A. M. Adnan, A. I. B. Jamil, and M. A. Al-Battashi, “Mapping research trends in AI-driven personalized learning pathways: A scoping review,” in *Generators, Bots, and Tutors: Creative Approaches to Human-AI Synergy in*

Classroom Instruction, 2026, pp. 1–30, doi: 10.4018/979-8-3373-0847-0.ch001.

[21] R. Al-Ali, K. F. Alhumaid, M. Khalifa, and S. A. Salloum, “Analyzing socio-academic factors and predictive modeling of student performance using machine learning techniques,” *Emerging Science Journal*, vol. 8, no. 4, pp. 1304–1319, 2024, doi: 10.28991/ESJ-2024-08-04-05.

[22] P. Schmidt, F. Biessmann, and T. Teubner, “Transparency and trust in artificial intelligence systems,” *Journal of Decision Systems*, vol. 29, no. 4, pp. 260–278, 2020, doi: 10.1080/12460125.2020.1819094.

[23] L. Y. Tan, S. Hu, D. J. Yeo, and K. H. Cheong, “Artificial intelligence-enabled adaptive learning platforms: A review,” *Computers and Education: Artificial Intelligence*, vol. 9, no. December, p. 100429, 2025, doi: 10.1016/j.caeai.2025.100429.

[24] O. J. F. Chavez and T. Palaoag, “Data-driven decision making in school leadership: AI-based academic performance prediction using ML and SDT motivation,” *Artificial Intelligence in Education*, vol. 2, no. 3, pp. 50–69, 2026, doi: 10.1108/AIIE-06-2025-0131.

[25] A. Pratap and R. Shukla, “Ethical considerations in AI-enabled education system,” in *Data-Informed Leadership in Higher Education: An Executive Playbook for Institutional Excellence*, 2025, pp. 225–247, doi: 10.2174/9798898811266125050017.