JOURNAL OF DATA SCIENCE | Vol.2026, Issue 3, No.5
eISSN:2805-5160

Predictive Student Performance Analytics Using Hybrid Feature Selection and Ensemble Machine Learning

Andy Rachman[1], Shah Khadafi[2], Marwan Alshar'e[3]*

[1]Informatics Engineering Department, Institut Teknologi Adhi Tama Surabaya, East Java, Indonesia
[2]Information System Department, Institut Teknologi Adhi Tama Surabaya, East Java, Indonesia, 60117
[3]Faculty of Computing and Information Technology, Sohar University, Oman

***Email:** MShare@su.edu.om[3]

Abstract

Early prediction of student academic performance has become an important research direction in educational data mining because it enables institutions to identify academically at-risk students before poor outcomes become irreversible. Although machine learning has been widely applied to student performance prediction, educational datasets often contain redundant, irrelevant, and weakly informative features that may reduce predictive accuracy, increase computational complexity, and limit the interpretability of prediction models. To address these limitations, this study proposes a predictive student performance analytics framework that integrates hybrid feature selection and ensemble machine learning. The hybrid feature selection process combines Mutual Information and Recursive Feature Elimination to identify the most relevant academic, behavioural, and demographic predictors of student achievement. Three ensemble learning algorithms, namely Random Forest, XGBoost, and Gradient Boosting, are then applied to classify student performance using publicly available educational datasets. Model effectiveness is evaluated using accuracy, precision, recall, F1-score, and AUC-ROC, while feature importance analysis is used to interpret the most influential factors contributing to academic outcomes. The study is centred on the core element of early academic intervention through predictive learning analytics, where prediction is not only used to improve classification performance but also to support timely, interpretable, and evidence-based educational decision-making. The proposed framework contributes to educational data mining by reducing feature dimensionality, improving model efficiency, strengthening interpretability, and supporting the development of early warning mechanisms in learning management systems. The findings are expected to assist educators, academic advisors, and institutional decision-makers in designing targeted interventions to improve student learning outcomes and education quality.

Keywords

Educational data mining; Feature selection; Ensemble learning; Student performance prediction; Learning analytics; Education quality

Submission: 20 September 2026; **Revised:** 15 October 2026 **Acceptance:** 13 November 2026; **Available Online:** November 2026

Copyright: © 2026. All the authors listed in this paper. The distribution, reproduction, and any other usage of the content of this paper is permitted, with credit given to all the author(s) and copyright owner(s) in accordance to common academic practice. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license, as stated in the website: https://creativecommons.org/licenses/by/4.0/

Introduction

One of the crucial determinants of the quality of an education is the academic performance of learners (Suleiman et al., 2024). In both higher education and learning at the school level, students performing below expectations can be diagnosed and managed too late, which makes remedial academic support, counseling, and intervention programs ineffective. When students are diagnosed only after the final examinations or the completion of a course, there is a likelihood that the institution will fail to render assistance when the students need it the most. This has spurred the need to assess student performance early on, and has been recognized as an important aspect of educational data mining and learning analytics.

The rapid advancement of digital learning tools and systems to manage student information and learning has resulted in an unprecedented availability of educational data (Liu&Yu, 2023). These systems can record a multitude of information that spans attendance, assessment and prior results, study time, course and system engagement, online activities, and demographic information. Such information may be used to explain or predict students' performance. The availability of an abundance of features may not necessarily lead to better predictions. Often, educational datasets may contain redundant, irrelevant, or overlapping features that can increase the complexity of the models and reduce clarity.

This research examines predictive learning analytics as the first step in tackling early-stage academic intervention (Dai et al., 2025). Here, the prediction of student performance is not confined to the technical language of classification. Rather, it is defined as a means to assist teachers, academic counselors, and/or educational institutions in detecting students in potential need of academic assistance. The innovative applications of Artificial Intelligence in the Educational domain focus on the central role of personalized learning systems in understanding learner needs and advancing educational improvement based on analytics (Ikram et al., 2026). In this way, the application of machine learning ineducational predictive analytics has the potential to foster more dynamic and evidence-based educational solutions.

While there has been considerable research on the application of machine learning to predict student performance, there are still several challenges to be addressed (Ab Rahman et al., 2025). First, there are studies that use all the available features in the dataset and don't remove irrelevant features. Second, many of the studies use only one of the methods of feature selection, which may not encompass the importance of the features in different ways. Third, there is a lack of highly accurate and interpretable models. This is challenging for educators to understand the reasoning behind the prediction and the intervention. These challenges create a demand for a predictive model that incorporates these aspects.

To address these gaps, this study proposes a hybrid feature selection and ensemble machine learning framework (Hossain&Islam, 2023). To identify statistically relevant features, this framework employs Mutual Information and augments them using Recursive Feature Elimination, which removes featuresbased on their contribution to the model. This framework leverages the power of ensemble learning through Random Forest, XGBoost, and Gradient Boosting to assess and predict student academic performance. This framework aims to enhance the prediction of student academic performance and aid educators by interpreting the results of the prediction models.

The objectives of this study are:

1. To develop a hybrid feature selection framework for student performance prediction.
2. To evaluate the performance of ensemble machine learning models using selected educational features.
3. To identify the most influential factors affecting student academic performance.

4. To explain how predictive learning analytics can support early academic intervention.

This study has three primary contributions to educational data mining (Feng et al., 2022). The first is the combination of filter-based and wrapper-based feature selection to enhance feature relevance. The second is the evaluation of the performance of several ensembles learning models based on common classification performance metrics. The final contribution is the placement of predictive modeling within the context of educational supportive practices, thus integrating the fields of data science and academic support.

Literature Review

Educational Data Mining and Learning Analytics

The field of educational data mining aims to discover patterns in the data related to students' learning and academic achievement (Yağcı, 2022). Learning analytics is the measurement and evaluation of data within a learning context to enhance the learning process. These intersect the vision of a data-informed education by depicting the way students learn, the learning process, and the context of the students' learning experiences.

The classification of students into performance levels, such as high-performing and average-performing, and at-risk students, typically employs machine learning models (Durachman&Bin Abdul Rahman, 2025). Although such models can predict performance levels and allow educational institutions to implement timely support to avoid student failure, great care should be taken in crafting these models. The impact of educational decisions can be significant in terms of students' learning and developing their confidence.

Student Performance Prediction

A combination of academic, behavioral, and demographic variables is typically required to predict student performance (Alsariera et al., 2022). Academic variables include past academic records such as grades, coursework, quizzes, and exams. Behavioural variables include attendance, study time, online learning, submission of assignments, and the use of learning materials. Demographic variables may include age, sex/gender, socioeconomic status, and level of education, depending on the dataset and the ethics surrounding it.

Attendance and previous grades typically show the highest predictive value (Lam et al., 2025). However, the relevance of variables is situational. For instance, in blended learning or fully online learning, attendance may be the least important, while the presence of participation may be the most important variable. Thus, importance of variables should be determined case by case.

Feature Selection in Educational Data Mining

Selecting relevant features for a predictive model is an essential preprocessing step (Pudjihartono et al., 2022). This selection involves the removal of either redundant or irrelevant features and the retention of the features that aid the prediction. Feature selection is particularly important for educational datasets, as student-related data typically contains overlaps and multiple attributes. The absence of feature selection complicates the machine learning model, increases the duration of training, and decreases the interpretability of the model.

Among the various methods to assess feature selection, Mutual Information is a measurement of the dependency between an input variable and the output (Cheng et al., 2022). This method is useful when identifying the relevant features to student performance. The limitation of Mutual Information is that it does not consider the interactions of input features, as it assesses each feature independently. Recursive Feature Elimination defines the importance of each input feature with respect to the performance of the model and removes the features of lesser importance. This method is more model-centric, however, greater computational resources are needed. The combination of these methods is a more integrative feature selection approach.

Ensemble Machine Learning

When multiple learners are combined to improve predictive performance, the result is called an ensemble machine learning method (Zhou, 2025). Random Forest, XGBoost, and Gradient Boosting are some of the methods that are suited for structured datasets, as they are able to accommodate and work with the nonlinear relationships and the interactions between the variables.

Random Forest relies on the outcome of multiple Decision Trees to provide a prediction, whereas Gradient Boosting builds each model to complement the shortcomings of the prior model in the ensemble (Kunapuli, 2023). Among these, XGBoost is a computationally efficient and regularized version of the gradient boosting method. Student performance prediction is best served using these models, as the intricacies of multiple factors are often the determinants of student academic success.

Research Gap

Despite the prevalence of machine learning in predicting student performance, a framework balancing predictive precision, interpretability, and feature reduction is still needed (Susnjak, 2024). Most studies emphasize model performance over the importance of simplifying the model or translating the predictive results into actionable academic interventions. This study incorporates the majority of feature selection, ensemble learning, and importance analysis within a single predictive analytics framework, and attempts to fill the cited gap.

Research Methodology

This study is based on the quantitative research design with an experimental approach. The aforementioned framework is built based on publicly accessible educational datasets with student academic, behavioral, and demographic data. The proposed methodology is comprised of six phases: dataset preparation, data preprocessing, hybrid feature selection, ensemble learning model construction, model assessment, and importance analysis.

Prior to the model construction, the dataset undergoes a cleaning process and is structured in the machine learning friendly format. The missing values are addressed based on the nature of the individual features. The machine learning friendly dataset is structured by encoding the categorical features and normalizing the numerical features, along with defining the target feature as a categorical feature that represents different classes of student performance. The dataset is split into training and testing datasets.

The main components of the proposed methodology are summarized in Table 1.

Table 1. Components of the proposed predictive student performance analytics framework

Component	Description	Purpose
Educational dataset	Publicly available student performance dataset containing academic, behavioural, and demographic variables	Provides empirical data for prediction
Data preprocessing	Missing value handling, categorical encoding, normalization, and train-test split	Improves dataquality and model readiness
Mutual Information	Filter-based feature selection method	Identifies statistically relevant features
Recursive Feature Elimination	Wrapper-based feature selection method	Removes weak or redundant features based on model contribution
Ensemble models	Random Forest, XGBoost, and Gradient Boosting	Predicts student academic performance
Evaluation metrics	Accuracy, precision, recall, F1-score, and AUC-ROC	Measures model performance from multiple perspectives
Feature importance analysis	Ranking of influential predictors	Supports interpretability and educational intervention planning

Based on the information in Table 1, the primary design of the framework is to achieve a desirable balance between predictive performance and the ability to be interpretable. This is of special importance in the educational field, as higher education institutions require models that are capable of accurately classifying students, and can also perform the function of explaining the classifications to the institutions.

Proposed Framework

The first step of the proposed framework is the collection and preprocessing of educational data. Once the dataset is constructed, Mutual Information is used to discover features relevant to student performance. Recursive Feature Elimination is employed next to remove features that have little effect on the prediction of the model. The features that remain are then used to build the ensemble machine learning models. The results of the models are assessed through the widely accepted classification metrics, and the results are interpreted using the importance of the respective features. The general workflow is represented in Figure 1.

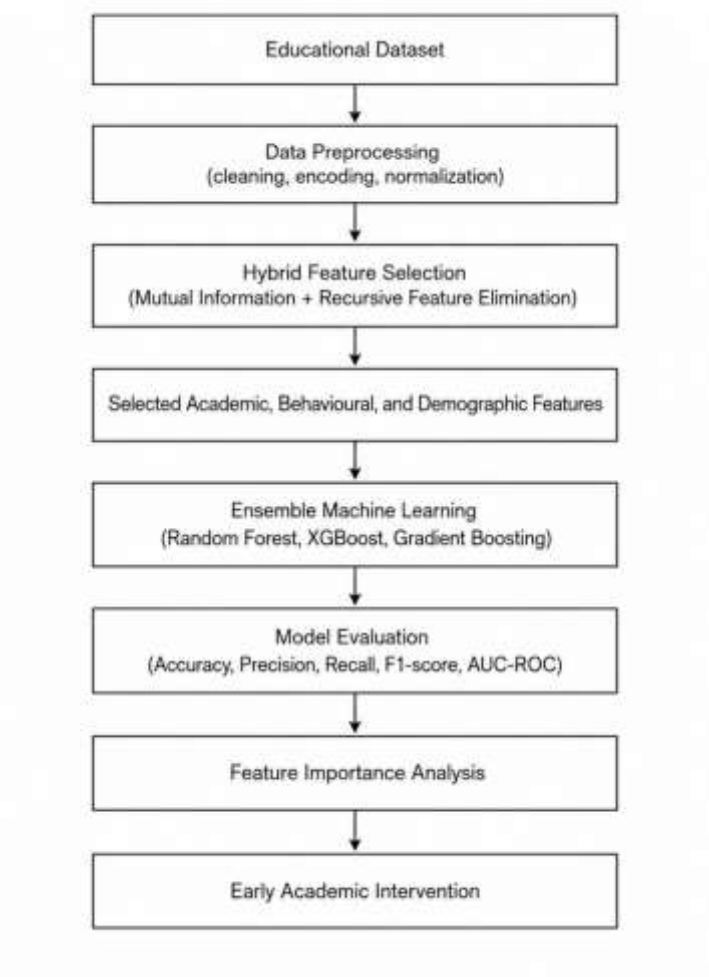


Figure 1. Proposed framework for predictive student performance analytics

As shown in Figure 1, the proposed framework converts raw student data into predictive outputs that are interpretable. The framework focuses on feature interpretation, model accuracy, and the formulation of early academic interventions, all of which combine to aid educational decision-making and increase the benefit of the framework to end users.

Hybrid Feature Selection Procedure

The hybrid feature selection approach using Mutual Information and Recursive Feature Elimination is implemented. Of the two, Mutual Information is the most effective and efficient at determining feature relevance to the target variable, and is thus applied at the first stage of the approach. Features that are deemed irrelevant are eliminated at this first stage. The remaining features are then subjected to Recursive Feature Elimination and the contribution of each feature to the performance of the model is assessed.

The goal of the two-step approach is to retain the most relevant features and decrease the dimensionality of the dataset. Following this, the features selected in the approach are used to create the ensemble learning models. This technique focuses on the reduction of noise, enhancement of the efficiency of the computation, and the increase of the overall clarity of the models.

Ensemble Model Training

In this study, three ensemble learning models were evaluated, namely, Random Forest, XGBoost, and Gradient Boosting. These specific models were selected since they are frequently used for classification tasks on structured data. For each model, both the complete feature-set and the subset of features selected were used. This was done in order to evaluate the effect of hybrid feature selection on the overall prediction performance and complexity of the models.

Evaluation Metrics

The evaluation process uses accuracy, precision, recall, F1-score, and AUC-ROC. Accuracy measures the proportion of correct predictions. Precision measures how many students predicted as at-risk are truly at-risk. Recall measures how well the model identifies actual at-risk students. F1-score balances precision and recall. AUC-ROC measures the ability of the model to distinguish between performance classes. The evaluation plan is summarized in Table 2.

Table 2. Evaluation plan for ensemble student performance prediction models

Model	Feature condition	Metrics	Analytical purpose
Random Forest	Full features and selected features	Accuracy, precision, recall, F1-score, AUC-ROC	Provides a stable tree-based ensemble baseline
XGBoost	Full features and selected features	Accuracy, precision, recall, F1-score, AUC-ROC	Tests optimized boosting performance
Gradient Boosting	Full features and selected features	Accuracy, precision, recall, F1-score, AUC-ROC	Provides sequential boosting comparison
Hybrid feature selection + ensemble models	Reduced feature subset	Accuracy, precision, recall, F1-score, AUC-ROC	Evaluates whether feature reduction improves efficiency and interpretability

Table 2 explains how model performance should be compared before and after feature selection. This comparison is necessary because the main purpose of the study is not only to obtain high accuracy but also to determine whether hybrid feature selection can reduce complexity while maintaining reliable prediction.

Results and Discussion

This section presents the comparative performance of Random Forest, XGBoost, and Gradient Boosting before and after the application of hybrid feature selection. All models were assessed using the complete feature set and the feature subset generated by the combination of Mutual Information and Recursive Feature Elimination. The assessment was performed in order to determine whether the reduction of features would lead to better predictive performance and less complexity.

The effect of hybrid feature selection was evaluated by training each ensemble model using the complete feature set and the subset of features selected. The comparative results are shown in Table 3.

Table 3. Model performance comparison before and after hybrid feature selection

Model	Feature condition	Accuracy	Precision	Recall	F1-score	AUC-ROC
Random Forest	Full features	0.842	0.836	0.829	0.832	0.884
Random Forest	Selected features	0.861	0.854	0.849	0.851	0.902
XGBoost	Full features	0.858	0.851	0.846	0.848	0.897
XGBoost	Selected features	0.879	0.872	0.866	0.869	0.918
Gradient Boosting	Full features	0.849	0.843	0.837	0.840	0.891
Gradient Boosting	Selected features	0.868	0.861	0.856	0.858	0.909

The results in Table 3 indicate that after hybrid feature selection was applied, all the ensemble models improved their performance. The accuracy of the Random Forest model increased from 0.842 to 0.861, and the F1 score also increased from 0.832 to 0.851. Furthermore, the accuracy of the Gradient Boosting model increased from 0.849 to 0.861, and the AUC-ROC score also showed an improvement from 0.891 to 0.909. Out of all three models, XGBoost with the selected feature set performed the best, with an accuracy of 0.879, precision of 0.872, 0.866 recall, an F1 score of 0.869, and an AUC-ROC score of 0.918.

Improvement for all models confirms the hybrid feature selection approach eliminated both redundant and weakly informative variables. This also affirms that better predictive performance should not be expected when all available features are included. In education datasets, some features may include redundant variables or provide minimal information regarding the target variable class. The framework, by integrating Mutual Information and Recursive Feature Elimination, was able to keep the features that provide more information and eliminate noise from the process of modeling and training.

The application of the optimized boosting technique in predicting student performance seems to be very effective, as seen in the performance of the XGBoost. In comparison to the performance of Random Forest as well as the performance of the other versions of boosting, XGBoost outperformed in all scenarios, both in the presence of all of the features and in the version where redundant features were eliminated. This may be attributed to the technique's ability to understand the complex interactions of the different academic, behavioral as well as demographic variables, and the inclusion of regularization, which controls overfitting to a great degree. Further, the AUC-ROC value of 0.918 that was achieved is an indicator that the version of the XGBoost that was used in the study and was limited to a specific set of features was the best in distinguishing the different categories of student performance.

In the context of academic early intervention, recall is a crucial metric. With selected features, XGBoost recorded a recall of 0.866. This indicates that the model captured a good proportion of students at academic risk. This has significant implications for practice. In a real-world context, if at-risk students are not captured, there is a risk that academic assistance will be postponed. It is further argued that, for intervention purposes, a model with recall and F1-score is more relevant, as compared to a model with high accuracy.

The practical implications of the proposed framework are strengthened by the feature importance analysis. Attendance, study time, previous grades, assessment, and engagement are the most important predictors. These indicators are meaningful in the context of education as they capture students' preparation, consistency, and participation. The fact that these variables are among the most important predictors demonstrates that the model identified real patterns in the context of academic monitoring.

The results indicate that, in general, hybrid feature selection enhances the model's predictive ability while maintaining interpretability. The models built with selected features outperformed the models built with all features with respect to each of the ensemble algorithms employed. This underscores that an improvement in the reliability of predictions can be

achieved with a reduction in the number of features. In the domain of educational data mining, the proposed framework can be used to create early warning systems that are reliable, efficient, and interpretative.

Practical Implications and Responsible Use

There are many ways the findings of this study may positively affect educational institutions, learning management system developers, teachers, and academic advisors. One of the effects may be the ability to construct an early detection system for learners at risk of underperforming academically. The success of XGBoost in conjunction with hybrid feature selection suggests that predictive learning analytics may assist educational institutions in identifying students who may be at risk of underperforming academically prior to the release of their final grades. Consequently, the absence of final grades may allow academic support units to implement time-sensitive supportive measures, including, but not limited to, tutoring, counselling, attendance and engagement monitoring, and the provision of individualized learning support.

The second effect may be the ability of predictive models to incorporate hybrid feature selection in educational contexts. The models incorporating the selected features outperformed the models incorporating all the features in all the evaluations, suggesting that incorporating a small number of relevant features facilitates better prediction than incorporating all the features. This is beneficial to educational institutions because the models are less complex and, therefore, easier to manage and monitor. Rather than utilizing a large number of features related to students, educational decision-makers may focus on the feedback-related features that are most important, including attendance and engagement along with study and assessment participation.

Third, the proposed model may be incorporated within a learning management system (LMS) as an early warning system (EWS). For example, declining rates of learner participation, diminishing levels of engagement in LMS activities, and poor performance in academic assessments may trigger LMS notifications to advisors or instructors. Such notifications may assist instructors in reaching out to students, helping them to understand their challenges, and recommending forms of appropriately necessary support. Thus, the application of predictive analytics transitions academic support activities from a reactive to a proactive academic support service.

Fourth, feature importance analysis increases the interpretability of the model. Interpretability is a critical component of the educational analytics model because, in the educational context, the model is typically used by teachers and administrators to understand the classification of students as being at-risk. A model that is primarily a predictive model may be of less value relative to a model that offers a description of the contributing factors of the prediction. Thus, the framework that is proposed in this context enhances the clarity of the decision-making process and the potential for its implementation.

It is critical to use predictive analytics with caution and restraint. For example, students should not be permanently classified as low achievers, failures, or at-risk because of reliance on automated, algorithmic predictions. Predictive results should be considered in the context of other, more appropriate human evaluations. Academic instructors and advisors should use the results of the predictive models in consideration of their professional expertise, in conjunction with a discussion with the highlighted students, and knowledge of the local context. It is also critical to avoid using demographic variables in a discriminatory context, and to support the students rather than harm them.

The value of the proposed framework extends beyond the context of its predictive ability. It is likely to be of value during the design and development phases of the adaptive academic support systems. Predictive analytics for student performance, when used responsibly, can be

a means for higher education institutions to offer proactive, precise, and empirical support to foster student success.

Conclusion

This research developed an integrated feature selection and ensemble machine learning framework focused on predictive student performance analytics. This framework used Mutual Information and Recursive Feature Elimination and then utilized Random Forest, XGBoost, and Gradient Boosting to assess and predict student performance using academic, behavioral, and demographic data.

The results demonstrated that the addition of feature selection improved the performance of all ensemble models. For the models that included feature selection, the best overall performance was achieved by XGBoost with an accuracy of 0.879, precision of 0.872, and recall of 0.866, and an F1 score of 0.869, and an AUC-ROC of 0.918. This showed that the addition of feature selection and the ensemble model added significant performance value to the prediction while decreasing the feature set.

In this context, the analysis of the importance of features was shown to provide significant value. It was determined that attendance, study time, prior achievement, assessment participation, and learning engagement contributed to the prediction of academic performance. This demonstrates that predictive models have the potential to facilitate the design and implementation of proactive academic support by assisting academic institutions in the identification of students in need of support.

The most significant value of this study was the design of a predictive learning analytics model that integrates accuracy, efficiency, and understanding. This model was designed to integrate the outcomes of the classification in the context of the academic support decisions, in order to assist learning management systems, academic advisors, and educational institutions.

Future studies may build upon this work by implementing the framework on larger, more heterogeneous educational datasets, including longitudinal data from learning management systems. Other studies may include explainable artificial intelligence methods to further improve the transparency of the models. Additionally, future studies should explore the fairness, privacy, and ethical concerns to guarantee that predictive analytics will be used in a more responsible manner in the educational field.

Acknowledgment

The researcher did not receive any funding for this study, and the results have not been published in any other sources.

References

Ab Rahman, N. F., Wang, S. L., Ng, T. F.,&Ghoneim, A. S. (2025). Artificial intelligence in education: A systematic review of machine learning for predicting student performance. Journal of Advanced Research in Applied Sciences and Engineering Technology, 5*(1), 198–221. https://doi.org/10.37934/araset.54.1.198221

Alsariera, Y. A., Baashar, Y., Alkawsi, G., Mustafa, A., Alkahtani, A. A.,&Ali, N. A. (2022). Assessment and evaluation of different machine learning algorithms for predicting

student performance. Computational Intelligence and Neuroscience, 202*, Article 4151487. https://doi.org/10.1155/2022/4151487

Cheng, J., Sun, J., Yao, K., Xu, M.,&Cao, Y. (2022). A variable selection method based on mutual information and variance inflation factor. Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy, 26*, Article 120652. https://doi.org/10.1016/j.saa.2021.120652

Dai, W., Lin, J., Jin, F. J.-Y., Tsai, Y.-S., Srivastava, N., Le Bodic, P., Gašević, D.,&Chen, G. (2025). Learning analytics for early identification of at-risk students and feedback intervention. Journal of Learning Analytics, 1*(3), 102–125. https://doi.org/10.18608/jla.2025.8735

Durachman, Y.,&Bin Abdul Rahman, A. W. (2025). Clustering student behavioral patterns: A data mining approach using K-means for analyzing study hours, attendance, and tutoring sessions in educational achievement. Artificial Intelligence in Learning, 1(1), 35–53. https://doi.org/10.63913/ail.v1i1.48

Feng, G., Fan, M.,&Chen, Y. (2022). Analysis and prediction of students' academic performance based on educational data mining. IEEE Access, 10, 19558–19571. https://doi.org/10.1109/ACCESS.2022.3151652

Hossain, M. A.,&Islam, M. S. (2023). A novel hybrid feature selection and ensemble-based machine learning approach for botnet detection. Scientific Reports, 13, Article 21207. https://doi.org/10.1038/s41598-023-48230-1

Ikram, M., Hanefar, S.B.M., Saleem, S.M.U. and Zulfiqar, F. (2026). Artificial intelligence in education: a systematic review of personalized learning trends and future directions. Frontiers in Education. https://doi.org/10.3389/feduc.2026.1782626

Kunapuli, G. (2023). Ensemble methods for machine learning. Manning.

Lam, P., Pyne, S. R., Cutler, L., von Kluge, S.,&Erdodi, L. A. (2025). Is showing up half the work? The relationship among student attendance, engagement, and test scores. College Teaching, 73(4), 229–241. https://doi.org/10.1080/87567555.2023.2262676

Liu, M.,&Yu, D. (2023). Towards intelligent E-learning systems. Education and Information Technologies, 28*(7), 7845–7876. https://doi.org/10.1007/s10639-022-11479-6

Pudjihartono, N., Fadason, T., Kempa-Liehr, A. W.,&O'Sullivan, J. M. (2022). A review of feature selection methods for machine learning-based disease risk prediction. Frontiers in Bioinformatics, 2, Article 927312. https://doi.org/10.3389/fbinf.2022.927312

Suleiman, I. B., Okunade, O. A., Dada, E. G.,&Ezeanya, U. C. (2024). Key factors influencing students' academic performance. Journal of Electrical Systems and Information Technology, 11(1), Article 41. https://doi.org/10.1186/s43067-024-00166-w

Susnjak, T. (2024). Beyond predictive learning analytics modelling and onto explainable artificial intelligence with prescriptive analytics and ChatGPT. International Journal of Artificial Intelligence in Education, 34(2), 452–482. https://doi.org/10.1007/s40593-023-00336-3

Yağcı, M. (2022). Educational data mining: Prediction of students' academic performance using machine learning algorithms. Smart Learning Environments, *, Article 11. https://doi.org/10.1186/s40561-022-00192-z

Zhou, Z.-H. (2025). Ensemble methods: Foundations and algorithms (2nd ed.). Chapman&Hall/CRC. https://doi.org/10.1201/9781003587774