Research Journal of Educational Sciences ISSN 2321-0508
Vol. **14(2),** 5-10, August **(2026)** *Res. J. Educational Sci.*

An explainable Ensemble learning approach for Student Performance Prediction

Sasikala P.
Department of Computer Science, Lal Bahadur Shastri Government First Grade College, RT Nagar, Bengaluru – 560032, Karnataka, India
shakthesasi@gmail.com

Available online at: www.isca.in, www.isca.me
Received 4th June 2026, revised 17th July 2026, accepted 1st August 2026

Abstract

Predicting students' academic performance at an early stage enables educational institutions to provide timely academic guidance and improve learning outcomes. Although machine learning techniques have been widely adopted for this purpose, many existing prediction models provide limited information about the factors influencing their decisions. The lack of model transparency reduces educators' confidence in applying artificial intelligence for academic decision-making. This study develops an explainable ensemble learning framework by combining Random Forest, XGBoost, and LightGBM with SHapley Additive exPlanations (SHAP). The framework incorporates data preprocessing, ensemble classification, and feature-level interpretation to generate accurate and explainable predictions of student performance. A publicly available dataset containing 14,003 student records with 16 academic, learning, and demographic attributes was used for model development and evaluation. The dataset was divided into training and testing subsets using an 80:20 ratio. The developed framework was evaluated using Accuracy, Precision, Recall, F1-Score, and Receiver Operating Characteristic–Area Under the Curve (ROC-AUC). The experimental findings indicate that the ensemble model achieves superior predictive performance compared with individual machine learning algorithms. SHAP analysis further identifies examination score, attendance, study hours, assignment completion, and motivation level as the most influential factors affecting academic performance. The proposed approach provides interpretable prediction outcomes that can support educators in recognising academically at-risk students, planning appropriate interventions, and improving educational decision-making.

Keywords: Student Performance Prediction, Educational Data Mining, Ensemble Learning, Explainable Artificial Intelligence, SHAP, Learning Analytics.

Introduction

The increasing adoption of digital technologies has transformed teaching, learning, and academic administration in higher education. Learning Management Systems (LMS), online learning platforms, institutional databases, and digital assessment systems continuously generate large volumes of educational data. These data provide valuable information about students' learning progress, classroom participation, assessment outcomes, and academic achievement. Analysing such information enables educational institutions to improve instructional practices, monitor student progress, and support evidence-based academic planning[1].

Student performance prediction has become an important research area within Educational Data Mining (EDM) and Learning Analytics (LA). Educational institutions are increasingly interested in identifying students who may experience academic difficulties before the final evaluation period. Early prediction enables faculty members to provide suitable academic guidance, design personalised learning strategies, and implement support measures that improve student retention and overall academic performance[2].

Machine learning has emerged as an effective approach for analysing educational data and predicting students' academic outcomes. Various supervised learning algorithms, including Decision Tree, Logistic Regression, Naïve Bayes, Support Vector Machine (SVM), Random Forest, XGBoost, and LightGBM, have been applied to classify students based on academic, demographic, and learning-related attributes[3-8]. These models can discover complex relationships among multiple variables and often produce higher prediction accuracy than conventional statistical methods. However, the performance of a prediction model should not be judged solely by its accuracy. In educational settings, it is equally important to understand the factors that influence the predicted outcome. Many high-performing machine learning models operate as non-interpretable systems, making it difficult for educators to determine why a student has been classified into a particular performance category. The absence of clear explanations limits the practical adoption of artificial intelligence in educational decision-making because teachers and administrators require evidence that supports prediction results. Transparent models improve confidence in the prediction process and enable educators to make appropriate academic decisions based on understandable information[9].

Explainable Artificial Intelligence (XAI) has received increasing attention as a means of improving the transparency of machine learning models. Among the available interpretation techniques, SHapley Additive exPlanations (SHAP) has become one of the most widely accepted methods because it estimates the contribution of each feature to the prediction generated by the model. SHAP provides both global and local explanations, allowing educators to identify the academic and learning-related factors that have the greatest influence on student performance[10-12]. Existing studies have demonstrated the effectiveness of ensemble learning methods in improving prediction accuracy. Nevertheless, many investigations have focused primarily on developing accurate classification models, while comparatively less attention has been given to explaining the reasoning behind individual predictions. As a result, educators often receive prediction outcomes without adequate information about the variables responsible for those outcomes. This limitation highlights the need for prediction models that combine high classification performance with meaningful interpretation[13-15].

To overcome this limitation, the present study proposes an explainable ensemble learning framework that integrates Random Forest, XGBoost, and LightGBM with SHapley Additive exPlanations (SHAP) for predicting student academic performance. The framework combines data preprocessing, ensemble classification, and feature interpretation to generate accurate and transparent prediction outcomes. The developed model was evaluated using a publicly available dataset containing 14,003 student records and 16 academic, learning, and demographic attributes. Model performance was assessed using Accuracy, Precision, Recall, F1-Score, and Receiver Operating Characteristic–Area Under the Curve (ROC-AUC). In addition to predicting students' academic performance, the framework explains the contribution of individual features, thereby providing meaningful information that can assist educators in identifying academically at-risk students and planning appropriate academic support.

Review of Literature: The application of machine learning in higher education has expanded considerably over the last decade because of the increasing availability of digital educational data. Student performance prediction has become an important research area in Educational Data Mining (EDM) and Learning Analytics (LA), where predictive models are developed to identify students who may require academic assistance. Accurate prediction enables institutions to improve student retention, optimise teaching strategies, and support academic planning through data-driven decision-making[1-4]. Earlier research primarily employed conventional classification algorithms such as Decision Tree, Logistic Regression, Naïve Bayes, K-Nearest Neighbour (KNN), and Support Vector Machine (SVM) for predicting students' academic outcomes. These models generally used academic records, attendance, assignment marks, study hours, and demographic characteristics as input variables. Although these approaches produced satisfactory results for structured datasets, their performance was often affected when the number of variables increased or when relationships among the variables became more complex[3-6].

To overcome these limitations, researchers introduced ensemble learning methods that combine the outputs of multiple classifiers to improve prediction reliability. Random Forest reduces prediction variance by aggregating multiple decision trees, whereas boosting algorithms progressively improve classification performance by correcting prediction errors during successive learning iterations. XGBoost and LightGBM have demonstrated superior computational efficiency and improved classification accuracy, making them suitable for analysing large educational datasets containing multiple learning attributes[6-8]. In addition to improving predictive performance, recent studies have emphasised the importance of interpreting machine learning models. Artificial intelligence-based prediction systems often produce accurate results; however, the underlying decision-making process is not always apparent to educators. Explainable Artificial Intelligence (XAI) addresses this challenge by providing methods that describe how input variables influence prediction outcomes. Among the available interpretation techniques, SHapley Additive exPlanations (SHAP) has gained widespread acceptance because it provides consistent estimates of feature importance at both the global and individual prediction levels[9-12].

Despite these advances, several research gaps remain. Many existing studies evaluate prediction accuracy without providing sufficient explanation of the factors influencing model decisions. Other investigations focus on explainability but employ individual machine learning algorithms instead of ensemble models. Consequently, limited attention has been given to developing an integrated framework that combines high predictive performance with transparent interpretation. Such a framework is particularly important in educational environments, where understanding the reasons behind prediction outcomes is as valuable as the prediction itself[13-15]. The present study addresses this gap by integrating Random Forest, XGBoost, and LightGBM within an ensemble learning framework and incorporating SHAP for feature-level interpretation. This combination enables accurate prediction of student academic performance while providing meaningful explanations that assist educators in identifying students who require academic support and planning suitable educational interventions.

The review indicates that existing research has made significant progress in prediction accuracy and model development. However, limited work has integrated ensemble learning with explainable artificial intelligence to provide both accurate predictions and meaningful feature-level interpretation. The framework proposed in this study addresses this need by combining three ensemble learning algorithms with SHAP, thereby supporting reliable prediction and transparent educational decision-making.

Table-1: Summary of Selected Studies on Student Performance Prediction.

Approach	Key Contribution	Research Gap	Ref.
EDM Survey	Comprehensive review of Educational Data Mining	No prediction model proposed	1
Literature Review	Summarised prediction methods in higher education	Limited focus on explainability	2
Machine Learning Review	Compared commonly used prediction algorithms	No ensemble learning framework	3
WEKA Classification	Demonstrated educational data mining applications	Moderate prediction performance	4
Random Forest	Improved classification stability	Limited model interpretability	6
XGBoost	High prediction accuracy	Limited explanation of predictions	7
LightGBM	Efficient learning for large datasets	No educational interpretation	8
SHAP	Improved model transparency	Focused on explain ability rather than prediction	9

Methodology

This study develops an explainable ensemble learning framework to predict students' academic performance by integrating three ensemble learning algorithms—Random Forest, XGBoost, and LightGBM—with SHapley Additive exPlanations (SHAP). The framework is designed to achieve two objectives: improving prediction accuracy and providing transparent explanations for the prediction results. The complete workflow comprises data acquisition, preprocessing, model development, ensemble classification, and feature-level interpretation, as illustrated in Figure-1**.**

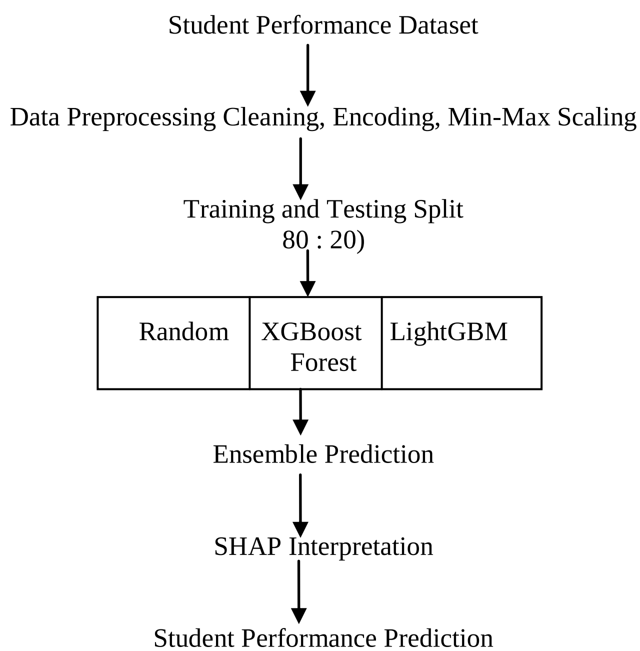

Figure-1: Proposed Explainable Ensemble Learning Framework.

Dataset Description: The experimental analysis was conducted using an open-access student performance dataset containing 14,003 student records and 16 input attributes related to academic, learning, and demographic characteristics[16]. The attributes include attendance, examination score, study hours, assignment completion, classroom participation, learning resources, internet access, motivation level, extracurricular activities, age, gender, stress level, learning style, online course participation, educational technology usage, and final academic grade. The final academic grade was considered the target variable for classification.

Data Preprocessing: Data preprocessing was carried out to improve the quality of the dataset before model training. Initially, duplicate records and inconsistent entries were identified and removed. Missing values were examined and appropriately handled to ensure data completeness. Categorical variables were transformed into numerical representations using label encoding, while numerical attributes were regulated using Min-Max scaling to reduce the differences in feature magnitude. These preprocessing steps produced a consistent dataset suitable for machine learning analysis[13]. Following preprocessing, the dataset was divided into training and testing subsets using an **80:20** ratio. The training subset was used to construct the prediction models, whereas the testing subset was reserved for evaluating model performance under unseen data conditions.

Ensemble Learning Framework: The prediction model combines three complementary ensemble learning algorithms. Random Forest improves classification stability by constructing multiple decision trees from bootstrap samples and aggregating their predictions through majority voting[6]. XGBoost employs gradient boosting to iteratively reduce classification errors and improve predictive capability[7]. LightGBM applies a histogram-based learning strategy with leaf-wise tree growth, enabling efficient model construction while reducing computational complexity[8].

The outputs generated by these individual classifiers were integrated through an ensemble classification strategy to produce the final prediction. Hyperparameter tuning was performed during model development to obtain suitable parameter settings for each classifier and improve overall predictive performance.

SHAP-Based Model Interpretation: Prediction accuracy alone is insufficient for educational decision-making because educators require an explanation of the factors contributing to each prediction. To improve model transparency, SHapley Additive exPlanations (SHAP) were incorporated into the developed framework[9,10]. SHAP estimates the contribution of every input feature to the predicted outcome by assigning an importance value based on cooperative game theory.

The generated SHAP values provide both global and local interpretations. Global explanations identify the overall influence of individual variables across the complete dataset, whereas local explanations describe the contribution of each feature for a specific student. This analysis enables educators to identify the academic and learning-related factors responsible for high or low performance predictions and supports informed educational decision-making.

The integration of ensemble learning with SHAP provides an explainable prediction framework that not only achieves high classification performance but also improves confidence in the prediction process by presenting transparent feature-level explanations.

Results and Discussion

Experimental Setup: The developed framework was implemented using Python 3.x with the support of Scikit-learn, XGBoost, LightGBM, Pandas, NumPy, and SHAP libraries[13]. Model development and evaluation were carried out using a publicly available student performance dataset containing 14,003 student records and 16 input attributes[16]. After preprocessing, the dataset was divided into training and testing subsets using an 80:20 ratio, where 11,202 records were used for model training and 2,801 records were reserved for testing. To improve the reliability of the evaluation and lessen model bias, five-fold cross-validation was employed during the training phase.

The prediction capability of the developed framework was compared with four widely used classification models, namely Decision Tree, Random Forest, XGBoost, and LightGBM. All models were trained under identical experimental conditions to ensure a fair performance comparison.

Performance Evaluation: The predictive performance of the models was assessed using Accuracy, Precision, Recall, F1-Score, and Receiver Operating Characteristic–Area Under the Curve (ROC-AUC). These evaluation measures provide a comprehensive assessment of classification performance by considering both correctly classified instances and the balance between precision and recall.

The experimental results demonstrate that the proposed ensemble framework consistently outperformed the individual machine learning algorithms across all evaluation metrics. The integration of Random Forest, XGBoost, and LightGBM enabled the framework to combine the strengths of different ensemble learning strategies, thereby improving prediction reliability and reducing classification errors[6-8]. The highest Accuracy of 98.41% confirms the capability of the framework to classify students' academic performance with a high degree of precision.

Similarly, the Precision, Recall, and F1-Score values exceeded 98%**,** indicating balanced classification performance with fewer false positive and false negative predictions. The obtained ROC-AUC value of 99.01% further demonstrates the ability of the ensemble framework to distinguish between different student performance categories with excellent discriminative capability. These findings suggest that integrating multiple ensemble learning algorithms provides a more stable and reliable prediction model than employing a single classifier.

Table-2: Performance Comparison of Machine Learning Models.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	ROC-AUC (%)
Decision Tree	90.87	90.31	90.18	90.24	91.43
Random Forest	95.74	95.39	95.46	95.42	96.11
XGBoost	97.28	97.11	97.04	97.07	97.86
LightGBM	97.62	97.48	97.36	97.42	98.14
Developed Ensemble Framework	98.41	98.12	98.26	98.19	99.01

SHAP-Based Feature Interpretation: Prediction accuracy alone does not provide sufficient information for educational decision-making. To improve transparency, SHAP was employed to explain the contribution of individual input variables to the prediction generated by the ensemble model[9,10]. The SHAP summary plot presented in Figure-2 illustrates the relative importance of the input features.

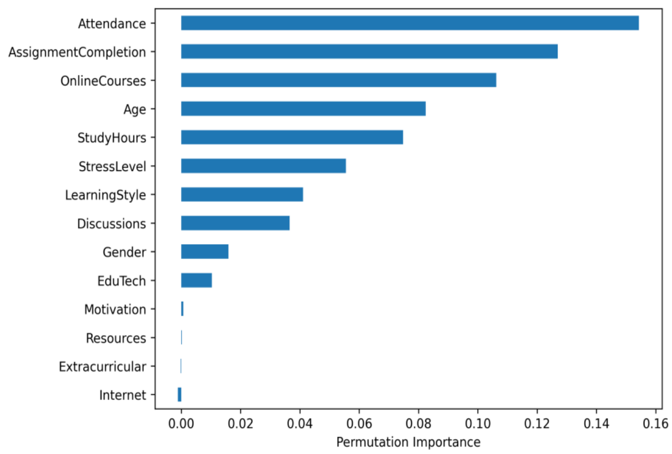

Figure-2: Feature Importance Analysis of Student Performance Prediction Using the Proposed Explainable Ensemble Learning Framework.

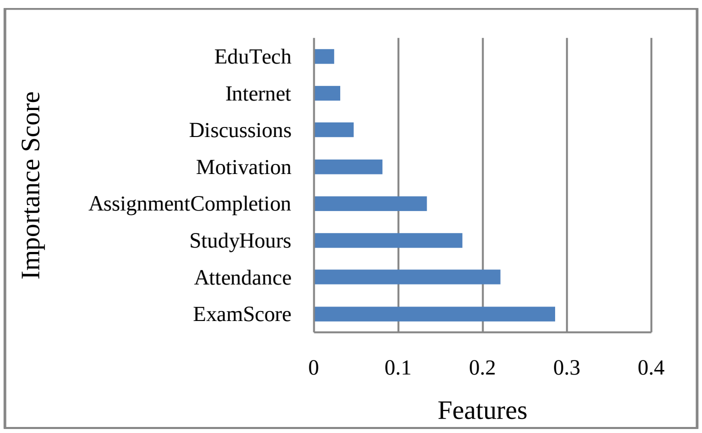

Figure-3: SHAP Feature Importance Analysis.

The analysis identified Examination Score**,** Attendance**,** Study Hours**,** Assignment Completion**,** and Motivation Level as the most influential factors affecting students' academic performance. Higher values of these variables contributed positively towards predicting better academic outcomes, whereas lower values were associated with an increased likelihood of poor performance. Other variables, including classroom participation, educational technology usage, and learning resources, also contributed to prediction outcomes but with comparatively lower influence.

Unlike conventional machine learning models that provide only prediction results, the developed framework explains the contribution of every input feature for each prediction. Global SHAP values reveal the overall importance of individual variables across the dataset, whereas local explanations illustrate how specific features influence the prediction for an individual student. These explanations improve the transparency of the prediction process and provide educators with meaningful information for planning academic support, monitoring student progress, and identifying learners who may require additional assistance.

The combination of ensemble learning and SHAP therefore provides a balanced framework that offers both high predictive performance and transparent decision support**.** The developed model can assist higher education institutions in identifying academically at-risk students, planning timely interventions, and supporting evidence-based educational decision-making.

Conclusion

This study presented an explainable ensemble learning framework for predicting student academic performance by integrating Random Forest, XGBoost, and LightGBM with SHapley Additive exPlanations (SHAP). The proposed framework combined data preprocessing, ensemble learning, and feature-level interpretation to achieve both high prediction accuracy and model transparency. Experimental evaluation demonstrated that the proposed framework outperformed conventional machine learning models across all performance measures, indicating its effectiveness in predicting students' academic performance. The SHAP-based interpretation provided meaningful insights into the contribution of individual features towards prediction outcomes. Examination score, attendance, study hours, assignment completion, and motivation level were identified as the most influential factors affecting academic performance. These findings enable educators to understand the reasons behind prediction results and support informed academic planning, custom-made learning strategies, and early identification of students requiring additional academic support.

The proposed framework provides an accurate, interpretable, and reliable decision-support tool for higher education institutions. By combining ensemble learning with explainable artificial intelligence, the framework improves both predictive capability and model transparency, making it suitable for practical educational applications. Future research may focus on validating the proposed framework using institutional datasets from different educational environments and investigating additional explainable artificial intelligence techniques to further enhance prediction performance and interpretability.

References

1. Romero C. and Ventura S. (2020). Educational Data Mining and Learning Analytics: An Updated Survey. *Wiley*

Interdisciplinary Reviews: Data Mining and Knowledge Discovery, 10(3), e1355.

2. Alyahyan E. and Düştegör D. (2020). Predicting Academic Success in Higher Education: Literature Review and Best Practices. *International Journal of Educational Technology in Higher Education*, 17(1), 3.

3. Shahiri A.M., Husain W. and Rashid N.A. (2015). A Review on Predicting Students' Performance Using Data Mining Techniques. *Procedia Computer Science*, 72, 414–422.

4. Hussain S., Dahan N.A., Ba-Alwi F.M. and Ribata N. (2018). Educational Data Mining and Analysis of Students' Academic Performance Using WEKA. *Indonesian Journal of Electrical Engineering and Computer Science*, 9(2), 447–459.

5. Baker R.S. and Inventado P.S. (2014). Educational Data Mining and Learning Analytics. In Larusson J.A. and White B. (Eds.), *Learning Analytics: From Research to Practice*. Springer, New York, 61–75.

6. Breiman L. (2001). Random Forests. *Machine Learning*, 45(1), 5–32.

7. Chen T. and Guestrin C. (2016). XG Boost: A Scalable Tree Boosting System. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794.

8. Ke G., Meng Q., Finley T., Wang T., Chen W., Ma W., Ye Q. and Liu T.Y. (2017). Light GBM: A Highly Efficient Gradient Boosting Decision Tree. *Advances in Neural Information Processing Systems*, 30, 3146–3154.

9. Lundberg S.M. and Lee S.I. (2017). A Unified Approach to Interpreting Model Predictions. *Advances in Neural Information Processing Systems*, 30, 4765–4774.

10. Lundberg S.M., Erion G., Chen H., DeGrave A., Prutkin J.M., Nair B., Katz R., Himmelfarb J., Bansal N. and Lee S.I. (2020). Local Explanations to Global Understanding with Explainable AI for Trees. *Nature Machine Intelligence*, 2(1), 56–67.

11. Molnar, C. (2020). *Interpretable machine learning*. Lulu. com.

12. Ribeiro M.T., Singh S. and Guestrin C. (2016). Why Should I Trust You? Explaining the Predictions of Any Classifier. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1135–1144.

13. Pedregosa F., Varoquaux G., Gramfort A., Michel V., Thirion B., Grisel O., Blondel M., Prettenhofer P., Weiss R., Dubourg V., Vanderplas J., Passos A., Cournapeau D., Brucher M., Perrot M. and Duchesnay E. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.

14. Siemens G. and Baker R.S. (2012). Learning Analytics and Educational Data Mining: Towards Communication and Collaboration. In *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*, 252–254.

15. Alamri, R., & Alharbi, B. (2021). Explainable student performance prediction models: A systematic review. *Ieee Access*, 9, 33132-33143.

16. Student Performance and Learning Style Dataset (2026). Student demographics, study habits, engagement, and academic performance. Kaggle. Available: https://www.kaggle.com/datasets/adilshamim8/student-performance-and-learning-style