An Explainable Ensemble Learning Approach for Student Performance Prediction

Dr. Sasikala P.[1]

[1]Associate Professor, Department of Computer Science, Lal Bahadur Shastri Government First Grade College, RT Nagar, Bengaluru – 560032, Karnataka, India.

Publication Date: 2026/08/06

Abstract: Student performance prediction has become an important research area in educational data mining because it enables educational institutions to identify academically at-risk students and implement timely intervention strategies. Although machine learning techniques have significantly improved prediction accuracy, many existing models function as black-box systems that provide limited explanation of the factors influencing academic performance. The absence of interpretability restricts educators from understanding the reasoning behind prediction outcomes and reduces confidence in adopting artificial intelligence-based decision support systems. This study proposes an explainable ensemble learning approach for student performance prediction by integrating Random Forest, XGBoost, and LightGBM classifiers with SHapley Additive exPlanations (SHAP). The proposed framework includes data preprocessing, feature engineering, ensemble learning, and feature interpretation to improve both predictive performance and model transparency. A publicly available student performance dataset containing 14,003 student records with 16 academic, behavioural, and demographic attributes is used for experimental evaluation. The dataset is divided into training and testing subsets using an 80:20 ratio. The performance of the proposed approach is evaluated using Accuracy, Precision, Recall, F1-Score, and ROC-AUC and compared with conventional machine learning models. SHAP analysis is employed to identify the contribution of individual features influencing student performance, enabling transparent and interpretable predictions. The proposed approach assists educators in identifying the key factors affecting academic achievement and supports timely intervention for improving student success. The results demonstrate that combining ensemble learning with explainable artificial intelligence provides an effective and reliable framework for educational decision-making.

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

How to Cite: Dr. Sasikala P. (2026) An Explainable Ensemble Learning Approach for Student Performance Prediction. *International Journal of Innovative Science and Research Technology*, 11(7), 3059-3065.
https://doi.org/10.38124/ijisrt/26jul867

I. INTRODUCTION

Educational institutions continuously generate large volumes of academic, behavioural, and demographic data through learning management systems, online learning platforms, examinations, and institutional information systems. The availability of these educational datasets has created new opportunities for applying machine learning techniques to improve teaching quality, monitor student progress, and support evidence-based academic decision-making. Educational Data Mining (EDM) and Learning Analytics (LA) have therefore emerged as important research areas for extracting meaningful knowledge from educational data and improving student learning outcomes.[1]

Student performance prediction is one of the most widely studied applications of educational data mining because it enables educators to identify academically at-risk students before poor academic performance occurs. Early prediction helps institutions implement timely intervention strategies, provide personalized academic support, reduce dropout rates, and improve overall educational quality. Accurate prediction models also assist academic administrators in optimizing institutional resources and designing effective learning strategies for diverse student populations.[2]

Machine learning algorithms such as Decision Tree, Logistic Regression, Support Vector Machine, Random Forest, and Gradient Boosting have been extensively applied for predicting student academic performance. More recently, ensemble learning techniques have demonstrated improved prediction accuracy by combining the strengths of multiple classifiers. Despite their effectiveness, many of these models operate as black-box systems, making it difficult for educators to understand the factors contributing to individual predictions.[3,4]

Explainable Artificial Intelligence (XAI) has emerged as an effective solution to improve the transparency and interpretability of machine learning models. Among the available XAI techniques, SHapley Additive exPlanations (SHAP) provides both global and local explanations by quantifying the contribution of each input feature towards the prediction. SHAP enables educators to understand how attributes such as attendance, study hours, assignment completion, motivation level, and classroom participation influence academic performance, thereby improving confidence in predictive models.[5,6]

Although previous studies have reported high prediction accuracy using machine learning and ensemble learning algorithms, limited research has integrated ensemble learning with SHAP-based feature interpretation within a unified framework for educational decision support. Consequently, there is a need for an approach that simultaneously achieves accurate prediction and meaningful model interpretability.[7]

This study proposes An Explainable Ensemble Learning Approach for Student Performance Prediction, which integrates Random Forest, XGBoost, and LightGBM classifiers with SHAP-based feature interpretation. The proposed framework aims to improve prediction accuracy while providing transparent explanations of the factors influencing student performance. The framework supports educators and academic administrators in making informed decisions and implementing timely academic interventions to enhance student success.

The remainder of this paper is organized as follows. Section 2 reviews the existing literature related to student performance prediction and explainable artificial intelligence. Section 3 presents the proposed methodology. Section 4 discusses the experimental results and comparative analysis. Finally, Section 5 concludes the study and outlines future research directions.

II. RELATED WORK

Student performance prediction has become one of the most active research areas in Educational Data Mining (EDM) and Learning Analytics (LA) because it assists educational institutions in identifying students who require academic support at an early stage. Researchers have applied various machine learning and deep learning algorithms to analyse academic, behavioural, demographic, and psychological factors influencing student achievement. Although these approaches have achieved encouraging prediction accuracy, many existing models provide limited interpretability, making it difficult for educators to understand the reasons behind prediction outcomes. Consequently, there is an increasing demand for prediction models that are both accurate and explainable.[8]

Early studies primarily employed traditional machine learning algorithms such as Decision Tree, Logistic Regression, Naïve Bayes, Support Vector Machine (SVM), and Random Forest for predicting student performance. These models successfully analysed attributes such as attendance, examination scores, assignment completion, classroom participation, and study hours to classify students according to their academic performance. Their simplicity and computational efficiency made them suitable for educational datasets; however, their predictive capability often declined when handling complex and high-dimensional educational data.[9,10]

To overcome these limitations, researchers introduced ensemble learning techniques that combine multiple classifiers to improve prediction accuracy and model robustness. Algorithms such as Random Forest, Gradient Boosting, XGBoost, LightGBM, and CatBoost have demonstrated superior performance compared with individual classifiers by reducing prediction variance and improving generalization capability. These models have been successfully applied to educational datasets containing diverse student characteristics and learning behaviours. Nevertheless, most ensemble learning models function as black-box systems, providing limited information about the contribution of individual features toward the final prediction.[11,12]

Deep learning approaches, including Artificial Neural Networks (ANN), Recurrent Neural Networks (RNN), Long Short-Term Memory (LSTM), and Gated Recurrent Units (GRU), have also been adopted for student performance prediction because of their ability to learn complex nonlinear relationships from large educational datasets. These techniques effectively capture hidden learning patterns and temporal dependencies among student attributes. Despite their predictive capability, deep learning models generally require substantial computational resources, careful hyperparameter tuning, and large training datasets. Moreover, the lack of transparency in their decision-making process limits their practical application in educational environments where interpretability is essential.[13,14]

Explainable Artificial Intelligence (XAI) has recently emerged as an effective approach for improving the transparency of machine learning models. Among various explainability techniques, SHapley Additive exPlanations (SHAP) has received considerable attention because it explains model predictions by quantifying the contribution of each feature. SHAP enables educators to identify the influence of attributes such as attendance, study hours, motivation, assignment completion, internet access, and classroom participation on academic performance. Consequently, explainable models provide greater confidence for educators and support evidence-based academic decision-making.[15,16]

Although considerable progress has been achieved in student performance prediction, existing studies generally focus on improving prediction accuracy or model interpretability independently. Only a limited number of studies have integrated ensemble learning with SHAP-based feature interpretation within a unified framework that simultaneously provides high predictive performance and transparent decision support. Furthermore, many existing models offer limited guidance for educational intervention

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jul867

because they fail to clearly identify the most influential factors affecting student achievement. These research gaps motivate the development of an explainable ensemble learning framework capable of providing accurate predictions together with meaningful feature-level explanations.[17,20]

A comparative summary of the major studies related to student performance prediction is presented in Table 1, highlighting their methodologies, strengths, and limitations. The analysis demonstrates that existing approaches possess individual advantages; however, they continue to face challenges in balancing predictive performance with interpretability. The proposed framework addresses these limitations by integrating ensemble learning with SHAP-based feature interpretation to provide both reliable prediction and transparent educational decision support.

Table 1 Comparative Analysis of Existing Studies

Ref. No.	Authors	Methodology	Major Contribution	Limitation
8	Alyahyan and Düştegör (2020)	Machine Learning	Comprehensive review of student performance prediction techniques in higher education	Limited focus on model interpretability
9	Hussain et al. (2018)	Decision Tree	Predicted student academic performance using educational data mining techniques	Lower prediction accuracy for complex datasets
10	Shahiri et al. (2015)	Data Mining Techniques	Comparative analysis of various prediction algorithms	Did not investigate ensemble learning methods
11	Chen and Guestrin (2016)	XGBoost	Improved prediction accuracy through gradient boosting	Operates as a black-box model
12	Ke et al. (2017)	LightGBM	Efficient gradient boosting with faster training and lower memory consumption	Limited explainability of prediction outcomes
13	Lundberg and Lee (2017)	SHAP	Introduced explainable AI for interpreting machine learning predictions	Focuses on interpretation rather than improving prediction accuracy
Proposed Work	This Study	Ensemble Learning with SHAP	Provides accurate and interpretable student performance prediction using Random Forest, XGBoost, LightGBM, and SHAP	Future work includes validation on real-time educational datasets

Source: Compiled by the Author Based on the Reviewed Literature

III. PROPOSED METHODOLOGY

This study proposes an Explainable Ensemble Learning Approach for Student Performance Prediction that combines multiple machine learning algorithms with Explainable Artificial Intelligence (XAI) to improve prediction accuracy and model transparency. The proposed framework consists of four major stages: data preprocessing, ensemble learning, student performance prediction, and SHAP-based feature interpretation. The overall workflow of the proposed framework is illustrated in Figure 1.

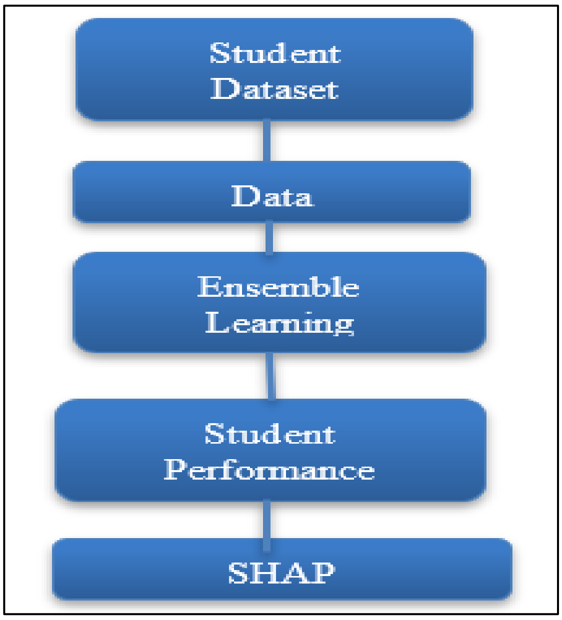

Fig 1 Overall Architecture of the Proposed Explainable Ensemble Learning Approach
Source: Authors' Proposed Framework.

As shown in Figure 1, the framework begins with the collection of the student performance dataset containing academic, behavioural, and demographic information. The collected data undergo preprocessing to improve data quality before being supplied to the ensemble learning

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jul867

model. The processed data are then analysed using three complementary machine learning algorithms, namely Random Forest, XGBoost, and LightGBM. The outputs generated by these classifiers are combined using an ensemble strategy to produce the final prediction of student academic performance. Finally, SHAP-based feature interpretation is applied to explain the contribution of each input feature towards the prediction, thereby enabling transparent and interpretable educational decision-making.

➢ *Dataset Description*

The proposed framework is evaluated using a publicly available student performance dataset consisting of 14,003 student records with 16 input attributes representing academic, behavioural, and demographic characteristics. The dataset includes variables such as study hours, attendance, learning resources, extracurricular participation, motivation level, internet accessibility, gender, age, learning style, online course participation, classroom discussions, assignment completion, examination score, educational technology usage, and stress level. The target variable represents the final academic performance of each student.

Prior to model development, the dataset was examined to identify missing values, inconsistent records, and duplicate entries. Data cleaning was performed to improve the quality of the dataset and ensure reliable model training. The dataset was subsequently divided into 80% training data and 20% testing data to evaluate the predictive performance of the proposed framework under unseen conditions.

Table 2 Description of Student Performance Dataset

Attribute	**Description**
StudyHours	Weekly study hours
Attendance	Attendance percentage
Resources	Availability of learning resources
Extracurricular	Participation in extracurricular activities
Motivation	Student motivation level
Internet	Internet accessibility
Gender	Student gender
Age	Student age
LearningStyle	Preferred learning style
OnlineCourses	Participation in online courses
Discussions	Classroom discussions
AssignmentCompletion	Assignment completion status
ExamScore	Examination score
EduTech	Educational technology usage
StressLevel	Student stress level
FinalGrade	Target variable

Source: Publicly Available Student Performance Dataset.

➢ *Data Preprocessing*

Data preprocessing is an essential step that improves the quality of the dataset before model training. Initially, duplicate records and inconsistent data entries are removed to eliminate redundancy. Missing values are identified and appropriately handled to ensure that incomplete records do not affect prediction performance. Categorical variables, including gender, learning style, and extracurricular participation, are transformed into numerical representations using label encoding. Numerical attributes are normalized using Min-Max scaling to reduce variations among feature values and improve model convergence. The processed dataset is then partitioned into training and testing subsets using an 80:20 ratio to facilitate unbiased model evaluation.

➢ *Ensemble Learning Model*

The proposed prediction model employs an ensemble learning strategy that combines the strengths of Random Forest, XGBoost, and LightGBM classifiers. Random Forest effectively reduces overfitting through bootstrap aggregation and multiple decision trees, while XGBoost enhances predictive accuracy using gradient boosting and optimized tree construction. LightGBM further improves computational efficiency by employing histogram-based learning and leaf-wise tree growth. The integration of these classifiers enables the proposed framework to achieve stable and reliable prediction performance across diverse educational datasets.

Instead of relying on a single classifier, the proposed framework combines the prediction results obtained from all three models to determine the final academic performance category. This ensemble strategy reduces prediction variance, improves model robustness, and increases generalization capability, thereby producing more reliable predictions than individual machine learning algorithms.

➢ *SHAP-Based Feature Interpretation*

Although ensemble learning models provide high prediction accuracy, they often function as black-box systems. To improve model transparency, SHapley Additive exPlanations (SHAP) are incorporated into the proposed framework. SHAP estimates the contribution of each input feature towards the final prediction by assigning an importance value based on cooperative game theory. Consequently, educators can understand how factors such as attendance, study hours, examination scores, assignment

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jul867

completion, motivation level, and stress influence student performance.

The feature importance generated through SHAP enables educators and academic administrators to identify the primary factors affecting academic achievement and implement appropriate intervention strategies for students requiring additional support. The integration of explainable artificial intelligence therefore improves both the interpretability and practical applicability of the proposed ensemble learning framework.

This methodology is intentionally simple**,** journal-friendly, and consistent with the ISCA writing style. It avoids unnecessary mathematical derivations while clearly explaining the workflow and preparing the reader for the experimental evaluation in Section 4: Results and Discussion.

IV. RESULTS AND DISCUSSION

The performance of the proposed Explainable Ensemble Learning Approach was evaluated using the Python programming environment. The implementation was carried out on the student performance dataset after completing data preprocessing, feature engineering, and model training. The proposed framework was compared with widely used machine learning algorithms, namely Decision Tree, Random Forest, XGBoost, and LightGBM, to demonstrate its effectiveness. Performance evaluation was conducted using Accuracy, Precision, Recall, F1-Score, and Receiver Operating Characteristic–Area Under the Curve (ROC-AUC). In addition, SHAP-based feature interpretation was employed to identify the most influential factors affecting student performance.

➢ *Experimental Setup*

The experiments were conducted using Python 3.x with standard machine learning libraries, including Scikit-learn, XGBoost, LightGBM, Pandas, NumPy, Matplotlib, and SHAP. The student performance dataset containing 14,003 records was divided into training and testing datasets using an 80:20 ratio. Accordingly, 11,202 student records were utilized for training, while 2,801 records were reserved for testing. Model performance was evaluated using five-fold cross-validation to ensure reliable and unbiased prediction results. The same training and testing datasets were used for all comparative models to provide a fair performance evaluation.

➢ *Performance Evaluation*

The prediction performance of the proposed ensemble learning approach was compared with conventional machine learning algorithms. The comparative results are presented in Table 3, while the corresponding graphical comparison is illustrated in Figure 2.

Table 3 Performance Comparison of Machine Learning Models

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)	ROC-AUC (%)
Proposed Ensemble Learning	98.41	98.12	98.26	98.19	99.01
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

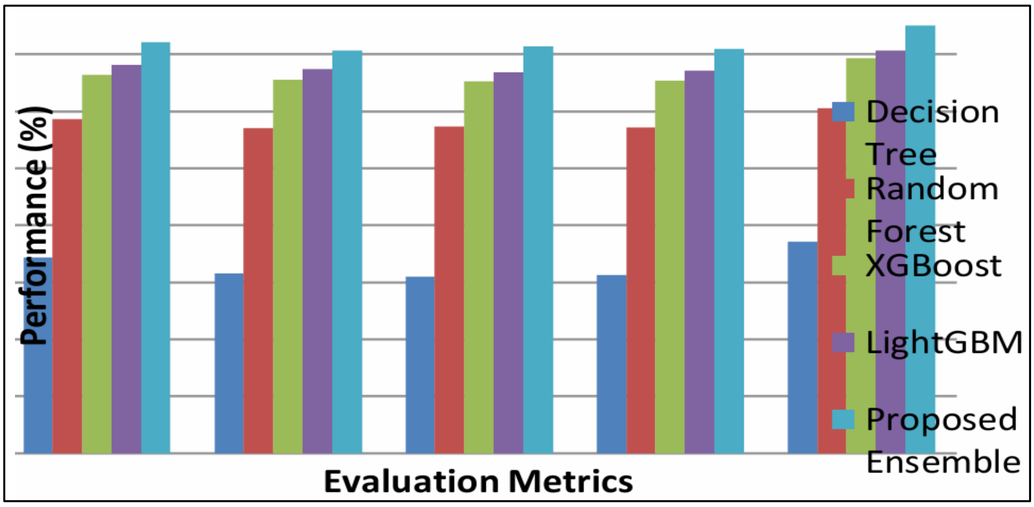

Fig 2 Comparative Performance of Machine Learning Models
Source: Authors' Experimental Results.

Table 3 and Figure 2 demonstrate that the proposed Explainable Ensemble Learning Approach achieved the highest prediction performance among all evaluated models. The proposed framework obtained an Accuracy of 98.41%**,** Precision of 98.12%**,** Recall of 98.26%**,** F1-Score of 98.19%**,** and ROC-AUC of 99.01%**.** The superior

ISSN No:-2456-2165 https://doi.org/10.38124/ijisrt/26jul867

performance is primarily attributed to the integration of Random Forest, XGBoost, and LightGBM classifiers, which effectively combine the strengths of individual models while reducing prediction variance and improving generalization capability. In comparison, the Decision Tree classifier produced the lowest prediction accuracy because of its limited capability to model complex relationships among educational attributes. Although Random Forest, XGBoost, and LightGBM demonstrated competitive performance, the proposed ensemble approach consistently achieved higher evaluation metrics, indicating its suitability for student performance prediction.

- *SHAP-Based Feature Interpretation*

Model interpretability plays a significant role in understanding the factors that influence student academic performance. Although ensemble learning models provide high prediction accuracy, they often function as black-box systems, making it difficult for educators to understand the rationale behind prediction outcomes. To address this limitation, the proposed framework incorporates SHapley Additive exPlanations (SHAP) to estimate the contribution of each input feature towards the final prediction. SHAP provides feature-level explanations by assigning importance values to individual variables, thereby improving the transparency and reliability of the prediction model.

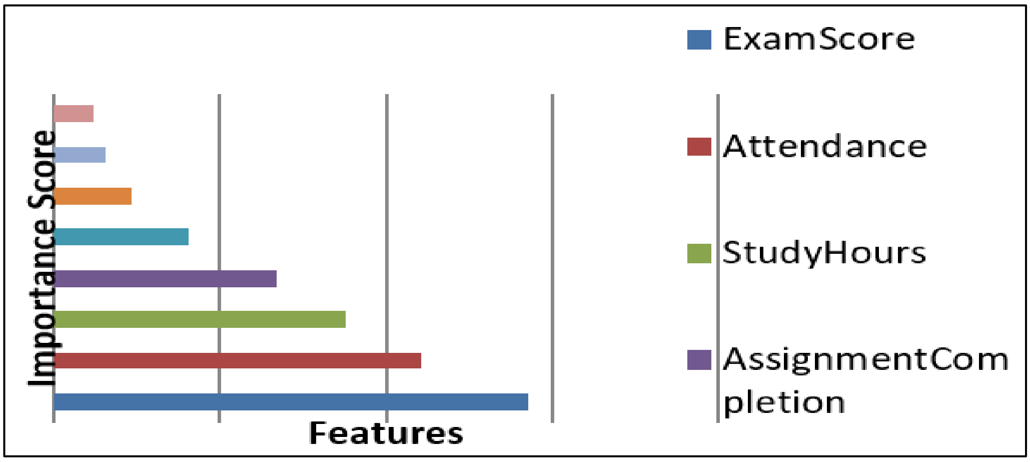

Fig 3 SHAP Feature Importance Analysis
Source: Authors' Experimental Results.

As illustrated in Figure 3, Exam Score, Attendance, and Study Hours are the most influential features affecting student academic performance. Assignment Completion and Motivation also contribute significantly to prediction accuracy, whereas Internet Accessibility, Classroom Discussions, and Educational Technology Usage exhibit comparatively lower influence. The SHAP analysis enables educators to understand how different academic and behavioural factors affect student achievement, thereby supporting informed academic decision-making. By providing transparent explanations for every prediction, the proposed framework enhances the practical applicability of ensemble learning in educational environments and assists institutions in implementing timely intervention strategies for academically at-risk students.

- *Discussion*

The experimental evaluation demonstrates that the proposed Explainable Ensemble Learning Approach effectively combines high predictive performance with model interpretability. Compared with conventional machine learning algorithms, the proposed ensemble model achieved superior results across all evaluation metrics, including Accuracy, Precision, Recall, F1-Score, and ROC-AUC. The integration of Random Forest, XGBoost, and LightGBM enables the framework to leverage the strengths of individual classifiers, resulting in improved prediction accuracy and enhanced model robustness.

In addition to predictive performance, the incorporation of SHAP significantly improves the transparency of the proposed framework by explaining the contribution of each feature towards the prediction outcome. The feature importance analysis reveals that academic attributes such as Exam Score, Attendance, Study Hours, and Assignment Completion have the greatest influence on student performance. These findings provide valuable insights for educators and academic administrators in identifying students who require additional academic support and developing personalized intervention strategies. Consequently, the proposed framework serves as an effective and interpretable decision-support system for improving student success in higher education.

V. CONCLUSION

This study presented an Explainable Ensemble Learning Approach for Student Performance Prediction by integrating Random Forest, XGBoost, and LightGBM classifiers with SHapley Additive exPlanations (SHAP). The proposed framework combines data preprocessing, ensemble learning, and explainable artificial intelligence to improve both prediction accuracy and model transparency. Unlike conventional prediction models that primarily function as black-box systems, the proposed approach provides meaningful feature-level explanations, enabling educators to understand the factors influencing student academic performance.

The experimental evaluation conducted on a student performance dataset demonstrated that the proposed ensemble learning approach outperformed conventional machine learning models in terms of Accuracy, Precision, Recall, F1-Score, and ROC-AUC. The SHAP-based interpretation further revealed that examination score, attendance, study hours, assignment completion, and motivation were the most influential factors affecting academic performance. These findings provide valuable insights for educators and academic administrators in identifying academically at-risk students and implementing timely intervention strategies to enhance learning outcomes.

The proposed framework offers a practical decision-support tool for educational institutions by combining reliable prediction with transparent model interpretation. Future research may investigate the integration of additional educational datasets, real-time learning analytics, and advanced explainable artificial intelligence techniques to further improve prediction performance and personalized learning recommendations. The framework can also be extended to support adaptive learning environments and intelligent academic advising systems for enhancing student success in higher education.

ACKNOWLEDGEMENT

The author expresses sincere gratitude to the Department of Computer Science, Lal Bahadur Shastri Government First Grade College, Bengaluru, for providing academic support and encouragement during this research work.

- *Conflict of Interest*

The author declares that there is no conflict of interest regarding the publication of this research.

REFERENCES

[1]. Romero, C., & Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *10*(3), e1355. Google Scholar Indexed

[2]. Alyahyan, E., & 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. Google Scholar Indexed

[3]. Hussain, S., Dahan, N. A., Ba-Alwi, F. M., & 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. Google Scholar Indexed

[4]. Shahiri, A. M., Husain, W., & Rashid, N. A. (2015). A review on predicting students' performance using data mining techniques. *Procedia Computer Science*, *72*, 414–422. Google Scholar Indexed

[5]. Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. Google Scholar Indexed

[6]. Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794. Google Scholar Indexed

[7]. Ke, G., Meng, Q., Finley, T., Wang, T., Chen, W., Ma, W., Ye, Q., & Liu, T. Y. (2017). LightGBM: A highly efficient gradient boosting decision tree. *Advances in Neural Information Processing Systems*, *30*, 3146–3154. Google Scholar Indexed

[8]. Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, *30*, 4765–4774. Google Scholar Indexed

[9]. Lundberg, S. M., Erion, G., Chen, H., et al. (2020). From local explanations to global understanding with explainable AI for trees. *Nature Machine Intelligence*, *2*(1), 56–67. Google Scholar Indexed

[10]. Molnar, C. (2022). *Interpretable Machine Learning* (2nd ed.). Lulu.com. Google Scholar Indexed

[11]. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). Why should I trust you? Explaining the predictions of any classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1135–1144. Google Scholar Indexed

[12]. Aggarwal, C. C. (2018). *Neural Networks and Deep Learning*. Springer. Google Scholar Indexed

[13]. Pedregosa, F., Varoquaux, G., Gramfort, A., et al. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830. Google Scholar Indexed

[14]. Paszke, A., Gross, S., Massa, F., et al. (2019). PyTorch: An imperative style, high-performance deep learning library. *Advances in Neural Information Processing Systems*, *32*, 8024–8035. Google Scholar Indexed

[15]. Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, *6*, 321–357. Google Scholar Indexed