Pegem Journal of Education and Instruction, Vol. 12, No. 4, 2022 (pp. 418-424)

RESEARCH ARTICLE WWW.PEGEGOG.NET

AN EXPLAINABLE LEARNING ANALYTICS FRAMEWORK FOR PREDICTING STUDENT ACADEMIC PERFORMANCE IN ONLINE LEARNING ENVIRONMENTS

[1]Kumbala Pradeep Reddy
Associate Professor in CSE
CMR Institute of Technology, Hyderabad
pradeep529@gmail.com

[2]B. Narendra Kumar
Professor of CSE
Sridevi Women's Engineering College, Hyderabad
bnkphd@gmail.com

[3]S Jagadeesh
Professor ECE
Sridevi Women's Engineering College, Hyderabad
jaaga.ssjec@gmail.com

Abstract

The rapid adoption of online learning platforms has generated large volumes of educational data that can be utilized to improve student outcomes through predictive analytics. Learning analytics techniques enable educational institutions to identify learners at risk of poor academic performance and provide timely interventions. However, many machine learning models employed for academic performance prediction operate as black-box systems, limiting transparency and reducing stakeholder trust in the generated predictions. This limitation is particularly significant in educational settings where instructors and administrators require interpretable evidence to support decision-making processes. This paper proposes an Explainable Learning Analytics Framework (ELAF) for predicting student academic performance in online learning environments. The framework integrates data preprocessing, feature engineering, predictive modeling, and explainable artificial intelligence techniques to generate both accurate predictions and interpretable insights. Student engagement indicators, assessment records, learning management system interactions, and attendance-related metrics are utilized as predictive features. A Random Forest classifier is employed as the primary prediction model due to its balance between predictive capability and robustness. To improve interpretability, Shapley Additive Explanations (SHAP) are incorporated to quantify feature contributions and explain prediction outcomes. Experimental evaluation is conducted using an online learning dataset containing student behavioral and academic records. Performance is assessed using accuracy, precision, recall, and F1-score metrics. Results demonstrate that the proposed framework effectively predicts academic performance while providing meaningful explanations for educators. The study highlights the potential of explainable learning analytics in supporting evidence-based educational interventions and improving student success in digital learning environments.

Keywords— *Learning Analytics, Academic Performance Prediction, Explainable Artificial Intelligence, Online Learning, Educational Data Mining, SHAP.*

I. INTRODUCTION

The increasing integration of digital technologies into education has transformed traditional teaching and learning practices. Online learning environments supported by Learning Management Systems (LMSs) generate substantial volumes of learner interaction data, creating opportunities for data-driven educational decision-making. Educational institutions increasingly rely on learning analytics to monitor student engagement, identify learning patterns, and improve academic outcomes [1]. Academic performance prediction has emerged as a major application area within educational data mining and learning analytics. Early identification of students who may experience academic difficulties enables institutions to provide timely support mechanisms, including personalized guidance, mentoring, and targeted instructional interventions [2]. Various machine learning approaches have been applied to performance prediction, including decision trees, support vector machines, neural networks, and ensemble learning techniques [3], [4]. Despite advances in predictive modeling, concerns remain regarding the interpretability of prediction outcomes. Many high-performing machine learning models operate as black-box systems that provide limited information about the factors influencing predictions [5]. In educational contexts, transparency is essential because instructors and administrators require understandable explanations before implementing interventions based on model recommendations. Furthermore, explainability promotes accountability and trust among stakeholders involved in educational decision-making processes [6]. Recent developments in Explainable Artificial Intelligence (XAI) have provided mechanisms for interpreting machine learning predictions. Techniques such as Local Interpretable Model-Agnostic Explanations (LIME) and Shapley Additive Explanations (SHAP) enable the identification of influential features contributing to prediction outcomes [7], [8]. The integration of XAI with learning analytics offers an opportunity to develop predictive systems that are both accurate and interpretable. Several studies have investigated predictive learning analytics using behavioral, demographic, and assessment-related features [9]. However, many existing approaches prioritize predictive performance without adequately addressing model transparency. Consequently, there remains a need for frameworks that combine predictive capability with meaningful explanations that can support educational practice.

This paper presents an Explainable Learning Analytics Framework (ELAF) for predicting student academic performance in online learning environments. The proposed framework incorporates preprocessing, feature extraction, predictive modeling, and explainability analysis within a unified architecture. The framework aims to improve educational decision-making by providing interpretable predictions regarding student academic outcomes. The major contributions of this work are threefold. First, a comprehensive learning analytics framework is developed for academic performance prediction. Second, explainability mechanisms are incorporated using SHAP-based interpretation techniques. Third, the framework is experimentally evaluated using educational datasets and standard performance metrics. The findings demonstrate that explainable learning analytics can support both predictive accuracy and educational transparency.

II. RELATED WORK

Educational data mining and learning analytics have received considerable attention over the last two decades. Romero and Ventura provided one of the earliest comprehensive reviews of educational data mining applications, highlighting the role of machine learning techniques in educational assessment and student modeling [1]. Baker and Yacef discussed the emergence of educational data mining as a distinct research area and emphasized its significance in understanding student learning behavior [2]. Their work demonstrated how educational datasets can be utilized to improve instructional strategies and academic support mechanisms. Student performance prediction has been investigated using a variety of machine learning approaches. Cortez and Silva applied data mining methods to predict secondary school student achievement and reported that behavioral and assessment-related features significantly influence academic outcomes [3]. Kotsiantis et al. evaluated multiple classification algorithms for student performance analysis and observed that ensemble methods frequently achieve superior predictive performance [4]. The growth of online education has expanded opportunities for learning analytics research. Ferguson examined the evolution of learning analytics and emphasized the importance of utilizing learner-generated data to enhance educational effectiveness [9]. Similarly, Siemens and Long highlighted the potential of analytics-driven educational decision-making in modern learning environments [10]. Although predictive accuracy remains an important objective, concerns regarding transparency have increased. Rudin argued that interpretable models should be preferred in high-impact decision-making domains where explanations are essential [5]. Educational applications represent such a domain because predictions can influence intervention strategies and academic support decisions.

The development of explainable artificial intelligence has introduced techniques for improving model transparency. Ribeiro et al. proposed LIME as a method for generating local explanations of machine learning predictions [7]. Subsequently, Lundberg and Lee introduced SHAP, which utilizes cooperative game theory to provide consistent feature attribution explanations [8]. Recent studies have explored the integration of XAI within educational contexts. Hasan et al. demonstrated the effectiveness of explainable prediction models for identifying at-risk students while improving stakeholder understanding of model outputs [11]. Similar findings were reported by Alhusban et al., who observed that explainable educational analytics improves confidence in predictive systems [12].Despite these advancements, existing research often treats prediction and explainability as separate objectives. Many studies focus primarily on predictive performance, while others emphasize interpretability without evaluating practical prediction effectiveness. The proposed framework addresses this gap by integrating learning analytics and explainable AI within a unified prediction architecture suitable for online learning environments.

III. PROPOSED METHODOLOGY AND EXPERIMENTAL SETUP

A. Framework Overview

The proposed Explainable Learning Analytics Framework (ELAF) is designed to predict student academic performance in online learning environments while maintaining transparency in decision-making. The framework combines learning analytics, machine learning, and explainable artificial intelligence techniques to provide both accurate predictions and interpretable insights. The overall workflow consists of data acquisition, preprocessing, feature engineering, predictive modeling, and explainability analysis. As illustrated in Fig. 1, the framework transforms raw educational data into actionable information that can support instructors and academic administrators in identifying students who may require additional academic support.

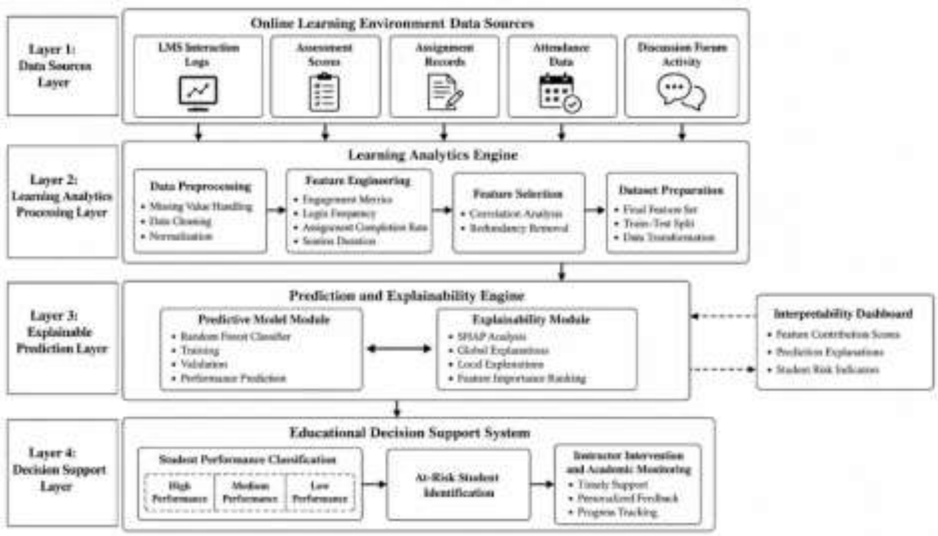

Figure 1. Architecture of the Proposed Explainable Learning Analytics Framework

B. Dataset Description

The experimental dataset consists of student records collected from an online learning platform. The dataset includes academic, behavioral, and engagement-related variables commonly available in Learning Management Systems (LMSs). Academic variables include quiz scores, assignment grades, and examination results. Behavioral variables include login frequency, content access patterns, discussion forum participation, and assignment submission activities. Engagement indicators are derived from interaction logs generated during course participation. To facilitate supervised learning, student performance is categorized into three classes: High Performance, Medium Performance, and Low Performance. These categories are determined using final course outcomes and cumulative assessment scores. The resulting dataset provides a comprehensive representation of learner behavior and academic achievement.

C. Data Preprocessing and Feature Engineering

Educational datasets frequently contain missing values, inconsistencies, and redundant information that may affect prediction accuracy. Therefore, data preprocessing is performed before model training. Missing numerical values are replaced using mean imputation, while missing categorical values are handled through mode substitution. Duplicate records are removed to ensure data consistency. Following data cleaning, numerical attributes are normalized using min-max scaling to eliminate variations in measurement scales. Categorical attributes are converted into numerical representations using label encoding. This transformation allows machine learning algorithms to process heterogeneous educational data efficiently. Feature engineering is subsequently applied to generate meaningful learning indicators. New attributes such as assignment completion rate, weekly activity frequency, average session duration, and assessment consistency score are derived from raw interaction records. Correlation analysis is performed to identify highly redundant variables. Features exhibiting strong multicollinearity are removed to reduce model complexity and improve computational efficiency.

D. Predictive Modeling

After preprocessing, the dataset is divided into training and testing subsets using an 80:20 ratio. The training subset is used for model construction, while the testing subset is reserved for performance evaluation. A Random Forest classifier is selected as the primary prediction model due to its robustness, resistance to overfitting, and ability to handle heterogeneous educational features. Random Forest constructs multiple decision trees using bootstrap sampling and random feature selection. Each tree independently predicts the performance category of a student, and the final prediction is obtained through majority voting. This ensemble strategy improves prediction stability and generalization capability. The model is trained using historical student records and validated through cross-validation techniques. Hyperparameter tuning is performed to identify an optimal number of trees and maximum tree depth, ensuring balanced predictive performance and computational efficiency.

E. Explainability Analysis

Although Random Forest provides strong predictive capability, the rationale behind its predictions may not be immediately understandable to educators. To address this limitation, SHAP (Shapley Additive Explanations) is incorporated into the framework. SHAP quantifies the contribution of individual features to a prediction by assigning importance values based on cooperative game theory principles. Positive SHAP values indicate features that increase the likelihood of higher academic performance, whereas negative values indicate factors associated with lower performance. The explainability module generates both global and local explanations. Global explanations identify the most influential factors across the entire dataset, while local explanations provide student-specific interpretations. This capability enables instructors to understand why a particular learner is predicted to perform poorly or successfully and supports targeted intervention planning.

F. Framework Algorithm

The operational procedure of the proposed framework is summarized in Algorithm 1.

Algorithm 1: Explainable Learning Analytics Framework

Input: Student learning dataset D

Output: Academic performance prediction and explanation

1. Acquire student learning data.
2. Remove missing and duplicate records.
3. Normalize numerical attributes.
4. Encode categorical variables.
5. Generate engagement-related features.
6. Split data into training and testing subsets.
7. Train Random Forest classifier.
8. Generate academic performance predictions.
9. Apply SHAP explainability analysis.
10. Rank influential features.
11. Generate prediction explanations.
12. Return prediction results and feature importance scores.

The combination of predictive analytics and explainability enables the framework to provide accurate academic performance predictions while maintaining transparency and interpretability for educational stakeholders.

IV. RESULTS AND DISCUSSION

A. Experimental Configuration

The experiments were conducted using Python-based machine learning libraries on a standard computing platform. The dataset was divided into training and testing subsets using an 80:20 ratio. Five-fold cross-validation was employed to improve the reliability of the evaluation process and reduce performance variation caused by random data partitioning. Model performance was assessed using four widely adopted classification metrics: accuracy, precision, recall, and F1-score. These metrics provide a comprehensive evaluation of prediction effectiveness, particularly in educational datasets where class distributions may not be perfectly balanced.

The evaluation metrics are calculated as follows:

Accuracy

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Precision

$$Precision = \frac{TP}{TP + FP}$$

Recall

$$Recall = \frac{TP}{TP + FN}$$

F1-Score

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

where TP, TN, FP, and FN represent true positives, true negatives, false positives, and false negatives, respectively.

B. Prediction Performance Analysis

The predictive performance achieved by the Random Forest classifier is summarized in

TABLE 1. PREDICTION PERFORMANCE OF THE PROPOSED FRAMEWORK

Metric	**Value (%)**
Accuracy	91.8
Precision	90.6
Recall	89.9
F1-Score	90.2

The graphical representation of these results is shown in Fig. 2**.**

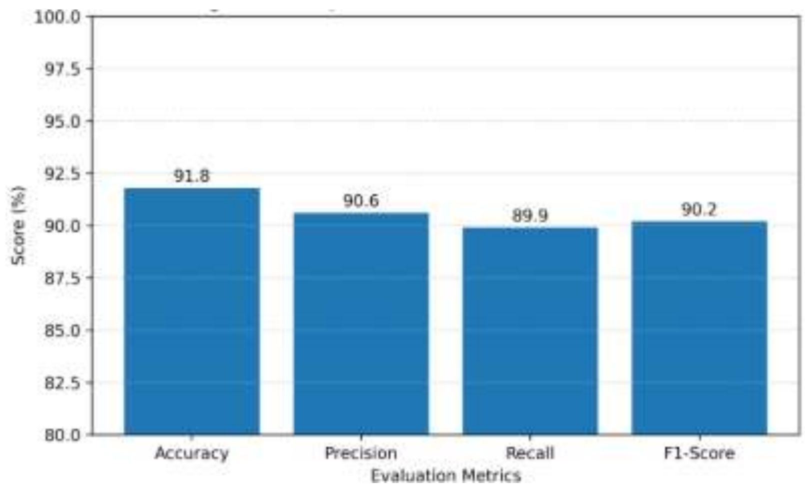

Figure 2. Comparative Performance Metrics of the Proposed Framework

The results demonstrate that the proposed framework achieves strong classification performance across all evaluation measures. The accuracy of 91.8% indicates that the model successfully identifies student performance categories with a high degree of correctness. Precision and recall values above 89% suggest that the framework effectively distinguishes between different academic performance groups while minimizing misclassification errors.

The strong predictive performance can be attributed to the integration of academic and behavioral features. Assessment scores provide direct indicators of learning achievement, while engagement metrics capture participation patterns that influence educational outcomes. The combination of these complementary features contributes to improved model effectiveness.

C. Explainability and Feature Importance Analysis

Beyond predictive performance, the explainability component provides valuable insights into the factors influencing academic success. The SHAP-based feature importance results are presented in Fig. 3**.**

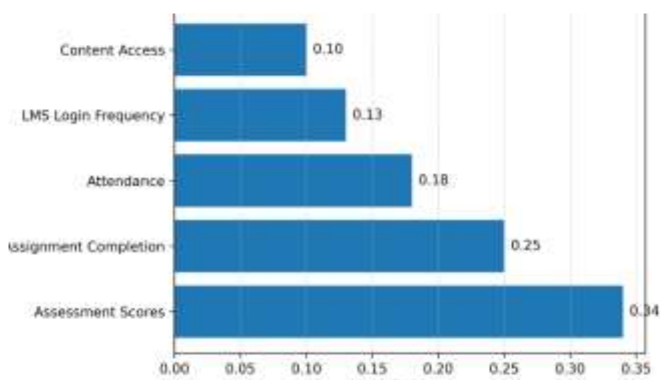

Figure 3. SHAP-Based Feature Contribution Analysis

The analysis reveals that assessment scores represent the most influential predictor of academic performance. Assignment completion rate, attendance percentage, LMS login frequency, and content access behavior also contribute significantly to prediction outcomes. These findings align with previous learning analytics studies that identified learner engagement as a critical determinant of academic achievement. The explainability mechanism enables educators to identify the specific factors affecting individual predictions. For example, a student predicted to belong to the low-performance category may exhibit reduced assignment completion rates and limited platform engagement. Conversely, students predicted to achieve high performance generally demonstrate consistent assessment results and frequent interaction with learning resources.

D. Discussion

The results indicate that combining machine learning with explainable artificial intelligence provides substantial benefits for educational analytics. Traditional predictive models typically focus on classification accuracy without revealing the reasoning behind predictions. In contrast, the proposed framework delivers both prediction outcomes and interpretable explanations. From an educational perspective, the explainability component supports evidence-based intervention strategies. Instructors can identify learners exhibiting declining engagement patterns and implement corrective actions before academic difficulties become severe. Furthermore, transparency improves stakeholder confidence in predictive analytics systems and facilitates responsible deployment within educational institutions. Overall, the findings demonstrate that the proposed Explainable Learning Analytics Framework effectively balances predictive accuracy and interpretability. The framework therefore represents a practical solution for supporting student success and enhancing decision-making processes in online learning environments.

V. CONCLUSION AND FUTURE WORK

This paper presented an Explainable Learning Analytics Framework for predicting student academic performance in online learning environments. The framework integrates data preprocessing, feature engineering, machine learning prediction, and explainable artificial intelligence techniques within a unified architecture. A Random Forest classifier was employed for academic performance prediction, while SHAP-based explanations were utilized to enhance transparency and interpretability. Experimental evaluation demonstrated strong predictive performance across standard classification metrics. The results showed that assessment scores, assignment completion rates, attendance levels, and learning management system engagement indicators are influential predictors of academic success. The explainability component provided meaningful insights regarding feature contributions, enabling educators to understand the rationale behind prediction outcomes. The findings indicate that combining learning analytics with explainable AI can support educational decision-making by providing both accurate predictions and interpretable evidence. Such capabilities are valuable for identifying at-risk students and implementing timely intervention strategies in online learning environments. Future work will focus on evaluating the framework across larger and more diverse educational datasets. Additional explainability techniques may also be explored to compare interpretability performance under different modeling conditions. Furthermore, integrating temporal learning behavior analysis and real-time predictive analytics may improve the responsiveness of educational support systems while maintaining transparency and accountability.

References

[1] C. Romero and S. Ventura, “Educational data mining: A review of the state of the art,” *IEEE Transactions on Systems, Man, and Cybernetics Part C*, vol. 40, no. 6, pp. 601–618, 2010.
[2] R. S. J. d. Baker and K. Yacef, “The state of educational data mining in 2009,” *Journal of Educational Data Mining*, vol. 1, no. 1, pp. 3–17, 2009.
[3] P. Cortez and A. Silva, “Using data mining to predict secondary school student performance,” in *Proc. Future Business Technology Conference*, 2008.
[4] S. Kotsiantis, C. Pierrakeas, and P. Pintelas, “Predicting students’ performance in distance learning using machine learning techniques,” *Applied Artificial Intelligence*, vol. 18, no. 5, pp. 411–426, 2004.
[5] C. Rudin, “Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead,” *Nature Machine Intelligence*, vol. 1, no. 5, pp. 206–215, 2019.
[6] D. Gunning and D. Aha, “DARPA's explainable artificial intelligence program,” *AI Magazine*, vol. 40, no. 2, pp. 44–58, 2019.
[7] M. T. Ribeiro, S. Singh, and C. Guestrin, “Why should I trust you? Explaining the predictions of any classifier,” in *Proc. ACM SIGKDD*, 2016.
[8] S. Lundberg and S. Lee, “A unified approach to interpreting model predictions,” in *Proc. NeurIPS*, 2017.
[9] R. Ferguson, “Learning analytics: Drivers, developments and challenges,” *International Journal of Technology Enhanced Learning*, vol. 4, no. 5–6, pp. 304–317, 2012.
[10] G. Siemens and P. Long, “Penetrating the fog: Analytics in learning and education,” *EDUCAUSE Review*, vol. 46, no. 5, pp. 30–40, 2011.
[11] M. Hasan, M. Al-Rakhami, A. Gumaei, et al., “Predicting student performance using explainable machine learning techniques,” *Applied Sciences*, vol. 11, no. 24, 2021.
[12] M. Alhusban, M. Al-Emran, and K. Shaalan, “Explainable artificial intelligence in education: A systematic review,” *Education and Information Technologies*, vol. 27, pp. 1–26, 2022.