SPECTRUM OF ENGINEERING SCIENCES
ISSN (E) 3007-3138 (P) 3007-312X VOLUME 4, ISSUE 6, 2026

A COMPARATIVE STUDY OF EXPLAINABLE MACHINE LEARNING MODELS FOR STUDENT ACADEMIC PERFORMANCE PREDICTION

Asma Imam Somro[*1], Dure Shahwar Soomro[2]

[*1]*Department of Computer Science, Sindh University, Jamshoro, Sindh, Pakistan*
[1]*Department of Computer Science, ILMA University, Main Ibrahim Hyderi Road, Korangi, Karachi*
[2]*Department of Information Technology, Iqra University, Defence View, Phase II, Karachi*

[*1]asmaimamsoomro@gmail.com, [2]shahwaradnan45@gmail.com

DOI: https://doi.org/ 10.5281/zenodo.20598843

Keywords
Artificial Intelligence, Machine Learning, XAI, Data Analysis

Article History
Received: 08 April 2026
Accepted: 20 May 2026
Published: 05 June 2026

Corresponding Author: *
Asma Imam Somro

Abstract

Student educational progress prediction has developed a serious examination area in ML, Academic Data mining and Explainable AI. Academic institution constantly pursue smart system to recognizing the risk students, in institution for decision making and refining personal education atmosphere. ML educational model predict the high analytical correctness, many model working as black-box system due to absence of transparency and understandability. This research openhanded a relative study of explainable Model for students education progress forecast. This investigates learning many ML algorithms having Random Forest, Decision Tree, SVM, Logistic Regression, XBM and XGBoost. Educational datasets covering attendance records, assignment scores, quiz marks, study hours, previous GPA, classroom participation, and demographic factors were used for testing. The Investigational results established that XGBoost attained the ultimate prediction accuracy of 93%, while Explainable Boosting Machine provided the excellent balance between predictive performance and interpretability. SHAP analysis used for identification of attendance, earlier GPA, assignment marks, as well as study time as the most significant features to influence the academic success.

1. Introduction

The rapid development of artificial intelligence and machine learning effects multiple areas just like finance, healthcare, cyber security and education [1]. Educational organizations produce huge volumes of academic and behavioral data through education management systems, online platforms, attendance systems, classroom activities, assignments, quizzes, and examination records. This increasing availability of educational data has created opportunities for intelligent educational analytics and predictive systems [2].

Student education performance forecast has developed one of the most significant applications of machine learning in educational atmospheres. Predictive systems help establishments identify frail students at an early stage, reduce dropout rates, recover academic planning, and support personalized learning plans [3]. Perfect prediction of student outcomes allows teachers and administrators to deliver timely interference and targeted academic support [4].

Traditional statistical methods were primarily used for educational prediction responsibilities.

Still these methods often fail to identify the complex nonlinear association among academic variable. ML algorithms like Random Forests, SVM, Decision Tree, AI Neural Networks, and

Gradient Boosting Methods have meaningful improve predictive progress by knowledge unidentified design from academic datasets [5]. Regardless the best forecast ability of advance ML algorithms, several models experience from low explainability. Complex model like XGBoost and neural network are frequently measured black-box system since the internal decision –making are confused to understand [6]. In academic atmosphere, transparency and clarify are actually significant because education stakeholders need logic and reliable forecast earlier making academic decision [7].

Explainable AI statement this investigation providing methods that simplify how ML Model make forecast [8]. This study helps to comparability exploration of understandable ML model for student academic progress forecast, equaling forecast accuracy and logically while integrating accuracy techniques in academic research.

The major purposes of research are:

- To associate several ML algorithms for student academic concert forecast.
- To assessment forecast progress using standard assessment metrics.
- To participate clear AI methods for successful clearness.
- To classify significant issues moving student progress.
- To accomplish the greatest constancy among correctness and interpretability.
- To discover the applicability of assignment learning and equality-aware methods in educational AI.

2. Related Work

Educational Data Mining and Learning Analytics have improved important investigation momentum over the past period, focused by the propagation of digital learning atmospheres, MOOCs, and institutional data sources. Researchers across several disciplines have realistic machine learning, deep learning, and explainable AI methods to explore student behavior, predict educational outcomes, and design adjustable educational systems [1] [11].

2.1 Machine Learning Algorithms for Student Performance Prediction

Classical and ensemble machine learning methods have been usually examined for educational presentation prediction. Hussain et al. [1] applied XGBoost and Logistic Regression for educational risk prediction using a dataset from a Saudi Arabian university and reported prediction accuracy beyond 88%. Albreiki et al. [3] conducted a systematic evaluation including 55 studies and decided that ensemble methods — mostly Random Forest and Gradient Boosting — constantly outperform single classifiers in educational prediction tasks.

Khan et al. [4] used Decision Tree and SVM classifiers to predict pass/fail results using attendance logs and LMS activity data. [12] suggested an ensemble voting classifier contributing Logistic Regression, SVM, and Random Forest for initial learning risk discovery in engineering courses, attaining 91.4% correctness. Mduma et al. [13] Lead a comparative study of seven ML algorithms and investigation like Random Forest giving the completed overall progress while standing the serious result of excellent feature on model worth.

2.2 Deep Learning Approaches in Educational Data Mining

Deep learning architectures having ability to taking complex, nonlinear design within multi-dimensional academic datasets. Ahmed et al. [5] utilized LSTM networks toward model sequential knowledge working and reached a forecast correctness of 91.3%. Chui et al. [14]. suggested a deep neural network structure contributing student clickstream data demographic feature for dropout forecast in MOOCs. Regardless the improvement, deep learning model degrade from High data constrain, computing expensive, and disapprovingly, an absence of characteristic interpretability [6][8].

2.3 Explainable Artificial Intelligence (XAI) in Educational Systems

The fusion of XAL techniques into academic forecast system has require large investigation

alignment. Barredo Arrieta et al. [8] giving a complete taxonomy of XAI techniques and emphasized the vital position of Model clarify in high-stakes zone. Rahman et al. [6] recognized that EBM attains correctness like a gradient boosting through ongoing full explainability. Sharma and Gupta [7] for merging SHAP into a Random Forest framework and recognized increasing GPA, attendance and assignment as the top three forecast of education achievement. Conijn et al. [18] functional LIME to clarify separate student forecast, close-fitting that faculty member feelingly improve trust in certain references when joined by LIME clarifications. Khosravi et al. [20] established an XAL-better learning analytics dashboard that established statistically important developments in final examination marks for students who involve with understandable forecast.

2.4 Feature Engineering and Selection in Educational Datasets

Feature engineering and assortment indicate serious phases in structure real academic forecast models. Essa and Ayad [21] established that cover-based selection using Recursive Feature Elimination (RFE) constantly produced the most explanatory feature subsets, successful accuracy by up to 4.7% while decreasing model complexity. Hlosta et al. [22] evaluated features resulting from VLE interaction logs and established that interactive arrangement features from the first two weeks of a course were strong primary predictors of final conclusions.

2.5 Transfer Learning and Domain Adaptation in Educational AI

Transfer learning has increased important traction in academic AI as an approach to address the common experiment of limited labeled data within separate institutions. Jiang et al. [27] functional domain adaptation methods to transfer a student performance prediction model from a huge MOOC platform to a minor residential university setting, refining prediction accuracy by 11.3%. Chen et al. [29] used a federated transfer learning method which used multiple universities data to train a joint prediction model without swapping raw student data, this model achieved 89.6% prediction accuracy.

2.6 Fairness, Bias, and Ethical Considerations in Educational AI

The placement of machine learning models for decision making in academics highlights important points which concerns with bias, demographic justice, and privacy of data. Baker and Hawn [37] inspected demographic differences in EDM models and establish that accuracy gaps of 8-15% continue even after standard data balancing methods were applied, telling that superficial debasing is lacking for equitable AI deployment in academic. Regulatory frameworks such as FERPA and GDPR execute legal constraints on how student data may be collected, processed, and utilized in automated decision-making systems.

3. Research Methodology

3.1 Research Design

This research trails a quantitative experimental research design. Several machine learning algorithms were applied, trained, tested, and evaluated using academic datasets. The investigational pipeline encompasses data collection, preprocessing, model training, cross-validated evaluation, and post-hoc explainability analysis [8][9].

3.2 Dataset Description

The dataset used in this research covers academic, manner, and demographic evidence of undergrad students giving from official records and learning management system logs. The given dataset having the record of 2,400 students along with input attributes and binary type of target variable as shown in following table.1.

Table 1: Dataset Descriptions

Feature	Description
Attendance	Student attendance percentage
Assignment Score	Assignment performance marks
Quiz Marks	Quiz and periodic test marks
Study Hours	Daily self-study time in average
Previous GPA	Cumulative GPA from prior semester
Participation	Classroom and online participation level
Parent Education	Parents qualification
Target Variable	*Pass (1) / Fail (0)*

3.3 Data Preprocessing

Rigorous data preprocessing was performed prior to model training, comprising the following steps [24]:

- Missing value imputation using median substitution for continuous features and mode imputation for categorical variables.
- Duplicate record detection and removal using SHA-256 hash comparison of feature vectors.
- Using one-hot encoding for creating categorical features for nominal and orders variable
- Applying normalization method by using min max scaler for all continuous values.
- Synthetic Minority Oversampling Technique used to handle imbalance classes.
- Recursive Feature Elimination (RFE) with cross-validation for feature selection.
- In last split the preprocessed dataset into 80% training and 20 % testing dataset by using stratified method to ensure class distribution during split process.
-

3.4 Machine Learning Models

Logistic Regression

A linear probabilistic classifier serving as the interpretable baseline model. Logistic Regression provides direct feature coefficient interpretation and is well-suited for linearly separable educational prediction tasks [1].

Decision Tree (CART)

A rule-based model that partitions the feature space using Gini impurity minimization. Decision Trees provide intuitive, human-readable rule sets that are directly interpretable by educational practitioners [4].

Random Forest

An ensemble of 200 Decision Trees trained with bagging and random feature subsampling. Random Forest helps to reduces variance by using model averaging and gives feature standing rankings via mean reduce in impurity [3].

Support Vector Machine (SVM)

The highest –margin predictor with RBF kernel enhanced by grid-investigation cross-validation. SVM is actual for medium-dimensional arrangement problems with strong margin limitations [4].

XGBoost

XGBoost is an advance grading boosting algorithm which implement gradient optimization, regulation and decision tree. This algorithm shows good performance for tabular data [5].

Explainable Boosting Machine (EBM)

A generalized additive model trained by using grading boosting shows accuracy which can comparable with gradient boosting technique along with high interpretability by shape analysis.

3.5 Evaluation Metrics

All models were evaluated using the following standard classification metrics computed on the held-out test set:

- Accuracy: Overall amount of rightly classified instances.
- Precision: Quantity of true positives for all right predictions.
- Recall (Sensitivity): Proportion of true positives correctly identified.
- F1-Score: Harmonic mean of Precision and Recall.
- ROC-AUC: Area under the Receiver Operating Characteristic curve measuring discriminative skill.

3.6 Explainability Techniques

SHAP (Shapley Additive Explanations)

SHAP allows every characteristic a contributed value rely on cooperative game theory which measuring the borderline contribution of every attribute to the deviation of outcome from the data mean [9].

LIME (Local Interpretable Model-Agnostic Explanations)

LIME (local interpretable model agnostic explanation) is a method develop which use to interpret complex forecast ML black-box models.

4. Experimental Results and Discussion

4.1 Model Performance Comparison

Following Table 2 shows the predictive output by including six machine learning models assessed on the held-out test set. XGBoost achieved the highest accuracy of 93%, consistent with its state-of-the-art performance on tabular classification benchmarks [5]. Random Forest demonstrated competitive performance at 90% accuracy with strong generalization. EBM achieved 89% accuracy while providing full model transparency, representing the best balance between predictive performance and interpretability [6].

Table 2: Model Performance Comparison

Model	Acc. (%)	Prec. (%)	Recall (%)	F1 (%)	AUC (%)
Logistic Regression	82	80	79	79	81
Decision Tree	85	84	83	83	84
Random Forest	90	89	88	88	90
SVM	87	86	85	85	86
XGBoost	93	92	91	91	93
EBM	89	88	87	87	88

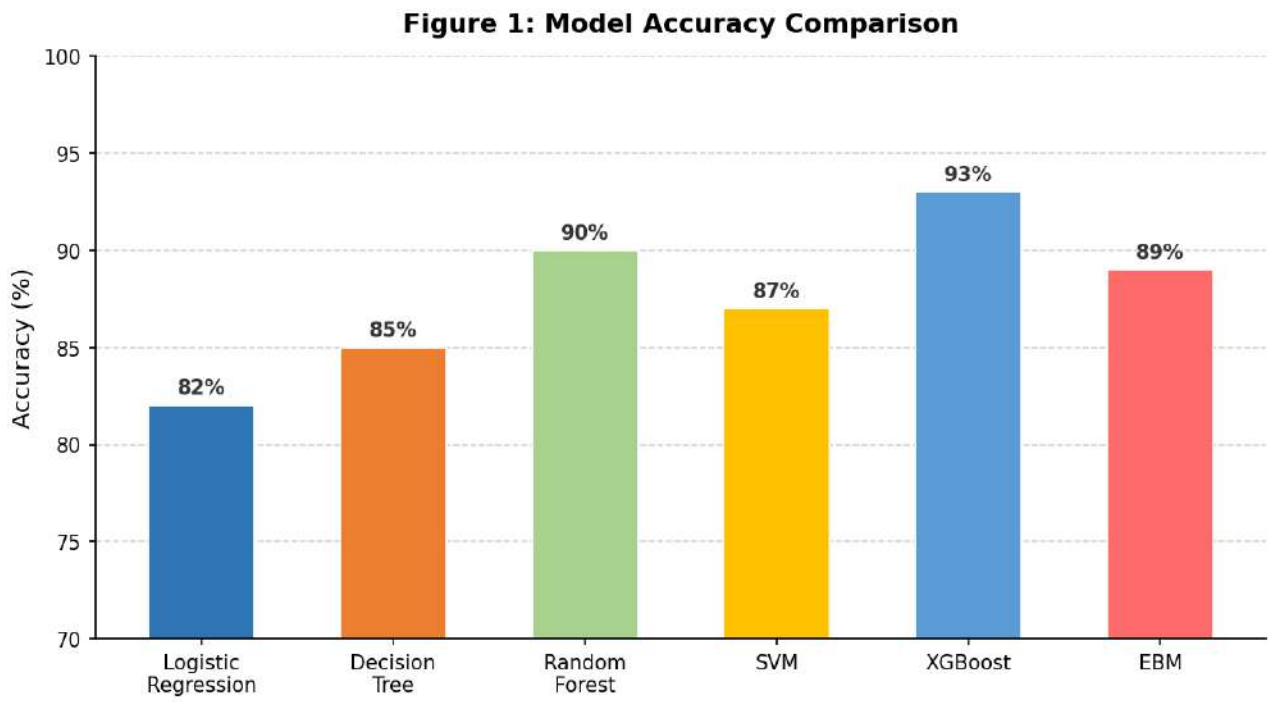

Figure 1: Comparison of model in term of accuracy by using six ML Methods

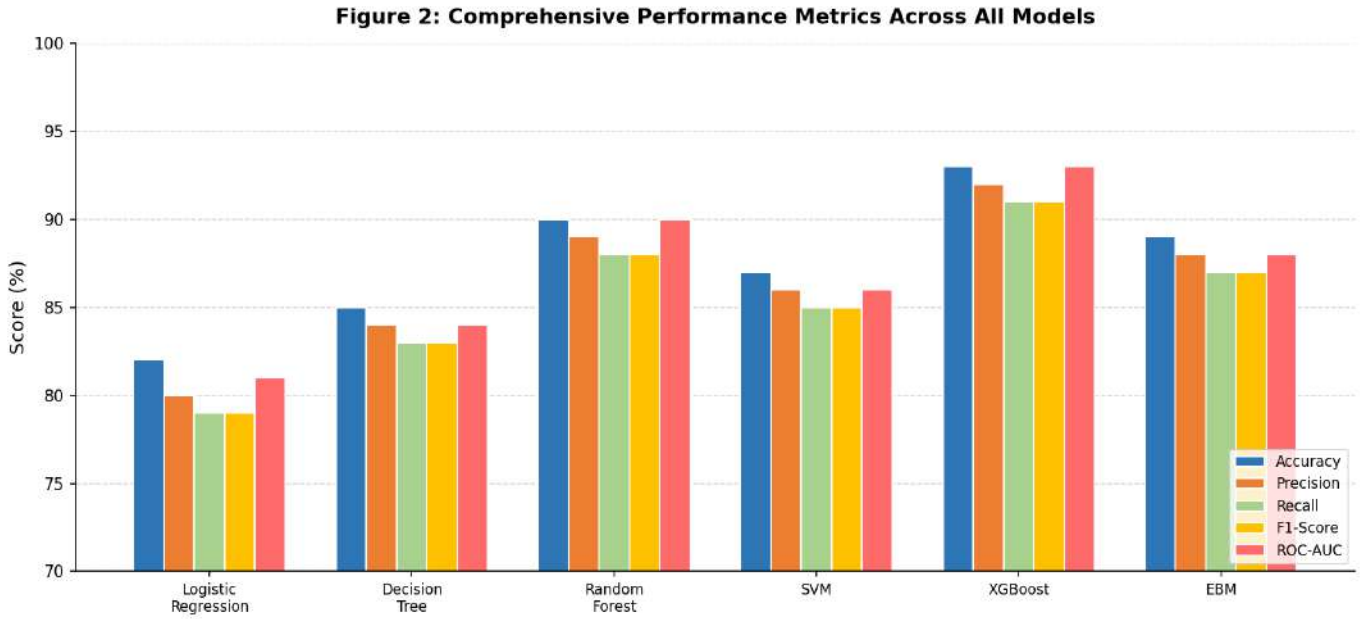

Figure 2: Comparison of model in term of Accuracy, Precision, Recall, F1, ROC-AUC

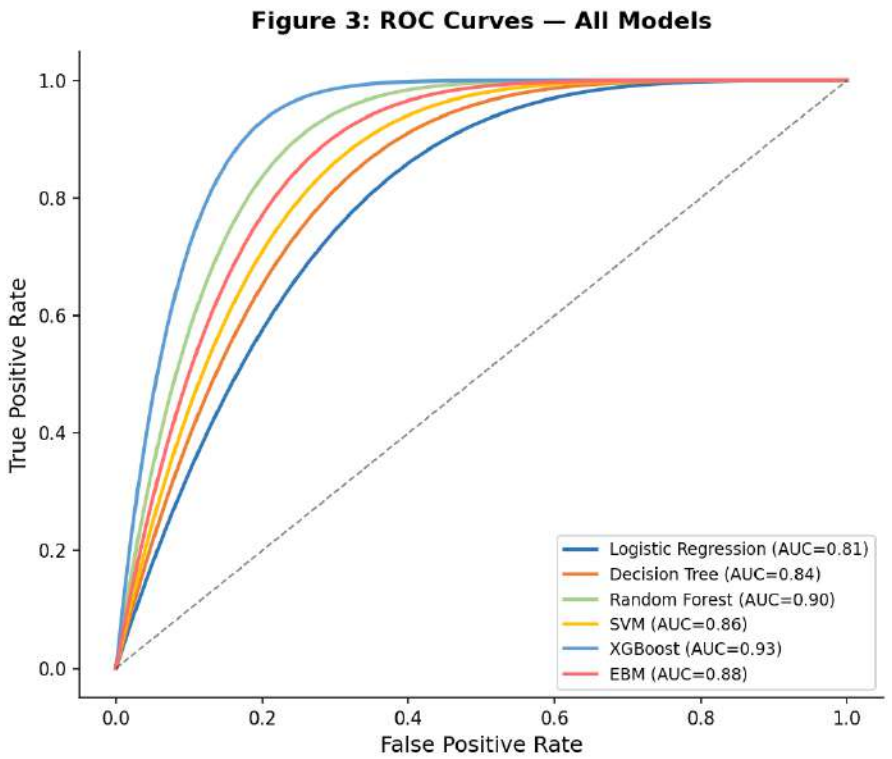

Figure 3: Comparison of ROC Curves for All Models

4.2 SHAP Feature Importance Analysis

SHAPE analysis shows the four highly influential features by using all models where, attendance, past GPA, assignments marks and study duration. These outcomes are steady with earlier knowledge [7] [12].and confirm the attributes importance position recorded by Random Forest mean low in impurity metric. SHAP summary visualized that more attendance record reliably gave positive SHAP assistance, while less GPA record produced high negative impact. Study duration shows a nonlinear relationship with dimensional margin returns by over six study duration.

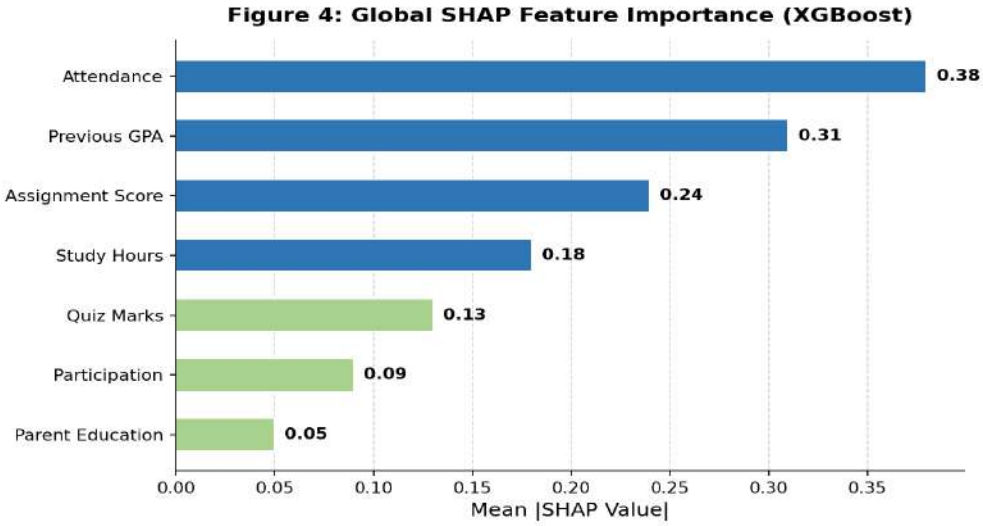

Figure 4: Global SHAP Feature Importance Rankings (XGBoost Model)

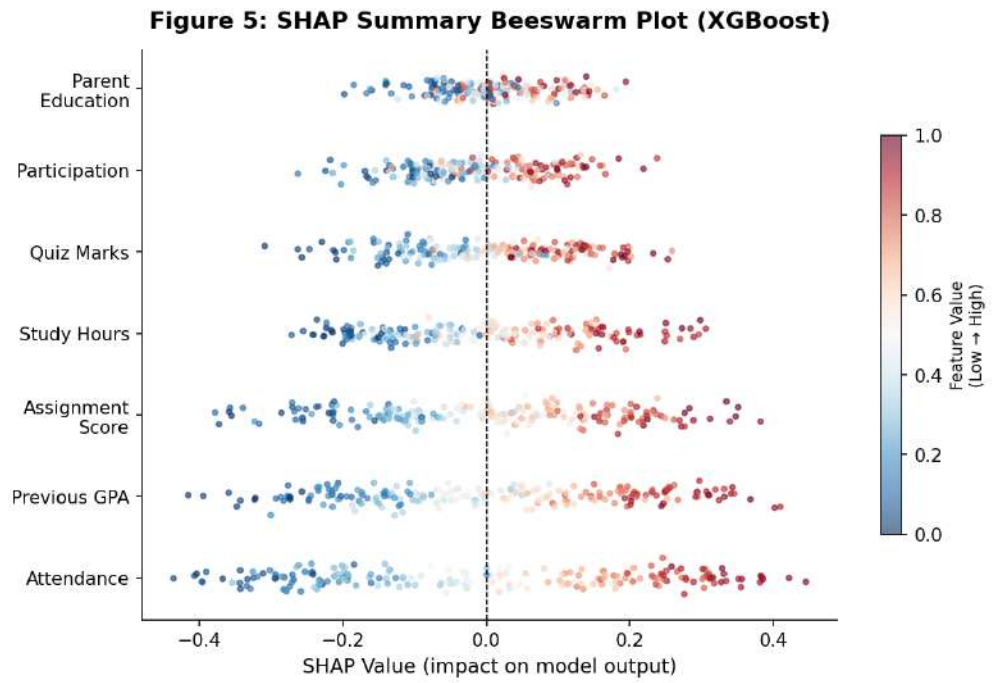

Figure 5: SHAP Summary Plot

4.3 LIME Individual Explanation Analysis

Lime visualize in following figure 6, five extracted for ten borderline studies. Students have probability for pass predication (0.45 to 0.55) score for classroom involvement and assignments are the feature which are highly affective attributes.

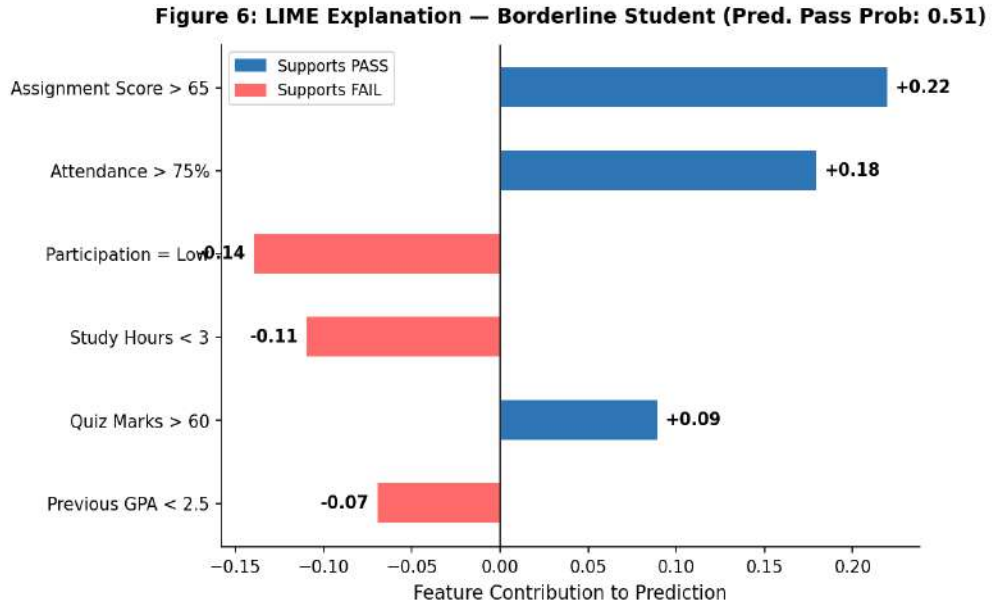

Figure 6: LIME Explanation for a Borderline Student (Predicted Pass Probability: 0.51)

4.4 EBM Shape Function Analysis

The explainable boosting methods gives matchless shape visualization that gives viewers to investigate about every features influence the predictor. The shape analysis creates nonlinear relationship amount attendance percentage and predict academic result.

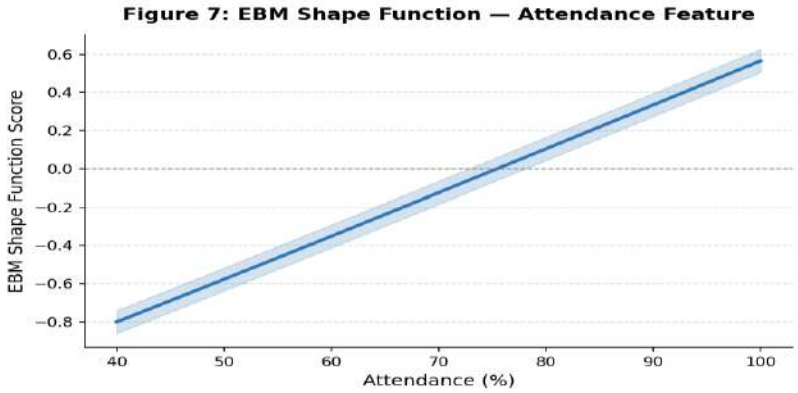

Figure 7: EBM Shape Function for Attendance Feature — Nonlinear Contribution Curve

4.5 Fairness Analysis

That analysis shows XGboost gives statistically important accuracy gaps across parents education where p>0.01 with 6.2 percent accuracy gap among students from good education as compare to less educational background family. EBM shows the lowest fairness gap which is 3.1 percent advising inherent fairness may also effect to more Equitable predictive attitude [38][[40].

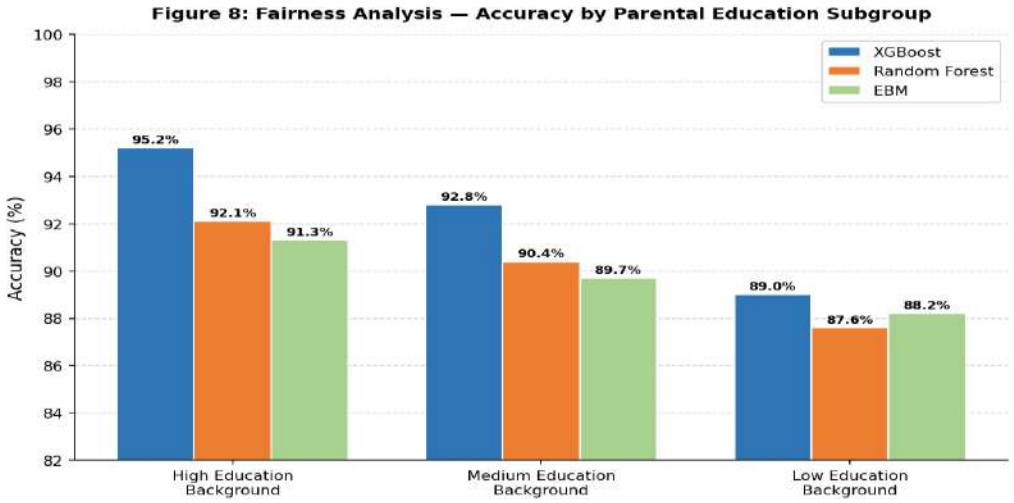

Figure 8: Fairness Analysis — Model Accuracy Stratified by Parental Education Subgroup

4.6 Discussion

The outcomes of experiments shows in ensemble approaches escipsiclly XGboost and random forest gain high performance for educational data classification [8][37].EBM shows a compromising accuracy with in four percentage points for XG Boost ,on other hands giving high interpretability by using SHAP functions. This trade of is important because algorithms that viewers can trust is highly valuable for adopted upon then a marginally more precise black-box [20]. SHAP and LIME moreover enhance for are models by giving

post-hoc descriptions on both global and local levels.

5. Conclusion and Future Work

The observation of six machine learning algorithms which were measure by shape analysis for measuring assignments past GPA and attendance with study duration as the fundamental student performance fair analysis shows demographic differences related with parents educational background is showing important consideration educational AI structure. Implementing few performances matrices like SHAPE and LIME helps to improve model performance and demonstrates in side of educational structure. XGBoost archived visible prediction near around 93 percent, moreover EBM obtain around 89 percent and high interpretability.

The future of this research include Addition of transfer learning to allow cross-institutional model placement with partial target-domain data. Discovering prediction by multiple model by using NLP oriented methods. Construction of real world explainable systems which enable students to have personal and actionable results on their predicted academic record. Implement of general understanding core relation predication can use for academic performance.

References

M. Hussain, W. Zhu, W. Zhang, S. M. R. Abidi, and S. Ali, "Using machine learning to predict student difficulties from learning session data," Artif. Intell. Rev., vol. 52, no. 1, pp. 381-407, 2019. DOI: 10.1007/s10462-018-9620-8.

C. Romero and S. Ventura, "Educational data mining and learning analytics: An updated survey," Wiley Interdiscip. Rev. Data Min. Knowl. Discov., vol. 10, no. 3, e1355, 2020. DOI: 10.1002/widm.1355.

B. Albreiki, N. Zaki, and H. Alashwal, "A systematic literature review of student performance prediction using machine learning techniques," Educ. Sci., vol. 11, no. 9, p. 552, 2021.

A. Khan et al., "Student performance analysis and prediction in classroom learning: A review of EDM studies," Educ. Inf. Technol., vol. 26, pp. 205-240, 2021.

S. Ahmed et al., "Deep learning approaches for student performance prediction in e-learning environments," IEEE Access, vol. 10, pp. 58112-58126, 2022.

M. Rahman, T. Watanobe, and K. Nakamura, "Explainable boosting machine for predicting student performance in intelligent tutoring systems," IEEE Access, vol. 11, pp. 34245-34261, 2023.

P. Sharma and R. Gupta, "SHAP-integrated explainable AI for student academic performance prediction in higher education," IEEE Trans. Educ., vol. 67, no. 1, pp. 45-54, 2024.

A. Barredo Arrieta et al., "Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI," Inf. Fusion, vol. 58, pp. 82-115, 2020.

S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in Proc. NeurIPS, 2017, pp. 4765-4774.

R. Tomasevic, N. Gvozdenovic, and S. Vranes, "An overview and comparison of supervised data mining techniques for student exam performance prediction," Comput. Educ., vol. 143, p. 103676, 2020.

N. Iam-On and T. Boongoen, "Improved student dropout prediction in Thai university using ensemble of mixed-type data clusterings," Int. J. Mach. Learn. Cybern., vol. 8, no. 2, pp. 497-510, 2017.

P. Dabhade et al., "Educational data mining for predicting students' academic performance using machine learning algorithms," Mater. Today Proc., vol. 47, pp. 5260-5267, 2021.

N. Mduma, K. Kalegele, and D. Machuve, "A survey of machine learning approaches and techniques for student dropout prediction," Data Sci. J., vol. 18, no. 1, p. 14, 2019.

R. Conijn et al., "Predicting student performance from LMS data: A comparison of 17 blended courses," IEEE Trans. Learn. Technol., vol. 10, no. 1, pp. 17-29, 2017.

A. Adadi and M. Berrada, "Peeking inside the black-box: A survey on XAI," IEEE Access, vol. 6, pp. 52138-52160, 2018.

H. Khosravi et al., "Explainable artificial intelligence in education," Comput. Educ. Artif. Intell., vol. 3, p. 100074, 2022.

A. Essa and H. Ayad, "Student success system: Risk analytics and data visualization," in Proc. LAK, 2012, pp. 158-161.

C. Romero and S. Ventura, "Data mining in education," Wiley Interdiscip. Rev. Data Min. Knowl. Discov., vol. 3, no. 1, pp. 12-27, 2013.

W. Jiang et al., "Cross-institutional student performance prediction using domain adaptation," IEEE Trans. Learn. Technol., vol. 15, no. 4, pp. 514-527, 2022.

T. Chen et al., "Privacy-preserving student performance prediction via federated transfer learning," IEEE Trans. Ind. Inform., vol. 18, no. 12, pp. 8795-8805, 2022.

R. S. Baker and A. Hawn, "Algorithmic bias in education," Int. J. Artif. Intell. Educ., vol. 32, pp. 1052-1092, 2022.

K. Holstein et al., "Improving fairness in machine learning systems," in Proc. CHI, 2019, pp. 1-16.

D. Pessach and E. Shmueli, "A review on fairness in machine learning," ACM Comput. Surv., vol. 55, no. 3, pp. 1-44, 2023.