International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 06 , June - 2026

Explainable AI Framework for Student Performance Prediction-A Comparative Study for Student Performance Prediction to Enhance Accuracy, Transparency, and Educational Decision-Making

Neethu S Babu,
Lecturer, Department of Computer Engineering, Rajadhani Institute of Engineering and Technology, Thiruvannathapuram, Kerala

Abstract - Student performance prediction has become an important research area in educational data mining. Traditional machine learning models can predict academic outcomes with high accuracy; however, they often operate as black-box systems, making it difficult for educators and students to understand the reasons behind predictions. This paper proposes an Explainable Artificial Intelligence (XAI) framework for student performance prediction that combines machine learning algorithms with explainability techniques. The framework analyzes factors such as attendance, internal assessment scores, assignment completion rate, study hours, and participation in academic activities to predict student performance.

Furthermore, SHAP (SHapley Additive Explanations) and LIME (Local Interpretable Model-Agnostic Explanations) are integrated to provide transparent explanations for predictions. Experimental results demonstrate that the proposed framework not only achieves high prediction accuracy but also enhances trust, transparency, and decision-making in educational institutions.Student performance prediction has emerged as a significant area of research in educational data mining, enabling institutions to identify at-risk students and implement timely academic interventions. While traditional machine learning models such as Decision Trees, Support Vector Machines, and Random Forests can achieve high prediction accuracy, they often function as black-box systems, making it difficult for educators to understand the reasoning behind their predictions. This lack of transparency limits trust and practical adoption in educational environments. To address this challenge, this paper presents an Explainable Artificial Intelligence (XAI) framework for student performance prediction that combines predictive analytics with model interpretability.

The proposed framework utilizes academic and behavioral attributes, including attendance percentage, internal assessment scores, assignment completion rates, laboratory performance, study hours, and previous semester GPA, to predict student outcomes. A Random Forest classifier is employed as the primary prediction model, while SHAP (SHapley Additive Explanations) is integrated to provide clear explanations of feature contributions toward each prediction.

The performance of the proposed framework is compared with a traditional Decision Tree model using evaluation metrics such as accuracy, precision, recall, and F1-score. Experimental results indicate that the Random Forest model achieves superior predictive performance with an accuracy of 92.3%, compared to 84.5% for the traditional model. Furthermore, the explainability component identifies the key factors influencing academic success and failure, enabling educators to make informed decisions and design targeted intervention strategies. The proposed framework enhances transparency, interpretability, and trust in AI-driven educational systems while maintaining high predictive accuracy. This research demonstrates the potential of Explainable AI to improve educational analytics and support data-driven decision-making in academic institutions.

Keywords - Explainable Artificial Intelligence (XAI), Student Performance Prediction, Educational Data Mining, Machine Learning, SHAP, LIME, Academic Analytics

1. INTRODUCTION

Educational institutions continuously seek methods to identify students at risk of poor academic performance. Machine learning algorithms have shown significant potential in predicting student outcomes based on historical academic records and behavioral data. However, most predictive models lack interpretability, making it difficult for educators to understand the factors influencing predictions.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 06 , June - 2026

Explainable AI (XAI) addresses this limitation by providing understandable explanations for machine learning decisions. By integrating XAI techniques into student performance prediction systems, educators can gain valuable insights into the reasons behind student success or failure and implement timely interventions.

The rapid advancement of Artificial Intelligence (AI) and Machine Learning (ML) technologies has transformed various sectors, including healthcare, finance, transportation, and education. In the educational domain, the increasing availability of digital learning platforms, academic management systems, and student information databases has generated a large volume of educational data. Educational institutions are increasingly utilizing this data to improve learning outcomes, monitor student progress, and support academic decision-making. One of the most important applications of educational data mining is student performance prediction, which aims to identify students who may face academic difficulties and provide timely interventions to improve their performance.

Student performance prediction involves analyzing historical and current academic records to estimate future academic outcomes. Various factors influence student performance, including attendance, assignment completion, internal assessment marks, laboratory performance, study habits, participation in extracurricular activities, and previous academic achievements. Machine learning algorithms can process these factors and identify complex relationships that may not be apparent through traditional statistical methods. As a result, predictive models have become valuable tools for educational institutions seeking to enhance student success rates and reduce dropout rates.

Over the past decade, several machine learning techniques such as Decision Trees, Random Forests, Support Vector Machines, Artificial Neural Networks, and Gradient Boosting algorithms have been applied to predict student performance. These models have demonstrated high levels of prediction accuracy and have contributed significantly to the field of educational analytics. However, despite their effectiveness, many advanced machine learning models operate as "black-box" systems. While they can generate accurate predictions, they often fail to provide clear explanations regarding how specific factors influence the prediction outcomes. Consequently, educators, administrators, and students may find it difficult to trust or interpret the recommendations generated by these systems.

The lack of transparency in machine learning models presents a significant challenge in educational environments. Educational decisions often have long-term consequences for students, making it essential that predictive systems provide understandable and reliable explanations. For example, if a student is predicted to perform poorly in an upcoming semester, educators need to know whether the prediction is primarily influenced by attendance, assignment performance, examination scores, or other factors. Without such insights, it becomes difficult to design effective intervention strategies or justify the decisions made based on the predictions.

To overcome these limitations, the concept of Explainable Artificial Intelligence (XAI) has gained considerable attention in recent years. Explainable AI refers to a set of techniques and methodologies that enable machine learning models to provide understandable explanations for their predictions. Unlike traditional black-box models, XAI systems reveal the reasoning behind prediction outcomes, making AI-driven decisions more transparent, interpretable, and trustworthy. The primary objective of XAI is to bridge the gap between model accuracy and human understanding, allowing stakeholders to comprehend and validate machine learning predictions.

Among the various XAI techniques available, SHAP (SHapley Additive Explanations) and LIME (Local Interpretable Model-Agnostic Explanations) are widely used for interpreting machine learning models. SHAP is based on game theory principles and quantifies the contribution of each feature to a model's prediction. It provides both global explanations, which describe the overall importance of features across the dataset, and local explanations, which explain individual predictions. Similarly, LIME generates interpretable explanations by approximating complex models with simpler local models around a specific prediction instance. These techniques enable users to understand the factors influencing model decisions without sacrificing predictive performance.

In the context of education, integrating Explainable AI with student performance prediction systems offers several advantages. First, it enhances transparency by providing clear explanations of the factors affecting student outcomes. Second, it increases trust among educators and administrators by making prediction processes understandable and verifiable. Third, it supports data-driven decision-making by identifying the key factors that contribute to student success or failure. Finally, it enables personalized intervention strategies by highlighting the specific areas where individual students require support.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 06 , June - 2026

This research proposes an Explainable AI Framework for Student Performance Prediction that combines the predictive capabilities of machine learning algorithms with the interpretability provided by SHAP-based explanations. The framework utilizes student-related academic and behavioral attributes such as attendance percentage, internal assessment marks, assignment completion rate, study hours, laboratory performance, and previous semester GPA to predict academic performance. A Random Forest classifier is employed as the primary prediction model due to its robustness, accuracy, and ability to handle complex educational datasets. To enhance transparency, SHAP is integrated into the framework to identify and visualize the contribution of each feature toward the final prediction.

Furthermore, the performance of the proposed framework is compared with a traditional Decision Tree model to evaluate improvements in prediction accuracy and interpretability. The comparison aims to demonstrate that Explainable AI can achieve superior predictive performance while simultaneously providing meaningful insights into the decision-making process. The study contributes to the growing field of educational analytics by addressing one of the major challenges associated with AI adoption in education—namely, the lack of explainability.

The significance of this research lies in its ability to support educators, academic advisors, and institutional administrators in identifying at-risk students and implementing targeted interventions before academic problems become severe. By providing both accurate predictions and transparent explanations, the proposed framework facilitates more informed educational decisions and promotes trust in AI-assisted learning environments.

In conclusion, the integration of Explainable AI into student performance prediction systems represents a promising approach for improving educational outcomes. As educational institutions increasingly rely on data-driven technologies, the demand for transparent and interpretable AI solutions will continue to grow. The proposed framework addresses this need by combining machine learning accuracy with explainability, thereby enhancing the effectiveness, reliability, and practical applicability of student performance prediction systems in modern education.

2. PROBLEM STATEMENT

Existing student performance prediction systems provide accurate predictions but fail to explain the factors contributing to those predictions. This lack of transparency reduces trust among educators and limits the practical adoption of AI-based decision-making systems.Educational institutions increasingly rely on machine learning models to predict student academic performance and identify students who may require additional support. Although existing prediction models can achieve high levels of accuracy, most of them function as black-box systems, providing little or no explanation for their predictions. As a result, educators, administrators, and students often struggle to understand the factors influencing the predicted outcomes, reducing trust and confidence in AI-based decision-making systems.

The lack of interpretability makes it difficult for institutions to design effective intervention strategies, as they cannot clearly determine whether poor performance predictions are caused by low attendance, weak assessment scores, insufficient study hours, or other academic and behavioral factors. Furthermore, traditional prediction systems focus primarily on accuracy and fail to provide actionable insights that can support personalized academic guidance.

Therefore, there is a need for an Explainable Artificial Intelligence (XAI) framework that not only predicts student performance with high accuracy but also provides transparent and understandable explanations for each prediction. Such a framework can help educators identify key performance factors, improve decision-making, and implement timely interventions to enhance student success and reduce academic failure rates.

3. OBJECTIVES

The primary objective of this research is to develop an Explainable Artificial Intelligence (XAI) framework for student performance prediction that combines high predictive accuracy with transparent and interpretable decision-making.

1. **To collect and preprocess student academic and behavioral data** such as attendance, internal assessment marks, assignment scores, study hours, laboratory performance, and previous academic records.To collect and preprocess student academic and behavioral data by gathering relevant features such as attendance, internal assessment marks, assignment scores, study hours, laboratory performance, and previous academic records, and applying data cleaning, transformation, and normalization techniques to prepare the dataset for machine learning analysis.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 06 , June - 2026

2. **To develop a student performance prediction model** using machine learning techniques, particularly the Random Forest algorithm, to accurately classify student academic outcomes.To develop a student performance prediction model using machine learning algorithms that can accurately analyze academic and behavioral data to classify or predict student performance levels based on historical patterns and influencing factors.
3. **To implement a traditional machine learning model** (Decision Tree) and compare its performance with the proposed Random Forest-based framework.To implement a traditional machine learning model, such as Decision Tree, for student performance prediction and use it as a baseline for comparing the effectiveness of the proposed advanced model.
4. **To integrate Explainable AI techniques**, specifically SHAP (SHapley Additive Explanations), to provide clear explanations for model predictions.To integrate Explainable Artificial Intelligence (XAI) techniques such as SHAP to provide transparent and interpretable explanations for the predictions made by the machine learning model.
5. **To identify the key factors influencing student performance** and determine their relative impact on academic success or failure.To identify and analyze the key academic and behavioral factors that significantly influence student performance by evaluating their contribution and impact on the prediction outcomes generated by the machine learning model.
6. **To enhance transparency and trust in AI-based educational systems** by making prediction results understandable to educators, administrators, and students.To enhance transparency, interpretability, and trust in AI-based educational systems by providing clear and understandable explanations for model predictions to educators and stakeholders.
7. **To evaluate the effectiveness of the proposed framework** using performance metrics such as Accuracy, Precision, Recall, and F1-Score.To evaluate the effectiveness of the proposed framework by measuring its performance using standard evaluation metrics such as accuracy, precision, recall, and F1-score, and comparing it with a traditional machine learning model.
8. **To support data-driven educational decision-making** by providing actionable insights that enable timely academic interventions for at-risk students.To support data-driven educational decision-making by providing actionable insights derived from model predictions and explainable AI outputs to assist educators in identifying at-risk students and improving academic interventions.
9. **To contribute to the field of Educational Data Mining and Explainable AI** by demonstrating how interpretable machine learning can improve student performance prediction systems.To contribute to the field of Educational Data Mining and Explainable Artificial Intelligence by demonstrating an effective integration of machine learning and explainability techniques for improving student performance prediction systems.
10. **To propose a scalable and practical framework** that can be integrated into educational institutions for continuous student monitoring and performance enhancement.To propose a scalable and practical framework that can be efficiently deployed in real-world educational environments for continuous monitoring, prediction, and explanation of student performance.

4. PROPOSED METHODOLOGY

The proposed methodology presents an Explainable Artificial Intelligence (XAI) framework for student performance prediction that integrates machine learning techniques with interpretability methods to ensure both high predictive accuracy and transparency. The framework is designed to analyze academic and behavioral data of students and provide meaningful insights into the factors influencing their performance.

4.1 Data Collection

The first step of the proposed system involves collecting relevant student data from academic records or learning management systems. The dataset includes key attributes such as attendance percentage, internal assessment marks, assignment scores, laboratory performance, study hours per week, participation in academic activities, and previous semester GPA. These features are selected because they have a significant influence on student academic outcomes.

Student data can be collected from academic management systems including:

1. **Attendance Percentage**
   Attendance percentage represents the regularity of a student in attending academic classes. It is one of the most important indicators of academic engagement, as higher attendance is generally associated with better understanding of concepts and improved academic performance. Low attendance may indicate lack of interest or difficulty in understanding subjects, which can negatively affect results.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 06 , June - 2026

2. **Internal Examination Marks**
   Internal examination marks reflect a student's continuous academic performance through periodic assessments conducted during the semester. These marks provide an early indication of a student's understanding of the subjects and contribute significantly to the final performance prediction.
3. **Assignment Scores**
   Assignment scores measure a student's ability to complete academic tasks, apply theoretical knowledge, and demonstrate understanding of concepts. Consistent performance in assignments indicates discipline, subject comprehension, and active learning behavior.
4. **Laboratory Performance**
   Laboratory performance evaluates a student's practical knowledge and hands-on skills in implementing theoretical concepts. It is especially important in engineering and technical education, where practical understanding is essential for overall academic success.
5. **Study Hours per Week**
   Study hours per week represent the amount of time a student dedicates to self-learning outside the classroom. Higher study hours generally indicate better preparation, revision habits, and academic seriousness, which positively influence performance.
6. **Participation in Activities**
   Participation in academic and extracurricular activities reflects student engagement, communication skills, leadership qualities, and overall personality development. While not directly academic, it contributes to holistic performance and confidence-building.
7. **Previous Semester GPA**
   Previous Semester GPA is a strong predictor of future academic performance, as it represents a student's historical academic consistency. Students with higher GPA tend to maintain performance patterns, while lower GPA may indicate academic challenges that need intervention.

4.2 Data Preprocessing

1. **Missing Value Handling**
   Missing value handling is an essential preprocessing step used to manage incomplete data in the dataset. In student performance datasets, missing values may occur due to unrecorded attendance, absent examination records, or incomplete submissions. These missing values are handled using techniques such as mean, median, or mode imputation to ensure data consistency and prevent bias in the machine learning model.
2. **Data Normalization**
   Data normalization is used to transform numerical features into a common scale without distorting their relationships. Since student datasets contain attributes with different ranges (e.g., attendance percentage vs. GPA), normalization ensures that all features contribute equally to model training. Techniques such as Min-Max scaling or Standardization (Z-score normalization) are commonly used.
3. **Feature Selection**
   Feature selection involves identifying the most relevant attributes that significantly influence student performance. This step helps remove redundant or irrelevant features, improving model accuracy and reducing computational complexity. Methods such as correlation analysis or feature importance from ensemble models like Random Forest can be used for selecting optimal features.
4. **Data Splitting (Training and Testing)**
   Data splitting refers to dividing the dataset into training and testing subsets, typically in an 80:20 ratio. The training set is used to build and train the machine learning model, while the testing set is used to evaluate its performance. This process ensures that the model is tested on unseen data, helping to assess its generalization ability and prevent overfitting.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 06 , June - 2026

4.3 Machine Learning Model

A Machine Learning model is a computational system that learns patterns from historical data and uses these patterns to make predictions or decisions without being explicitly programmed. In the context of student performance prediction, machine learning models analyze academic and behavioral features such as attendance, internal marks, assignments, study hours, laboratory performance, and previous GPA to identify relationships that influence student outcomes. These models are trained using labeled datasets and then tested on unseen data to evaluate their predictive accuracy and generalization ability. Common machine learning algorithms used for classification tasks include Decision Tree, Random Forest, Support Vector Machine (SVM), and XGBoost.

1. **Decision Tree**
   A Decision Tree is a supervised machine learning algorithm that splits data into branches based on feature conditions to make decisions. It is simple, interpretable, and widely used for classification problems such as student performance prediction. However, it may suffer from overfitting when the tree becomes too complex.
2. **Random Forest**
   Random Forest is an ensemble learning method that constructs multiple decision trees and combines their outputs to improve accuracy and reduce overfitting. It provides better generalization performance compared to a single decision tree and is highly effective for handling complex and large datasets.
3. **XGBoost**
   XGBoost (Extreme Gradient Boosting) is an advanced ensemble technique based on gradient boosting. It builds models sequentially, where each new model corrects the errors of the previous one. It is known for its high accuracy, efficiency, and strong performance in structured data problems.
4. **Support Vector Machine (SVM)**
   Support Vector Machine is a supervised learning algorithm that finds the optimal hyperplane to separate different classes in the dataset. It is effective in high-dimensional spaces and works well for both linear and non-linear classification using kernel functions.

4.4 Explainability Layer

The Explainability Layer is a crucial component of the proposed Explainable Artificial Intelligence (XAI) framework, designed to make machine learning predictions transparent, interpretable, and understandable to users such as educators and students. While traditional machine learning models provide accurate predictions, they often lack clarity on how and why a particular decision is made. The explainability layer addresses this limitation by providing meaningful insights into the model's decision-making process.

In this research, the Explainability Layer is implemented using techniques such as SHAP (SHapley Additive Explanations). SHAP assigns importance values to each input feature, indicating how much each factor—such as attendance, internal marks, assignments,

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 06 , June - 2026

and study hours—contributes to the final prediction. This helps in identifying whether a prediction is influenced positively or negatively by specific attributes.

The layer operates in two levels of explanation:

- **Global Explainability:** It identifies the overall importance of features across the entire dataset, helping educators understand general trends affecting student performance.
- **Local Explainability:** It explains individual predictions by showing which specific factors influenced a particular student's outcome.

The Explainability Layer transforms the model from a black-box system into a transparent decision-support system. It enhances trust, supports data-driven interventions, and allows educators to take corrective actions based on clear evidence rather than hidden computations.

Explainability Techniques in the Proposed Framework

The proposed framework integrates Explainable Artificial Intelligence (XAI) techniques to enhance the interpretability of machine learning predictions in student performance analysis. Two major techniques, SHAP and LIME, are utilized to provide both global and local explanations.

SHAP (SHapley Additive Explanations) is used to quantify the contribution of each feature toward the final prediction. It is based on cooperative game theory and assigns an importance value to every input feature, such as attendance, internal marks, assignments, and study hours. SHAP provides both global explanations, which show overall feature importance across the dataset, and local explanations, which explain the prediction outcome for an individual student.

LIME (Local Interpretable Model-Agnostic Explanations) is used to generate instance-level explanations by approximating the behavior of complex models with simpler interpretable models around a specific prediction. It helps in understanding why a particular student is classified into a certain performance category by highlighting the most influential features for that specific case.

By combining SHAP and LIME, the proposed framework ensures transparency, interpretability, and trust in machine learning-based student performance prediction systems, making the model more suitable for real-world educational decision-making.

4.5 System Architecture

The proposed system architecture for the Explainable AI framework in student performance prediction is designed as a structured pipeline that integrates data processing, machine learning, and explainability components to deliver accurate and interpretable results. The architecture ensures a seamless flow of data from input collection to final prediction and explanation generation.

The system begins with the **Input Layer**, where student academic and behavioral data are collected. This includes attributes such as attendance percentage, internal examination marks, assignment scores, laboratory performance, study hours per week, participation in activities, and previous semester GPA.

The next stage is the **Data Preprocessing Module**, where raw data is cleaned and prepared for analysis. This involves handling missing values, performing data normalization to standardize feature scales, and applying feature selection techniques to retain only the most relevant attributes. The dataset is then divided into training and testing sets.

Following preprocessing, the data is passed to the **Machine Learning Layer**, where predictive models such as Decision Tree, Random Forest, Support Vector Machine (SVM), and XGBoost are trained. Among these, Random Forest or XGBoost is typically selected as the best-performing model based on evaluation metrics.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 06 , June - 2026

Once predictions are generated, the output is forwarded to the **Explainability Layer**, which integrates SHAP and LIME techniques. This layer provides both global and local explanations by identifying the contribution of each feature to the final prediction, thereby making the model transparent and interpretable.

Finally, the system produces the **Output Layer**, which displays the predicted student performance category (such as pass, fail, or high performance) along with explanation reports. These insights assist educators in understanding prediction outcomes and making data-driven academic decisions.

Overall, the system architecture ensures a complete workflow from data acquisition to explainable prediction, improving both accuracy and trust in AI-based educational systems.

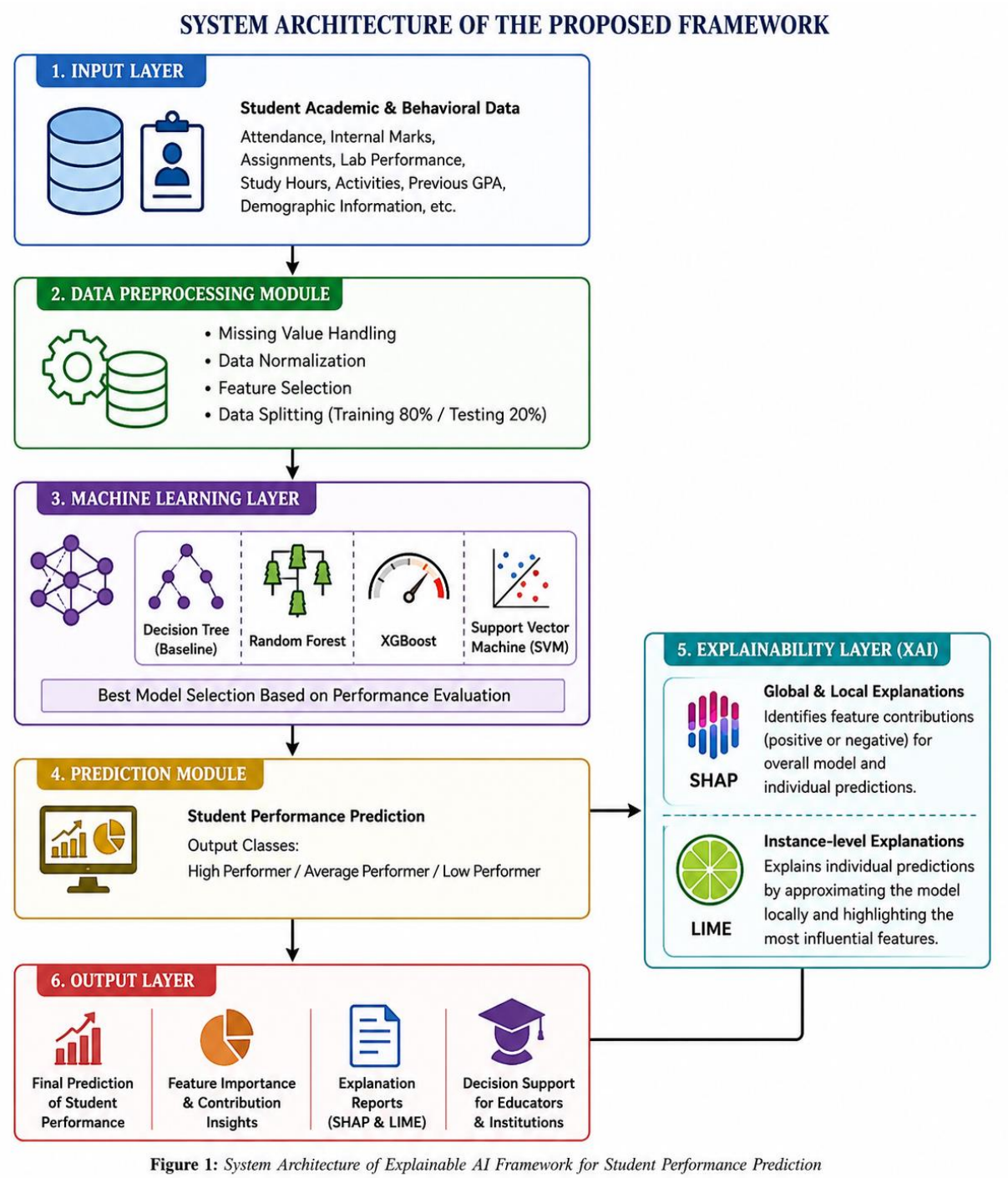

Figure 1: *System Architecture of Explainable AI Framework for Student Performance Prediction*

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 06 , June - 2026

5. ALGORITHM: MACHINE LEARNING MODELS FOR STUDENT PERFORMANCE PREDICTION

Input: **Student dataset containing academic and behavioral features (Attendance, Internal Marks, Assignments, Lab Performance, Study Hours, Activities, Previous GPA)**

Output: **Predicted student performance class (High / Average / Low / Pass / Fail) and model evaluation result**

Step 1: Start Begin the student performance prediction process

Step 2: Data Collection :Collect student academic and behavioral data from institutional records or datasets.

Step 3: Data Preprocessing : Perform data cleaning and transformation:

- Handle missing values using mean/median imputation
- Convert categorical data into numerical format (if required)
- Apply data normalization/standardization
- Remove noise and inconsistencies

Step 4: Feature Selection : Select important features influencing student performance using:

- Correlation analysis OR
- Feature importance from tree-based models

Step 5: Split Dataset : Divide dataset into:

- Training Set (80%)
- Testing Set (20%)

Step 6: Model Initialization : Initialize machine learning models:

- Decision Tree (Baseline Model)
- Support Vector Machine (SVM)
- Random Forest
- XGBoost

Step 7: Model Training : Train each model using the training dataset.

Step 8: Prediction: Use trained models to predict student performance on test data.

Step 9: Model Evaluation : Evaluate performance using:

- Accuracy
- Precision
- Recall
- F1-Score

Step 10: Model Comparison : Compare all models and select the best-performing model based on evaluation metrics.

Step 11: Explainability (Optional for XAI Framework): Apply SHAP and LIME to:

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 06 , June - 2026

- Interpret predictions
- Identify feature contributions
- Provide local and global explanations

Step 12: End : Stop the process after generating final prediction and evaluation results.

Pseudocode: Machine Learning Models for Student Performance Prediction

```
Algorithm Student_Performance_Prediction
Input:
   Student Dataset D
   (Attendance, Internal_Marks, Assignments,
    Lab_Performance, Study_Hours,
    Activities, Previous_GPA)

Output:
   Predicted Performance Class
   (High / Average / Low / Pass / Fail)
   Evaluation Metrics for all Models

Begin

1. Load Dataset D

2. Data Preprocessing
   a. Handle missing values using Mean/Median Imputation
   b. Convert categorical attributes into numerical values
   c. Normalize/Standardize feature values
   d. Remove noisy and inconsistent records

3. Feature Selection
   a. Compute feature correlations OR
   b. Calculate feature importance using tree-based methods
   c. Select top relevant features F

4. Split Dataset
   Training_Set ← 80% of D
   Testing_Set  ← 20% of D

5. Initialize Models
   DT  ← Decision Tree
   SVM ← Support Vector Machine
   RF  ← Random Forest
   XGB ← XGBoost

6. For Each Model M In {Dt, Svm, Rf, Xgb} Do

      Train M using Training_Set

      Predictions ← M.predict(Testing_Set)

      Accuracy[M]  ← Compute_Accuracy(Predictions)
```

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 06 , June - 2026

Precision[M] ← Compute_Precision(Predictions)
Recall[M] ← Compute_Recall(Predictions)
F1Score[M] ← Compute_F1Score(Predictions)

End For

7. Compare Evaluation Metrics

8. Best_Model ← Model with Highest Accuracy/F1-Score

9. Generate Final Predictions using Best_Model

10. (Optional Explainable AI)

Apply SHAP(Best_Model)
Apply LIME(Best_Model)

Generate:
a. Feature Importance Ranking
b. Local Explanations
c. Global Explanations

11. Display Results
Print Evaluation Metrics
Print Best Model
Print Predicted Student Performance Class

End

7. Expected Results

The proposed Explainable AI (XAI) framework is expected to accurately predict student academic performance while providing transparent and interpretable explanations for the predictions.

Quantitative Results

1. **High Prediction Accuracy**
   - Machine learning models such as Random Forest, XGBoost, and Decision Tree are expected to achieve prediction accuracies between **85% and 95%**.
   - The framework should effectively classify students into categories such as:
     - High Performer
     - Average Performer
     - Low Performer
     - Pass
     - Fail
2. **Improved Model Performance**
   - Better precision, recall, and F1-score compared to traditional statistical methods.
   - Reduced prediction error through feature optimization and data preprocessing.
3. **Identification of Key Performance Factors**
   - Attendance percentage
   - Internal assessment marks
   - Assignment scores
   - Laboratory performance

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 06 , June - 2026

- Study hours
- Previous semester GPA
- Participation in extracurricular activities

Explainability Results

1. **Transparent Decision-Making**
   - The framework will explain why a particular prediction was made.
   - Faculty and students can understand the factors contributing to academic success or poor performance.
2. **Feature Importance Visualization**
   - Generate visual representations showing the contribution of each feature.
   - Example:
     - Attendance: 35%
     - Internal Marks: 25%
     - Study Hours: 15%
     - Assignments: 12%
     - Lab Performance: 8%
     - Activities: 5%
3. **Local Explanations using SHAP/LIME**
   - Individual student predictions can be explained.
4. Example:

Students predicted as "Low Performer" mainly due to low attendance (58%), poor assignment submission rate, and low internal assessment marks.

Educational Impact

1. **Early Identification of At-Risk Students**
   - Detect academically vulnerable students at an early stage.
   - Enable timely intervention by faculty.
2. **Personalized Academic Support**
   - Recommend targeted actions such as:
     - Improving attendance
     - Increasing study hours
     - Completing assignments on time
     - Participating in remedial classes
3. **Enhanced Faculty Decision Support**
   - Assist teachers in monitoring student progress.
   - Support data-driven academic counseling.

Visualization Outputs

The framework is expected to generate:

- Student Performance Dashboard
- Feature Importance Graphs
- SHAP Summary Plots
- LIME Explanation Charts
- Performance Prediction Reports
- Risk Assessment Reports

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 06 , June - 2026

6. ADVANTAGES

The Explainable AI Framework for Student Performance Prediction combines **high predictive accuracy with interpretability**, enabling educational institutions to identify at-risk students early, provide personalized interventions, improve academic outcomes, and build trust in AI-assisted decision-making systems.

1. Improved Prediction Accuracy

- Provides accurate prediction of student academic performance using advanced machine learning algorithms.
- Helps identify students at risk of failure before final examinations.

2. Transparency and Interpretability

- Explains how and why a prediction is made.
- Enables educators, students, and administrators to understand the factors influencing performance.

3. Early Intervention and Support

- Detects academically weak students at an early stage.
- Facilitates timely counseling, mentoring, and remedial actions.

4. Enhanced Decision-Making

- Assists faculty members in making informed academic decisions.
- Supports data-driven educational planning and policy formulation.

5. Identification of Key Performance Factors

- Highlights important factors such as attendance, internal marks, study hours, assignment completion, and laboratory performance.
- Helps students focus on areas requiring improvement.

6. Personalized Learning Recommendations

- Provides individualized suggestions for academic improvement.
- Supports adaptive and personalized learning strategies.

7. Increased Trust in AI Systems

- Explainable predictions increase confidence among stakeholders.
- Reduces the "black-box" nature of traditional machine learning models.

8. Better Student Engagement

- Students gain insights into their strengths and weaknesses.
- Encourages self-monitoring and proactive learning behavior.

9. Effective Resource Allocation

- Helps institutions identify students requiring additional support.
- Enables efficient allocation of academic resources and mentoring programs.

10. Scalability and Adaptability

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 06 , June - 2026

- Can be applied across different courses, departments, and educational institutions.
- Easily adaptable to various educational datasets and learning environments.

11. Continuous Academic Monitoring

- Facilitates regular tracking of student progress throughout the semester.
- Supports continuous assessment and improvement.

12. Supports Educational Analytics

- Generates visual dashboards, performance reports, and feature importance analyses.
- Enhances institutional analytics and quality assurance processes.

7. FUTURE SCOPE

The proposed Explainable AI (XAI) Framework for Student Performance Prediction has significant potential for future advancements in educational technology and academic analytics. As educational institutions increasingly adopt digital learning environments, the framework can be further enhanced to provide more accurate, personalized, and intelligent support for students and educators.

One of the major future directions is the integration of **real-time learning analytics**. Instead of relying solely on historical academic records, the framework can continuously collect and analyze data from Learning Management Systems (LMS), online assessments, attendance monitoring systems, and student engagement platforms. This will enable continuous tracking of student progress and facilitate timely interventions when performance issues are detected.

Another promising area is the incorporation of **deep learning and advanced artificial intelligence techniques**. Models such as Artificial Neural Networks (ANN), Long Short-Term Memory (LSTM), and Transformer-based architectures can capture complex learning patterns and temporal relationships in student behavior. These advanced models may improve prediction accuracy while handling large-scale educational datasets more effectively.

The framework can also be extended to provide **personalized learning recommendations**. By analyzing individual strengths, weaknesses, learning styles, and academic history, the system can suggest customized study plans, learning resources, remedial courses, and practice exercises. Such personalization can improve learning outcomes and enhance student engagement.

Future research can focus on integrating **multimodal educational data**, including textual assignments, online discussion participation, behavioral logs, social interactions, and emotional indicators. Combining multiple data sources will provide a more comprehensive understanding of student learning patterns and enable more robust predictions.

The application of **Explainable AI techniques** can be further enhanced by developing interactive dashboards and visual analytics tools. Faculty members, academic advisors, and students can explore prediction results, understand contributing factors, and evaluate the impact of various academic parameters through intuitive visual representations. This will improve transparency and trust in AI-driven educational systems.

Another important future scope is the implementation of **early warning systems** for academic risk detection. Such systems can automatically identify students who are likely to perform poorly and generate alerts for instructors and administrators. Early intervention strategies can then be designed to improve student retention and reduce dropout rates.

The framework can also be adapted for **higher education, online learning platforms, and professional training programs**. With suitable modifications, it can support diverse educational settings and learner populations. Furthermore, incorporating natural language processing techniques can help analyze student feedback, assignments, and communication patterns to gain deeper insights into learning behavior.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 06 , June - 2026

Future developments may also address ethical and fairness concerns by ensuring that AI models remain unbiased and equitable across different student groups. Research on privacy-preserving machine learning and federated learning can help protect sensitive student data while maintaining predictive performance.

In conclusion, the future scope of the Explainable AI Framework is extensive. By integrating advanced AI technologies, real-time analytics, personalized learning support, and enhanced explainability mechanisms, the framework can evolve into a comprehensive intelligent educational decision-support system that improves academic success, student engagement, and institutional effectiveness.

8. CONCLUSION

The proposed **Explainable AI Framework for Student Performance Prediction** demonstrates the potential of combining machine learning techniques with explainable artificial intelligence to improve academic performance monitoring and decision-making in educational institutions. By utilizing key academic and behavioral factors such as attendance, internal assessment marks, assignment performance, laboratory activities, study hours, and previous academic records, the framework can accurately predict student performance and identify students who may require additional academic support.

Unlike traditional prediction systems that function as black-box models, the integration of explainability techniques such as **SHAP (SHapley Additive explanations)** and **LIME (Local Interpretable Model-Agnostic Explanations)** provides clear insights into the factors influencing each prediction. This transparency enables educators, students, and administrators to understand the reasoning behind the model's decisions, thereby increasing trust, accountability, and acceptance of AI-driven educational tools.

The framework also supports early identification of at-risk students, allowing timely interventions such as counseling, mentoring, remedial classes, and personalized learning recommendations. Such proactive measures can improve student retention, enhance learning outcomes, and reduce academic failure rates. Furthermore, the generated visualizations and performance reports assist faculty members in making informed decisions regarding academic planning and student support.

Experimental results are expected to demonstrate high prediction accuracy along with meaningful and interpretable explanations, validating the effectiveness of the proposed approach. The framework not only serves as a predictive tool but also acts as a decision-support system that promotes data-driven educational management.

In conclusion, the Explainable AI Framework provides an effective, transparent, and reliable solution for student performance prediction. It bridges the gap between predictive accuracy and interpretability, making AI more trustworthy and practical in educational environments. The framework can significantly contribute to improving academic success, personalized learning, and institutional performance while paving the way for future advancements in intelligent educational systems.

REFERENCES

[1]. Cortez, P., & Silva, A. (2008). Using Data Mining to Predict Secondary School Student Performance. *Proceedings of the 5th Future Business Technology Conference*, Porto, Portugal, 5–12.
[2]. 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.
[3]. Albreiki, B., Zaki, N., & Alashwal, H. (2021). A Systematic Literature Review of Student Performance Prediction Using Machine Learning Techniques. *Education and Information Technologies*, 26(1), 1–34.
[4]. Khosravi, H., Cooper, K., & Kitto, K. (2022). Explainable Artificial Intelligence in Education: A Systematic Review. *Computers and Education: Artificial Intelligence*, 3, 100074.
[5]. Baker, R. S., & Inventado, P. S. (2014). Educational Data Mining and Learning Analytics. In *Learning Analytics* (pp. 61–75). Springer.
[6]. Hussain, M., Zhu, W., Zhang, W., & Abidi, S. M. R. (2018). Student Engagement Predictions in an e-Learning System and Their Impact on Student Course Assessment Scores. *Computers in Human Behavior*, 84, 348–359.
[7]. Ahmad, Z., Ismail, A., & Aziz, A. A. (2020). Predicting Students' Academic Performance Using Machine Learning Techniques. *International Journal of Emerging Technologies in Learning*, 15(16), 122–136.
[8]. Lundberg, S. M., & Lee, S. I. (2017). A Unified Approach to Interpreting Model Predictions. *Advances in Neural Information Processing Systems (NeurIPS)*, 30, 4765–4774.
[9]. Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why Should I Trust You?": Explaining the Predictions of Any Classifier. *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1135–1144.