Explainable Ai For Predicting Student Academic Performance

V.Sathiyapriya*,B.Nisha**, B.Yamunadevi**, K.Sujitha**,

*Assistant Professor Computer science and Engineering, A.V.C College of Engineering,
**UG Students - Computer science and Engineering, A.V.C College of Engineering,Mannampandal-609 305,

Abstract

Words Predicting student academic performance is an important task in the education system, as it helps identify students who may need additional support. In recent years, Artificial Intelligence (AI) and Machine Learning (ML) techniques have been widely used to analyze student data such as attendance, marks, assignment scores, and study habits. These models can provide accurate predictions about student performance, such as pass, fail, or grade level. However, most traditional AI models act as "black boxes," meaning they do not clearly explain how the predictions are made.This lack of transparency makes it difficult for teachers and institutions to trust and effectively use these systems. To overcome this issue, Explainable Artificial Intelligence (XAI) is introduced. XAI provides clear and understandable explanations for the predictions made by AI models. It helps identify the key factors that influence student performance, such as low attendance or poor assignment scores.In this study, AI models are used to predict student performance, and XAI techniques such as Feature Importance, LIME, and SHAP are applied to explain the results. This approach improves both accuracy and transparency. By combining prediction with explanation, the system supports better decision-making and early intervention.Overall, Explainable AI plays a crucial role in creating a more reliable, transparent, and student-centered educational system.

Keywords: Explainable AI, Student Performance Prediction, Machine Learning, Feature Importance, Academic Analytics, Decision Support Systems.

1. Introduction

In today's modern education system, predicting student academic performance has become an essential task for improving learning outcomes and ensuring student success. Educational institutions generate and store large volumes of data related to students, including attendance records, internal examination marks, assignment scores, study habits, and classroom participation. Analyzing this data manually is not only time-consuming but also prone to errors and inconsistencies. Therefore, there is a growing need for intelligent systems that can efficiently analyze such data and provide accurate predictions.

Artificial Intelligence (AI) and Machine Learning (ML) techniques play a significant role in addressing this challenge. These technologies can process large datasets quickly, identify hidden patterns, and generate meaningful insights. AI-based models are widely used to predict student performance, such as determining whether a student is likely to pass, fail, or achieve a specific grade. Such predictions help educators identify students who may be at risk of poor performance at an early stage. Early identification enables teachers to provide personalized guidance, additional support, and targeted interventions, ultimately improving student outcomes and overall academic success.

Despite their advantages, traditional AI models have a major limitation. Many of these models, especially complex ones like Neural Networks, operate as “black boxes.” This means they provide predictions without explaining how the results are derived. As a result, educators may find it difficult to trust these systems, as they cannot clearly understand the reasons behind a particular prediction. For instance, if a student is predicted to perform poorly, the model may not specify whether it is due to low attendance, insufficient study time, or poor assignment performance. To overcome this limitation, Explainable Artificial Intelligence (XAI) has been introduced. XAI focuses on making AI models more transparent and interpretable by providing clear explanations for their predictions. It helps in identifying the key factors influencing student performance, allowing educators to better understand the reasoning behind the results. By combining accurate prediction with meaningful explanations, XAI enhances trust, improves decision-making, and supports the development of a more effective, reliable and student-centered education system.

2. Literature Review

In recent years, the application of Artificial Intelligence (AI) and Machine Learning (ML) in the field of education has grown significantly, particularly in predicting student academic performance. Researchers have explored various data-driven approaches that utilize student-related information such as attendance records, previous academic results, assignment scores, participation, and study habits. These factors help in identifying patterns and trends that can be used to predict student outcomes accurately. Early prediction of performance enables educators to take timely actions to support students and improve learning outcomes.

Several machine learning algorithms have been widely used for this purpose, including Decision Trees, Support Vector Machines (SVM), Naive Bayes, K-Nearest Neighbors (KNN), and Artificial Neural Networks (ANN). Among these, Decision Trees are highly preferred due to their simplicity and interpretability, as they generate clear rules and conditions that are easy for educators to understand. For instance, they can illustrate relationships such as how low attendance or poor assignment performance can lead to lower academic results. On the other hand, advanced models like Neural Networks and ensemble methods tend to provide higher prediction accuracy but often lack transparency.

One of the major challenges identified in existing studies is the lack of interpretability in complex models. These models are commonly referred to as “black box” systems because their internal decision-making processes are not easily understandable. As a result, educators may find it difficult to trust or rely on these predictions, which limit their practical application in real-world educational environments. Additionally, the absence of clear explanations makes it challenging to identify the key factors influencing student performance.

To overcome these limitations, researchers have introduced Explainable Artificial Intelligence (XAI), which focuses on making AI models more transparent and interpretable. XAI techniques provide insights into how predictions are generated, allowing users to understand the contribution of different features. Popular methods such as LIME (Local Interpretable Model-Agnostic Explanations) and SHAP (Shapley Additive explanations) are widely used to highlight important factors affecting predictions.

Recent studies indicate that integrating XAI with traditional AI models not only improves prediction accuracy but also enhances user trust and decision-making. By providing clear and understandable explanations, XAI enables educators to identify at-risk students early and take appropriate interventions. Thus, Explainable AI plays a crucial role in developing reliable, transparent, and student-centered educational systems.

3. Methodology

3.1. Data Collection

Student data is collected from educational institutions to analyze academic performance. The dataset includes important features such as attendance, internal marks, assignment scores, study hours, and class participation. These features are selected because they directly influence student performance. The main input features used in this study are listed in below.

Student Performance Features:

- Attendance percentage
- Internal exam marks
- Assignment scores
- Study hours
- Class participation

3.2. Data Preprocessing

After collecting the data, preprocessing is performed to prepare it for analysis. This step includes handling missing values, removing incorrect data, and converting data into a suitable format. Data normalization is also applied to improve model performance. Proper preprocessing ensures better accuracy and reliability of the model.

3.3. Model Development

A machine learning model such as Decision Tree or Random Forest is used to build the prediction system. These models are chosen because they provide good accuracy and are relatively easy to interpret. The dataset is divided into training and testing sets. The model is trained using the training data and evaluated using the testing data.

3.4. Performance Prediction

Once the model is trained, it is used to predict student academic performance. The system can classify students into categories such as pass, fail, or grade levels. This helps teachers identify students who may face academic difficulties at an early stage.

3.5. Explainable AI Techniques

To make the predictions understandable, Explainable AI (XAI) techniques are applied. Methods such as Feature Importance, LIME, and SHAP are used to explain the results. These techniques identify the key factors that influence the prediction, such as low attendance or poor assignment scores.

3.6. System Workflow

The overall workflow of the system is shown in Figure 1. It includes data input, model processing, prediction, and explanation. This methodology helps in combining accurate prediction with clear explanation, making the system more transparent and useful for teachers and students.

4. Results and Discussion

4.1. Model Performance

The proposed Artificial Intelligence (AI) model was evaluated using a dataset containing various student-related attributes such as attendance, internal marks, assignment scores, and study habits. The main objective of the model was to predict student academic performance, including whether a student is likely to pass, fail, or achieve a specific grade. The model demonstrated strong performance with good accuracy and consistency across different test cases.

Machine learning algorithms such as Decision Tree and Random Forest were applied to analyze the data and identify patterns. Among these, the Random Forest algorithm provided better performance due to its ability to handle complex data and reduce overfitting. The model was trained and tested using standard evaluation techniques, and the results indicated that it can reliably predict student outcomes. This proves that AI-based systems can effectively process large volumes of educational data and generate meaningful predictions that can support academic decision-making.

4.2. Important Features Analysis

The feature analysis revealed that certain factors have a significant impact on student performance. Among all the features, attendance, internal marks, assignment scores, and study hours were identified as the most influential. These features contributed heavily to the model's prediction accuracy.

Students with low attendance were more likely to perform poorly in exams, as they tend to miss important classroom learning and discussions. Similarly, students with low assignment scores often showed weak understanding of the subject, leading to lower overall performance. On the other hand, students who maintained regular attendance, actively participated in class activities, and spent more time studying consistently achieved better results.

This analysis highlights that both academic factors (marks, assignments) and behavioral factors (attendance, study habits) play a crucial role in determining student success. Understanding these factors helps educators focus on areas where students need improvement and provide timely support.

4.3. Explainability of Results

One of the key strengths of this study is the use of Explainable Artificial Intelligence (XAI) techniques to interpret the model's predictions. Methods such as Feature Importance, LIME (Local Interpretable Model-agnostic Explanations), and SHAP (Shapley Additive explanations) were used to provide clear explanations for each prediction made by the model.

These techniques help in identifying how each feature contributes to the final outcome. For example, the model can explain that a student is predicted to fail mainly due to low attendance and poor assignment performance. This level of transparency makes the model more trustworthy and easier to understand for teachers and educational administrators.

The explanation process is visually represented in Figure 2, where both prediction results and their corresponding reasons are shown. This ensures that the system is not a "black box" but an interpretable tool that provides both outcomes and justifications.

4.4. Discussion

The results clearly demonstrate that integrating AI with XAI techniques enhances both the accuracy and transparency of student performance prediction systems. While AI provides powerful predictive capabilities, XAI ensures that these predictions are understandable and actionable.

Teachers can use these insights to identify at-risk students at an early stage and take necessary actions such as providing extra coaching, mentoring, or academic support. This proactive approach can significantly improve student outcomes and reduce failure rates.

Moreover, the system supports better decision-making by providing data-driven insights rather than relying only on intuition. It also promotes a personalized learning environment, where students receive support based on their individual needs.

In conclusion, the combination of AI and XAI offers a highly effective solution for modern education systems. It not only improves prediction accuracy but also builds trust by explaining the reasoning behind each prediction. Therefore, this approach has great potential to enhance teaching strategies, support student success, and create a more efficient and transparent educational environment.

5. Conclusion and Future Work

This study presents the use of Explainable Artificial Intelligence (XAI) for predicting student academic performance. The proposed system uses student data such as attendance, internal marks, assignment scores, and study hours to make predictions. Machine learning models like Decision Tree and Random Forest are used to analyze the data and identify patterns. The results show that the model can accurately predict student performance and help in identifying students who may need additional support.

An important advantage of this system is the use of explainability. XAI techniques such as Feature Importance, LIME, and SHAP provide clear reasons for each prediction. This helps teachers and educators understand how different factors affect student performance. Instead of just giving a result, the system explains why a student is likely to perform in a certain way. This increases trust in the model and makes it more useful in real-world educational environments.

Overall, the integration of AI and XAI improves both prediction accuracy and transparency. It supports better decision-making and helps in improving the quality of education by enabling early intervention for students at risk.

Although the proposed system shows promising results, there are several areas for improvement in the future. First, more features can be included to improve prediction accuracy, such as socio-economic background, learning behavior, and extracurricular activities. Including more diverse data can provide a deeper understanding of student performance.

Second, advanced machine learning and deep learning models can be explored to further enhance prediction performance. Techniques such as neural networks may provide better results when working with large and complex datasets.

Third, the system can be developed into a real-time application that can be used by schools and colleges. This would allow teachers to monitor student performance continuously and take immediate action when needed.

Finally, improving the visualization of explanations can make the system more user-friendly. Better graphical representations can help teachers and students easily understand the results. Therefore, future work will focus on improving accuracy, usability, and real-world implementation of Explainable AI systems in education.

6. References

Pedro Domingos, "A Few Useful Things to Know About Machine Learning," Communications of the ACM, 2012.

Leo Breiman, "Random Forests," Machine Learning Journal, 2001.

Marco Tulio Ribeiro, "Why Should I Trust You? Explaining the Predictions of Any Classifier," KDD, 2016.

Scott M. Lundberg and Su-In Lee, "A Unified Approach to Interpreting Model Predictions," NeurIPS, 2017.

Cristianini Nello and John Shawe-Taylor, An Introduction to Support Vector Machines, Cambridge University Press, 2000.

Tom M. Mitchell, Machine Learning, McGraw-Hill, 1997.

Ian Goodfellow, Yoshua Bengio, and Aaron Courville, Deep Learning, MIT Press, 2016.

Trevor Hastie, Robert Tibshirani, and Jerome Friedman, The Elements of Statistical Learning, Springer, 2009.

Thomas Dietterich, "Ensemble Methods in Machine Learning," International Workshop, 2000.

IEEE, "Explainable Artificial Intelligence in Education," IEEE Access, 2020.

Elsevier, "Applications of AI in Education," Computers & Education Journal, 2019.

Springer, "Educational Data Mining and Learning Analytics," 2018.

Association for Computing Machinery, "Learning Analytics and Student Performance Prediction," 2021.