Journal of Information System Exploration and Research

https://shmpublisher.com/index.php/joiser

p-ISSN 2964-1160 | e-ISSN 2963-6361

Explainable Artificial Intelligence-Based Model for Student Academic Performance Prediction

Wildan Hidayatulloh[1*], Fathoni Mahardika[2], Dani Indra Junaedi[3]

[1,2,3] *Study Program of Informatics, Universitas Sebelas April, Indonesia*

DOI: https://doi.org/10.52465/joiser.v4i1.624

Received 08 October 2025; Accepted 21 January 2026; Available online 26 January 2026

Article Info

Keywords:

Machine learning;
Explainable AI;
SHAP;
LIME;
Random forest;
XGBoost

Abstract

This study focuses on predicting student academic performance while emphasizing model interpretability through Explainable Artificial Intelligence (XAI). The main objective is to identify potential academic risks using machine learning models and provide transparent explanations for their decisions. Historical student academic data were used to train and evaluate two classification models: Random Forest and XGBoost. The results show that both models exhibit strong predictive performance. Random Forest achieved an accuracy of 90.77% and a precision of 0.7500 for the risk class, while XGBoost attained a higher recall of 0.7000 with an accuracy of 89.23% and a precision of 0.6364. Both models achieved an identical F1-score of 0.6667 for the risk class. The application of XAI methods, namely SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations), revealed the main features influencing the predictions. Globally, G2 (previous period's final grade), failures (number of failed courses), and absences were identified as the most critical factors. Local interpretations from SHAP and LIME also clarified individual predictions, both correct and misclassified. The study contributes to developing an accurate and transparent decision-support system to enable more personalized, effective, and data-driven academic interventions.

This is an open-access article under the CC BY-SA license.

1. Introduction

A decline in academic performance or even probationary status has become a significant challenge faced by higher education institutions. Although a large amount of student academic data is available, its proactive utilization to detect and prevent these risks remains limited. Commonly used analytical approaches tend to be descriptive, offering neither accurate predictions nor deep interpretation of individual student conditions. Consequently, early intervention efforts in academic management are often hindered, even though historical data analysis should ideally serve as a foundation for more targeted actions.

* ***Corresponding Author:***

Wildan Hidayatulloh,
Study Program of Informatics,
Universitas Sebelas April,
Angkrek Situ No.19, Situ, Kec. Sumedang Utara, Kabupaten Sumedang, Jawa Barat 45323.
Email: wildanhidayatulloh73@gmail.com

With the rapid advancement of technology in machine learning and data analytics, predicting students' academic performance has become a widely explored topic. Various classification models such as Decision Tree, Random Forest, Support Vector Machine (SVM), and Gradient Boosting have been successfully applied to predict academic outcomes ranging from on-time graduation and probation status to final grade estimation [1], [2], [3]. Despite their predictive accuracy, these models often operate as "black boxes," making them difficult to interpret for end users particularly in higher education contexts involving non-technical stakeholders such as lecturers, program coordinators, and university management.

This situation creates an urgent need for systems capable not only of prediction but also of providing transparent explanations of their predictive results. In this regard, the Explainable Artificial Intelligence (XAI) approach offers a promising solution by bridging the gap between the complexity of predictive models and the interpretability required by users [4], [5], [6] [4]. XAI enables machine learning models to be evaluated not only based on predictive performance but also on transparency and clarity in decision-making. As a result, analysis outcomes become more trustworthy and effectively support evidence-based decision making.

Two popular and widely adopted XAI methods are SHAP (SHapley Additive exPlanations) [7] and LIME (Local Interpretable Model-agnostic Explanations) [8]. SHAP provides interpretation based on Shapley value theory from game theory by calculating each feature's contribution to the model output consistently and additively [9], [10], [11]. Meanwhile, LIME builds a simple local model around the instance being explained, offering intuitive and human-understandable explanations [11]. Although these methods have been successfully implemented in various domains such as finance, healthcare, and e-commerce their optimal application in education, particularly for interpreting models that predict students' academic performance, remains relatively limited.

Several studies in Indonesia have attempted to apply SHAP and LIME for explaining classification model predictions, for instance, in social media sentiment analysis [12], MSME credit feasibility evaluation, and disease detection [9]. These studies demonstrated that XAI approaches effectively identify influential features both globally and individually. However, the application of XAI to academic data, especially for detecting risks such as probation or general performance decline, remains underexplored and requires further investigation [5], [13], [14], [15].

In higher education, transparency in evaluation and intervention processes is essential. Academic advisors and program coordinators not only need to know who is at risk but, more importantly, why those students are predicted to be at risk. Such insights provide a foundation for more personalized and targeted interventions, such as offering additional academic guidance, adjusting study loads, or revising learning strategies. Therefore, implementing XAI could be an essential step toward more efficient and equitable data-driven academic management [2], [4].

Based on this urgency, this study aims to develop a classification model to predict students' academic risk and explain its predictive outcomes using Explainable AI approaches based on SHAP and LIME. Using historical student academic data, the model will be evaluated from both predictive accuracy and interpretability perspectives. Specifically, this study makes three main contributions to the literature on educational data mining. First, it proposes an integrated framework that combines ensemble learning algorithms (Random Forest and XGBoost) with Explainable Artificial Intelligence (XAI) to predict student academic risk, effectively addressing the trade-off between predictive accuracy and model transparency. Second, unlike previous studies that often focus solely on global feature importance, this research provides a dual-layer interpretability analysis using SHAP and LIME, allowing for the identification of risk factors at both the population (global) and individual student (local) levels. Third, the study demonstrates how these interpretable insights can be translated into actionable decision support for academic advisors, enabling more personalized and timely interventions for students identified as "at risk."

2. Literature Review

Machine learning (ML) has been widely used to predict student academic performance, employing algorithms such as Decision Tree, Random Forest, Support Vector Machine (SVM), and Gradient Boosting [1], [3], [9], [16], [17], [18]. These models have shown strong predictive capabilities in identifying students at risk of poor academic outcomes or potential dropout based on historical factors such as attendance, grades, and demographics. However, many studies prioritize predictive accuracy over interpretability, leaving the reasoning behind predictions largely inaccessible to educators and

administrators [2], [19]. This lack of transparency limits the practical adoption of predictive analytics in educational decision-making, where interpretability is critical for trust and accountability

Explainable Artificial Intelligence (XAI) has emerged as an effective solution to address this "black-box" issue in complex machine learning models [4], [5], [6]. Among the most prominent XAI methods, SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-Agnostic Explanations) have been widely utilized to interpret model predictions in domains such as finance, healthcare, and e-commerce [7], [8], [9], [10]. SHAP, grounded in cooperative game theory, quantifies each feature's contribution to model output, providing consistent and additive explanations across instances [19].

Meanwhile, LIME constructs simplified local surrogate models to generate human-readable, instance-level explanations [1], [2], [20]. These approaches have significantly enhanced transparency in automated systems, increasing stakeholder trust in AI-assisted decision-making [3], [4], [5], [16], [20].

In the educational context, applications of XAI remain relatively limited despite its potential benefits. Several studies have implemented SHAP and LIME to explain predictive models for student dropout and exam performance, successfully identifying influential features such as previous grades, attendance, and behavioral patterns [2], [10]. However, most of these works focus primarily on global-level explanations, providing limited insight into individual student predictions that could inform personalized interventions, other studies have focused on optimizing algorithmic accuracy or efficiency while overlooking interpretability, which is essential for practical application in education [7], [10], [12], [20].

The novelty of this study lies in integrating SHAP and LIME within a classification framework for predicting academic risk using student performance data [5], [9]. Unlike previous research that primarily targets prediction accuracy, this study emphasizes both predictive performance and interpretability. By combining the explanatory power of XAI with machine learning classification models, this work bridges the gap between accuracy and transparency. This approach enables educators and administrators to understand why certain predictions occur, supporting responsible AI adoption and fostering more transparent, equitable, and data-driven decision-making in higher education [10], [11], [12].

3. Method

This study employed a quantitative computational experiment approach focusing on the development and interpretation of machine learning–based models for predicting student academic performance using Explainable Artificial Intelligence (XAI). As illustrated in Figure 1, the research procedure consisted of four primary stages: data collection and preprocessing, model training and evaluation, implementation of XAI methods, and visualization of interpretability results [21], [22], [23].

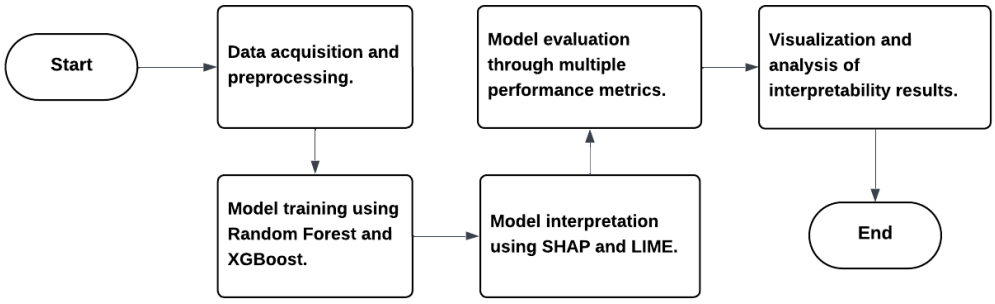


Figure 1. Research Workflow of the Proposed Academic Risk Prediction Model.
The workflow outlines the major stages of the study, including data preprocessing, model training using Random Forest and XGBoost, model evaluation, and interpretability analysis with SHAP and LIME.

The data used in this study were obtained from a publicly available academic dataset that was curated and adjusted to match the higher education context. It contained various attributes, including gender, age, study program, current semester, daily study habits, scholarship status, class attendance, previous GPA (SGPA), and current GPA (CGPA). These attributes served as independent variables contributing to the prediction of academic risk levels. To provide a clearer overview of the dataset structure, a sample of the records used in this study is presented in Table 1.

Table 1. Sample of Student Academic Dataset and Attributes

Gender	Age	Program	Semester	Study Hours	Scholarship	Attendance	SGPA	CGPA	Risk Status
Female	21	BCSE	2	7	No	100	3.40	3.40	Not at Risk
Male	25	BCSE	13	2	Yes	90	2.40	2.41	At Risk
Male	23	BCSE	8	2	Yes	90	3.94	3.90	Not at Risk
Male	21	BCSE	1	5	Yes	95	3.30	3.30	Not at Risk
Male	20	BCSE	2	4	Yes	85	2.68	3.15	Not at Risk

In this study, academic risk was defined as a condition where a student's cumulative GPA (CGPA) fell below 2.50. Consequently, CGPA values were categorized into two target classes: "At Risk" (CGPA < 2.50) and "Not at Risk" (CGPA ≥ 2.50).

Prior to model development, an ETL (Extract, Transform, Load) process was carried out to ensure data quality. During extraction, the dataset was obtained from Excel or CSV sources. The transformation stage involved cleaning missing values, encoding categorical attributes (e.g., gender, study program) using label or one-hot encoding, and applying normalization or standardization where necessary. Finally, the cleaned data were loaded into a machine learning pipeline for model training. These preprocessing steps ensured data consistency, reduced bias, and minimized potential errors during training.

In this study, Random Forest and XGBoost were selected based on their theoretical superiority in handling structured tabular data compared to single decision trees or deep neural networks.

Random Forest was chosen as a representative of the Bagging (Bootstrap Aggregating) technique. Theoretically, Random Forest reduces the model's variance by constructing multiple independent decision trees on different data subsets and averaging their predictions. This characteristic makes it highly robust against overfitting, which is critical when dealing with educational datasets that often contain noise from behavioral factors.

Complementing this, XGBoost (Extreme Gradient Boosting) was selected to leverage the Boosting framework. Unlike Random Forest which builds trees in parallel, XGBoost builds trees sequentially, where each new tree aims to correct the errors (residual errors) of the previous ones. This approach effectively reduces the model's bias. Furthermore, XGBoost is theoretically distinct due to its built-in regularization parameters (L1 and L2), which control model complexity. This feature provides a significant advantage in handling the class imbalance observed in this study (At Risk vs. Not At Risk), ensuring that the model does not bias its predictions solely toward the majority class.

Model performance was evaluated using four standard metrics accuracy, precision, recall, and F1 score to provide a balanced assessment of predictive capability, especially considering the class imbalance in academic risk prediction. This multi-metric evaluation ensured that both the correctness and sensitivity of the models were adequately represented.

To enhance interpretability, two Explainable Artificial Intelligence (XAI) methods, SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-Agnostic Explanations), were applied. SHAP quantified the contribution of each feature to the model's output based on Shapley values, offering both global and local interpretability. Meanwhile, LIME provided intuitive, instance-level explanations by constructing simplified local models. The integration of these methods enabled transparent, trustworthy predictions, supporting informed academic decision-making.

All computational experiments were performed in Google Colaboratory using Python 3.10 and key libraries such as scikit-learn, XGBoost, pandas, matplotlib, seaborn, SHAP, and LIME. The cloud-based environment ensured efficient and reproducible implementation without requiring complex local configurations.

4. Results and Discussion

4.1 Classification Model Performance Analysis

This section presents the results of the computational experiments conducted to predict students' academic risk using two well-established machine learning algorithms: Random Forest and XGBoost Classifier. The evaluation focuses not only on the predictive accuracy of the models but also on their interpretability through Explainable Artificial Intelligence (XAI).

Both models were trained and evaluated using the prepared academic dataset, with an 80:20 ratio for training and testing data, respectively. The performance of each model was assessed using four

metrics: accuracy, precision, recall, and F1-score. The results of this evaluation are summarized in Table 1.

Table 2. Comparison of Classification Model Performance Metrics

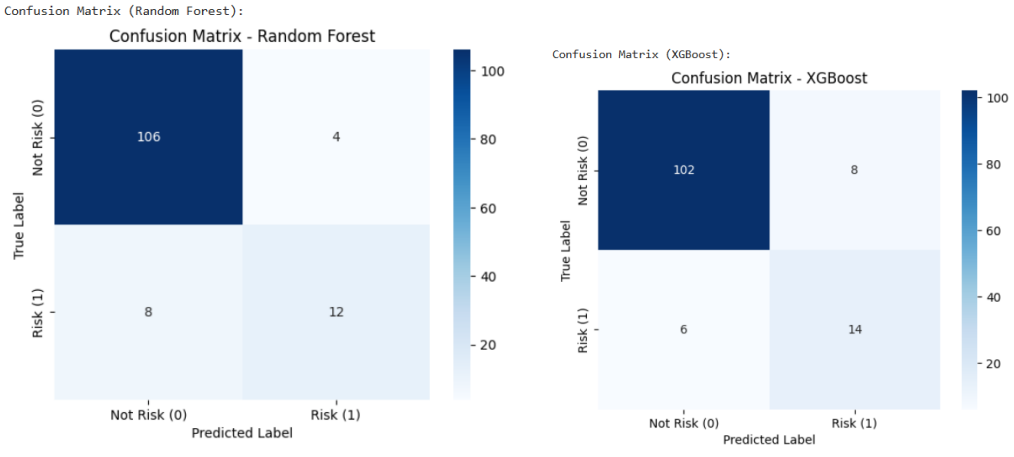
Metric	Random Forest	XGBoost
Accuracy	0.9077	0.8923
Precision (Risk Class)	0.7500	0.6364
Recall (Risk Class)	0.6000	0.7000
F1-Score (Risk Class)	0.6667	0.6667

As shown in Table 1, the Random Forest model achieved an accuracy of 90.77%, slightly higher than that of XGBoost, which achieved 89.23%. Although both models demonstrated high overall accuracy, this metric alone may not fully represent performance, especially when dealing with imbalanced datasets. In academic risk prediction, the number of "at-risk" students (minority class) is often significantly smaller than the number of "not-at-risk" students (majority class), making metrics such as precision, recall, and F1-score more informative for evaluating the models.

The Random Forest model obtained a higher precision (0.7500), indicating that when it predicts a student as "at risk," three out of four predictions are correct. This shows that the model minimizes false positives and provides more reliable alerts, avoiding unnecessary interventions for students who are not actually at risk. In contrast, the XGBoost model demonstrated a higher recall (0.7000), meaning it was more effective in detecting students who were genuinely at risk. However, this improvement in recall came at the cost of lower precision (0.6364), resulting in more false-positive predictions.

Both models achieved an identical F1-score of 0.6667 for the at-risk class, suggesting that their overall balance between precision and recall is comparable. This implies that while Random Forest performs slightly better in terms of accuracy and precision, XGBoost is more sensitive in identifying actual at-risk students. The choice between these models, therefore, depends on the institutional priority whether to minimize false positives (favoring Random Forest) or to maximize early detection of at-risk students (favoring XGBoost).

The confusion matrices presented in Figure 2 and Figure 3 further illustrate the distribution of correct and incorrect classifications for each model. The Random Forest model correctly classified 106 students as not at risk (TN) and 12 students as at risk (TP), with 4 false positives (FP) and 8 false negatives (FN). Similarly, the XGBoost model correctly identified 102 TN and 14 TP, with 8 FP and 6 FN. These results confirm that both models exhibit complementary strengths: Random Forest is slightly more precise, while XGBoost provides better sensitivity toward the minority class.



(a) (b)

Figure 2. (a) Confusion Matrix of the Random Forest model for academic risk prediction. (b) Confusion Matrix of the XGBoost model for academic risk prediction.

The matrix displays true and false classifications for at-risk and not-at-risk students in the test dataset.

4.2 Global Model Interpretation Using SHAP

To understand how each feature contributes to the overall prediction of academic risk, the study employed the SHAP (SHapley Additive exPlanations) method as part of the Explainable Artificial Intelligence (XAI) approach. SHAP provides a consistent and theoretically grounded way to interpret

complex machine learning models by attributing a contribution value, known as the Shapley value, to each feature for every prediction. These contribution values can be aggregated to reveal how individual variables influence the model globally across all instances.

In this study, the SHAP Summary Plot was generated to visualize the global importance and direction of influence of each feature within the XGBoost Classifier, as shown in Figure 4. Due to technical constraints in the interaction between the SHAP library and the Random Forest configuration within the computational environment, the visualization for Random Forest could not be generated optimally. Therefore, the global interpretation focuses on the XGBoost model, which successfully produced interpretable SHAP visualizations.

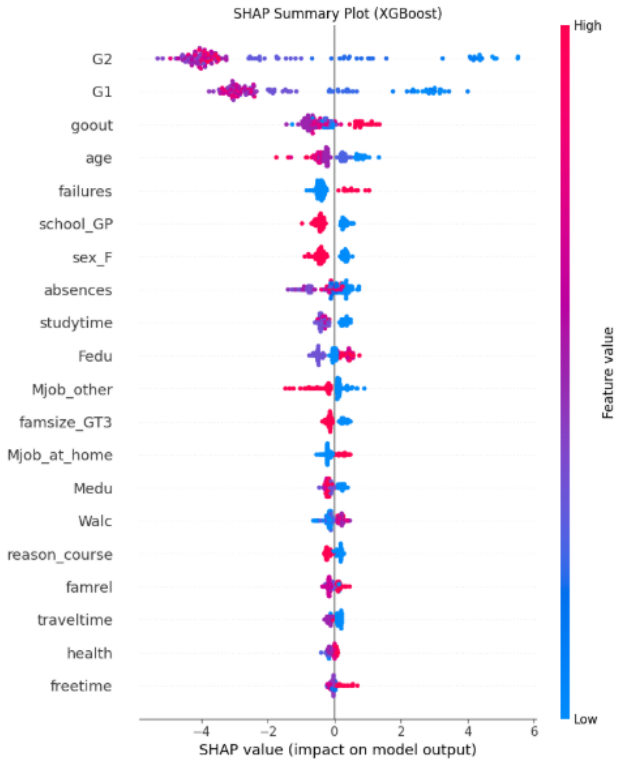


Figure 4. SHAP Summary Plot for the XGBoost model.

The plot illustrates the most influential features contributing to academic risk predictions, where red points indicate high feature values that increase risk and blue points indicate low feature values that decrease risk.

The visual analysis in Figure 4 reveals that G2 (Previous Period's Final Grade) is the most critical feature, appearing at the top of the importance ranking. As observed in the plot, the x-axis represents the SHAP value, where positive values indicate a push towards the "At Risk" prediction and negative values indicate a push towards "Not at Risk." For G2, the red points (representing high grades) are concentrated on the negative side (left), while the blue points (representing low grades) are on the positive side (right). This visual pattern confirms that students with higher past grades are strongly protected against academic risk, whereas those with lower grades are driven closer to the risk threshold.

Conversely, the features failures (number of failed courses) and absences (number of class absences) show the opposite pattern. For these features, the red points stretch toward the right side (positive SHAP values). This explicitly indicates that a higher number of failures and increased absenteeism are strong contributors pushing the model to classify a student as "At Risk."

Interestingly, behavioral variables such as health (self-reported health status) and goout (frequency of going out) also demonstrated notable contributions. As shown in the middle section of Figure 4, the red points for these features tend to cluster on the positive side. This suggests that students who report poor health conditions (higher value on the scale) or frequent social outings are visually correlated with an increased likelihood of academic risk. These findings provide valuable insights for educational institutions, highlighting that student performance is influenced not only by academic history but also by well-being and lifestyle factors.

While the SHAP visualization for the Random Forest model could not be presented, the key features identified by XGBoost were consistent with the built-in feature importance ranking of the Random Forest model. This consistency reinforces the reliability of the findings and indicates that G2, failures, and absences are robust predictors of academic risk. The global interpretability offered by SHAP thus enables a more transparent understanding of the underlying patterns in student performance, which can be leveraged for early intervention and policy development in academic management.

4.3 Local Model Interpretation Using SHAP Force Plot and LIME

Local interpretation was conducted to provide instance-level explanations for model predictions using SHAP Force Plot and LIME. These methods help visualize why the model categorized individual students as “at risk” or “not at risk,” by identifying which features most strongly influenced each decision.

Representative examples from the test set illustrate typical model behavior. For correctly classified students (True Positives and True Negatives), both SHAP and LIME confirmed that previous performance indicators such as G2 (previous term final grade) played a dominant role. High G2 values consistently pushed the prediction toward “Not at Risk,” while lower G2 values, combined with a higher number of failures and absences, increased the likelihood of being predicted “At Risk.”

Conversely, in misclassified cases (False Positives and False Negatives), both models exhibited overreliance on single features such as G2, sometimes ignoring compensating factors like attendance improvement or health status. For example, several students were incorrectly labeled as “At Risk” despite showing recent positive academic trends. These findings highlight that, while XGBoost and Random Forest achieve strong predictive performance, interpretability analysis reveals opportunities to improve model balance through feature weighting and temporal dynamics.

Figure 5 illustrates a typical LIME explanation for a correctly classified student, where G2 and absences exerted the largest local impact. Meanwhile, Figure 6 shows a SHAP Force Plot depicting how both positive and negative feature contributions shaped the final prediction.

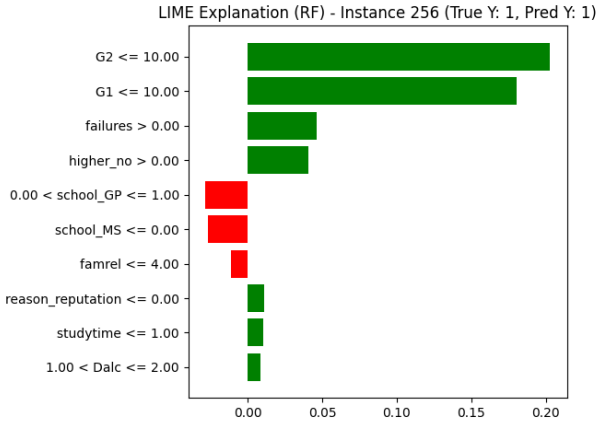


Figure 5. LIME Explanation for Student ID 256 (True Y: 1, Pred Y: 1) by the Random Forest model. Green bars represent features that support the “At Risk” prediction, while red bars represent features that support the “Not At Risk” prediction.

.

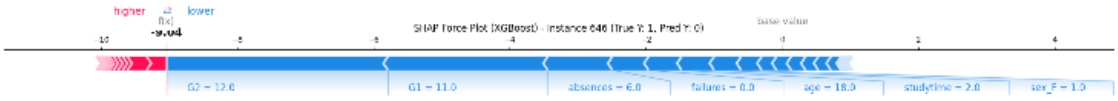


Figure 6. SHAP Force Plot for Student ID 646 (True Y: 1, Pred Y: 0) by the XGBoost model. Red arrows indicate features pushing the prediction toward “At Risk,” while blue arrows indicate features pushing toward “Not At Risk.”

Overall, both SHAP and LIME provide consistent and complementary insights. SHAP offers a mathematically grounded, consistent contribution for each feature, while LIME provides intuitive local explanations that can be easily interpreted by non-technical stakeholders such as academic advisors and program coordinators. Together, they enhance transparency, accountability, and trust in the use of predictive analytics within educational settings.

4.4 Discussion of Implications and Research Contributions

This study successfully developed and interpreted two classification models Random Forest and XGBoost Classifier for predicting student academic risk, enhanced through the integration of Explainable Artificial Intelligence (XAI) methods, namely SHAP and LIME. The results demonstrate that both algorithms possess strong predictive capabilities, with Random Forest exhibiting higher precision (0.7500) and XGBoost achieving higher recall (0.7000). These complementary strengths highlight the trade-off between minimizing false positives and maximizing true risk detection. Ultimately, the choice of model depends on institutional priorities whether to ensure efficient intervention targeting (favoring precision) or to avoid missing potential at-risk students (favoring recall).

These findings align with previous studies such as [1]and [17], which also reported that ensemble learning methods like Random Forest and XGBoost outperform single decision trees in educational data mining tasks. However, a critical advancement in this study compared to [17] and [18] is the shift from black-box prediction to transparent decision-making. While prior works primarily prioritized accuracy metrics (Precision/Recall), this study demonstrates that high predictive performance can be achieved without sacrificing interpretability. Furthermore, unlike [2] which focused heavily on global explanations using SHAP, our approach integrates LIME to provide granular, instance-level insights. This dual-layer interpretation addresses the gap highlighted by [10], ensuring that the model is practically useful for identifying specific risk factors for individual students, not just general trends.

The primary contribution of this research lies in incorporating XAI to improve the transparency and interpretability of traditionally black-box machine learning models. The global interpretation using SHAP Summary Plot consistently identified G2 (Previous Period's Final Grade), failures (Number of Failed Courses), and absences (Number of Absences) as the most influential predictors of academic risk. These findings provide empirical evidence of the critical role of past academic performance and behavioral engagement in determining student outcomes. The consistency of these features across both Random Forest and XGBoost models reinforces their robustness as key indicators for early academic risk detection.

Furthermore, local interpretability through SHAP Force Plot and LIME Explanations confirmed the models' ability to justify predictions at the individual level. Correct predictions (True Positive and True Negative cases) reflected logical relationships between performance indicators and outcomes, while misclassified cases (False Positive and False Negative) revealed valuable insights into model limitations. For instance, False Positive cases often resulted from overemphasis on low prior grades despite subsequent improvement, while False Negative cases occurred when relatively strong historical performance masked emerging academic decline. These patterns highlight the importance of continuously refining models and incorporating dynamic or longitudinal data to better capture academic trajectories.

From a practical perspective, the integration of XAI in academic analytics offers substantial benefits for data-driven decision-making within higher education. Transparent and interpretable models enable stakeholders such as academic advisors, program heads, and institutional policymakers—to understand the rationale behind each prediction. This understanding fosters trust and allows more personalized interventions, such as targeted academic counseling, workload adjustment, or curriculum enhancement. By transforming predictive results into actionable insights, educational institutions can move toward a more proactive and equitable system of academic management.

In summary, this study contributes a methodological framework that bridges predictive accuracy with interpretability, promoting the responsible adoption of AI technologies in education. By combining the predictive strength of ensemble learning algorithms with the explanatory power of XAI, this research demonstrates how machine learning can be used not only to identify students at risk but also to explain why they are at risk. Such interpretability is essential for ensuring fairness, accountability, and human-centered decision-making in educational data analytics.

5. Conclusion

This study aimed to analyze and predict student academic performance with a particular focus on model interpretability through Explainable Artificial Intelligence (XAI). Using historical academic data, two classification algorithms Random Forest and XGBoost Classifier were developed to identify students at potential academic risk. The evaluation results revealed that both models demonstrated strong predictive performance. Random Forest achieved higher precision (0.7500), effectively reducing

false positives, while XGBoost obtained higher recall (0.7000), allowing for the detection of a larger proportion of at risk students. Although their strengths differ, both models yielded an identical F1-score of 0.6667 for the at-risk class, indicating balanced overall performance.

The integration of XAI methods, specifically SHAP and LIME, transformed the black-box nature of the models into transparent and explainable systems. Global interpretation using SHAP Summary Plot identified G2 (Previous Period's Final Grade), failures (Number of Failed Courses), and absences (Number of Absences) as the most influential predictors of academic risk. Meanwhile, local interpretation using SHAP Force Plot and LIME Explanations provided deeper insights into individual predictions, clarifying why certain students were classified as "at risk" or "not at risk." These findings not only validate the models' reasoning but also demonstrate how interpretability can enhance trust and usability for non-technical stakeholders.

In essence, the study contributes to the development of an academic decision-support framework that combines predictive accuracy with model transparency. Such systems can empower educational institutions to make data driven and personalized interventions aimed at improving student outcomes.

Despite these contributions, this study acknowledges several limitations. First, the dataset used is derived from a specific higher education context, which may limit the direct generalizability of the findings to other institutions with different student demographics or academic policies. Second, the current model relies on static historical data, potentially overlooking real-time changes in student behavior during the semester.

For future work, it is recommended to expand the dataset by including more diverse academic records from multiple institutions to address the generalizability issue. Additionally, future research could incorporate temporal or longitudinal data to capture dynamic changes in student behavior over time. Exploring more advanced ensemble or deep learning techniques—interpreted through XAI—could also provide richer insights. Finally, practical implementation of the model in real academic settings should be investigated, including the development of user-friendly dashboards and evaluation of the impact of XAI-driven interventions on student performance.

References

[1] E. Ben George, "Explainable AI Methods for Predicting Student Grades and Improving Academic Success," *J. Inf. Syst. Eng. Manag.*, vol. 10, no. 23s, pp. 117–126, 2025, doi: 10.52783/jisem.v10i23s.3680.

[2] F. T. Johora, M. N. Hasan, A. Rajbongshi, M. Ashrafuzzaman, and F. Akter, "An explainable AI-based approach for predicting undergraduate students academic performance," *Array*, vol. 26, no. February, p. 100384, 2025, doi: 10.1016/j.array.2025.100384.

[3] M. Madeniyetov, "ANALYZING E-LEARNING STUDENT PERFORMANCE AND FEATURE IMPACT USING XGBOOST AND SHAP M. Madeniyetov," pp. 12–19.

[4] A. M. Salih *et al.*, "A Perspective on Explainable Artificial Intelligence Methods: SHAP and LIME," *Adv. Intell. Syst.*, vol. 7, no. 1, pp. 1–8, 2025, doi: 10.1002/aisy.202400304.

[5] J. Prentzas and A. Binopoulou, "Explainable Artificial Intelligence Approaches in Primary Education: A Review," *Electron.*, vol. 14, no. 11, 2025, doi: 10.3390/electronics14112279.

[6] D. Adi and N. Nurdin, "Explainable Artificial Intelligence (XAI) towards Model Personality in NLP task," *IPTEK J. Eng.*, vol. 7, no. 1, p. 11, 2021, doi: 10.12962/j23378557.v7i1.a8989.

[7] S. M. Lundberg and S. I. Lee, "A unified approach to interpreting model predictions," *Adv. Neural Inf. Process. Syst.*, vol. 2017-Decem, no. Section 2, pp. 4766–4775, 2017.

[8] M. T. Ribeiro, S. Singh, and C. Guestrin, "'Why Should I Trust You?' Explaining the Predictions of Any Classifier," *NAACL-HLT 2016 - 2016 Conf. North Am. Chapter Assoc. Comput. Linguist. Hum. Lang. Technol. Proc. Demonstr. Sess.*, pp. 97–101, 2016, doi: 10.18653/v1/n16-3020.

[9] A. S. Iffadah, Trimono, and Dwi Arman Prasetya, "Shapley Additive Explanations Interpretation of the XGBoost Model in Predicting Air Quality in Jakarta," *J. Ris. Inform.*, vol. 7, no. 3, pp. 119–127, 2025, doi: 10.34288/jri.v7i3.366.

[10] P. S. R. Aditya and M. Pal, "Local Interpretable Model Agnostic Shap Explanations for machine learning models," no. c, 2022, [Online]. Available: http://arxiv.org/abs/2210.04533

[11] M. T. Syamkalla, S. Khomsah, and Y. S. R. Nur, "Implementasi Algoritma Catboost Dan Shapley Additive Explanations (SHAP) Dalam Memprediksi Popularitas Game Indie Pada Platform Steam," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 11, no. 4, pp. 777–786, 2024, doi: 10.25126/jtiik.1148503.

[12] A. F. Hadi, A. F. Zulva, M. L. Hakim, M. D. Saputra, and H. Sadiyah, “Implementasi Explainable Machine Learning: Visualisasi Global Explainability and Local Interpretability pada Analisis Sentimen dengan SHAP dan LIME,” *BIAStatistics J. Stat. Teor. dan Apl. Biomed. Ind. Bus. Soc. Stat.*, vol. 17, no. 1, pp. 1–18, 2023, [Online]. Available: https://biastatistics.statistics.unpad.ac.id/?journal=biastatistics&page=article&op=view&path%5B%5D=219

[13] W.-C. Choi, C.-T. Lam, P. C.-I. Pang, and A. J. Mendes, “A Systematic Literature Review of Explainable Artificial Intelligence (XAI) for Interpreting Student Performance Prediction in Computer Science and STEM Education,” pp. 221–227, 2025, doi: 10.1145/3724363.3729027.

[14] M. El Jihaoui, O. E. K. Abra, and K. Mansouri, “Predicting and Interpreting Student Academic Performance: A Deep Learning and Shapley Additive Explanations Approach,” *SHS Web Conf.*, vol. 214, p. 01001, 2025, doi: 10.1051/shsconf/202521401001.

[15] E. Kalita, H. El, A. Kukkar, S. Hussain, T. Ali, and S. Gaftandzhieva, “LSTM - SHAP based academic performance prediction for disabled learners in virtual learning environments : a statistical analysis approach,” 2025.

[16] H. Tao, Y. Wen, R. Yu, Y. Xu, and F. Yu, “Predictive model establishment for forward-head posture disorder in primary-school-aged children based on multiple machine learning algorithms,” *Front. Bioeng. Biotechnol.*, vol. 13, no. May, pp. 1–12, 2025, doi: 10.3389/fbioe.2025.1607419.

[17] A. Khosravi and A. Azarnik, “Leveraging Educational Data Mining : XGBoost and Random Forest for Predicting Student Achievement”.

[18] A. Ridwan, A. Mudi, and L. Ningsih, “Journal of Education and Predict Students ’ Dropout and Academic Success with XGBoost,” pp. 1–8.

[19] E. Raditya and R. Indraswari, “Integration of LIME Explainable AI to Enhance Interpretability of Deep Learning Models in Box Palette Classification,” *Ilk. J. Comput. Sci. Appl. Informatics*, vol. 6, no. 2, pp. 87–95, 2024, doi: 10.28926/ilkomnika.v6i2.653.

[20] A. Munna and E. Zuliarso, “Interpretasi model Stacking Ensemble untuk analisis sentimen ulasan aplikasi pinjaman online menggunakan LIME,” *Aiti*, vol. 21, no. 2, pp. 183–196, 2024, doi: 10.24246/aiti.v21i2.183-196.

[21] M. Afzaal *et al.*, “Explainable AI for Data-Driven Feedback and Intelligent Action Recommendations to Support Students Self-Regulation,” *Front. Artif. Intell.*, vol. 4, no. November, pp. 1–20, 2021, doi: 10.3389/frai.2021.723447.

[22] L. C. Nnadi, Y. Watanobe, M. Rahman, and A. M. John-otumu, “applied sciences Prediction of Students ’ Adaptability Using Explainable AI in Educational Machine Learning Models,” 2024.

[23] F. Nugraha, W. Widowati, and A. Sugiharto, “Machine Learning Methods for Academic Achievement Prediction : A Bibliometric Review,” vol. 02, pp. 221–226, 2025.