Journal of Soft Computing Exploration
Homepage: htttps://shmpublisher.com/index.php/joscex
p-ISSN: 2746-7686
e-ISSN: 2746-0991

.

An interpretable hybrid ensemble model for early academic risk detection with shap-based explainability

Rita Wahyuni Arifin[1], Sumardiono[2], Shalahuddin[3], Yeffry Handoko Putra[4], Khoem Sambath[3]
[1]Department of Informatics Management, Universitas Bina Insani, Indonesia
[2]Department of System Information, Universitas Bina Insani, Indonesia
[3]Department of Computer Science, Universitas Bina Insani, Indonesia
[4]Department of Postgraduate of Information System, Universitas Komputer Indonesia, Indonesia
[5]Department of Computer Science, National Polytechnic of Cambodia, Cambodia

Article Info

Article history:

Received April 10, 2026
Revised May 19, 2026
Accepted July 1, 2026

Keywords:

Academic risk
Educational data mining
Ensemble learning
Interpretable machine learning
Prediction

ABSTRACT

Student dropout remains a major challenge in educational institutions, affecting both academic achievement and institutional performance. Early identification of at-risk students is essential to support timely intervention and improve retention rates. This study proposes an interpretable hybrid ensemble model for early academic risk detection by integrating stacking ensemble learning, Bayesian optimization, and SHAP-based interpretability. The study used a public dataset from Kaggle containing 10,000 student records with demographic, behavioral, and academic performance attributes. Exploratory analysis indicated that at-risk students generally had lower GPA and attendance rates, while higher stress levels were associated with increased dropout risk. Due to the moderately imbalanced class distribution in the dataset, SMOTE was applied only to the training data after train–test splitting to improve minority class representation while avoiding data leakage during evaluation. Experimental results demonstrate that the proposed hybrid model outperformed individual baseline classifiers and standard stacking ensemble methods, achieving an accuracy of 0.93 ± 0.01, precision of 0.92 ± 0.01, recall of 0.91 ± 0.01, and F1-score of 0.91 ± 0.01 under stratified 5-fold cross-validation. The integration of Bayesian Optimization improved model stability and classification consistency, while SHAP-based analysis identified GPA, CGPA, attendance rate, study hours, and stress index as the most influential contributors to prediction outcomes.

Corresponding Author:

Rita Wahyuni Arifin
Department of Informatics Management
Universitas Bina Insani, Bekasi, Jawa Barat, Indonesia
Email: ritawahyuni@binainsani.ac.id
https://doi.org/10.52465/joscex.v7i3.50

1. INTRODUCTION

Student dropout [1] remains a significant challenge in modern education systems, as it directly impacts student success and institutional performance. High dropout rates not only reduce graduation outcomes but also reflect inefficiencies in academic support systems. Therefore, early identification of students at risk is essential to enable timely intervention and improve retention. With the increasing availability of educational

data, institutions are now able to leverage data-driven approaches to better understand student behavior [2], and academic performance [3], [4], [5].

Learning analytics and Educational Data Mining (EDM) have become important research areas for analyzing student behavior and predicting academic outcomes using machine learning techniques [6], [7]. Various classification algorithms, including Logistic Regression, Decision Tree, and Random Forest, have been widely applied for academic risk prediction and student performance analysis [8], [9], [10], [11], [12]. These approaches demonstrate the potential of machine learning to support early intervention strategies in educational environments.

However, individual machine learning models often experience limitations when handling complex and heterogeneous educational datasets. Educational data commonly contain nonlinear relationships, behavioral variability, and class imbalance problems that may reduce prediction accuracy and model stability. To address these limitations, ensemble learning techniques have increasingly been adopted because they combine multiple classifiers to improve robustness and generalization capability [13], [14]. Among ensemble approaches, stacking ensemble methods are particularly effective because they integrate heterogeneous base learners capable of capturing more complex educational patterns.

Recent studies have also incorporated explainable artificial intelligence (XAI) approaches to improve model transparency in educational prediction systems. Methods such as SHAP and LIME have been used together with machine learning algorithms including Support Vector Machine, Logistic Regression, and Extreme Gradient Boosting to explain prediction outcomes and feature contributions [15], [16], [17], [18]. Explainability is essential in educational contexts because prediction results should be interpretable for academic stakeholders and decision-making processes.

Despite these advancements, only limited studies have integrated class imbalance handling, stacking ensemble learning, Bayesian hyperparameter optimization, and explainable AI within a single predictive educational analytics framework. Most previous studies focused either on improving predictive performance or on model interpretability independently.

Therefore, this study proposes an interpretable hybrid ensemble framework that integrates SMOTE, stacking ensemble learning, Bayesian Optimization [19], and SHAP-based explainability for early academic risk detection. The proposed framework aims to improve predictive accuracy, model robustness, minority class detection, and interpretability simultaneously. By utilizing demographic, behavioral, and academic-related features, this research contributes to the development of an intelligent and explainable educational decision support system for identifying students at academic risk [20], [21].

2. METHOD

This study adopts a structured research methodology to develop an interpretable hybrid ensemble model for early academic risk detection. This study adopts the Cross-Industry Standard Process for Data Mining (CRISP-DM) framework [22], which consists of business understanding, data understanding, data preparation, modeling, evaluation, and deployment stages. The overall research framework is illustrated in Figure 1.

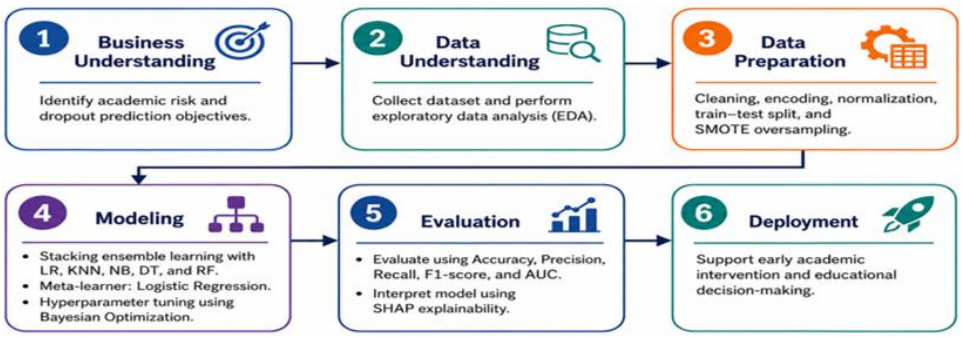

Figure 1. Proposed research framework for early academic risk prediction

Figure 1 presents the proposed research framework based on the CRISP-DM methodology. The process begins with business understanding to define academic risk prediction objectives, followed by data understanding through dataset collection the dataset is collected from data public (Kaggle.com) with 19

features and exploratory data analysis. The data preparation stage includes cleaning, encoding, normalization, train–test splitting, and SMOTE oversampling [23]. In the modeling stage, a stacking ensemble model is developed using several base learners and Logistic Regression as the meta-learner, with Bayesian Optimization applied for hyperparameter tuning. The evaluation stage assesses model performance using Accuracy, Precision, Recall, F1-score, and AUC, while SHAP explainability is used to interpret feature contributions. Finally, the deployment stage supports early academic intervention and educational decision-making. The results of the model can assist educational institutions in identifying at-risk students and developing timely support strategies.

Research Design and Identification Problem

This study adopts a quantitative experimental approach using machine learning techniques to develop an interpretable predictive model for early academic risk detection. The research framework follows Cross-Industry Standard Process for Data Mining (CRISP-DM) methodology, which consists of problem understanding, data understanding, data preprocessing, modelling, evaluation, and interpretation stages [3]. This framework is selected due to its flexibility in handling real-world data and its suitability for educational data mining applications.

Dataset Collection

The dataset used in this study was obtained from Kaggle under the title "Student Dropout Prediction Dataset (https://www.kaggle.com/datasets/meharshanali/student-dropout-prediction-dataset). The dataset consists of 10,000 student records designed for educational analytics and academic risk prediction research, total 19 columns. It contains demographic, behavioral, and academic performance variables, including age, gender, attendance rate, GPA, study hours, stress index, and family income. The dataset was accessed in January 2026 and was used as a publicly available benchmark dataset for educational data mining experiments.

The dataset includes demographic, behavioral, and academic performance features. Demographic attributes include gender, parental education, and department. Previous studies have identified academic performance indicators, attendance patterns, and behavioral variables as important predictors of student academic risk [24] and dropout tendencies [25].

The target variable represents student academic risk, indicating whether a student is at risk of dropout or not. The diversity of features enables comprehensive modeling of student performance and supports early detection of academic risk. The target variable represents academic risk [26], defined as a binary classification problem.

$$y = \begin{cases} 1 \\ 0 \end{cases} \tag{1}$$

where 1 indicates at-risk and 0 otherwise.

Data Preprocessing

Data preprocessing is conducted to ensure data quality and compatibility with machine learning models. Categorical variables, including Gender, Internet Access, and Extracurricular Participation, were encoded using label encoding. Meanwhile, numerical variables such as GPA, CGPA, Attendance Rate, Study Hours per Day, Stress Index, Assignment Delay Days, Family Income, and Travel Time Minutes were normalized using Min–Max scaling to ensure comparable feature ranges during model training.

Exploratory Data Analysis (EDA) is performed to analyze data distribution and identify patterns, including class imbalance and feature relationships. These preprocessing steps improve model stability and enhance predictive performance. Before model development, several preprocessing steps were performed to improve data quality and model compatibility.

Handling missing values and class imbalance

Missing values in numerical features were handled using median imputation, while missing categorical attributes were replaced using mode imputation. Categorical variables were transformed into numerical representations using label encoding, and numerical features were normalized using Min-Max Scaling.

Because the dataset exhibited moderate class imbalance between at-risk and non-risk students, the Synthetic Minority Oversampling Technique (SMOTE) was applied to the training dataset after train-test splitting. SMOTE was implemented only on the training data to avoid data leakage and preserve the integrity of model evaluation. During cross-validation, SMOTE was independently applied to each training fold before model fitting. The oversampling process was implemented using the imbalanced-learn library with k=5 nearest neighbors. This approach helps improve minority class representation and enhances the model's ability to correctly identify at-risk students.

Proposed Hybris Model

The stacking approach combines the predictions of base learners into a meta-feature vector, which is then processed by a meta-learner using Logistic Regression [10]. This approach enables the model to learn optimal combinations of base predictions and improves generalization performance.

This study proposes a hybrid ensemble learning model based on stacking. Five machine learning algorithms are used as base learners: Logistic Regression (LR), K-Nearest Neighbors (KNN), Naive Bayes (NB), Decision Tree (DT), and Random Forest (RF). Each base learner $h_j(x)$, for $j = 1,2,\dots,5$, produces a prediction:

$$z_j = h_j(x) \tag{2}$$

where $h_j(x)$is the prediction function of the j-th base learner, $z_j$is the predicted output (class probability or label) from the j-th model. The outputs of all base learners are combined into a meta-feature vector:

$$\mathbf{z} = [h_1(x), h_2(x), h_3(x), h_4(x), h_5(x)] \tag{3}$$

where $\mathbf{z} \in \mathbb{R}^5$is the meta-feature vector, each element represents the prediction of a base learner This approach enables the model to learn optimal combinations of base predictions and improves generalization performance[48], [49]. The meta-learner uses Logistic Regression to generate the final prediction:

$$\hat{y} = \sigma\left(\beta_0 + \sum_{j=1}^{5} \beta_j z_j\right) \tag{4}$$

where $\hat{y}$is the final predicted probability of academic risk, $\beta_0$is the bias (intercept) term, $\beta_j$is the coefficient corresponding to the j-th base learner, $z_j$is the input from the base learner, $\sigma(\cdot)$is the sigmoid activation function defined as:

$$\sigma(t) = \frac{1}{1+e^{-t}} \tag{5}$$

To further enhance model performance, Bayesian optimization is applied to automatically tune hyperparameters. The Expected Improvement (EI) acquisition function is used to efficiently explore the parameter space and identify optimal configurations, reducing the limitations of manual tuning methods [20]. Let $f(x)$denote the objective function representing model performance (e.g., F1-score). The optimization objective is:

$$x^* = \arg\max_{x \in X} f(x) \tag{6}$$

where x represents a set of hyperparameters, $\mathcal{X}$is the hyperparameter search space, $x^*$is the optimal hyperparameter configuration. The Expected Improvement (EI) acquisition function is used to efficiently explore the parameter space and identify optimal configurations, reducing the limitations of manual tuning methods [6]. The Expected Improvement (EI) acquisition function is defined as:

$$EI(x) = \mathbb{E}[\max(0, f(x) - f(x^+))] \tag{7}$$

where $EI(x)$measures the expected improvement at point x, $f(x^+)$is the best observed performance so far $\mathbb{E}[\ ]$denotes the expectation operator. This formulation enables efficient exploration and exploitation of the hyperparameter space, leading to improved model performance.

Bayesian Optimization was employed to automatically identify optimal hyperparameter configurations as shown in Table 1 for the proposed ensemble model. The optimization process used the Expected Improvement (EI) acquisition function to efficiently balance exploration and exploitation within the search space.

Table 1. Hyperparameter configurations

Model	Hyperparameter	Search Space
Random Forest	n_estimators	50–300
Random Forest	max_depth	3–20
KNN	n_neighbors	3–15
Logistic Regression	C	0.001–10

Table 1 presents the hyperparameter configurations and search spaces used during the Bayesian Optimization process. Several important parameters from the base learning algorithms were optimized to improve predictive performance and model stability. For the Random Forest model, the number of estimators (n_estimators) was optimized within the range of 50 to 300, while the maximum tree depth (max_depth) was explored between 3 and 20 to control model complexity and prevent overfitting. In the K-Nearest Neighbors (KNN) model, the number of neighbors (n_neighbors) was optimized within the range of 3 to 15 to determine the most suitable neighborhood size for classification. In addition, the Logistic Regression model optimized the regularization parameter (C) within the range of 0.001 to 10 to balance model flexibility and generalization capability. These search spaces were selected to provide sufficient exploration of model configurations while maintaining computational efficiency during the optimization process.

Model Evaluation

Model evaluation was conducted using stratified k-fold cross-validation to reduce evaluation bias and assess model robustness. Performance metrics were calculated as the average values across folds along with their standard deviations. Several evaluation metrics are used, including Accuracy, Precision, Recall, F1-score, and Area Under the Curve (AUC). These metrics provide a comprehensive assessment of classification performance and allow comparison between models. The F1-score is defined as:

$$F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} \tag{8}$$

To further validate the effectiveness of the proposed model, a comparative analysis is conducted between individual classifiers, stacking ensemble, and the hybrid model with Bayesian optimization.

Model Interpretation

To enhance model transparency and support explainable AI in education, Shapley Additive exPlanations (SHAP) are employed to interpret the model predictions [27]. SHAP values quantify the contribution of each feature to the prediction outcome, enabling identification of key factors influencing academic risk. SHAP values are based on cooperative game theory and can be expressed as:

$$\phi_i = \sum_{S \subseteq N \setminus \{i\}} \frac{|S|!(|N|-|S|-1)!}{|N|!} [f(S \cup \{i\}) - f(S)] \tag{9}$$

where $\phi_i$represents the contribution of feature i. This interpretability analysis provides valuable insights for educational stakeholders and supports data-driven decision- making [28] for early intervention strategies.

3. RESULTS AND DISCUSSIONS

This section presents the results of the proposed interpretable hybrid ensemble model for early academic risk detection, along with a comprehensive discussion of the findings. The analysis follows a structured data driven approach, beginning with data exploration and preprocessing, followed by model development and evaluation, and concluding with model interpretation using SHAP [29].

Problem Identification and Data Characteristics

The study begins with the identification of academic risk as a critical issue in educational systems, particularly related to student dropout. To address this problem, a dataset consisting of 10,000 student records is analyzed to understand its characteristics. Initial exploratory analysis reveals that the dataset contains a mix of demographic, behavioral, and academic features. The distribution of the target variable indicates the presence of class imbalance as shown in Figure 2, where the number of at-risk students is lower compared to non-risk students at.

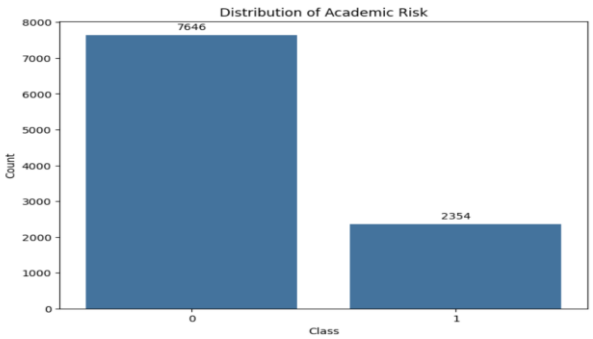

Figure 2. Distribution of academic risk classes

Figure 2 illustrates the distribution of academic risk classes in the dataset. The results show that the majority of students are categorized as non-risk (class 0), with a total of 7,646 instances, while 2,354 instances are classified as at-risk students (class 1). The exploratory analysis indicates observable differences between at-risk and non-risk students across several academic and behavioral variables. These patterns suggest potential associations between academic performance, behavioral factors, and academic risk prediction. This condition justifies the use of ensemble learning to improve robustness. Table 2 presents the characteristics and descriptions of the variables used in this study.

Table 2. Dataset characteristics

No	Feature	Description	Type
1	student_id	unique student identifier	integer
2	age	student age	numerical
3	gender	student gender	categorical
4	family_income	family income level	numerical
5	internet_access	access to internet	categorical
6	study_hours_per_day	daily study duration	numerical
7	attendance_rate	attendance percentage	numerical
8	assignment_delay_days	delay in assignment submission	integer
9	travel_time_minutes	travel time to campus	numerical
10	part_time_job	student employment status	categorical
11	scholarship	scholarship status	categorical
12	stress_index	student stresslevel	numerical
13	gpa	grade point average	numerical
14	semester_gpa	gpa per semester	numerical
15	cgpa	cumulative gpa	numerical
16	semester	current semester	categorical
17	department	academic department	categorical
18	parental_education	parent education level	categorical
19	dropout	academic risk (target variable)	integer

As shown in Table 2, the dataset consists of 10,000 student records with 19 features, including demographic, behavioral, and academic attributes. Consists of 3 integer data types, 7 categorical data types, and 9 numerical data types. However, the proportion of missing data is relatively small. Missing numerical values were handled using median imputation, while categorical missing values were replaced using mode imputation during preprocessing. The target variable, Dropout, represents academic risk and is used for classification. The target variable, Dropout, represents academic risk and is used for classification. The dataset shows moderate class imbalance, with 7,646 non-dropout instances (76.46%) and 2,354 dropout instances (23.54%) out of 10,000 records. To improve minority class representation and reduce class bias, SMOTE was applied during the training phase.

Distribution Numeric Feature

To better understand the characteristics of the dataset, an exploratory visualization of numerical variables was conducted prior to model development as shown in Figure 3a and 3b. This analysis aims to examine the distribution, variability, and potential outliers of the numerical features used in the study.

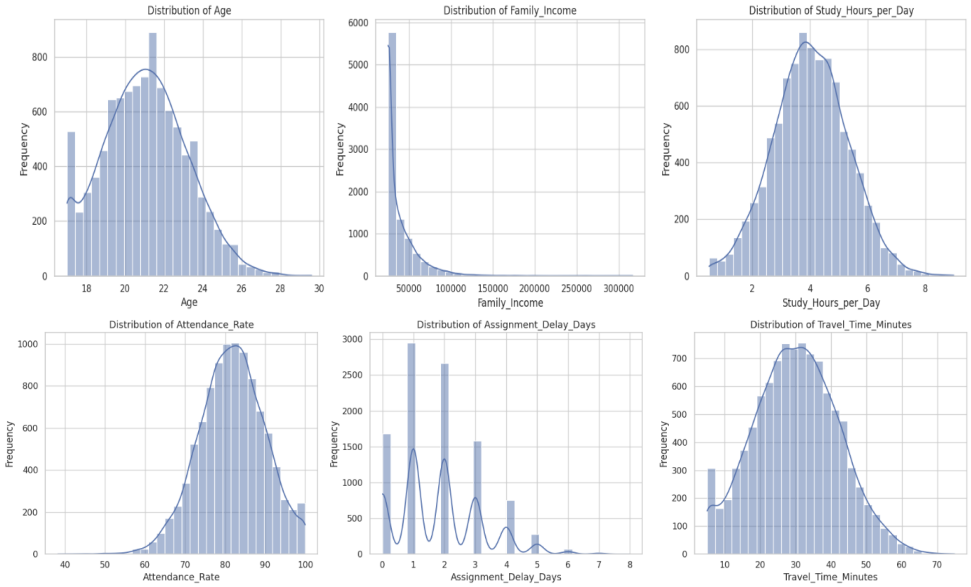

Figure 3a. Distribution of selected numerical variables for exploratory data analysis

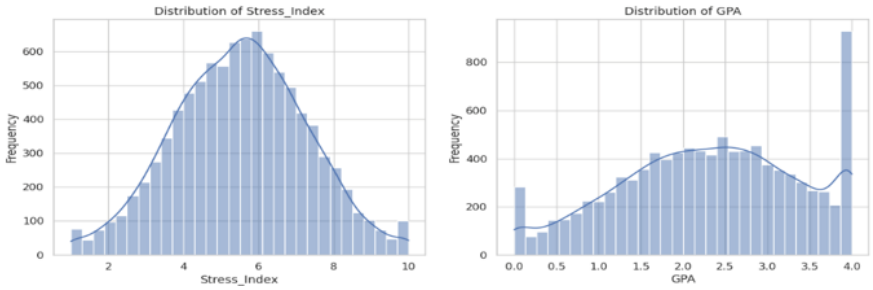

Figure 3b. Distribution of selected numerical variables for exploratory data analysis

Figure 3a and 3b presents the distribution of selected numerical features used in the dataset, including age, family_income, study_hours_per_day, attendance_rate, assignment_delay, travel_time_minutes, stress_index, GPA. The visualization is intended to provide an overview of data distribution, variability, and potential outliers prior to model development and preprocessing. Only numerical variables were visualized in this stage because distribution-based visualization techniques such as histograms are more appropriate for continuous numerical attributes compared to categorical variables. Here is the summary as shown in Table 3 showing the average GPA, attendance rate, study hours, assignment delay, travel time, and stress index for students who did not drop out (0) and those who did (1):

Final Preprocessed Dataset

After completing normalization, cleaning, and feature engineering, the final preprocessed dataset was constructed. It contains both transformed and original features that are ready to be used for model training. Table 5 provides a snapshot of the final dataset after preprocessing.

Table 3. Comparison of behavioral features by dropout status

Dropout	Average GPA	Average Attendance Rate	Average Study Hours per Day	Average Assignment Delay Days	Average Travel Time Minutes	Average Stress Index
0	2.58	82.48	4.07	1.73	29.99	5.25
1	1.43	79.31	3.80	1.99	30.78	6.31

Based on Table 3 students who did not drop out have a notably higher average GPA and a slightly higher average attendance rate compared to those who did drop out. In addition, behavioral factors such as attendance and study hours show similar patterns, where lower attendance rates and fewer study hours increase the probability of dropout. Conversely, the stress index exhibits a positive relationship with academic risk, indicating that higher stress levels contribute significantly to the likelihood of dropout. Overall, these findings suggest that academic risk is influenced by a combination of academic, behavioral, and psychological factors, providing a strong foundation for subsequent predictive modeling. The observed relationships between features

and dropout risk further support the use of ensemble learning methods and interpretability techniques, such as SHAP, to ensure both predictive performance and model transparency.

Model Development Results

The performance results presented in Table 4 represent the average scores obtained from stratified 5-fold cross-validation along with the corresponding standard deviations. The relatively low standard deviation values indicate that the proposed hybrid model demonstrates stable and consistent predictive performance across different validation folds.

Table 4. Performance comparison of model configurations across cross-validation folds

Model Configuration	Description	Accuracy (Mean ± Std)	Precision (Mean ± Std)	Recall (Mean ± Std)	F1-score (Mean ± Std)
Logistic Regression (LR)	Individual baseline classifier	0.82 ± 0.03	0.81 ± 0.03	0.80 ± 0.04	0.80 ± 0.03
K-Nearest Neighbors (KNN)	Individual baseline classifier	0.85 ± 0.02	0.84 ± 0.02	0.83 ± 0.03	0.83 ± 0.02
Naive Bayes (NB)	Individual baseline classifier	0.84 ± 0.03	0.83 ± 0.03	0.82 ± 0.03	0.82 ± 0.03
Decision Tree (DT)	Individual baseline classifier	0.86 ± 0.02	0.85 ± 0.02	0.84 ± 0.03	0.84 ± 0.02
Random Forest (RF)	Individual baseline classifier	0.89 ± 0.02	0.88 ± 0.02	0.87 ± 0.03	0.87 ± 0.02
Stacking Ensemble	Combination of base learners using Logistic Regression as meta-learner	0.91 ± 0.01	0.90 ± 0.01	0.89 ± 0.01	0.89 ± 0.01
Hybrid Model (Stacking + BO)	Stacking ensemble optimized using Bayesian Optimization	0.93 ± 0.01	0.92 ± 0.01	0.91 ± 0.01	0.91 ± 0.01

Table 4 presents the performance comparison of individual classifiers, stacking ensemble, and the proposed hybrid model using stratified cross-validation. The evaluation results are reported as the mean performance together with the corresponding standard deviation across folds to provide a more comprehensive assessment of model stability and generalization capability.

Among the individual baseline classifiers, Random Forest achieved the best performance with an accuracy of 0.89 ± 0.02 and an F1-score of 0.87 ± 0.02. The stacking ensemble further improved predictive performance by combining multiple base learners through a Logistic Regression meta-learner, achieving an accuracy of 0.91 ± 0.01 and an F1-score of 0.89 ± 0.01.

Furthermore, the proposed hybrid model integrating stacking ensemble learning with Bayesian Optimization achieved the highest overall performance, obtaining an accuracy of 0.93 ± 0.01 and an F1-score of 0.91 ± 0.01. The relatively low standard deviation values across folds indicate that the proposed model demonstrates stable and consistent performance during cross-validation.

These findings suggest that the integration of ensemble learning and automated hyperparameter optimization contributes positively to predictive robustness and classification effectiveness in academic risk prediction.

Effect of Bayesian Optimization

The effect of Bayesian Optimization (BO) on model performance was evaluated by comparing the stacking ensemble model before and after hyperparameter tuning. The results indicate that the application of BO provides consistent improvements across evaluation metrics, particularly in terms of F1-score and model stability.

In this study, Bayesian Optimization was applied to optimize several important hyperparameters of the base learning algorithms. For the Random Forest model, the number of estimators (n_estimators) was optimized within the range of 50–300, while the maximum tree depth (max_depth) was explored between 3–20. For the K-Nearest Neighbors (KNN) model, the number of neighbors (n_neighbors) was optimized within the range of 3–15. The Logistic Regression meta-learner optimized the regularization parameter (C) within the range of 0.001–10 using the Expected Improvement (EI) acquisition strategy. As shown in Table 4, the stacking ensemble model improved from an F1-score of 0.89 ± 0.01 before optimization to 0.91 ± 0.01 after applying Bayesian Optimization, representing an improvement of approximately 2%. The optimized model also showed lower variance across cross-validation folds, indicating improved robustness and generalization capability. Overall, Bayesian Optimization enhanced predictive performance, model stability, and classification consistency without requiring exhaustive manual hyperparameter tuning.

Model Evaluation (ROC & Confusion Matrix)

.

The model performance is further evaluated using ROC curves and confusion matrix analysis to assess classification effectiveness (Figure 4). In addition, the confusion matrix (Figure 5) provides detailed insight into classification outcomes, showing that the model correctly identifies most at-risk and non-risk students with a balanced error distribution. This is particularly important for minimizing false negatives in early academic risk detection.

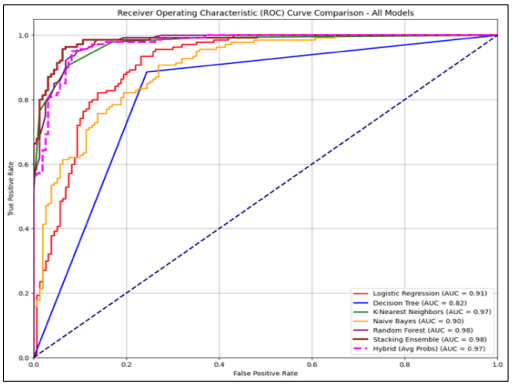

Figure 4. ROC curve comparison all models

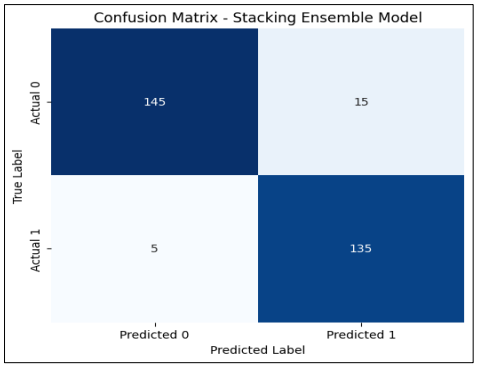

Figure 5. Confusion matrix -stacking ensemble

As result shown in Figure 4 Logistic Regression (AUC = 0.91), this means that there is a 91% chance that the Logistic Regression model will rank a randomly chosen positive instance higher than a randomly chosen negative instance. A high AUC indicates good separability between the positive and negative classes. The Decision Tree model has an AUC of 0.82. This indicates that it performs reasonably well in distinguishing between the two classes, but it's not as strong as some of the other models. K-Nearest-Neighbors (AUC = 0.97), this is a very strong AUC score. It suggests that the KNN model is highly effective at separating positive and negative examples, indicating excellent classification performance. With an AUC of 0.90, the Naive Bayes model also demonstrates strong predictive power, being able to differentiate between the classes with a high degree of accuracy Random Forest (AUC = 0.98) this is the highest AUC among the individual base learners. A score of 0.98 signifies outstanding classification performance, meaning the Random Forest model is extremely good at ranking positive cases above negative ones.

According to Figure 5, the Stacking Ensemble model demonstrates high levels of precision and sensitivity. Out of a total of 300 test samples, the model correctly predicted 280 samples, consisting of 145 True Negatives (TN) and 135 True Positives (TP). The overall accuracy of the model is recorded at 93.3%.

Further analysis of the error distribution reveals that the model produced only 5 False Negatives (FN), indicating an exceptional recall capability in identifying the positive class. Although there were 15 False Positives (FP), the low number of False Negatives suggests that the model is highly reliable in minimizing the risk of missed detections for the target class. The combination of a high AUC value and a matrix distribution dominated by the main diagonal confirms that the Stacking Ensemble is the most stable and reliable model for implementation in this study. The model is very good at minimizing false negatives (risk of missed detection). Highly reliable for early detection.

Model Interpretation Using SHAP

To enhance model transparency, SHAP analysis is employed to interpret the contribution of each feature to the prediction results. The analysis indicates that academic performance variables, such as GPA and semester GPA, are among the most influential factors in predicting academic risk. In addition, behavioral features, including attendance rate, study habits, and stress index, also contribute significantly to the model's

decisions. These findings suggest that academic risk is influenced by a combination of academic achievement and behavioral patterns. The use of SHAP enables the model to provide interpretable insights, supporting its applicability in educational decision-making contexts. The use of SHAP enables the model to provide interpretable insights, supporting its applicability in educational decision-making contexts.

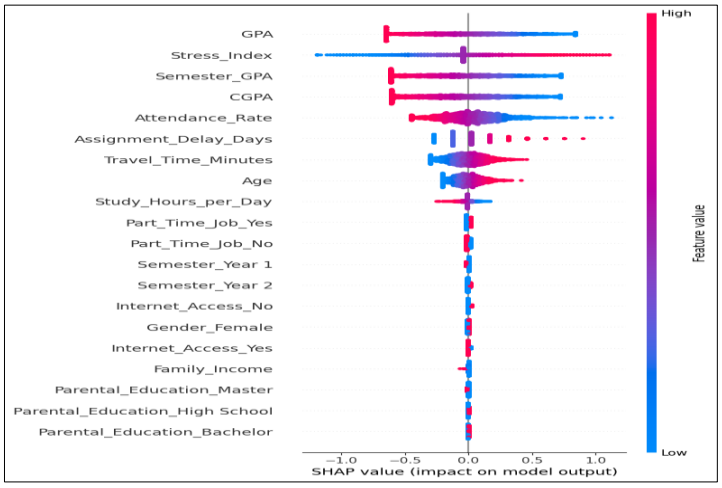

Figure 6. SHAP summary plot for feature importance

Figure 6 presents the SHAP summary plot, which illustrates the contribution of each feature to the model's prediction outcomes. The SHAP analysis shows that academic performance variables, including GPA, CGPA, and Semester GPA, have the largest contribution ranges, with SHAP values approximately spanning from −0.8 to 1.2. In the SHAP framework, positive values indicate stronger contribution toward at-risk predictions, while negative values indicate contribution toward non-risk predictions. Higher GPA-related values are generally associated with negative SHAP values, whereas lower GPA values tend to correspond with positive SHAP values, indicating stronger contribution toward academic risk predictions.

Behavioral variables such as Stress Index, Attendance Rate, Study Hours per Day, and Assignment Delay Days also demonstrate meaningful contributions to prediction outcomes. The Stress Index exhibits a relatively wide SHAP distribution (approximately −1.0 to 1.3), where higher stress values are more frequently associated with positive SHAP values. Conversely, higher attendance rates and longer study hours tend to correspond with negative SHAP values, indicating lower contribution toward at-risk predictions. Assignment Delay Days generally show positive SHAP contributions, with several instances exceeding 0.5 SHAP value, suggesting relatively strong contribution toward academic risk predictions.

In addition, contextual variables such as Travel Time Minutes and Family Income exhibit relatively smaller SHAP ranges centered near zero, indicating more moderate influence on model predictions compared to academic and behavioral variables. Overall, the SHAP analysis demonstrates that academic performance and behavioral variables contribute more prominently to the prediction process. However, these findings should be interpreted as feature contributions to model predictions rather than direct causal relationships with student dropout or academic risk. Table 5 summarizes the feature contribution insights obtained from the SHAP analysis, providing transparency for data-driven academic risk prediction.

Table 5. Feature importance insight based on SHAP analysis

Feature	Impact Direction	Interpretation
GPA / CGPA	Negative	Higher GPA reduces dropout risk
Attendance Rate	Negative	Better attendance lowers risk
Study Hours	Negative	More study time reduces risk
Stress Index	Positive	Higher stress increases dropout risk
Assignment Delay	Positive	Late submissions increase risk
Travel Time	Positive	Longer travel slightly increases risk

Discussion

The findings of this study demonstrate that the proposed hybrid ensemble [30] framework is effective for early academic risk prediction in educational datasets.

The integration of SMOTE improves minority class representation and enhances the model's ability to identify at-risk students more effectively, particularly in moderately imbalanced educational datasets. The stacking ensemble framework enables the model to capture heterogeneous and nonlinear educational patterns more effectively than individual classifiers. Bayesian Optimization improves model robustness and stability by automatically identifying suitable hyperparameter configurations without exhaustive manual search. Previous papers have discussed many ML models but the result black-box models, which are difficult to explain. This study uses SHAP to explain feature contributions, the direction of those contributions, and the interpretability of predictions.

The main contribution of this study lies in the integration of SMOTE, stacking ensemble learning, Bayesian Optimization, and SHAP-based explainability within a unified predictive educational analytics framework. The proposed approach not only improves predictive performance and model stability but also enhances interpretability and minority class detection for early academic risk prediction [31]. Unlike previous studies that focused primarily on either predictive performance or explainability independently, this study integrates imbalance handling, ensemble learning, automated hyperparameter optimization, and explainable AI within a single framework for educational risk prediction.

Despite the promising results, this study has several limitations. First, the research relies on a publicly available dataset that may not fully represent real institutional academic environments. Second, the dataset is cross-sectional and does not capture longitudinal student learning behavior over time. In addition, the study focuses primarily on structured tabular data and does not incorporate learning management system (LMS) interaction data or textual educational records that may provide additional predictive insights.

4. CONCLUSION

This study proposed an interpretable hybrid ensemble framework integrating stacking ensemble learning, Bayesian Optimization, and SHAP-based explainability for early academic risk prediction. The experimental results demonstrate that the proposed approach achieves improved predictive performance and stable classification capability across stratified cross-validation.

The findings indicate that academic performance variables, including GPA, CGPA, and Semester GPA, together with behavioral factors such as attendance rate, study hours, and stress index, contribute substantially to the model's prediction outcomes. In addition, the SHAP-based interpretability analysis provides transparent insights into feature contributions, supporting data-driven academic monitoring and early intervention strategies.

Despite the promising results, this study has several limitations. The research relies on a publicly available dataset that may not fully represent real institutional academic environments, and the dataset does not capture longitudinal student learning behavior over time. Furthermore, the study primarily focuses on structured tabular data without incorporating learning management system (LMS) interaction data or other educational behavioral records.

Future work may extend this research by utilizing real institutional datasets, integrating temporal educational data, and exploring advanced deep learning or multimodal learning approaches to further improve predictive capability and interpretability in academic risk prediction systems.

ACKNOWLEDGEMENTS

I would like to thanks Bina Insani University and NPIC for supporting us, also to the LPPM Unit (Lembaga Penelitian dan Pengabdian kepada Masyarakat / Research and Community Service Institution) for serving as a medium for the transfer of knowledge between the Lecturers and the Speakers/Resource Persons. The researcher used ChatGPT as an AI tool to generate Python commands while conducting exploratory data analysis (EDA) on Google Colab and creating a research flowchart.

CREDIT AUTHORSHIP CONTRIBUTION STATEMENT

Rita Wahyuni Arifin: Conceptualization, Methodology Visualization, Writing Original Draft, Sumardiono: Validation, Methodology, Writing review and Editing. Yeffry Handoko Putra: Investigation, Methodology, Resources, Validation, Formal Analysis. Khoem Sambath: Data Curation, Visualization, Validation.

DECLARATION OF COMPETING INTERESTS

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

DATA AVAILABILITY

Data will be made available on request.

REFERENCES

[1] L.-V. Poellhuber, B. Poellhuber, M. Desmarais, C. Leger, N. Roy, and M. Manh-Chien Vu, "Cluster-Based Performance of Student Dropout Prediction as a Solution for Large Scale Models in a Moodle LMS," in *LAK23: 13th International Learning Analytics and Knowledge Conference*, New York, NY, USA: ACM, Mar. 2023, pp. 592–598. doi: 10.1145/3576050.3576146.

[2] Z. Chi, S. Zhang, and L. Shi, "Analysis and Prediction of MOOC Learners' Dropout Behavior," *Appl. Sci.*, vol. 13, no. 2, p. 1068, Jan. 2023, doi: 10.3390/app13021068.

[3] S. Batool, J. Rashid, M. W. Nisar, J. Kim, H.-Y. Kwon, and A. Hussain, "Educational data mining to predict students' academic performance: A survey study," *Educ. Inf. Technol.*, vol. 28, no. 1, pp. 905–971, Jan. 2023, doi: 10.1007/s10639-022-11152-y.

[4] Y. Chen and L. Zhai, "A comparative study on student performance prediction using machine learning," *Educ. Inf. Technol.*, vol. 28, no. 9, pp. 12039–12057, Sep. 2023, doi: 10.1007/s10639-023-11672-1.

[5] H. Kaur and T. Kaur, "A Prediction Model for Student Academic Performance Using Machine Learning-Based Analytics," 2023, pp. 770–775. doi: 10.1007/978-3-031-18461-1_50.

[6] S. M. Dol and D. P. M. Jawandhiya, "A Review of Data Mining in Education Sector," *J. Eng. Educ. Transform.*, vol. 36, no. S2, pp. 13–22, Jan. 2023, doi: 10.16920/jeet/2023/v36is2/23003.

[7] R. Trakunphutthirak and V. C. S. Lee, "Application of Educational Data Mining Approach for Student Academic Performance Prediction Using Progressive Temporal Data," *J. Educ. Comput. Res.*, vol. 60, no. 3, pp. 742–776, Jun. 2022, doi: 10.1177/07356331211048777.

[8] W. Xiao, P. Ji, and J. Hu, "A survey on educational data mining methods used for predicting students' performance," *Eng. Reports*, vol. 4, no. 5, May 2022, doi: 10.1002/eng2.12482.

[9] O. Iatrellis, I. K. Savvas, P. Fitsilis, and V. C. Gerogiannis, "A two-phase machine learning approach for predicting student outcomes," *Educ. Inf. Technol.*, vol. 26, no. 1, pp. 69–88, Jan. 2021, doi: 10.1007/s10639-020-10260-x.

[10] F. I. Kurniadi, M. A. Dewi, D. F. Murad, S. G. Rabiha, and A. Romli, "An Investigation into Student Performance Prediction using Regularized Logistic Regression," in *2023 IEEE 9th International Conference on Computing, Engineering and Design (ICCED)*, IEEE, Nov. 2023, pp. 1–6. doi: 10.1109/ICCED60214.2023.10425782.

[11] H. R. Mkwazu and C. Yan, "Grade Prediction Method for University Course Selection Based on Decision Tree," in *Proceedings of the 2020 International Conference on Aviation Safety and Information Technology*, New York, NY, USA: ACM, Oct. 2020, pp. 593–599. doi: 10.1145/3434581.3434691.

[12] O. Jiménez, A. Jesús, and L. Wong, "Model for the Prediction of Dropout in Higher Education in Peru applying Machine Learning Algorithms: Random Forest, Decision Tree, Neural Network and Support Vector Machine," in *2023 33rd Conference of Open Innovations Association (FRUCT)*, IEEE, May 2023, pp. 116–124. doi: 10.23919/FRUCT58615.2023.10143068.

[13] L. D. Yulianto, Satriawan Desmana, S. Sutikman, and W. Winarsih, "Binary Classification of Academic Outcomes Using Ensemble Learning and Neural Networks: A Case Study on OULAD," *J. Info Sains Inform. dan Sains*, vol. 15, no. 01 SE-Articles, pp. 151–163, Jul. 2025, [Online]. Available: https://ejournal.seaninstitute.or.id/index.php/InfoSains/article/view/7005

[14] N. A. Butt, Z. Mahmood, K. Shakeel, S. Alfarhood, M. Safran, and I. Ashraf, "Performance Prediction of Students in Higher Education Using Multi-Model Ensemble Approach," *IEEE Access*, vol. 11, pp. 136091–136108, 2023, doi: 10.1109/ACCESS.2023.3336987.

[15] K. M. Hasib, F. Rahman, R. Hasnat, and M. G. R. Alam, "A Machine Learning and Explainable AI Approach for Predicting Secondary School Student Performance," in *2022 IEEE 12th Annual Computing and Communication Workshop and Conference (CCWC)*, IEEE, Jan. 2022, pp. 0399–0405. doi: 10.1109/CCWC54503.2022.9720806.

[16] Y. Jang, S. Choi, H. Jung, and H. Kim, "Practical early prediction of students' performance using machine learning and eXplainable AI," *Educ. Inf. Technol.*, vol. 27, no. 9, pp. 12855–12889, Nov. 2022, doi: 10.1007/s10639-022-11120-6.

[17] E. Tiukhova *et al.*, "Explainable Learning Analytics: Assessing the stability of student success prediction models by means of explainable AI," *Decis. Support Syst.*, vol. 182, p. 114229, Jul. 2024, doi: 10.1016/j.dss.2024.114229.

[18] A. Al Maruf, R. Ara Rumy, R. I. Sony, and Z. Aung, "Predictive Analysis of Bangladeshi Students' Academic Performances Using Ensemble Machine Learning with Explainable AI Techniques," in *2024 27th International Conference on Computer and Information Technology (ICCIT)*, IEEE, Dec. 2024, pp. 1200–1205. doi: 10.1109/ICCIT64611.2024.11021990.

[19] Kevin Fong-Rey Liu and Jia-Shen Chen, "Prediction and assessment of student learning outcomes in calculus a decision support of integrating data mining and Bayesian belief networks," in *2011 3rd International Conference on Computer Research and Development*, IEEE, Mar. 2011, pp. 299–303. doi: 10.1109/ICCRD.2011.5764024.

[20] S. Deepan, D. Rajendran, N. Dhaliwal, K. Praveena, M. Lourens, and A. R. J, "Design of Intelligent Decision Support System With Ensembled Machine Learning to Predict Students' Performance," in *2024 International Conference on Communication, Computer Sciences and Engineering (IC3SE)*, IEEE, May 2024, pp. 1586–1590. doi: 10.1109/IC3SE62002.2024.10593641.

[21] S. TK and Midhunchakkravarthy, "Academic Performance Prediction of At-Risk Students using Machine Learning Techniques," in *2023 3rd International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)*, IEEE, May 2023, pp. 1222–1227. doi: 10.1109/ICACITE57410.2023.10183199.

[22] M. Agarwal and B. B. Agarwal, "Methodological Implementation for Predicting Student Performance Using Data Mining Classifiers and Machine Learning," in *2023 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS)*, IEEE, Nov. 2023, pp. 243–248. doi: 10.1109/ICCCIS60361.2023.10425150.

[23] A. Yaqin, M. Rahardi, and F. F. Abdulloh, "Accuracy Enhancement of Prediction Method using SMOTE for Early Prediction Student's Graduation in XYZ University," *Int. J. Adv. Comput. Sci. Appl.*, vol. 13, no. 6, 2022, doi: 10.14569/IJACSA.2022.0130652.

[24] H. Karalar, C. Kapucu, and H. Gürüler, "Predicting students at risk of academic failure using ensemble model during pandemic in a distance learning system," *Int. J. Educ. Technol. High. Educ.*, vol. 18, no. 1, p. 63, Dec. 2021, doi: 10.1186/s41239-021-00300-y.

[25] C. Blundo, V. Loia, and F. Orciuoli, "A Time-Aware Approach for MOOC Dropout Prediction Based on Rule Induction and Sequential Three-Way Decisions," *IEEE Access*, vol. 11, pp. 113189–113198, 2023, doi: 10.1109/ACCESS.2023.3323202.

[26] K. Fahd, S. Venkatraman, S. J. Miah, and K. Ahmed, "Application of machine learning in higher education to assess student

academic performance, at-risk, and attrition: A meta-analysis of literature," *Educ. Inf. Technol.*, vol. 27, no. 3, pp. 3743–3775, Apr. 2022, doi: 10.1007/s10639-021-10741-7.

[27] J. G. C. Krüger, A. de S. Britto, and J. P. Barddal, "An explainable machine learning approach for student dropout prediction," *Expert Syst. Appl.*, vol. 233, p. 120933, Dec. 2023, doi: 10.1016/j.eswa.2023.120933.

[28] S. Bhoite, C. H. Patil, S. Thatte, V. J. Magar, and P. Nikam, "A Data-Driven Probabilistic Machine Learning Study for Placement Prediction," in *2023 International Conference on Intelligent Data Communication Technologies and Internet of Things (IDCIoT)*, IEEE, Jan. 2023, pp. 402–408. doi: 10.1109/IDCIoT56793.2023.10053523.

[29] M. Nagy and R. Molontay, "Interpretable Dropout Prediction: Towards XAI-Based Personalized Intervention," *Int. J. Artif. Intell. Educ.*, vol. 34, no. 2, pp. 274–300, Jun. 2024, doi: 10.1007/s40593-023-00331-8.

[30] J. Niyogisubizo, L. Liao, E. Nziyumva, E. Murwanashyaka, and P. C. Nshimyumukiza, "Predicting student's dropout in university classes using two-layer ensemble machine learning approach: A novel stacked generalization," *Comput. Educ. Artif. Intell.*, vol. 3, p. 100066, 2022, doi: 10.1016/j.caeai.2022.100066.

[31] B. Prenkaj, P. Velardi, G. Stilo, D. Distante, and S. Faralli, "A Survey of Machine Learning Approaches for Student Dropout Prediction in Online Courses," *ACM Comput. Surv.*, vol. 53, no. 3, pp. 1–34, May 2021, doi: 10.1145/3388792.