LEMBAGA PENELITIAN DAN PENGABDIAN KEPADA MASYARAKAT
UNIVERSITAS BINA SARANA INFORMATIKA
Alamat : Jl. Kramat Raya No. 98, Senen. Jakarta Pusat 10450
Telp. (021) 23231170 Fax (021) 21236158 Ext. 704, 705 e-mail: lppm@bsi.ac.id

SURAT TUGAS

1972/B.01/LPPM-UBSI/X/2025

Tentang

PENELITIAN YANG DIPUBLIKASIKAN DALAM PROSIDING
Periode September 2025 - Februari 2026

Menulis Pada International Conference on Advanced Information Scientific Development (ICAISD) 2025
4-5 November 2025 | Universitas Bina Sarana Informatika

Judul Makalah :

Predicting Academic Performance Through Student Study Habits : A Comparative Analysis of Machine Learning Classification Algorithms

Menimbang :
1. Bahwa perlu di adakan pelaksanaan Kegiatan dalam rangka Seminar.
2. Untuk keperluan tersebut, pada butir 1 (satu) di atas, maka perlu dibentuk kelompok Seminar.

MEMUTUSKAN

Pertama : Menugaskan kepada saudara

1. 199409117 Supriyanta
2. 202204065 Eka Rahmawati
3. 200810954 Candra Agustina

Sebagai Penulis yang mempublikasikan Penelitianya.

Kedua : Mempunyai tugas sbb:
Melaksanakan Tugas yang diberikan dengan penuh rasa tanggung jawab.

Ketiga : Keputusan ini berlaku sejak tanggal ditetapkan, dengan ketentuan apabila dikemudian hari terdapat kekeliruan akan diubah dan diperbaiki sebagaimana mestinya.

Jakarta, 1 Oktober 2025
LPPM Universitas Bina Sarana Informatika
Ketua

Agus Junaidi, M. Kom

Tembusan

- Ka. LPPM Universitas Bina Sarana Informatika
- Arsip
- Ybs

ICAISD 2025
International Conference on Advanced Information Scientific Development

Lembaga Penelitian dan Pengabdian Masyarakat
Universitas Bina Sarana Informatika
Jl. Kramat Raya No. 98, Senen, Jakarta Pusat 10450
Telp. (021) 23231170 ext. 704, Fax. (021) 21236158
Email : lppm@bsi.ac.id

LETTER OF ACCEPTANCE

No: 203/ICAISD-UBSI/IX/2025

Dear Authors:
Supriyanta; Eka Rahmawati and Candra Agustina
Bina Sarana Informatika University, Indonesia

Congratulations!

It is a great pleasure to inform you that, after the peer review process, your paper:

No Paper : 1571183629
Entitled : **Predicting Academic Performance Through Student Study Habits: A Comparative Analysis of Machine Learning Classification Algorithms**

Has been officially accepted for oral presentation at the International Conference on Advanced Information Scientific Development (ICAISD-2025), with the topic *Artificial Intelligence: Advancing Research and Computational Innovations for Global Welfare*, held on November, 4-5th 2025, in Jakarta.

Thank you for submitting your work to the International Conference on Advanced Information Scientific Development (ICAISD-2025). We hope you will submit your paper to the next conference.

Jakarta, September 23rd, 2025
Chair of the 2025 ICAISD Committee

(Dr. Sumanto, M. Kom)

Contact Person:
Sumanto +62 877-7838-9464
Dwi Puji Hastuti +62 877-7838-9464
Email: icaisd@bsi.ac.id

ICAISD 2025
International Conference on Advanced Information Scientific Development

Lembaga Penelitian dan Pengabdian Masyarakat
Universitas Bina Sarana Informatika
Jl. Kramat Raya No. 98, Senen, Jakarta Pusat 10450
Telp. (021) 23231170 ext. 704, Fax. (021) 21236158
Email : lppm@bsi.ac.id

Attachment:

Please read the instructions below carefully regarding the revisions and presentation:

1. The final manuscript should follow the standard IEEE conference templates with Latex or US letter formats that can be found at: http://www.ieee.org/conferences_events/conferences/publishing/templates.html. Papers should be written in English with a minimum length of four (4) and a maximum paper length of six (6) printed pages, including figures, tables, and pictures.
2. The accepted similarity level is a maximum of 20%, which you may check using Turnitin or another similar plagiarism check.
3. The final manuscript PDF file must be validated by IEEE PDF eXpress before submission to ensure it is IEEE Xplore-compliant.
   Go to https://ieee-pdf-express.org/ and use conference ID: 68166X to log in to the website. The deadline for submission at PDF-Express is October 2, 2025.
4. Upload the final manuscript (PDF) previously obtained from PDF Express in EDAS, and the EDAS page of a submitted paper has the same list of authors, title, and abstract (minor wording differences in the abstract are okay).
5. Authors are required to submit their manuscript in a MS Word FILE (.doc / docx), and zip all the pictures or figures in high resolution by October 2, 2025. Create a 10-minute video presentation of their research (compulsory). Both upload at https://bit.ly/2025icaisd-cameraready
6. Presenters are required to complete the registration fee no later than October 4th, 2025. Local/International authors can register themselves and pay the fees in Rupiahs/USD via (G-Forms) https://bit.ly/2025icaisd-registration

   Student Fees: USD 200 / Rp. 2.400.000;
   IEEE Member: USD 200 / Rp. 2.600.000;
   International / Local Presenter: USD 225 / Rp. 2.800.000;
   Second Registered Paper: USD 100 / Rp. 1.500.000;

   Please add three digits at the end of the nominal (098), such as Rp. 2.800.098 (Local Transfer)

   Local/International presenters can directly transfer (Domestic Interbank Transfer for Local) / (International Wire Transfer for International) the fees for registration payment in Rupiahs/USD to the bank account:
   Beneficiary:
   Beneficiary Account Number: 701129970
   BBeneficiary Account/Name: Yayasan Bina Sarana Informatika
   Beneficiary Address: Jl. Dewi Sartika No.77, Cawang, Kramat Jati District, East Jakarta, 13630, Indonesia

Lembaga Penelitian dan Pengabdian Masyarakat
Universitas Bina Sarana Informatika
Jl. Kramat Raya No. 98, Senen, Jakarta Pusat 10450
Telp. (021) 23231170 ext. 704, Fax. (021) 21236158
Email : lppm@bsi.ac.id

Beneficiary Bank:
Bank Name: Bank Permata
Bank Address: M.H. Thamrin St. No.3, Gambir, Gambir District, Jakarta, 10110, Indonesia
Swift Code: BBBAIDJA

Use the following transfer description: ICAISD2025_Paper No EDAS_Name.

7. Submission of IEEE Copyright: The IEEE Copyright should be submitted through EDAS. You should go to the EDAS page of your paper, select the “Copyright Form” button, and follow the instructions.
8. No-show policy: All authors should be aware that the accepted paper must be presented at the conference to be included in the final proceedings.

Predicting Academic Performance Through Student Study Habits: A Comparative Analysis of Machine Learning Classification Algorithms

1st Supriyanta
Faculty of Engineering and Informatics
Bina Sarana Informatika University
Jakarta, Indonesia
supriyanta.spt@bsi.ac.id

2nd Eka Rahmawati
Faculty of Engineering and Informatics
Bina Sarana Informatika University
Jakarta, Indonesia
eka.eat@bsi.ac.id

3rd Candra Agustina
Faculty of Engineering and Informatics
Bina Sarana Informatika University
Jakarta, Indonesia
candra.caa@bsi.ac.id

***Abstract*—Enhancing student performance begins with a clear understanding of the factors that shape academic outcomes. This study introduces a predictive classification model designed to identify students' academic status—whether they pass or fail—by analyzing their study habits alongside key socio-demographic factors. To achieve this, we applied four machine learning algorithms: Support Vector Machine (SVM), Naïve Bayes, logistic regression, and XGBoost. These models were tested using a dataset comprising 500 student records to evaluate their predictive accuracy and overall effectiveness. The original numeric target, final grade, was binarized into pass ($\geq$ 60) and fail ($\leq$ 60) classes. StandardScaler preprocessing was applied, and the evaluation metrics included Accuracy, Precision, Recall, F1-score, AUC, and Confusion Matrix. Among the models tested, Logistic Regression delivered the best performance, achieving an accuracy of 85**

***Index Terms*—machine learning, educational data mining, study habits, student performance prediction, classification**

I. Introduction

Student academic performance serves as a key benchmark for measuring educational success and plays a crucial role in guiding institutional assessments and intervention strategies [1]. In today's educational landscape, where quality assurance and accountability are more critical than ever, gaining a thorough understanding of the factors that shape student learning outcomes is essential. Academic achievement is not the outcome of a single element but rather the outcome of a complex interplay of behavioral and socio-demographic influences, ranging from study habits, attendance patterns, and sleep quality to parental education levels, access to the Internet, and participation in extracurricular activities [2] [3].

Traditional methods of evaluating academic performance often depend on fixed metrics, such as test scores or GPA. While useful, these measures offer limited predictive power and overlook the dynamic behaviors that contribute to a student's overall success. Moreover, conventional statistical techniques frequently struggle to capture the complex underlying connections among the many interconnected factors that shape educational outcomes.

Advances in artificial intelligence technology have opened avenues for machine learning (ML) methods to convert educational data into predictive and actionable insights. ML techniques facilitate the modeling of nonlinear interactions and the automatic identification of patterns that mirror students' potential for academic success or risk. Most prior studies in educational data mining have concentrated on regression models to numerically forecast grades or GPAs. However, the binary classification method, which predicts whether a student will pass or fail, holds greater practical significance as an early warning system that educators and educational stakeholders can use directly.

Unlike previous studies that focused on single models or restricted feature sets, this study provides a comparative evaluation of four widely used algorithms—Logistic Regression, SVM, Naïve Bayes, and XGBoost—while emphasizing interpretability, which is essential for educational decision-making. Furthermore, outcomes-based education (OBE) highlights the importance of aligning learning outcomes with measurable performance indicators. Integrating predictive models with OBE principles supports the early identification of at-risk students and enables timely interventions that improve educational achievement.

The main contribution of this study lies not only in benchmarking widely used algorithms but also in emphasizing interpretability and early warning potential. This is particularly important in education, where decision-makers need transparent insights rather than black-box predictions This study aimed to evaluate the effectiveness of four widely used classification algorithms in the context of academic status prediction: Support Vector Machine (SVM), Naïve Bayes, logistic regression, and XGBoost. A comparative analysis was conducted on the four models using a comprehensive set of evaluation metrics to identify the most optimal approach for

Identify applicable funding agency here. If none, delete this.

classifying student learning outcomes based on learning habits and demographic characteristics.

Previous research has determined the influence of various factors on student achievement. Machine learning-based research has been conducted to determine student learning achievements. Jin used machine learning analysis to determine the influence of parental and school efforts [4]. The findings of this study can help schools emphasize gender-specific interventions to improve the academic achievement of all students. Machine learning has also been applied to predict educational outcomes using a multilayer perceptron neural network combined with a backpropagation algorithm. This approach was employed to classify average grades, academic retention, and graduation rates based on data from 655 students at a private university [5]. The outcomes demonstrated strong predictive accuracy, revealing that learning strategies were the most significant factor in forecasting academic performance, whereas coping strategies played a key role in predicting graduation outcomes. In contrast, background information emerged as the most reliable indicator for identifying students at risk of dropping out of school. A study [6] proposed the use of AutoML to improve the accuracy of predicting student performance before the academic program begins to support early intervention. Predictions by other parameters were also conducted using three parameters: midterm exam scores, department data, and faculty data [7]. This study contributes to advancing learning analytics in higher education by supporting data-driven decision-making, particularly in identifying at-risk students and selecting the most effective machine learning techniques.

Previous research has also explored student performance prediction using models such as Generative Adversarial Networks (GAN) and Support Vector Machines (SVM) [8]. The outcomes of this research can improve the accuracy of student academic performance predictions, especially in the context of learning at school and tutoring at home. The Support Vector Machine (SVM) demonstrated superior performance in predicting student academic outcomes compared to ten other traditional machine learning algorithms. This was particularly evident in tests conducted on a publicly available dataset comprising 1,000 student records [9]. Student academic performance has also been predicted by classifying grades using an optimized Naive Bayes Classifier (NBC). The findings revealed that enhancing the NBC model using a nature-inspired optimization algorithm significantly boosted the accuracy of the predictions, demonstrating the potential of such hybrid approaches in educational data mining [10]. Logistic Regression has also been utilized to help identify the most suitable types of extracurricular activities for first-year students, aiming to support and enhance their academic success [11]. More accurate and interpretative predictions of student academic achievement were also made using the XGBoost algorithm, yielding great accuracy and enabling further analysis for the development of learning implementation strategies [12]. The role of machine learning in predicting student learning success is important for determining the key factors. Other factors that may trigger success also need to be considered

This study aims to evaluate the effectiveness of four widely used classification algorithms in the context of academic status prediction: Support Vector Machine (SVM), Naive Bayes, Logistic Regression, and XGBoost. A comparative analysis was conducted on the four models utilizing a comprehensive set of evaluation metrics to identify the most optimal approach for classifying student learning outcomes based on learning habits and demographic characteristics.

II. Methodology

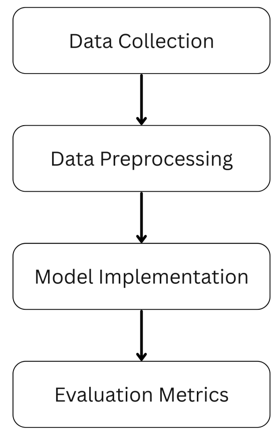

Fig. 1: Research Methodology.

Figure 1 shows the four stages of the research: data collection, data preprocessing, model implementation, and evaluation metrics. The dataset was obtained from Kaggle (https://www.kaggle.com/datasets/prekshad2166/student-study-habits), comprising 500 student records, each featuring normalized variables on study behavior (e.g., weekly study hours, sleep hours), academic attributes (attendance, assignments), and socio-demographics (parental education, internet access, part-time job status). The target variable, `final_grade`, was transformed into a binary class: 1 (pass) if $final\_grade \geq 60$ or 0 (fail) otherwise.

A. Data Collection

The dataset utilized in this study was obtained from Kaggle, a widely recognized open-source data repository commonly used for machine learning research and experiments. The dataset contains 500 student records and includes a variety of features related to study habits and sociodemographic factors that are known to influence academic performance.

The key attributes in the dataset included weekly study hours, sleep duration, class attendance, part-time employment status, Internet access, parental education levels, and involvement in extracurricular activities. The original target variable, the final grade, was transformed into a binary classification label: pass and fail, aligning with the objective of predicting academic status.

This dataset was selected because of its relevance to educational data mining research, well-structured format, and absence of missing values, which ensured a clean and reliable foundation for model training. The data underwent a preprocessing phase before being used to train and evaluate the various machine learning classification models.

B. Data Preprocessing

The preprocessing stage was performed to prepare the data for processing to obtain the maximum outcomes. Several tasks are performed at this stage.

a) Feature scaling was performed utilizing StandardScaler: Feature scaling was performed utilizing StandardScaler, a standardization method that transforms the data distribution so that it has a mean = 0 and standard deviation = 1. Standardizing numerical features is a crucial step to ensure that they share a consistent scale, particularly when applying machine learning algorithms such as SVM or Logistic Regression, which are highly sensitive to variations in feature magnitude.

b) The dataset contained no missing values: Since the dataset contains no missing values, there is no need for additional imputation or data cleaning. This not only streamlines the analysis process but also minimizes the risk of bias or inaccuracies that can arise from handling incomplete data.

c) One-hot encoded categorical variables were already incorporated: To prepare the data for analysis, categorical variables were converted into numerical values using one-hot encoding, which creates individual binary columns for each category. For example, the "parent's education" variable, that includes categories like "great school," "bachelor's degree," and "master's degree," was transformed into three separate binary features. This ensures that machine learning algorithms can process the data effectively, without mistakenly assuming any hierarchy or numerical connection among the categories.

d) The data was split into 80% training and 20% testing: The dataset was divided into two segments, by 80% used for training the model and the remaining 20% set aside for testing. This strategy ensures that the model learns from a substantial portion of the data while its performance is evaluated on data it has not encountered before. This split provides a more reliable and unbiased measure of the model's accuracy and its ability to generalize to real-world scenarios.

C. Models Employed

a) Support Vector Machine (SVM): The Support Vector Machine is a classification algorithm that seeks the optimal hyperplane to separate data into two classes [13]. This study employed a non-linear kernel, allowing the Support Vector Machine (SVM) to project data into a greater-dimensional space using kernel functions such as the Radial Basis Function (RBF). Additionally, probability estimates were incorporated to generate class probabilities, providing a more interpretable understanding of classification outcomes.

b) Naive Bayes: Naïve Bayes is a probabilistic classification technique based on Bayes' theorem that operates under the simplifying assumption that all input features are independent of one another—an approach commonly referred to as the " naïve " assumption [14]. The model assumes that each feature follows a Gaussian or normal distribution, making it well-suited for handling continuous numerical data.

c) Logistic Regression: Logistic Regression is a linear classification algorithm that models the probability of an observation belonging to a class using a logistic (sigmoid) function [15]. This algorithm is equipped with regularization (L1 or L2) to avoid overfitting and improve model generalization. Logistic regression was also performed because of its great interpretability.

d) XGBoost Classifier: XGBoost is a gradient boosting-based ensemble learning algorithm that builds decision trees iteratively to minimize prediction errors. XGBoost is known for its efficiency, speed, and excellent performance in various data mining competitions [16]. This algorithm also has built-in capabilities for handling missing values and regularizing the model.

D. Evaluation Metrics

The model classification performance was evaluated using six commonly used metrics in machine learning to ensure the reliability and accuracy of the predictions, as follows:

a) Accuracy: It measures the proportion of correct predictions to the total number of predictions. This metric provides an overall picture of the frequency with which the model correctly predicts the class.

b) Precision: indicates the proportion of positive predictions that are actually positive. This metric is important in contexts where false positives should be minimized.

c) Recall: It measures how well the model identifies all actual positive cases in the dataset. A high recall value indicates that the model can detect most cases that should be predicted as positive.

d) F1-score: The harmonic mean of precision and recall. This metric provides a balance between the two and is particularly useful when there is an imbalance in the number of data points in each class.

e) AUC: It measures the model's ability to distinguish among classes. The greater the AUC value (maximum 1), the better the model is at separating the positive and negative classes.

f) Confusion Matrix: It provides a visual representation of the number of correct and incorrect predictions for each class in the form of four categories: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). This metric helps to identify common types of errors.

III. Results and Discussion

After all models were developed and tested using the processed dataset, a performance evaluation was conducted to compare the effectiveness of each classification algorithm in predicting students' academic status (pass/fail). The evaluation was based on six main metrics: accuracy, precision, recall, F1-score, area under the curve (AUC), and confusion matrix. These metrics were chosen because they provide a comprehensive overview of the predictive ability, classification accuracy, and sensitivity of the model to imbalanced data. The evaluation outcomes are shown in Table 1.

TABLE I: Evaluation outcomes of classification models

Model	*Accuracy*	*Precision*	*Recall*	*F1-score*	*AUC*
SVM	0.81	0.823	0.929	0.872	0.877
Naïve Bayes	0.75	0.846	0.786	0.815	0.797
Logistic Regression	**0.85**	**0.887**	**0.90**	**0.894**	**0.895**
XGBoost	0.76	0.829	0.829	0.829	0.812

Table 1 reveals the performance evaluation outcomes of the four classification algorithms used, namely, Support Vector Machine (SVM), Naïve Bayes, logistic regression, and XGBoost. Each model was evaluated based on five main metrics: Accuracy, Precision, Recall, F1-score, and AUC.

In general, Logistic Regression revealed the best performance among all models, with the highest accuracy of 0.85, precision of 0.887, recall of 0.900, F1-score of 0.894, and AUC of 0.895. These outcomes suggest that the model is not only accurate but also highly reliable in identifying students who successfully passed. Among the models tested, SVM achieved the highest recall score (0.929), demonstrating its strong ability to correctly detect nearly all students who graduated. However, its precision was slightly lower than that of Logistic Regression, indicating that it produced a greater number of false positives in its predictions.

Meanwhile, Naive Bayes and XGBoost revealed relatively lower performances than the other two models. Naïve Bayes excelled in precision (0.846) but had lower recall and AUC, whereas XGBoost produced balanced metrics but did not stand out, with F1-score and recall of 0.829, respectively. Based on these outcomes, it can be concluded that Logistic Regression is the most optimal model for classifying academic status based on students' learning habits in this study, given the balance between accuracy and interpretability. Further evaluation was conducted using a Confusion Matrix.

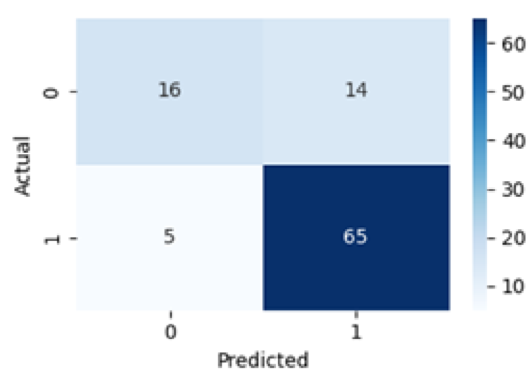

Fig. 2: Confusion Matrix of SVM.

Figure 2 presents the confusion matrix for the Support Vector Machine (SVM) model in predicting students' academic status. This model successfully classified 65 students who passed (true positive) and 16 students who did not pass (true negative). However, there were 14 students who should have failed but were predicted to pass (false positive) and five students who should have passed but were predicted to fail (false negative). These outcomes indicate that SVM performs very well in identifying students who will pass but still produces some errors in classifying failing students.

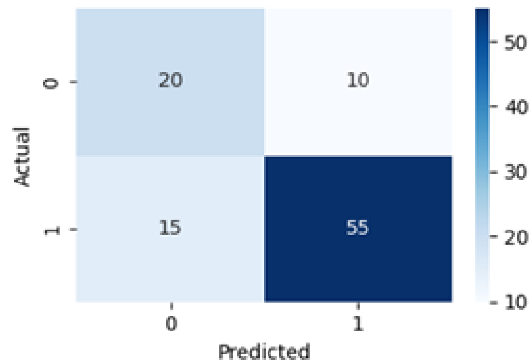

Fig. 3: Confusion Matrix of Naïve Bayes.

Figure 3 shows the confusion matrix for the Naïve Bayes model in classifying student graduation status. This model correctly identified 55 students who passed (true positives) and 20 students who did not pass (true negatives). However, there were 15 students who should have passed but were predicted to fail (false negatives) and 10 students who were predicted to pass but should have failed (false positives). These outcomes reflect the balanced performance of Naïve Bayes, although it still reveals weaknesses in distinguishing students who truly passed from those who did not.

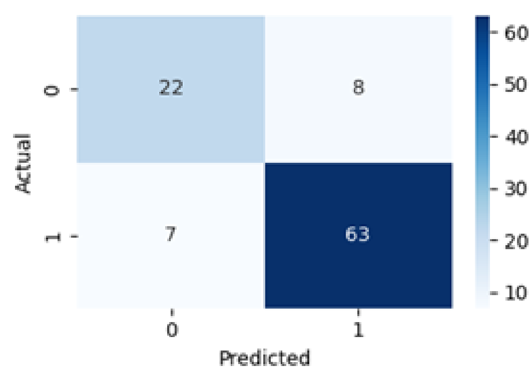

Fig. 4: Confusion Matrix of Logistic Regression.

Figure 4 shows the confusion matrix for the Logistic Regression model in predicting students' academic status. This model successfully classified 63 students who passed (true positives) and 22 students who failed (true negatives). Meanwhile, there were eight false-positive cases, where students who should not have passed were predicted to pass, and seven false-negative cases, where students who should have passed were predicted to fail. Overall, these outcomes indicate that Logistic Regression has balanced and accurate predictive performance in distinguishing between students who passed and those who did not.

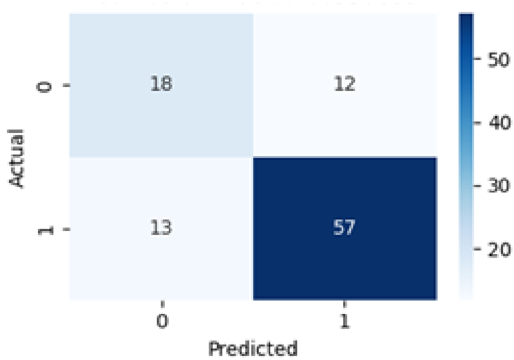

Fig. 5: Confusion Matrix of XGBoost.

Figure 5 shows the confusion matrix for the XGBoost model in predicting student graduation outcomes. The model correctly identified 57 students who passed (true positives) and 18 who failed (true negatives). However, it also misclassified 13 students who passed as having failed (false negatives), and 12 students who failed were incorrectly predicted to have passed (false positives). Although these outcomes reveal that XGBoost delivers a generally strong performance, some misclassifications remain in both categories.

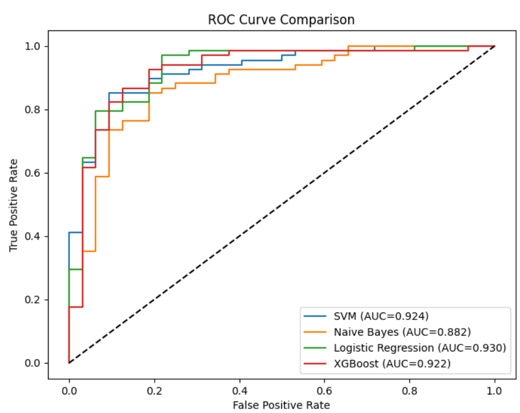

Fig. 6: ROC Curve Comparison.

The ROC curve analysis (Fig. 6) provides a threshold-independent view of the model performance. Logistic Regression achieved the highest AUC, followed by SVM, Naïve Bayes, and XGBoost. This confirms the superiority of Logistic Regression in balancing sensitivity and specificity for early warning applications.

IV. Conclusion

This study assessed the performance of four classification algorithms—Support Vector Machine (SVM), Naïve Bayes, Logistic Regression, and XGBoost—in predicting students' academic status based on their study habits and demographic characteristics. The findings indicate that Logistic Regression delivered the most consistent and well-balanced performance across all evaluation metrics, making it the most effective model for graduation status prediction in this context. While SVM excelled in recall, and both Naïve Bayes and XGBoost demonstrated strengths in precision and metric balance, Logistic Regression emerged as the most reliable choice overall. To build on these insights, future research could incorporate additional psychological factors, such as learning motivation, academic anxiety, and individual learning styles, which play a significant role in student achievement. Moreover, exploring the use of longitudinal datasets, deep learning approaches, and hybrid modeling techniques could further enhance predictive accuracy. Such advancements would contribute to the development of more adaptive and personalized intervention strategies and strengthen decision-making in data-driven educational systems.

References

[1] M. Weatherton and E. E. Schussler, "Success for all? A call to re-examine how student success is defined in higher education," *CBE Life Sci Educ*, vol. 20, no. 1, pp. 1–13, 2021, doi: 10.1187/cbe.20-09-0223.

[2] P. Peng and R. A. Kievit, "The Development of Academic Achievement and Cognitive Abilities: A Bidirectional Perspective," *Child Dev Perspect*, vol. 14, no. 1, pp. 15–20, Mar. 2020, doi: 10.1111/cdep.12352.

[3] H. Abuhassna et al., "Development of a new model on utilizing online learning platforms to improve students' academic achievements and satisfaction," *Int. J. Educ. Technol. High. Educ.*, vol. 17, no. 1, Dec. 2020, doi: 10.1186/s41239-020-00216-z.

[4] X. Jin, "Predicting academic success: machine learning analysis of student, parental, and school efforts," *Asia Pacific Educ. Rev.*, 2023, doi: 10.1007/s12564-023-09915-4.

[5] M. F. Musso, C. F. R. Hernández, and E. C. Cascallar, "Predicting key educational outcomes in academic trajectories: a machine-learning approach," *High Educ.*, vol. 80, no. 5, pp. 875–894, Nov. 2020, doi: 10.1007/s10734-020-00520-7.

[6] H. Zeineddine, U. Braendle, and A. Farah, "Enhancing Prediction of Student Success: Automated Machine Learning Approach," 2020. [Online]. Available: https://www.researchgate.net/publication/346393467

[7] M. Yağcı, "Educational data mining: prediction of students' academic performance using machine learning algorithms," *Smart Learn. Environ.*, vol. 9, no. 1, Dec. 2022, doi: 10.1186/s40561-022-00192-z.

[8] S. Sarwat et al., "Predicting Students' Academic Performance with Conditional Generative Adversarial Network and Deep SVM," *Sensors*, vol. 22, no. 13, Jul. 2022, doi: 10.3390/s22134834.

[9] N. Naicker, T. Adeliyi and J. Wing, "Linear Support Vector Machines for Prediction of Student Performance in School-Based Education," *Math. Probl. Eng.*, vol. 2020, 2020, doi: 10.1155/2020/4761468.

[10] X. Zheng and C. Li, "Predicting Students' Academic Performance Through Machine Learning Classifiers: A Study Employing the Naive Bayes Classifier (NBC)," 2024. [Online]. Available: www.ijacsa.thesai.org

[11] A. M. Kulp, A. B. Pascale, and M. Grandstaff, "Types of Extracurricular Campus Activities and First-Year Students' Academic Success," *J. Coll. Stud. Ret.*, vol. 23, no. 3, pp. 747–767, Nov. 2021, doi: 10.1177/1521025119876249.

[12] S. Wang and B. Luo, "Academic achievement prediction in higher education through interpretable modeling," *PLoS One*, vol. 19, no. 9, Sep. 2024, doi: 10.1371/journal.pone.0309838.

[13] M. Jabardi, "Support Vector Machines: Theory, Algorithms, and Applications," *Infocommunications Journal*, vol. 17, no. 1, pp. 66–75, 2025, doi: 10.36244/ICJ.2025.1.8.

[14] O. Peretz, M. Koren, and O. Koren, "Naive Bayes classifier – An ensemble procedure for recall and precision enrichment," *Eng. Appl. Artif. Intell.*, vol. 136, Oct. 2024, doi: 10.1016/j.engappai.2024.108972.

[15] E. C. Zabor, C. A. Reddy, R. D. Tendulkar, and S. Patil, "Logistic Regression in Clinical Studies," *Int. J. Radiat. Oncol. Biol. Phys.*, vol. 112, no. 2, pp. 271–277, Feb. 2022, doi: 10.1016/j.ijrobp.2021.08.007.
[16] S. Wang and B. Luo, "Academic achievement prediction in higher education through interpretable modeling," *PLoS One*, vol. 19, no. 9, Sep. 2024, doi: 10.1371/journal.pone.0309838.