International Journal Of Health, Engineering And Technology (IJHET) *E-ISSN 2829 - 8683*
Volume 5, Number 1, May 2026, *Page. 780 - 785*
Email : editorijhess@gmail.com

Student Achievement Prediction Comparison Of Naïve Bayes And Svm Using Ai Optimization In Smpn 5

Azzahra Pematasari Suwandi[1)], Sudin Saepudin[2)], Gina Syabani Yuda[3)]*
[1,2,3)] Information system / , Nusa Putra University

*Corresponding Author
Email : azzahra.permatasari_si22@nusaputra.ac.id

Abstract

This study aims to build a predictive model of student learning achievement by comparing the performance of Naive Bayes algorithm and Support Vector Machine (SVM) optimized using Synthetic Minority Over-sampling Technique (SMOTE) and Grid Search. Methods used in this study include data collection, preprocessing, data sharing, application of SMOTE to handle data imbalances, as well as parameter optimization using Grid Search. Next, the model was built using Naive Bayes algorithm and SVM, then evaluated using accuracy metrics to determine the best performance. The results showed that the accuracy of Naive Bayes algorithm before SMOTE application was 71%, but decreased to 56% after SMOTE application. Meanwhile, the SVM algorithm showed stable results with an accuracy of 68% both before and after the application of SMOTE. This shows that optimization techniques do not always improve the performance of the model, depending on the characteristics of the data used. Thus, SVM models are considered more consistent, while Naive Bayes is more sensitive to data changes. The resulting Model can be used as an aid in decision-making in the field of education to more accurately identify the level of student achievement.

Keywords: Keywords: Learning Achievement, Naive Bayes, Support Vector Machine, Prediction, Machine Learning.

INTRODUCTION

Student learning achievement is the main indicator in assessing the success of the educational process.(Zheng & Li, 2024) However, the evaluation process carried out in schools still tends to be conventional and relies on manual analysis, (A. Y. Imran et al., 2025) this makes it less effective at identifying patterns from large and complex academic data. This condition causes limitations in early detection of students who have the potential to experience a decrease in achievement, so that the interventions provided are less than optimal. (Fitriani et al., 2022) Technological developments, especially in the field of Educational Data Mining (EDM), allow the use of educational data to produce more accurate and meaningful information.(Darmawan et al., 2023) Machine learning methods such as Naive Bayes and Support Vector Machine (SVM) are widely used in predictive research because of their ability to classify data.(Riska Rismaya et al., 2025) Naive Bayes has advantages in simplicity and efficiency, while SVM is known to have good performance in handling data with high complexity.(Science et al., 2022) Nevertheless, the performance of both methods is strongly influenced by the data conditions and the selection of parameters used.(Parameswari et al., 2025) Problems that often arise in the modeling process is the imbalance of data and not optimal parameters used, so as to reduce the level of accuracy of the model. Therefore, optimization techniques such as Synthetic Minority Over-sampling Technique (SMOTE)are needed(M. Imran et al., n.d.) to handle data imbalances and Grid Search to find the best parameters. The application of this technique is expected to significantly improve the performance of predictive models.(Fathurahman et al., 2025) Based on these problems, this study aims to build a predictive model of student learning achievement by comparing the performance of the Naive Bayes algorithm and Support Vector Machine optimized using SMOTE and Grid Search techniques. (Jannah et al., 2025)This study is expected to produce a model with the best level of accuracy as a decision support in the educational environment. (A. Y. Imran et al., 2025)

RESEARCH METHODS

This study uses a quantitative approach with an experimental design to build a predictive model of student learning achievement. This study applies two test scenarios, namely a model without the application of the Synthetic Minority Over-sampling Technique (SMOTE) as a baseline and a model with the application of SMOTE as an optimization technique to handle data imbalances(A. Y. Imran et al., 2025). The Data used is student academic data obtained from SMPN 5, totaling 910 data points.(Index @ Smpn5kotasukabumi.Sch.Id, n.d.)The Data consists of several attributes such as assignment grades, daily tests, UAS, attendance, and subject grades, which are as independent variables (features), while the dependent variable is a classification of student learning achievement divided into high, medium, and low categories.(Darmawan et al., 2023) The classification is based on score ranges, where the high category is defined as scores ≥ 85, the medium category ranges from 70–84, and the low category is defined as scores < 70. The data were collected from secondary data available in schools. The Data were collected from secondary data available in schools.(Syaputra et al., 2024)

The process of data processing and model building is carried out using the Python programming language with the help of the Scikit-learn library(Agyemang et al., 2024). The stages of research include data preprocessing , which includes data cleaning, normalization, and labeling of data. The dataset was divided into training and testing sets using an 80:20 ratio.(Darmawan et al., 2023). Tuning parameters is a crucial stage to improve the performance of the model. In this study, Grid Search Cross-Validation (Grid Search CV) technique is applied to systematically search for the optimal combination of parameters.(Ismanto et al., 2023) Grid Search works by testing all possible combinations of parameters that have been determined at a certain range of values, then selecting the set of parameters that produce the highest accuracy value through a cross-validation process. This technique ensures that the model is not overfitting and has good generalizability to new data.(Maulana et al., 2024)**.**

The data splitting process was performed randomly to ensure that both sets represent the overall data distribution. The parameter optimization process using Grid Search was carried out by testing several parameter combinations. For the Support Vector Machine (SVM), the parameters tested include the regularization parameter (C), kernel type (linear and radial basis function), and gamma values. The optimal parameter combination was selected based on the best model performance.(Reza et al., 2025) To overcome the imbalance of data on optimization scenarios, the SMOTE technique. A prediction Model is built using two machine learning algorithms, namely Naive Bayes and Support Vector Machine (SVM), and parameter optimization using Grid Search to obtain the best model configuration.(Fajriah & Kurniawan, 2024).

Model evaluation was done using accuracy metrics by comparing the performance of each algorithm in both test scenarios.(Jannah et al., 2025) The analysis was conducted to determine the most effective and consistent model in predicting student achievement.(Nalatissifa et al., 2021) The research process as a whole is shown in Figure 1.

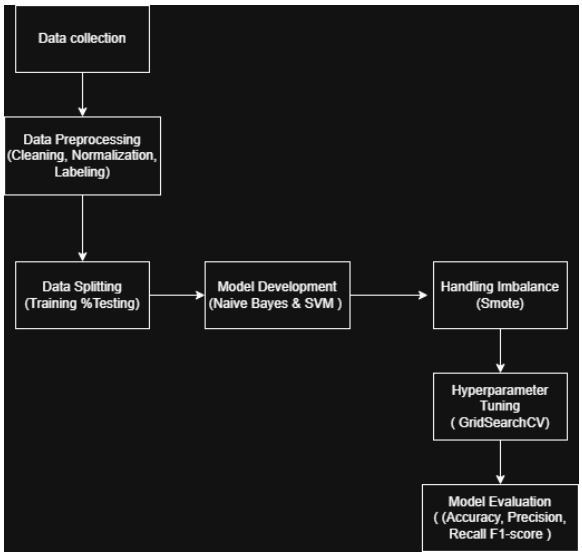

Figure 1. Proposed research methods

RESULTS AND DISCUSSION

The results of model evaluation showed that the application of optimization techniques have different effects on each algorithm. In general, the performance of the model is influenced by the characteristics of the data and the methods used in the learning process.(Cholissodin et al., 2020) Detailed evaluation results can be seen in Table 1.

Tabel 1. algorithm model comparison results

Model	Method	Accuracy	Precision	Recall	F1-Score
Naive Bayes	Without SMOTE	0.71	0.74	0.71	0.72
Naïve Bayes	With SMOTE	0.56	0.70	0.66	0.60
SVM	Without Optimization	0.68	0.73	0.68	0.70
SVM	With Optimization	0.68	0.75	0.68	0.65

In Naive Bayes algorithm, it can be seen that the application of SMOTE actually decreases the level of accuracy compared to models without optimization. This suggests that Naive Bayes is quite sensitive to changes in the distribution of data. The addition of synthetic data through SMOTE can cause a change in the probability pattern used by this algorithm, resulting in a decrease in the performance of the model in performing the classification.

In contrast to Naive Bayes, the Support Vector Machine (SVM) algorithm showed stable performance both before and after the application of SMOTE. This indicates that SVM has a better ability to handle variations in data distribution and is less affected by the addition of synthetic data. This stability is an advantage of SVM in building consistent predictive models.

Based on these results, it can be concluded that the application of the SMOTE technique does not necessarily improve the performance of the model, but depends on the characteristics of the algorithm used. In this study, SVM proved to be more consistent, while Naive Bayes showed better performance in the original data conditions without optimization. Therefore, the selection of the right method is an important factor in building a predictive model of student achievement.

To strengthen the results of the evaluation, a comparative visualization of the performance of the models shown in Figure 2 was performed. This visualization makes it clear that SVM has more consistent performance than Naive Bayes, especially after the application of SMOTE technique.

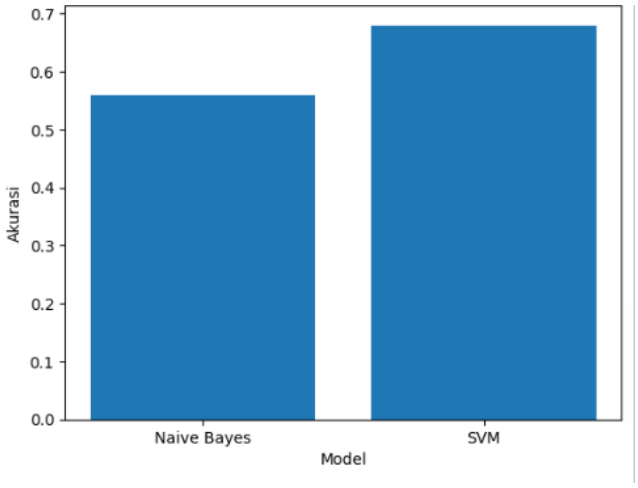

Figure 2. graph of model comparison results

Based on the comparison results, Support Vector Machine (SVM) showed better performance than Naive Bayes, with an accuracy of 0.68, while Naive Bayes 0.56. In addition, the value of precision, recall, and f1-score on SVM is also higher, so the model is more consistent in classifying data. Furthermore, to see the classification performance in more detail, confusion matrix is used in the best model, namely SVM, shown in Figure 3.

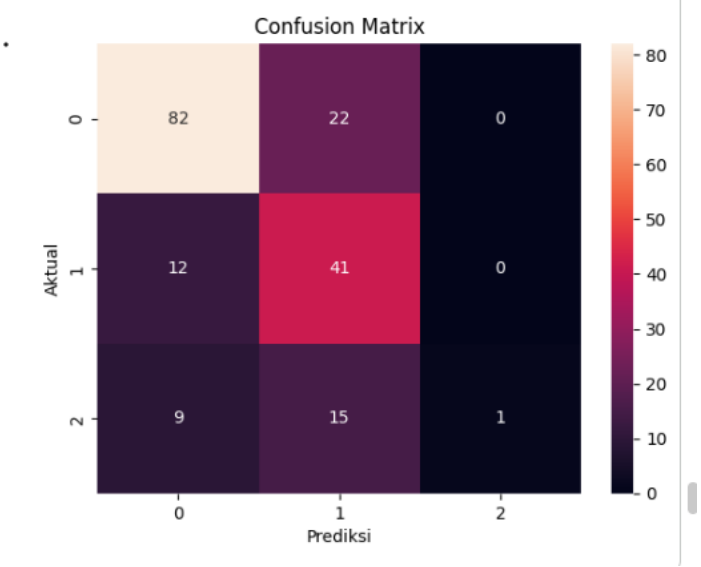

Figure 3. SVM confusion matrix

Based on the confusion matrix, SVM models are able to classify most of the data well in each category. However, there are still some misclassifications in certain classes, which indicate the similarity of characteristics between classes. Nevertheless, the model still showed stable performance in predicting student achievement. The results showed that the application of SMOTE has a different effect on each algorithm. SVM tend to be more stable, whereas Naive Bayes shows better performance on data without optimization.

CONCLUSION

This study aims to build a predictive model of student learning achievement by comparing the naive Bayes algorithm and the support vector machine (SVM) using SMOTE and Grid Search optimization techniques. The results showed that the application of the SMOTE technique gives a different effect on each algorithm. Naive Bayes tends to experience a decrease in performance after the application of SMOTE, while SVM shows more stable performance in the face of data imbalances. Therefore, it can be concluded that the selection of algorithms is very influential on the prediction results, and in this study SVM becomes a more effective and consistent method in predicting student achievement. The resulting Model is expected to be used as a tool in decision- making in the field of Education.

REFERENCES

Agyemang, E. F., Mensah, J. A., Ampomah, O., & Agyekum, L. (2024). Predicting Students' Academic Performance Via Machine Learning Algorithms: An Empirical Review and Practical Application. *Computer Engineering and Intelligent Systems*, *15*(1), 86–102. https://doi.org/10.7176/ceis/15-1-09

Darmawan, A., Yudhisari, I., Anwari, A., & Makruf, M. (2023). Pola Prediksi Kelulusan Siswa Madrasah Aliyah Swasta dengan Support Vector Machine dan Random Forest. *Jurnal Minfo Polgan*, *12*(1), 387–400. https://doi.org/10.33395/jmp.v12i1.12388

Fathurahman, M. H., Firmansyah, H., & Asriyani, W. (2025). *Prediksi Performa Akademik Mahasiswa Menggunakan Support Vector Machine (SVM) pada RapidMiner Prediction of Students ' Academic Performance Using Support Vector Machine (SVM) in RapidMiner*. 19525–19529.

Fitriani, E., Susilo, P. H., & Budi, A. S. (2022). Sistem Cerdas Prediksi Prestasi Belajar Menggunakan Algoritma Naive Bayes di MA Sains Roudlotul Qur'an Lamongan. *Generation Journal*, *6*(1), 58–67. https://doi.org/10.29407/gj.v6i1.16118

Imran, A. Y., Sanjaya, M. R., Bayu Wijaya Putra, & Gabriel Ekoputra Hartono Cahyadi. (2025a). Optimization of Sentiment Analysis on Tokopedia User Reviews Using Gridsearchcv and Smote with Machine Learning Algorithms. *INOVTEK Polbeng - Seri Informatika*, *10*(3), 1634–1644. https://doi.org/10.35314/5ax8km80

Imran, M., Latif, S., Mehmood, D., & Shah, M. S. (n.d.). *Student Academic Performance Prediction using Supervised Learning Techniques*. *d*, 92–104. *index @ smpn5kotasukabumi.sch.id*. (n.d.). Retrieved https://smpn5kotasukabumi.sch.id/

Ismanto, E., Ghani, H. A., Izrin, N., & Saleh, B. (2023). *A comparative study of machine learning algorithms for virtual learning environment performance prediction*. *12*(4), 1677–1686. https://doi.org/10.11591/ijai.v12.i4.pp1677-1686

Jannah, M. D., Fauzi, A., Emilia, C., & Hikmayanti, H. (2025). Klasifikasi Penyakit Hipertensi Menggunakan Support Vector Machine dan Naive Bayes. *Jurnal Algoritma*, *22*(1), 905–913. https://doi.org/10.33364/algoritma/v.22-1.2316

Maulana, B. A., Fahmi, M. J., Imran, A. M., & Hidayati, N. (2024). Analisis Sentimen Terhadap Aplikasi Pluang Menggunakan Algoritma Naive Bayes dan Support Vector Machine (SVM). *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, *4*(2), 375–384. https://doi.org/10.57152/malcom.v4i2.1206

Parameswari, S. D., Lubis, M., Suakanto, S., Ramadhan, Y. Z., Amanah, R. N., & Dila, R. A. (2025). Studi Perbandingan Naïve Bayes dan Support Vector Machine (SVM) dalam Analisis Sentimen Pengguna Metaverse. *Jurnal Teknologi Dan Manajemen Industri Terapan*, *4*(3), 1059–1065. https://doi.org/10.55826/jtmit.v4i3.1122

Peling, I. B. A., Arnawan, I. N., Arthawan, I. P. A., & Janardana, I. G. N. (2017). Implementation of Data Mining To Predict Period of Students Study Using Naive Bayes Algorithm.*International Journal of Engineering and Emerging Technology*, *2*(1), 53.https://doi.org/10.24843/ijeet.2017.v02.i01.p11

Riska Rismaya, Dwi Yuniarto, & David Setiadi. (2025). Penerapan Algoritma Machine Learning dalam Prediksi Prestasi Akademik Mahasiswa. *Router : Jurnal Teknik Informatika Dan Terapan*, *3*(1), 15–23. https://doi.org/10.62951/router.v3i1.389

Syaputra, R., Siswa, T. A. Y., & Pranoto, W. J. (2024). Model Optimasi SVM Dengan PSO-GA dan SMOTE Dalam Menangani High Dimensional dan Imbalance Data Banjir. *Teknika*, *13*(2), 273–282. https://doi.org/10.34148/teknika.v13i2.876

Zheng, X., & Li, C. (2024). *Predicting Students ' Academic Performance Through Machine Learning Classifiers : A Study Employing the Naive Bayes Classifier (NBC)*. *15*(1), 994–1008.