E-ISSN: 2964-6219, P-ISSN: 2964-6200
Volume 5 Number 1 February 2027.
pp: 123-131. DOI: https://doi.org/10.58477/cj.v5i1.595.

Computer Journal

RESEARCH ARTICLE Open Access

Evaluating the Performance of Naïve Bayes and Support Vector Machine Classifiers for Diabetes Classification Using Cross-Validation

Zainudin [1*], **Andri Fahmi** [2]

[1*,2] Department of Informatics Engineering, Universitas Pamulang, South Tangerang City, Banten Province, Indonesia.

Correspondence email:
fahmi.andri@unpam.ac.id.

Article History
Received: 19 October 2026.
Revised: 23 October 2026.
Accepted: 14 November 2026.
Available Online: 5 December 2026
Published: 1 February 2027.

Funding Information
Universitas Pamulang.

Abstract

This study aims to compare the performance of the Naïve Bayes and Support Vector Machine algorithms in diabetes classification using the Pima Indians Diabetes Database. The dataset consists of 768 observations with eight predictor attributes and one target attribute, namely Outcome. The research methodology includes data understanding, preprocessing, model training, and evaluation using Stratified 10-Fold Cross-Validation. Preprocessing encompasses handling zero values on certain attributes, median imputation, and feature standardization for SVM. Evaluation was carried out using accuracy, precision, recall, F1-score, ROC-AUC, and a confusion matrix. The test results showed that SVM obtained an average accuracy of 0.7474, precision of 0.6805, recall of 0.5373, F1-score of 0.5969, and ROC-AUC of 0.8256. Meanwhile, Naïve Bayes obtained an average accuracy of 0.7460, precision of 0.6610, recall of 0.6010, F1-score of 0.6222, and ROC-AUC of 0.8128. Based on the confusion matrix, Naïve Bayes correctly classified 161 of the 268 diabetes samples, whereas SVM correctly classified 144 samples. The results indicate that SVM achieved slightly higher accuracy and ROC-AUC, whereas Naïve Bayes provided better recall and F1-score values, demonstrating a superior capability in recognizing diabetes samples in this test. Consequently, algorithm selection must consider classification objectives and evaluation metrics, particularly when detecting the diabetes class is a priority.

Keywords: Diabetes; Naïve Bayes; Support Vector Machine; Stratified K-Fold Cross-Validation; Classification.

Abstrak

Penelitian ini bertujuan membandingkan kinerja algoritma Naïve Bayes dan Support Vector Machine dalam klasifikasi diabetes menggunakan Pima Indians Diabetes Database. Dataset terdiri atas 768 observasi dengan delapan atribut prediktor dan satu atribut target, yaitu Outcome. Metode penelitian meliputi pemahaman data, preprocessing, pelatihan model, dan evaluasi menggunakan Stratified 10-Fold Cross-Validation. Preprocessing mencakup penanganan nilai nol pada atribut tertentu, imputasi median, serta standardisasi fitur untuk SVM. Evaluasi dilakukan menggunakan akurasi, presisi, recall, F1-score, ROC-AUC, dan confusion matrix. Hasil pengujian menunjukkan bahwa SVM memperoleh rata-rata akurasi sebesar 0,7474, presisi 0,6805, recall 0,5373, F1-score 0,5969, dan ROC-AUC 0,8256. Sementara itu, Naïve Bayes memperoleh rata-rata akurasi sebesar 0,7460, presisi 0,6610, recall 0,6010, F1-score 0,6222, dan ROC-AUC 0,8128. Berdasarkan confusion matrix, Naïve Bayes mengklasifikasikan 161 dari 268 sampel diabetes dengan benar, sedangkan SVM mengklasifikasikan 144 sampel dengan benar. Hasil tersebut menunjukkan bahwa SVM memiliki akurasi dan ROC-AUC yang sedikit lebih tinggi, sedangkan Naïve Bayes memberikan nilai recall dan F1-score yang lebih baik serta lebih mampu mengenali sampel diabetes pada pengujian ini. Dengan demikian, pemilihan algoritma perlu mempertimbangkan tujuan klasifikasi dan metrik evaluasi yang digunakan, terutama ketika kemampuan mendeteksi kelas diabetes menjadi prioritas.

Kata Kunci: Diabetes; Naïve Bayes; Support Vector Machine; Stratified K-Fold Cross-Validation; Klasifikasi.

© The Author(s) 2027. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

1. Introduction

Diabetes mellitus is a chronic metabolic disease characterized by increased blood glucose levels due to impaired insulin production, insulin function, or both. This condition can cause various complications if not managed properly, making early identification an essential aspect of diabetes control efforts. The rapid development of data processing technology provides an opportunity to support this identification process through machine learning. This approach enables systems to learn patterns from medical data to classify individuals into specific categories based on available characteristics (Karo Karo & Hendriyana, 2022). Research has shown that the application of machine learning for diabetes prediction has evolved through the utilization of diverse datasets, algorithms, training strategies, and evaluation metrics. However, these algorithms still require rigorous evaluation so that model performance can be assessed objectively rather than relying solely on specific test results. One of the datasets widely used in diabetes classification research is the Pima Indians Diabetes Database. This dataset consists of 768 samples, eight medical predictor attributes, and one output attribute indicating diabetes status. These characteristics enable the development of a binary classification model to distinguish between diabetic and non-diabetic status. Nevertheless, this dataset presents several challenges, including class imbalance and the presence of zero values in several attributes that are medically implausible to be zero. Gupta and Goel (2023) suggest that handling unknown and isolated values through different preprocessing strategies can significantly affect classification results. Furthermore, Merdin Shamal Salih (2024) applied data processing stages—including outlier handling, imputation, normalization, and feature reduction—before evaluating several algorithms. Their findings indicate that the quality of data processing is a crucial component in developing robust diabetes classification models. Therefore, handling zero values must rely on the characteristics of relevant attributes to ensure valid information is neither omitted nor improperly altered.

Various machine learning algorithms have been applied to address diabetes classification problems, including Naïve Bayes and Support Vector Machines (SVM). Naïve Bayes is a probabilistic classification algorithm based on Bayes' Theorem, which assumes conditional independence between attributes given the class. This algorithm is computationally simple and often serves as a baseline approach in classification tasks. Meanwhile, SVM determines optimal separating boundaries between classes based on data characteristics and can utilize kernel functions to handle non-linear relationship patterns. These differing working principles provide a solid basis for evaluating the comparative capabilities of both algorithms on an identical dataset. Such a comparison is essential because algorithm performance is determined not only by model characteristics but also by data distribution, attribute processing, and applied evaluation procedures (Jannah *et al.*, 2025). Previous research has provided extensive overviews regarding the application of various algorithms for diabetes classification. Research has evaluated models such as Random Forest, Multilayer Perceptron, and Logistic Regression using the Pima Indians Diabetes dataset, reporting that Multilayer Perceptron achieved an accuracy of 86.08% in their experiments. Chang *et al*. (2023) also examined the application of several machine learning algorithms for diabetes classification using the Pima dataset. Furthermore, Tong *et al*. (2024) compared multiple approaches, including Naïve Bayes and SVM, using both the Pima Indians Diabetes and NHANES datasets, where Random Forest attained 79% accuracy for binary classification on the Pima dataset using a 60:40 train-test split. Ahmed *et al*. (2025) similarly compared several algorithms, including Naïve Bayes, in predicting diabetes among a female population using the Pima dataset and reported superior performance using Random Forest.

These collective findings demonstrate that top-performing algorithms can vary depending on experimental design, data characteristics, and processing procedures. Consequently, performance figures derived from studies employing disparate testing schemes cannot be directly compared as measures of universal algorithm excellence. In addition to primary experimental studies, several systematic investigations have evaluated the evolution of machine learning applications in diabetes research. Olusanya *et al*. (2022) conducted a systematic review and meta-analysis of classification models for type 2 diabetes mellitus prediction, revealing performance variations across the examined approaches. Katiyar *et al*. (2024) reviewed machine learning and deep learning methods for diabetes detection, emphasizing datasets, algorithms, evaluation metrics, and research limitations. Meanwhile, Khokhar *et al*. (2025) reviewed 53 articles on diabetes prediction, discussing datasets, training strategies, and validation techniques such as cross-validation. Although these studies underscore the importance of appropriate evaluation procedures, their diverse scopes, datasets, and objectives preclude direct application of their results to determine the optimal algorithm for a specific experimental design. Building upon previous research, the research gap addressed in this study does not stem from a complete lack of prior literature involving Naïve Bayes and SVM for diabetes classification, as both algorithms have been extensively studied. Instead, the primary gap lies in

the necessity for a comparative evaluation utilizing consistent data processing procedures, shared validation folds, and rigorous performance measurement to ensure a fairer comparison between the two algorithms. Prior studies implemented varied experimental designs—such as distinct training-test splits, feature selection methods, and preprocessing combinations—making performance interpretations difficult when evaluated solely on reported accuracy figures.

To bridge this gap, this study evaluates Naïve Bayes and SVM on the identical dataset using Stratified 10-Fold Cross-Validation. Class proportions are maintained across each fold, and both algorithms are evaluated using identical fold divisions. Furthermore, preprocessing steps are embedded within a pipeline so that imputation and standardization parameters are learned exclusively from the training data of each fold. This approach mitigates data leakage and yields a more representative performance estimate than evaluations relying on a single data split. This study aims to evaluate and compare the performance of Naïve Bayes and SVM algorithms in diabetes classification using the Pima Indians Diabetes Database via Stratified 10-Fold Cross-Validation, ensuring consistent data partitioning while preserving class proportions. In the preprocessing stage, zero values on specific medical attributes inconsistent with valid measurements are treated as missing values and imputed using the median of the training data. Imputation and standardization are executed within the pipeline to derive preprocessing parameters strictly from training folds rather than the entire dataset. Standardization supports SVM learning, whereas Naïve Bayes employs preprocessing steps suited to its probabilistic characteristics. Model performance is assessed using accuracy, precision, recall, F1-score, and the area under the receiver operating characteristic curve (ROC-AUC), alongside confusion matrix analyses. These metrics facilitate a comprehensive evaluation of classification capabilities, particularly in balancing overall prediction accuracy against class-specific sample recognition. The primary contribution of this study lies in conducting a controlled comparative evaluation of Naïve Bayes and SVM using a consistent validation design and pipeline-based preprocessing on the Pima Indians Diabetes dataset. The implementation of Stratified 10-Fold Cross-Validation, shared fold partitions, and pipeline data processing establishes a reproducible and controlled experimental framework. This study does not assume absolute algorithmic superiority; rather, it analyzes performance differentials based on empirical test results across established metrics. The findings are intended to provide empirical insights into the performance characteristics of both algorithms and serve as a reference for future medical data classification research.

2. Method

This study employs an experimental approach to compare the performance of the Naïve Bayes algorithm and Support Vector Machines (SVM) in diabetes classification. The dataset utilized is the Pima Indians Diabetes Database obtained from Kaggle. The research stages consist of dataset collection, data understanding, stratified cross-validation, preprocessing, model training and evaluation, and performance comparison. All stages are designed to ensure that both algorithms are evaluated using identical data and testing procedures, as illustrated in Figure 1.

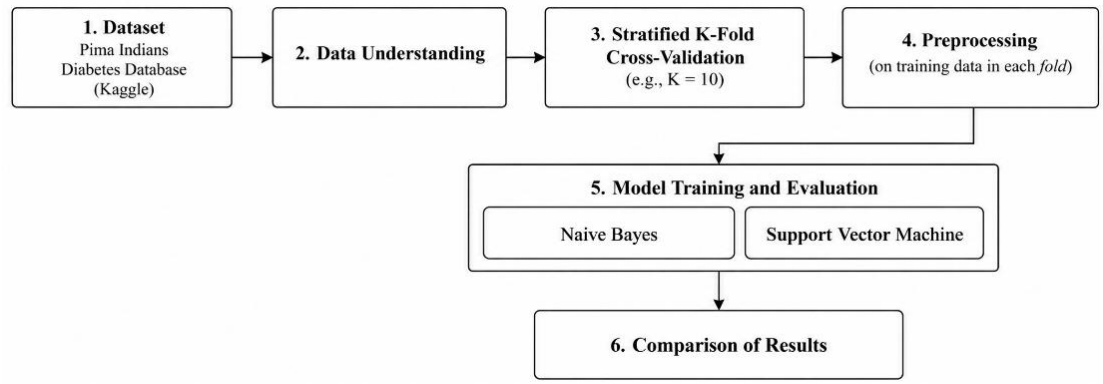


Figure 1. Stages of Research Method

1) Dataset

This study utilizes the Pima Indians Diabetes Database, which comprises 768 observations, eight predictor attributes, and one target attribute. The predictor attributes include the number of pregnancies, glucose levels, blood pressure, skin fold thickness, insulin levels, body mass index (BMI), diabetes pedigree

function, and age. The Outcome attribute serves as the classification target, where a value of 0 indicates a non-diabetic class and a value of 1 indicates a diabetic class. The dataset contains 500 observations in class 0 and 268 observations in class 1.

2) Data Understanding

The data understanding stage aims to determine the characteristics and quality of the dataset prior to the modeling process. This examination encompasses the volume of data, attribute types, missing values, target class distribution, and the presence of zero values across each attribute. Special attention is directed toward the Glucose, BloodPressure, SkinThickness, Insulin, and BMI attributes, as zero values in these fields indicate unavailable measurements. Conversely, a zero value in the Pregnancies attribute represents a clinically valid condition. The findings from this examination govern the treatment of data during the preprocessing stage (Kibria *et al.*, 2022; Olisah *et al.*, 2022).

3) Stratified K-Fold Cross-Validation

Model evaluation is conducted using Stratified 10-Fold Cross-Validation. This method partitions the dataset into ten subsets while preserving the proportion of classes within each section. In each iteration, nine subsets are utilized for model training, while the remaining subset serves as test data. This process is repeated ten times, ensuring that every subset acts as test data once. Data partitioning is executed via controlled randomization to enable experiment replication using identical configurations. Furthermore, stratified cross-validation preserves class proportions across both training and validation sets (Prusty *et al.*, 2022; Szeghalmy & Fazekas, 2023).

4) Preprocessing

Preprocessing is performed independently within each fold following the separation of training and test data. At this stage, zero values occurring in the Glucose, BloodPressure, SkinThickness, Insulin, and BMI attributes are treated as missing values and imputed using the median of each attribute derived exclusively from the training data. The median is selected because it is less sensitive to extreme values compared to the mean. For the Gaussian Naïve Bayes model, feature standardization is omitted in this experiment. The entire imputation and standardization pipeline ensures that processing parameters are not computed using test data. This protocol is critical to prevent data leakage—namely, the infiltration of information from the test set into the training procedure (Butt *et al.*, 2021; Rosenblatt *et al.*, 2024).

5) Model Training and Evaluation

This study compares Gaussian Naïve Bayes and an SVM classifier utilizing the Radial Basis Function (RBF) kernel. Gaussian Naïve Bayes is a probabilistic algorithm that assumes continuous features follow a Gaussian distribution for each class and remain conditionally independent given the class (Shams *et al.*, 2025). Meanwhile, SVM determines optimal separating boundaries between classes and employs the RBF kernel to handle non-linear separation patterns. The SVM parameters are configured with a C value of 1.0 and a gamma setting of scale. Both models are trained on training folds and evaluated on the corresponding test fold, establishing a consistent experimental setup (Olisah *et al.*, 2022).

6) Comparison of Results

The final stage involves comparing the performance evaluations of Naïve Bayes and SVM based on the mean and standard deviation of each metric. The comparative analysis incorporates accuracy, precision, recall, F1-score, ROC-AUC, and the confusion matrix. Accuracy evaluates overall prediction correctness, whereas precision, recall, and F1-score assess classification efficacy for the positive class. Additionally, ROC-AUC provides insight into class separation capability.

3. Results and Discussion

3.1 Class Distribution on Dataset

This study utilized the Pima Indians Diabetes Database obtained from Kaggle, comprising a total of 768 samples consisting of eight predictor attributes and one target attribute, namely Outcome. The eight predictor attributes include Pregnancies, Glucose, BloodPressure, SkinThickness, Insulin, BMI, DiabetesPedigreeFunction, and Age, whereas the Outcome attribute indicates diabetes status, with a value

of 0 representing the non-diabetic category and a value of 1 representing the diabetic category. Based on the class distribution, there were 500 non-diabetic samples (65.10%) and 268 diabetic samples (34.90%). This distribution indicates class imbalance, as detailed in Table 1. Meanwhile, correlation analysis was conducted to establish a preliminary overview of the relationships among variables in the dataset. As illustrated in the correlation heatmap in Figure 2, the Glucose attribute exhibited the highest positive correlation with Outcome at 0.47, followed by BMI at 0.29, Age at 0.24, and Pregnancies at 0.22. Conversely, BloodPressure and SkinThickness each correlated with Outcome by 0.07. Additionally, a correlation of 0.54 was observed between Pregnancies and Age, and 0.44 between SkinThickness and Insulin. These findings indicate varying levels of correlation among attributes, with Glucose demonstrating the strongest linear relationship with the target variable among the analyzed features.

Table 1. Dataset Class Distribution

Category	Quantity	Percentage
No Diabetes	500	65.10%
Diabetes	268	34.90%
Total	768	100.00%

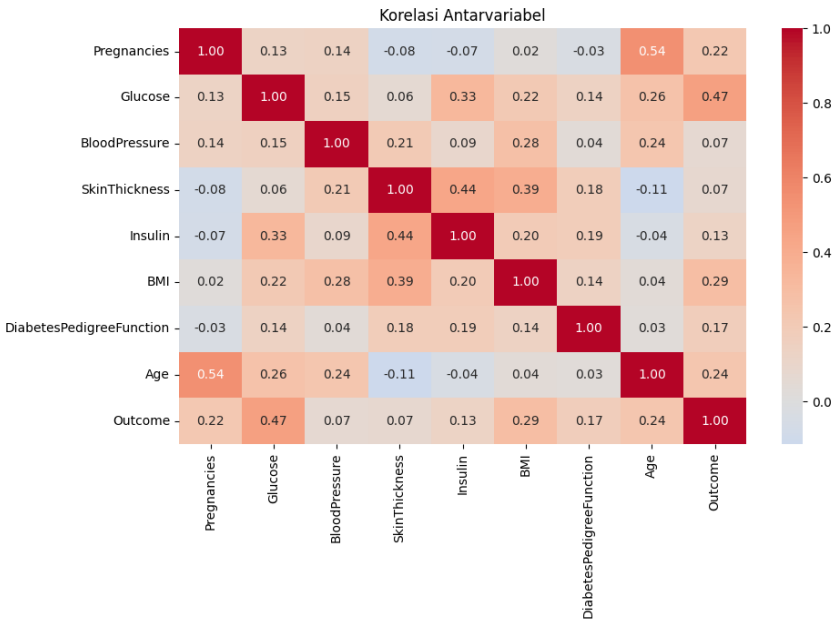


Figure 2. Intervariable Correlation Heatmap

3.2 Model Performance

Evaluation Results Evaluation was conducted using Stratified 10-Fold Cross-Validation to maintain class proportions across each fold. In each iteration, nine folds served as training data and one fold as validation data, repeating until every fold functioned as validation data. Preprocessing was executed within the pipeline for each fold. Zero values in the Glucose, BloodPressure, SkinThickness, Insulin, and BMI attributes were treated as unavailable values and imputed using the median of the training data from the corresponding fold. Feature standardization was applied exclusively to the SVM model, whereas Gaussian Naïve Bayes was implemented without standardization. This configuration was maintained consistently throughout the evaluation to mitigate the risk of information leakage from validation data into the training process.

Table 2. Comparison of Naïve Bayes and SVM Performance Based on Stratified 10-Fold Cross-Validation

Metric	Naïve Bayes (Mean ± SD)	SVM (Mean ± SD)
Accuracy	0.7460 ± 0.0638	0.7474 ± 0.0410
Precision	0.6610 ± 0.1288	0.6805 ± 0.0947
Recall	0.6010 ± 0.1143	0.5373 ± 0.0727
F1-Score	0.6222 ± 0.0928	0.5969 ± 0.0654
ROC-AUC	0.8128 ± 0.0602	0.8256 ± 0.0411

SVM achieved an average accuracy of 74.74%, slightly outperforming Naïve Bayes at 74.60%. Given the marginal difference of 0.14 percentage points, the performance of both models is relatively comparable in terms of accuracy. In terms of precision, SVM obtained a score of 68.05%, whereas Naïve Bayes achieved 66.10%, indicating a higher average precision for SVM. Conversely, for the recall metric, Naïve Bayes attained 60.10%, surpassing SVM's 53.73%. Naïve Bayes also yielded a higher F1-score of 62.22% compared to SVM's 59.69%. These results demonstrate that Naïve Bayes provided a better balance between precision and recall based on the F1-score and exhibited a superior ability to recognize positive samples based on average recall values. Meanwhile, SVM recorded a higher ROC-AUC value of 82.56% compared to Naïve Bayes' 81.28%, reflecting superior class discrimination capability during testing. Furthermore, SVM demonstrated a lower standard deviation in accuracy (0.0410) compared to Naïve Bayes (0.0638).

3.3 Confusion Matrix Analysis

Confusion matrix analysis was performed to evaluate the capability of the Naïve Bayes and SVM models in classifying diabetic and non-diabetic samples. This analysis identifies the exact count of correct predictions and misclassifications for each model to enable a detailed performance comparison. A summary of the confusion matrix results for both models is presented in Table 3, while visual representations are provided in Figure 3.

Table 3. Confusion Matrix Summary for Naïve Bayes and SVM Models

Model	True Negative	False Positive	False Negative	True Positive
Naïve Bayes	412	88	107	161
SVM	430	70	124	144

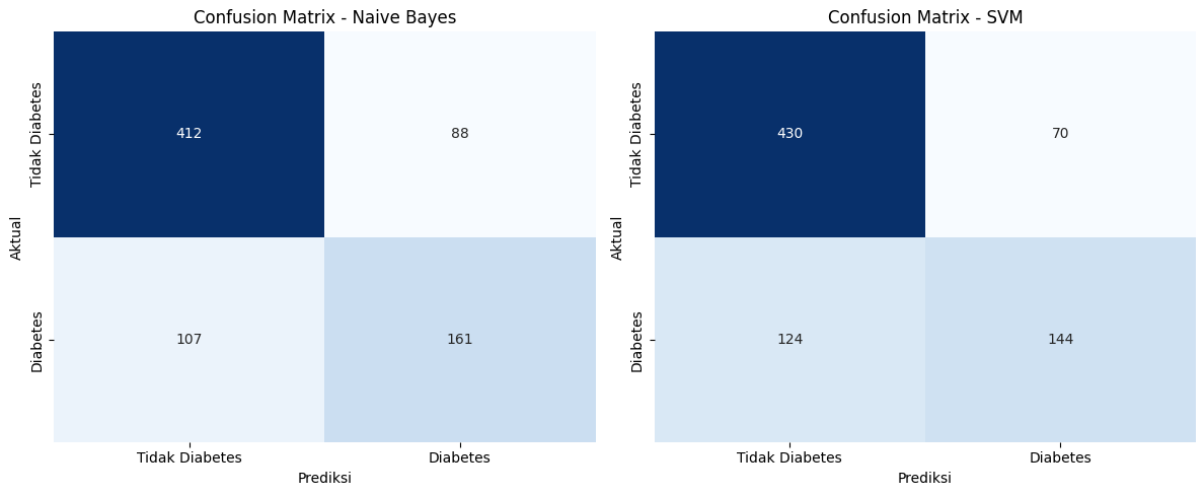


Figure 3. Confusion Matrix of Naïve Bayes and SVM

As shown in Table 3 and Figure 3, the Naïve Bayes model successfully classified 412 non-diabetic samples and 161 diabetic samples correctly. However, it generated 88 misclassifications where non-diabetic samples were incorrectly predicted as diabetic, and 107 misclassifications where diabetic samples were incorrectly predicted as non-diabetic. In contrast, the SVM model correctly classified 430 non-diabetic samples and 144 diabetic samples, producing 70 false positive errors and 124 false negative errors. These results indicate that SVM achieved a higher number of correct predictions for the non-diabetic class, as evidenced by its true negative count of 430 compared to 412 for Naïve Bayes, alongside a lower false positive count (70 versus 88). Conversely, Naïve Bayes demonstrated superior capability in identifying diabetes samples, achieving 161 correct predictions compared to 144 for SVM, as well as a lower false negative count (107 versus 124). In diabetes classification, minimizing false negatives is critical because false negative instances represent individuals who possess the condition but are incorrectly predicted as non-diabetic, potentially preventing patients from receiving timely further medical examination. Based on these confusion matrix outcomes, Naïve Bayes exhibited a stronger ability to recognize diabetes samples, whereas SVM excelled in classifying non-diabetic samples. These findings confirm that model selection should not rely solely on accuracy but must also account for misclassification types and specific clinical application objectives.

3.4 Comparison of the Two Algorithms

Performance evaluation was further synthesized to compare Naïve Bayes and SVM across five evaluation metrics: accuracy, precision, recall, F1-score, and ROC-AUC, as illustrated in the performance comparison graph in Figure 4.

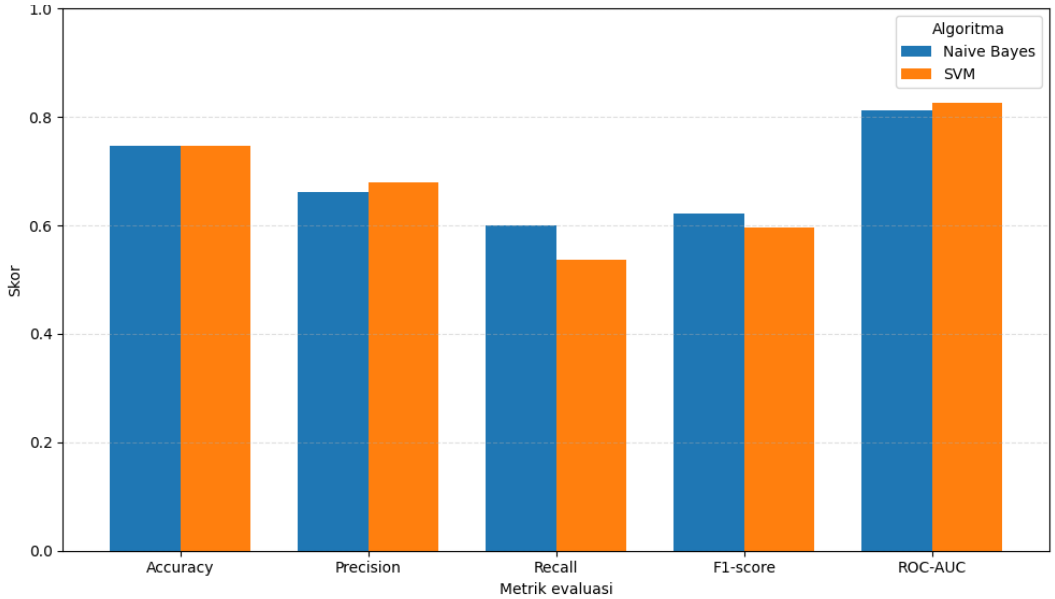


Figure 4. Naïve Bayes and SVM Performance Comparison

As depicted in Figure 4, Naïve Bayes and SVM demonstrated comparable performance in terms of accuracy, recording 0.7460 and 0.7474 respectively, with a minimal difference of 0.14 percentage points. Regarding precision, SVM achieved 0.6805, outperforming Naïve Bayes (0.6610). In the recall metric, Naïve Bayes achieved 0.6010, whereas SVM recorded 0.5373, highlighting the superior capability of Naïve Bayes in recognizing diabetes samples. Naïve Bayes maintained this advantage in the F1-score metric with 0.6222 compared to SVM's 0.5969, confirming a more balanced trade-off between precision and recall. Meanwhile, SVM attained a slightly higher ROC-AUC value of 0.8256 compared to Naïve Bayes' 0.8128, indicating enhanced class discrimination capability under the ROC curve.

4. Conclusion

This study compared the performance of Gaussian Naïve Bayes and Support Vector Machine with the RBF kernel in diabetes classification using the Pima Indians Diabetes Database through the Stratified 10-Fold Cross-Validation method. The test results showed that SVM obtained an average accuracy of 74.74%, a precision of 68.05%, and a ROC-AUC of 82.56%. Meanwhile, Naïve Bayes obtained an average accuracy of 74.60%, a precision of 66.10%, and a ROC-AUC of 81.28%. In the recall and F1-score metrics, Naïve Bayes obtained higher average scores of 60.10% and 62.22%, respectively, compared to SVM's 53.73% and 59.69%. The evaluation of the confusion matrix also showed that Naïve Bayes produced 161 correct predictions in the diabetes class with 107 false negatives, while SVM produced 144 true predictions and 124 false negatives. The study demonstrated distinct performance characteristics between the two algorithms under the evaluated experimental configuration. SVM achieved higher average scores in accuracy, precision, and ROC-AUC, whereas Naïve Bayes achieved higher scores in recall and F1-score. However, these performance differences remain descriptive and cannot be declared as statistically significant advantages before formal statistical significance testing is conducted. Furthermore, the findings are limited to a single dataset comprising 768 samples from a specific female population of Pima heritage, meaning generalizations to other populations should be approached with caution. Future research should evaluate the effects of class weighting and decision threshold tuning, perform formal statistical tests, and incorporate external datasets to assess the cross-population consistency and robustness of both algorithms.

References

Ahmed, A., Khan, J., Arsalan, M., Ahmed, K., Shahat, A. A., Alhalmi, A., & Naaz, S. (2025). Machine learning algorithm-based prediction of diabetes among female population using PIMA dataset. *Healthcare*, *13*(1), Article 37. https://doi.org/10.3390/healthcare13010037

Butt, U. M., Letchmunan, S., Ali, M., Hassan, F. H., Baqir, A., & Sherazi, H. H. R. (2021). Machine learning based diabetes classification and prediction for healthcare applications. *Journal of Healthcare Engineering*, *2021*, Article 9930985. https://doi.org/10.1155/2021/9930985

Chang, V., Bailey, J., Xu, Q. A., & Sun, Z. (2023). Pima Indians diabetes mellitus classification based on machine learning (ML) algorithms. *Neural Computing and Applications*, *35*(22), 16157–16173. https://doi.org/10.1007/s00521-022-07049-z

Gupta, S. C., & Goel, N. (2023). Predictive modeling and analytics for diabetes using hyperparameter tuned machine learning techniques. *Procedia Computer Science*, *218*, 1257–1269. https://doi.org/10.1016/j.procs.2023.01.104

Jannah, M. D., Fauzi, A., Emilia, C., & Hikmayanti, H. (2025). Klasifikasi penyakit hipertensi menggunakan Support Vector Machine dan Naïve Bayes. *Jurnal Algoritma*, *22*(1). https://doi.org/10.33364/algoritma/v.22-1.2316

Karo Karo, I. M., & Hendriyana, H. (2022). Klasifikasi penderita diabetes menggunakan algoritma machine learning dan Z-Score. *Jurnal Teknologi Terpadu*, *8*(2), 105–112. https://doi.org/10.54914/jtt.v8i2.564

Katiyar, N., Thakur, H. K., & Ghatak, A. (2024). Recent advancements using machine learning & deep learning approaches for diabetes detection: A systematic review. *E-Prime - Advances in Electrical Engineering, Electronics and Energy*, *9*, Article 100661. https://doi.org/10.1016/j.prime.2024.100661

Khokhar, P. B., Gravino, C., & Palomba, F. (2025). Advances in artificial intelligence for diabetes prediction: Insights from a systematic literature review. *Artificial Intelligence in Medicine*, *164*, Article 103132. https://doi.org/10.1016/j.artmed.2025.103132

Kibria, H. B., Nahiduzzaman, M., Goni, M. O. F., Ahsan, M., & Haider, J. (2022). An ensemble approach for the prediction of diabetes mellitus using a soft voting classifier with an explainable AI. *Sensors*, *22*(19), Article 7268. https://doi.org/10.3390/s22197268

Merdin Shamal Salih. (2024). Diabetic prediction based on machine learning using PIMA Indian dataset. *Communications on Applied Nonlinear Analysis*, *31*(5s), 138–156. https://doi.org/10.52783/cana.v31.1008

Olisah, C. C., Smith, L., & Smith, M. (2022). Diabetes mellitus prediction and diagnosis from a data preprocessing and machine learning perspective. *Computer Methods and Programs in Biomedicine*, *220*, Article 106773. https://doi.org/10.1016/j.cmpb.2022.106773

Olusanya, M. O., Ogunsakin, R. E., Ghai, M., & Adeleke, M. A. (2022). Accuracy of machine learning classification models for the prediction of type 2 diabetes mellitus: A systematic survey and meta-analysis approach. *International Journal of Environmental Research and Public Health*, *19*(21), Article 14280. https://doi.org/10.3390/ijerph192114280

Prusty, S., Patnaik, S., & Dash, S. K. (2022). SKCV: Stratified K-fold cross-validation on ML classifiers for predicting cervical cancer. *Frontiers in Nanotechnology*, *4*, Article 972421. https://doi.org/10.3389/fnano.2022.972421

Rosenblatt, M., Tejavibulya, L., Jiang, R., Noble, S., & Scheinost, D. (2024). Data leakage inflates prediction performance in connectome-based machine learning models. *Nature Communications*, *15*(1), Article 1928. https://doi.org/10.1038/s41467-024-46150-w

Shams, M. Y., Tarek, Z., & Elshewey, A. M. (2025). A novel RFE-GRU model for diabetes classification using PIMA Indian dataset. *Scientific Reports*, *15*(1), Article 1120. https://doi.org/10.1038/s41598-024-82420-9

Szeghalmy, S., & Fazekas, A. (2023). A comparative study of the use of stratified cross-validation and distribution-balanced stratified cross-validation in imbalanced learning. *Sensors*, *23*(4), Article 2333. https://doi.org/10.3390/s23042333

Tong, H. L., Ng, H., & Arul Ananthan, H. (2024). Predicting diabetes mellitus with machine learning techniques. *Journal of Engineering Technology and Applied Physics*, *6*(1), 91–99. https://doi.org/10.33093/jetap.2024.6.1.12