KOMTEKINFO JOURNAL
LPPM Universitas Putra Indonesia YPTK Padang
Jl. Raya Lubuk Begalung, Padang, West Sumatera, Indonesia, Zip Code: 25221
Volume: 13, Issue: 2, Page: 56 - 65, June 30th, 2026, e-ISSN: 2502-8758
Available online at website: https://jkomtekinfo.org/ojs/index.php/komtekinfo

Analysis of Strategies for Improving Learning Quality Based on Naive Bayes and Support Vector Machines

Fadhila Putri Sani[1]✉, Syafri Arlis[2], Agung Rahmadhanu[3]
[1,2,3] Department of Informatics Engineering (Master's Program), Faculty of Computer Science, Universitas Putra Indonesia "YPTK", Padang, Indonesia, 25224
fadhilaputrisani@gmail.com [1] , syafri_arlis@upiyptk.ac.id [2]., agung_ramadhanu@upiyptk.ac [3]

Abstract

Islamic educational institutions, particularly Islamic boarding schools, face increasing challenges in improving the quality of learning. The learning quality in Islamic boarding schools should be analyzed in depth to support effective improvement strategies. Based on this background, this study aims to classify strategies for enhancing learning quality using the Naïve Bayes and Support Vector Machine (SVM) algorithms. Naïve Bayes with a Gaussian distribution is widely recognized for its simplicity and accuracy in data classification. Meanwhile, Support Vector Machines (SVM) with a linear kernel are effective for linearly separable and high-dimensional data, enabling stable and efficient modeling in the context of data-driven analysis of learning quality in formal education. The data were collected through questionnaires distributed to 100 female students and 100 teachers. The variables examined include teacher competence, infrastructure, school management, student participation, and learning quality level. The analysis results indicate that the Naïve Bayes algorithm achieved superior performance with an accuracy of 90%, precision of 95.65%, recall of 83.33%, and an F1-score of 86.56%. In contrast, the Support Vector Machine (SVM) obtained an accuracy of 80%, precision of 58.97%, recall of 66.67%, and an F1-score of 62.32%.These findings demonstrate that Naïve Bayes provides more stable classification performance across all learning quality categories. Conversely, the Support Vector Machine (SVM) shows less optimal performance in the low-quality class due to the limited number of data samples. This study contributes effectively to the classification of learning quality levels in Islamic boarding schools.

Keywords: Learning quality; Islamic boarding school; Data mining; Naïve Bayes; Support Vector Machine (SVM).

1. Introduction

The development of information technology has brought significant changes to modern educational practices at various levels and institutional contexts. Educational institutions now have the ability to manage, store, and analyze large amounts of data, which were previously difficult to manage manually [1]. This data processing allows for more systematic, objective, and accountable evaluation of learning based on empirical evidence. Furthermore, advances in information technology open up broad opportunities for the use of artificial intelligence as a foundation for data-driven decision-making and improving the quality of learning [2].

As Islamic educational institutions, Islamic boarding schools also face increasingly complex challenges in sustainably improving the quality of learning. Traditional learning methods used are often deemed inadequate to meet the needs and characteristics of today's digital generation [3]. According to Sobari et al. (2025), the quality of learning in Islamic boarding schools is not solely determined by formal academic aspects. Religious activities such as memorization, discipline, and student attendance also contribute significantly to learning success [4].

In the context of Islamic boarding schools, learning quality is not solely determined by formal academic aspects. It is also influenced by non-academic variables such as memorization, discipline, and attendance, which contribute significantly to students' success. Therefore, evaluation models need to accommodate these unique characteristics [5]. Several studies have shown that technology-based learning innovations in boarding school environments can improve learning outcomes compared to traditional approaches, because the use of digital strategies (e.g., guided discovery with media) strengthens student engagement and facilitates understanding of complex concepts through more structured learning activities. [6].

On a national scale, research on learning loss detection in Islamic Religious Education also shows that certain boosting models can be more stable on large datasets than conventional methods. Therefore, learning quality evaluations should ideally consider data characteristics (size, features, class imbalance) before establishing the final algorithm [7].

Submitted: 12-03-2026 | Revised: 15-04-2026 | Accepted: 11-05-2026 | Published: 30-06-2026

Comparative research in the educational context shows that algorithm selection significantly impacts prediction quality. For example, when multiple models are tested to predict graduation/learning success, performance across algorithms can vary, requiring institutions to empirically test models and not rely on a single algorithm as the sole standard [8].

This situation drives the need for a data-driven approach to understand learning dynamics more objectively and measurably. One rapidly developing approach is Educational Data Mining (EDM), which utilizes machine learning algorithms to uncover hidden patterns in educational data. [9] Through Educational Data Mining (EDM), educational institutions can evaluate learning outcomes and detect the risk of academic failure early. This approach also enables the formulation of more targeted learning strategies based on student characteristics. [10]

The application of Educational Data Mining (EDM) enables the simultaneous and integrated analysis of both academic and non-academic data. Through systematic analysis, educational institutions can obtain a comprehensive understanding of the factors that influence learning quality [11]. The ability of Educational Data Mining (EDM) to identify hidden patterns makes it an effective tool for supporting data-driven decision making. This approach assists educational institutions in designing more effective and sustainable interventions [12].

Among the various algorithms used in Educational Data Mining (EDM), Naïve Bayes and Support Vector Machines (SVM) are among the most widely applied in educational research. The popularity of these algorithms is driven by their effectiveness in handling complex and diverse educational data [13]. Naïve Bayes is a probabilistic classification algorithm based on Bayes' Theorem, which assumes independence among features. This assumption allows each attribute to be considered as contributing independently to the resulting predicted class [14].

Naïve Bayes is well known for its advantages in computational efficiency, ease of implementation, and stable performance on small to medium-sized datasets [15]. Research on student admission selection in Islamic boarding schools has shown that this algorithm can achieve an accuracy rate of 88.89%. An Area Under the Curve (AUC) value of 1.000 indicates excellent classification performance. Similar results have also been reported in text-based sentiment analysis of educational services in modern Islamic boarding schools [16].

In addition to Naïve Bayes, Support Vector Machines (SVM) are machine learning algorithms that operate by constructing an optimal hyperplane to separate data into specific classes. The primary advantage of Support Vector Machines (SVM) lies in their ability to handle high-dimensional data and complex nonlinear patterns [17]. Numerous studies have shown that Support Vector Machines (SVM) can produce more accurate predictions of academic performance compared to traditional methods. Their application in state Islamic higher education institutions has achieved accuracy rates above 90%, with AUC values exceeding 0.9.

Based on this background, this study aims to systematically design strategies for improving learning quality at Syafa'aturrasul Islamic Boarding School. The proposed approach combines the Naïve Bayes and Support Vector Machines (SVM) algorithms to obtain a more balanced analysis. The application of these two algorithms is expected to produce a more measurable and objective learning evaluation model. The results of this evaluation can serve as a foundation for pesantren policymakers in formulating data-driven strategies to enhance learning quality.

2. Methods

This study employs a quantitative approach with a systematic research design, enabling a clear and structured implementation of the research process. The research framework used in this study is presented in the Figure 1.

Figure 1 illustrates the research framework, which depicts the stages carried out in this study. These stages are further divided into sub-stages to facilitate the research process. The following section provides a detailed explanation of each stage undertaken in this study. The focus of this study is limited to junior and senior high school–level students at Syafa'aturrasul Islamic Boarding School, particularly female students, as well as teachers. The research population consists of approximately 100 female students and 100 active teachers. The primary data were obtained through the distribution of questionnaires to the respondents. The questionnaire items were designed based on relevant learning indicators within the pesantren environment, including: (X1) Teacher Competence, (X2) Facilities and Infrastructure, (X3) School Management, and (X4) Student Engagement. In addition, learning quality (Y) was measured through a combination of academic data and an overall assessment of students' learning outcomes.

The four input variables (X1–X4) were identified as the main factors influencing learning quality in the pesantren. Meanwhile, the output variable (Y) represents learning quality categories labeled as "high," "medium," and "low," based on the overall scores obtained. These categories were used as class labels for the classification task. Data collection was conducted through preliminary observation and survey methods. The preliminary observation aimed to understand the actual conditions of learning quality in the pesantren including teachers' instructional methods

and variations in learning quality across classes. Subsequently, questionnaires were distributed to both teachers and students. The research instruments were developed in two versions: one for students, assessing variables X1, X2, X4, and learning quality (Y) from the students' perspective, and another for teachers, assessing variable X3 and learning quality (Y) from the teachers' perspective. All questionnaire items were measured using a five-point Likert scale (1–5) for each indicator.

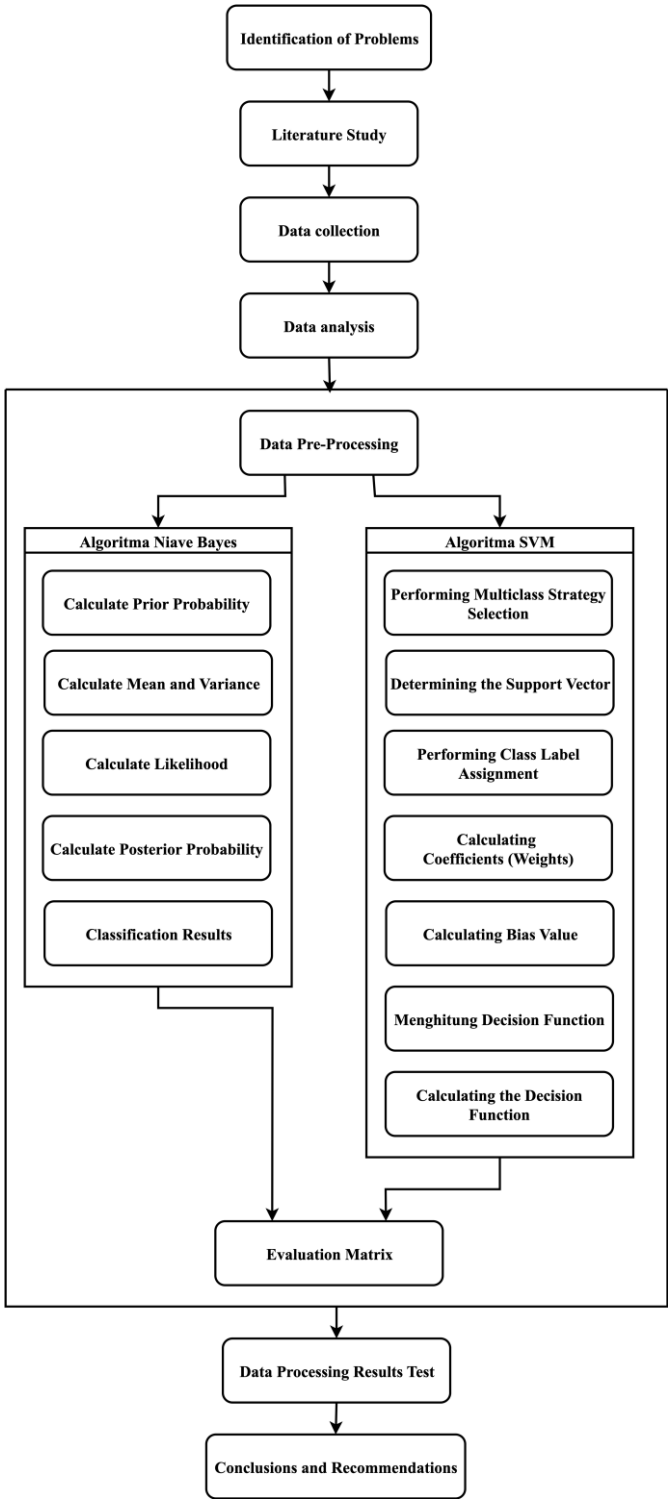

Figure 1. Research Framework

2.1 Naïve Bayes

The prior probability is used to determine the initial likelihood that a document belongs to a particular class. This value is obtained from the proportion of instances in each class relative to the total number of instances across all classes.

a. Computing the Prior Probability

The prior probability is used to determine the initial likelihood that a document belongs to a particular class. This value is obtained from the proportion of instances in each class relative to the total number of instances across all classes.

b. Calculating the Mean and Variance

The probability of each feature within a class is calculated to determine the contribution of each feature to the document classification process. This calculation is performed separately for each feature using the available data.

c. Calculating the Posterior Probability

The final probability of a data instance belonging to a particular class is calculated by combining the prior probability of the class and the likelihood of each feature.

2.2 Support Vector Machine (SVM)

After understanding the theoretical concepts and fundamental components of the Support Vector Machine (SVM), its implementation needs to be described in the form of a systematic procedure. The application of SVM in solving classification problems requires a structured sequence of steps to ensure the construction of an optimal model. In general, the working steps of this algorithm can be outlined as follows:

a. Multiclass Strategy Selection

This stage determines the strategy used to address classification problems involving more than two classes in the SVM algorithm. In the One-vs-All strategy, each class is compared against the combination of all other classes, resulting in k models for k classes.

b. Determining Support Vectors

Support vectors are the most important data points that define the position and margin of the SVM hyperplane.

c. Determining Class Label Assignment

Each data instance is assigned a label for model training according to the selected multiclass strategy. The target class is labeled +1, while all other classes are labeled −1, so that the model remains a binary SVM.

d. Calculating the Bias Value

The bias determines the position of the hyperplane relative to the origin and is calculated using the support vectors. The bias formula is given.

e. Performing Classification

Each model built using the One-Against-One strategy provides a vote for class i or j based on its output value. The results of all models are then combined, typically through a majority voting mechanism, to determine the final class label of the input data x The classification process for each class pair is carried out by evaluating the decision.

2.3 Model Performance Evaluation

Model performance measurement is an evaluation process used to assess how well a model performs categorization tasks. Commonly used metrics in this evaluation include precision, recall, F-measure, and accuracy. The values of True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN) are used to calculate these metrics, representing the prediction outcomes and model performance for each class.

3. Results and Discussions

This section describes the research workflow, starting from the preprocessing of questionnaire data to the classification of learning quality using two algorithms, namely Naïve Bayes and Support Vector Machine (SVM). In general, the raw questionnaire data were validated, transformed into summarized values (average scores), and then divided into training and testing datasets to build and evaluate the models.

All processes were carried out to ensure that the assessment of learning quality at Syafa'aturrasul Islamic Boarding School could be conducted in a more objective and measurable manner. The main emphasis at this stage is on computational consistency and the validation of model performance through cross-platform evaluation using Microsoft Excel and Python.

3.1 Preprocessing

Preprocessing aims to prepare the data so that they are suitable for classification modeling. In this study, questionnaire responses were examined for completeness and consistency, followed by the removal of invalid or missing values. Subsequently, data transformation and the separation of training and testing datasets were performed. The transformed data for five sample respondents are presented in Table 1 below:

Tabel 1 Data Transformation

No	Responden	X1	X2	X3	X4	Y
1	Responden 1	4,83	4,833	4,12	4,667	1
2	Responden 2	4,66	4,667	4,25	4,833	1
3	Responden 3	4,5	5	4,5	5	1

4	Responden 4	5	4,5	5	4,5	0
5	Responden 5	3,83	4,167	3,75	4,5	1

Table 1 presents the preprocessed data, showing that each respondent is represented by average numerical values for each research variable, namely data into quantitative data that are ready to be processed by classification algorithms.teacher competence (X1), facilities and infrastructure (X2), school management (X3), and student engagement (X4). This process aims to simplify Likert scale–based questionnaire

In addition, variable Y in the table represents the learning quality category, where a value of 0 indicates high quality and a value of 1 indicates medium quality. With this data format, each respondent has a clear attribute pattern, facilitating the analysis and classification process using both the Naïve Bayes and Support Vector Machine methods.

3.2 Naïve Bayes Method Analysis

This study employs the Naïve Bayes algorithm to analyze the data. This algorithm was selected due to its ability to perform classification using a simple probabilistic approach. The method predicts learning quality categories based on the predefined variables.

a. Calculating the Prior Probability

The prior probability is calculated based on the class distribution in the training dataset. This value represents the initial probability of each learning quality category before considering the predictor variables. The complete recap of the mean and variance values is presented in Table 2.

Table 2. Results of Mean and Variance Calculation

No	Variabel	Mean			Variasi		
		Class 0	Class 1	Class 2	Class 0	Class 1	Class 2
1	X1	4,952	4,365	3,833	0,01	0,137	0,292
2	X2	4,941	4,153	3,833	0,022	0,279	0,028
3	X4	4,893	4,529	4,375	0,031	0,195	0,27
4	X5	4,964	4,427	4,083	0,017	0,121	0,035

Table 2 presents a summary of the statistical parameters for each predictor variable. The mean values represent the central tendency of the data, while the variance indicates the degree of dispersion within each learning quality category. These parameters serve as key components in the probability density function used for Naïve Bayes classification.

b. Calculating the Likelihood

The next step involves calculating the likelihood values for each variable in each class. The likelihood represents the probability that a particular variable value occurs within a specific class. This calculation is performed using the Gaussian probability density function. The likelihood computation for all test data produces a large set of probabilistic values, which represent how likely an observation is to belong to each learning quality category. These calculations are carried out for each predictor variable across all test data samples.

Class label assignment is required to train each binary classifier. The labels are converted to +1 for the target class and −1 for the remaining classes using Equation (2.7). This conversion transforms the multiclass problem into three binary classification problems. The complete details of the sample class label assignment are presented in the following Table 3.

Table 3. Sample Class Label Assignment

No	Baris	X1	X2	X3	X4	Y	Label
0 vs All							
1	1	4,833	4,833	4,125	4,667	1	-1
2	3	4,5	5	4,5	5	1	-1
3	4	5	4,5	5	4,5	0	1
4	10	4,833	5	4,875	5	0	1
5	13	4,667	4,667	5	5	0	1

The table 3 presents a subset of the training data that serve as support vectors along with their newly assigned binary labels. This table visualizes how each sample is allocated during the training of the three separate SVM models. These data form the computational basis for identifying the optimal hyperplane for each classifier.

c. Calculating the Posterior Probability

The posterior probability is calculated by multiplying the prior probability by the product of all likelihood values. The resulting posterior value represents the model's confidence in the class membership of each sample. The posterior probability calculation results for each respondent in the test dataset were then compiled. These values serve as the basis for determining the final prediction of the learning quality class. The complete summary of posterior probabilities is presented in the following Table 4.

Table 4. Posterior Probability Calculation Results

No	Responden	Class 0	Class 1	Class 2
1	Responden 71	8,61751E-30	0,267550356	0,093858595
2	Responden 72	1,62265E-14	0,010589231	7,57943E-11
3	Responden 73	0,002953656	0,063326196	1,35196E-10
4	Responden 74	1,08618E-05	0,114931805	6,16399E-08
5	Responden 75	3,68226E-13	0,167595761	0,000504143

Table 4 presents the posterior probability values for the three learning quality categories across all test data. Each row displays three comparative probability values for a single respondent. The class with the highest probability value is selected as the final prediction of the Naïve Bayes model.

5. Classification Results

The classification results are obtained by comparing the posterior probability values of each class. The decision rule follows Equation (2.6) by selecting the class with the highest probability value. The complete summary of the classification results is presented in Table 4.

Table 5. presents the posterior probability

No	Responden	Class 0	Class 1	Class 2	Prediksi
1	Responden 71	8,61751E-30	0,267550356	0,093858595	1
2	Responden 72	1,62265E-14	0,010589231	7,57943E-11	1
3	Responden 73	0,002953656	0,063326196	1,35196E-10	1
4	Responden 74	1,08618E-05	0,114931805	6,16399E-08	1
5	Responden 75	3,68226E-13	0,167595761	0,000504143	1

Table 5 presents the predicted learning quality categories for five test data samples as examples. The prediction column indicates the class with the highest posterior probability in each row. This table serves as the basis for calculating the accuracy of the Naïve Bayes model by comparing the predicted classes with the actual classes.

3.3 Support Vector Machine (SVM)

The second algorithm employed in this study is the Support Vector Machine (SVM) with a linear kernel. This method aims to find an optimal hyperplane that separates samples from different classes. The application of linear SVM is intended to compare its performance with that of the Naïve Bayes algorithm.

a. Multiclass Strategy Selection

The multiclass classification problem in SVM is addressed using the One-vs-Rest strategy. This strategy trains one binary classifier for each class. The details of the implementation of this strategy are presented in Table 5.

Table 6 Presents Three Binary

No	Strategy	Information
1	0 vs All	Binary SVM for detecting Class 0
2	1 vs All	Binary SVM for detecting Class 1
3	2 vs All	Binary SVM for detecting Class 2

Table 6 presents three binary strategies within the One-vs-Rest approach. Each strategy focuses on distinguishing one specific class from the combination of all other classes. This approach enables the linear SVM to predict three learning quality categories.

b. Determining the Support Vectors

Support vectors are identified for each binary classifier in the One-vs-Rest strategy. The row indices of the training data that serve as support vectors are then recorded. The results of support vector identification for all three strategies are presented in Table 6 below.

Table 7. Strategi One vs Rest

No	Strategy	Data Row
1	0 vs All	[1, 3, 4, 10, 13, 14, 18, 28, 31, 33, 36, 39, 47, 51, 56, 66, 67]
2	1 vs All	[1, 2, 3, 4, 6, 7, 8, 9, 10, 12, 13, 14, 17, 18, 19, 22, 25, 26, 28, 29, 30, 31, 32, 33, 34, 36, 37, 39, 40, 41, 43, 45, 47, 49, 51, 52, 54, 55, 56, 59, 61, 65, 66, 67, 68, 70]
3	2 vs All	[5, 6, 7, 9, 19, 20, 21, 37, 42, 43, 44, 48, 49, 50, 54, 57, 63, 65, 68, 69]

Table 7 shows the row indices of the training data that function as support vectors. Each binary classification strategy has a different set of support vectors. These results demonstrate the complexity and unique characteristics of the class separation for each strategy.

c. Class Label Assignment

Class label assignment is required to train each binary classifier. The labels are converted to +1 for the target class and −1 for the other classes using Equation (2.7). This conversion transforms the multiclass problem into three binary classification problems.

d. Calculating the Coefficients (Weights)

The SVM model calculates coefficients, or weights, for each feature. These weights determine the orientation and position of the separating hyperplane. The computation aims to maximize the margin between different classes. The weighting results of the SVM for

the 0 vs All, 1 vs All, and 2 vs All strategies indicate that each strategy produces a different weight vector. These differences reflect the magnitude and direction

e. Calculating the Bias Value

The bias value is computed to complete the SVM decision function. The bias determines the shift of the hyperplane from the origin in the feature space. This calculation utilizes the support vectors and the previously obtained weight values. The bias value calculation in the Support Vector Machine (SVM) model is performed to complement the decision function in the One-vs-Rest strategy, so that each hyperplane can be optimally positioned with respect to the training data. The bias value is obtained by calculating the difference between the class label and the result of multiplying the weights by the feature vector in each support vector, then averaging them to obtain the final bias. This process ensures that the decision function produces outputs that are consistent with the class labels of each data, while maintaining a balanced separation between classes in the feature space.

The final bias values are calculated for all three binary classifiers in the One-vs-Rest strategy. Each classifier has a unique bias value that adjusts its corresponding hyperplane. The results of the bias value calculations for all models are presented in Table 7 below.

Table 8. Results of Bias Value Calculation

No	Strategy	Bias Value
1	0 vs All	-26,838
2	1 vs All	9,722
3	2 vs All	-0,125

Table 8 presents the final bias values for each classification strategy. The differences in bias values reflect the relative positions of each hyperplane in the feature space. These parameters constitute an essential component of the equation used to predict the class of new data.

f. Calculating the Decision Function

The decision function is calculated to determine the predicted class of the test data. This calculation combines the attribute weights, feature values, and bias. The decision function results for all test data are then compiled. This table records the output values produced by each classifier for every respondent. The complete summary of the decision function calculations is presented in Table 8 below.

Table 9. Results of Decision Function Calculation

No	Responden	0 vs All	1 vs All	2 vs All
1	Responden 71	-2,912	0,474	-0,212
2	Responden 72	-1,827	0,829	-0,213
3	Responden 73	-0,265	-0,292	-0,22
4	Responden 74	-0,745	-0,085	-0,219
5	Responden 75	-1,537	0,149	-0,216

Table 9 shows the decision function values produced by the three classifiers for 30 test data samples. Each column represents the confidence of a binary model in classifying a sample into its target class. The class with the highest score determines the final predicted class under the One-vs-Rest strategy.

g. Classification Results

The final classification results are obtained by comparing the decision function values from the three models. The One-vs-Rest strategy selects the class with the highest decision function value. This decision rule is applied in accordance with Equation (2.11). A summary of the classification results for five respondents using the SVM model is presented in Table 9.

Table 10. SVM Classification Results

No	Responden	0 vs	1 vs All	2 vs All	Y
1	Responden 71	-2,912	0,474	-0,212	1
2	Responden 72	-1,827	0,829	-0,213	1
3	Responden 73	-0,265	-0,292	-0,22	1
4	Responden 74	-0,745	-0,085	-0,219	1
5	Responden 75	-1,537	0,149	-0,216	1

Table 10 presents the predicted learning quality categories generated by the SVM algorithm. The prediction column indicates the class with the highest decision function value in each row. This table serves as the basis for evaluating the accuracy of the SVM model by comparing the predicted results with the actual classes.

3.4 Evaluation Process

The evaluation process is conducted by comparing the predictions produced by both algorithms with the actual class labels. This comparison illustrates the performance of each model in classifying the test data. The prediction results of both algorithms are presented in Table 10 below.

Table11. Prediction Results of the Naïve Bayes and SVM Algorithms

No	Responden	X1	X2	X3	X4	Y Aktual	Y Naïve Bayes	Y SVM
1	Responden 71	4,167	4	4,875	4,167	1	1	1
2	Responden 72	4,667	4,5	3,625	5	1	1	1
3	Responden 73	4,667	4,5	4,875	5	1	1	1
4	Responden 74	4,833	4,667	4,875	4,333	2	1	1
5	Responden 75	4,667	4,333	5	4,167	1	1	1

Table 11 presents a direct comparison between the actual class labels, the Naïve Bayes predictions, and the SVM predictions. These data serve as the basis for constructing the confusion matrix for each algorithm. Further analysis is conducted using confusion matrices in graphical form, as shown in Figure 2.

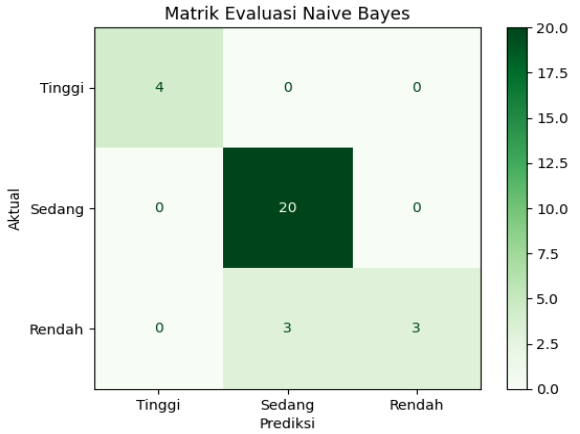

Figure 2. Naïve Bayes Evaluation Matrix Results

Figure 2 illustrates the visualization of the evaluation matrix of the Naïve Bayes model. This matrix summarizes the number of correct and incorrect predictions for each class and displays the test data along with their predicted class labels. The visualization highlights the correspondence between the actual values and the predicted values. This summary also includes the model's accuracy, precision, recall, and F1-score. The model's performance, with all evaluation metrics ranging approximately from 80% to 90%. These results indicate that the model successfully classified the test data with high accuracy. The optimal performance demonstrates the suitability of the Naïve Bayes algorithm for the characteristics of the dataset used. Meanwhile, the evaluation matrix of the SVM model is required to measure the performance of the constructed classification model. This evaluation provides a comprehensive overview of prediction accuracy and misclassification errors. The analysis is conducted by comparing the predicted labels with the actual labels, as shown in Figure 3.

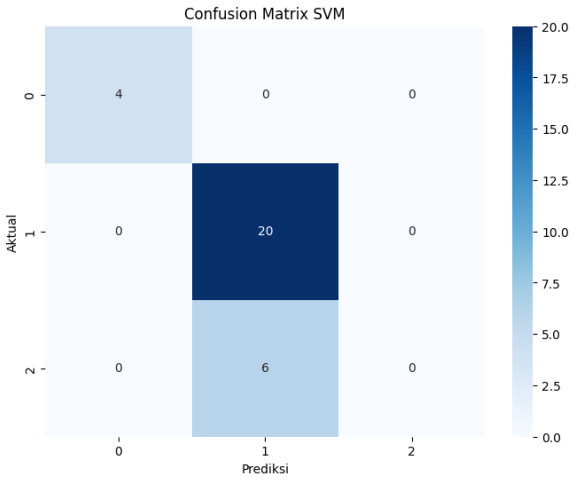

Figure 3. SVM Evaluation Matrix Results

Figure 3 illustrates the visualization of the evaluation matrix of the SVM model. This matrix summarizes the number of correct and incorrect predictions for each class. The detailed evaluation serves as the basis for assessing the reliability of the model and displays the test data along with their predicted class labels. This visualization highlights the correspondence between the actual values and the predicted values. The summary includes the model's accuracy, precision, recall, and F1-score.

The evaluation of the implementation results shows that the Naïve Bayes algorithm achieves higher performance than SVM for this dataset. The accuracy metric of Naïve Bayes reaches 90%, outperforming SVM, which achieves an accuracy of 80%. These findings confirm the suitability of the dataset characteristics with the feature independence assumption underlying the Naïve Bayes algorithm.

Cross-platform validation further demonstrates full consistency between the implementations in Microsoft Excel and Python. Both environments produce identical confusion matrix values and evaluation metrics. This consistency strengthens the reliability of the computational procedures and the validity of the developed models. Based on these final results, Naïve Bayes is recommended as the preferred algorithm for classifying data with similar characteristics. This model offers higher accuracy with relatively more efficient computation. The implementation of the algorithm has proven to be robust and consistent in classifying learning quality at Syafa'aturrasul Islamic Boarding School.

These differences indicate that Naïve Bayes is more effective in capturing data patterns in the context of this study. Furthermore, the consistency of the results obtained from both Excel and Python computations reinforces the conclusion that the Naïve Bayes algorithm is a more accurate and reliable method for

measuring learning quality at Pondok Pesantren Syafa'aturrasul.

4. Conclusions

Based on the results of the algorithm performance evaluation, it can be concluded that both the Naïve Bayes and Support Vector Machine (SVM) algorithms are capable of measuring and classifying learning quality using quantifiable evaluation metrics. However, Naïve Bayes demonstrates superior performance, achieving an accuracy of 90%, precision of 95.65%, recall of 83.33%, and an F1-score of 86.56%. In comparison, SVM yields an accuracy of 80%, precision of 58.97%, recall of 66.67%, and an F1-score of 62.32%.

References

[1] A. Amobonye, J. Lalung, G. Mheta, and S. Pillai, "Writing a Scientific Review Article: Comprehensive Insights for Beginners," *The Scientific World Journal*, vol. 2024, pp. 1–13, Jan. 2024, doi: 10.1155/2024/7822269.

[2] S. Nundy, A. Kakar, and Z. A. Bhutta, "How to Write the Introduction to a Scientific Paper?," in *How to Practice Academic Medicine and Publish from Developing Countries?*, Singapore: Springer Nature Singapore, 2022, pp. 193–199. doi: 10.1007/978-981-16-5248-6_17.

[3] C. Busse and E. August, "How to Write and Publish a Research Paper for a Peer-Reviewed Journal," *Journal of Cancer Education*, vol. 36, no. 5, pp. 909–913, Oct. 2021, doi: 10.1007/s13187-020-01751-z.

[4] S. Sobari et al., "Meningkatkan Model Prediksi Kelulusan Santri Tahfidz … Menggunakan Random Forest," JITET, 2025.

[5] D. N. Asy'ari et al., "Implementation of Digital Guided Discovery Learning at Ibrahimy Boarding School," IEEE WAIE, 2024.

[6] R. M. Sapdi et al., "Exploring Classification Algorithms for Detecting Learning Loss in Islamic Religious Education: A Comparative Study," International Journal on Informatics Visualization, 2024.

[7] L. F. Azevedo, F. Canário-Almeida, J. Almeida Fonseca, A. Costa-Pereira, J. C. Winck, and V. Hespanhol, "How to write a scientific paper—Writing the methods section," *Rev Port Pneumol*, vol. 17, no. 5, pp. 232–238, Sep. 2011, doi: 10.1016/j.rppneu.2011.06.014.

[8] S. K. Arora and D. Shah, "Writing methods: How to write what you did?," *Indian Pediatr*, vol. 53, no. 4, pp. 335–340, Apr. 2016, doi: 10.1007/s13312-016-0847-7.

[9] S. R. N. Reis and A. I. Reis, "How to write your first scientific paper," in *2013 3rd Interdisciplinary Engineering Design Education Conference*, IEEE, Mar. 2013, pp. 181–186. doi: 10.1109/IEDEC.2013.6526784.

[10] Yuhefizar, R. Watrianthos, and R. Komalasari, "An LDA Analysis for Topic Modeling of the RESTI Journal," in *2024 2nd International Symposium on Information Technology and Digital Innovation (ISITDI)*, IEEE, Jul. 2024, pp. 46–52. doi: 10.1109/ISITDI62380.2024.10796703.

[11] Ronal Watrianthos and Y. Yuhefizar, "Exploring Research Trends and Impact: A Bibliometric Analysis of RESTI Journal from 2018 to 2022," *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, vol. 7, no. 4, pp. 970–981, Aug. 2023, doi: 10.29207/resti.v7i4.5101.

[12] IEEE, "IEEE Editorial Style Manual for Authors," *IEEE Editorial*, 2019.

[13] O. Tur, A. Tur, V. Shabunina, and E. Chernaia, "The IEEE Style: Peculiarities of the Format and Application Prospects," in *2020 IEEE Problems of Automated Electrodrive. Theory and Practice (PAEP)*, IEEE, Sep. 2020, pp. 1–4. doi: 10.1109/PAEP49887.2020.9240790.

[14] D. J. S. Montagnes, E. I. Montagnes, and Z. Yang, "Finding your scientific story by writing backwards," *Mar Life Sci Technol*, vol. 4, no. 1, pp. 1–9, Feb. 2022, doi: 10.1007/s42995-021-00120-z.

[15] D. R. Hess, "How to Write an Effective Discussion," *Respir Care*, vol. 68, no. 12, pp. 1771–1774, Dec. 2023, doi: 10.4187/respcare.11435.

[16] A. Ramadhanu, H. Hendri, F. Hadi, D. Guswandi, D. M. Putra, R. Hardianto, and S. D. Rizki, "Hybrid decision support system and image processing for classifying priority applications in the Padang Government," *CSRID (Computer Science Research and Its Development Journal)*, vol. 18, no. 1, pp. 178–191, 2026.

[17] A. Ramadhanu, H. Hendri, and F. Hadi, "Organic fertilizer content detection based on image segmentation and texture analysis," in *Proc. 2025 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS)*, Dec. 2025, pp. 50–55.

Biographies of Authors

Fadhila Putri Sani Fadhila Putri Sani is a graduate student in the Master's Program in Informatics Engineering at the Faculty of Computer Science, Universitas Putra Indonesia "YPTK" Padang. She was born in Baserah on August 16, 1994. She is currently pursuing her studies in the field of Informatics Engineering. She can be contacted via email at fadhilaputrisani@gmail.com

Syafri Arlis Syafri Arlis was bornin Padang on October 23, 1986. His undergraduate study was completed in 2009 at Universitas Putra Indonesia YPTK. He completed his master'sdegree at Putra Indonesia University, YPTK Padang. He currently serves as a lecture in the Informatics Engineering study program at the Universitas Putra Indonesia YPTK Padang. Teaching history that has been carried out starting from 2011 until now, such as databasesand digital image processing. Published research history places more emphasis on the atificial

	neural network and digital image processing. He can be contacted at email: syafri_arlis@upiyptk.ac.id.
	**Agung Ramadhanu** works at Universitas Putra Indonesia YPTK Padang as a lecturer and researcher. He was born on April 15, 1991, in Muaro Bungo. He was majoring in computer science and pursuing a bachelor's, master's, and doctoral degree at Universitas Putra Indonesia YPTK Padang. That university awarded him a S.Kom., M.Kom., and Dr. in computer science. He can be reached by email at agung_ramadhanu@upiyptk.ac .id. His office is located in Lubuk Begalung Street, Padang, Sumatera Barat, Indonesia, at Universitas Putra Indonesia YPTK Padang. His areas of specialisation are image processing and artificial intelligence.