

Machine Learning Algorithm for Questionnaire-Based Student Learning Style Classification

Herman Yuliansyah¹, Agung Tri Lestari², Anton Yudhana³

¹Department of Informatics, Universitas Ahmad Dahlan, Bantul, Indonesia, 55191

²Master Program of Informatics, Universitas Ahmad Dahlan, Yogyakarta, Indonesia, 55164

³Department of Electrical Engineering, Universitas Ahmad Dahlan, Bantul, Indonesia 55191

Article Info

Article history:

Received Feb 09, 2026

Revised Mar 12, 2026

Accepted Mar 23, 2026

Keywords:

Classification Data

Comparative Analysis

K-fold Cross Validation

Machine Learning

Student Learning Styles

ABSTRACT

Identifying students' learning styles is an important factor in supporting adaptive and data-driven learning. However, conventional methods based on manual questionnaires still have limitations in terms of efficiency and accuracy for data processing. This study presents a comparative analysis of machine learning algorithms to classify student learning styles based on questionnaire data. The dataset used consists of 1,170 student data with three learning style classes, namely visual, auditory, and kinesthetic. The four supervised learning algorithms used are Naïve Bayes, Decision Tree, Random Forest, and K-Nearest Neighbors. Model performance evaluation was conducted using 5-fold (80:20) and 10-fold (90:10) cross-validation with accuracy, precision, recall, and F1-score metrics. The results of the experiment show that the Naïve Bayes algorithm has the most optimal and stable performance with the highest accuracy value of 90.60% in both validation scenarios. These findings indicate that machine learning-based classification approaches, particularly Naïve Bayes, are effective for identifying student learning styles and have the potential to support the development of adaptive and personalized learning systems.

This is an open access article under the [CC BY-SA](#) license.



Corresponding Author:

Herman Yuliansyah,

Department of Informatics,

Universitas Ahmad Dahlan,

Jl. Ringroad Selatan, Kragilan, Tamanan, Kec. Banguntapan, Kabupaten Bantul, Daerah Istimewa Yogyakarta 55191.

Email: herman.yuliansyah@tif.uad.ac.id

1. INTRODUCTION

Learning style significantly affects students' learning effectiveness, motivation, and academic achievement by shaping their interaction with instructional materials [1]–[3]. Identifying students' learning styles is the first step in improving the quality of education. Specifically, understanding individual learning preferences can help in designing a more appropriate and effective curriculum [4].

Several previous studies have conducted research on learning styles using paper-based (conventional) media [5]. Differences in learning preferences, such as visual, auditory, and kinesthetic, require an adaptive and individual-characteristic-based learning approach. However, the process of identifying student learning styles at the secondary school level is still dominated by conventional methods based on manual questionnaires, which have limitations in terms of efficiency, data processing accuracy, and flexibility of use. In addition, there are various machine learning algorithms with different characteristics, but their accuracy still shows relatively limited performance. Algorithmically advanced detection systems have not had a significant implementative impact if they are difficult for students and teachers to operate. Hasan et al. [6], conducted research with positive results on the environment in terms of student engagement, which

showed great interest in future educational applications. Agung Nugraha & Budiyo [7], conducted research with results from the use of questionnaires, which showed that student learning styles were dominated. Based on previous studies, most existing research on learning style classification has focused on limited datasets or specific educational levels, with relatively simple validation strategies. In addition, comprehensive comparative studies using large-scale questionnaire data at the junior high school level are still limited. Therefore, there is a need for a systematic evaluation that not only compares classical machine learning algorithms but also analyzes their stability and generalization performance in real-world educational environments.

The main challenge in identifying students' learning styles today lies in the inefficiency of conventional paper-based [5], methods, which are still predominantly used at the junior high school level. These methods are considered irrelevant because they require a long time to process data, carry a high risk of documentation errors, and have limited accessibility that cannot be flexibly implemented through various digital devices. On the other hand, although machine learning algorithms offer great potential for classification automation, there is still ambiguity regarding which algorithm is the most optimal because each method has different performance characteristics and accuracy levels when faced with patterns of student behavior data. The problem is further complicated by the fact that a system with a highly sophisticated algorithm has no real impact if it is difficult for users to operate.

Despite the increasing use of machine learning techniques for learning style identification, several limitations remain in existing studies. Many previous works focus on relatively small datasets or specific educational contexts, which limits the generalizability of the obtained models. In addition, several studies evaluate classification performance using only a single validation strategy, making it difficult to assess the stability and robustness of the algorithms when applied to different data partitions. As a result, there is still limited empirical evidence regarding which machine learning algorithm provides the most reliable and consistent performance for questionnaire-based learning style classification in real educational environments.

To address these limitations, this study conducts a systematic comparative analysis of several classical machine learning algorithms using a large-scale questionnaire dataset and multiple validation scenarios. The novelty of this research lies not in proposing a new algorithm, but in providing a comprehensive empirical evaluation framework that emphasizes both classification accuracy and model stability across different validation schemes.

The main contributions of this study can be summarized as follows:

1. Large-scale educational dataset evaluation – This study utilizes a relatively large dataset consisting of 1,170 student questionnaire responses, enabling more reliable empirical analysis of learning style classification compared to many previous studies with limited samples.
2. Robust model validation strategy – The performance of four widely used machine learning algorithms (Naïve Bayes, Decision Tree, Random Forest, and K-Nearest Neighbors) is systematically evaluated using two cross-validation scenarios (5-fold and 10-fold) to analyze both accuracy and performance stability.
3. Empirical evidence for algorithm suitability – The study provides comprehensive experimental results demonstrating that Naïve Bayes offers the most stable and accurate performance for questionnaire-based learning style data, highlighting the effectiveness of probabilistic approaches for categorical educational datasets.

Through this framework, the study contributes to the field of educational data mining by providing a clearer understanding of the suitability of different machine learning algorithms for learning style classification and by supporting the development of data-driven adaptive learning systems.

2. RESEARCH METHOD

The research began with data collection by conducting a comparative experiment on machine learning algorithms, namely Naive Bayes (NB), Decision Tree (DT), Random Forest (RF), and K-Nearest Neighbors (KNN). The dataset was collected from 1,170 student data using the instruments in Table 1. To ensure the validity and objectivity of the accuracy results, the testing

process was carried out by applying the k-fold cross-validation method using two testing scenarios, namely 5-fold (80:20) and 10-fold (90:10).

The research instrument used was a learning style questionnaire consisting of 30 statements designed to identify students' learning preferences and characteristics. These statements were compiled to represent students receive, process, and understand information in the learning process as the basis for machine learning-based learning style classification. The questionnaire was grouped into five main aspects, namely Self-Character, Interests and Talents, Behaviorism, Emotions, and Cognition. The Self-Character and Interests and Talents aspects describe students' tendencies and interests in certain learning methods, while the Behavioristic and Emotional aspects reflect students' behavior patterns and affective conditions during learning. The Cognitive aspect represents the students understand and process information in completing learning tasks.

Students' responses to each statement were coded and processed into categorical data used as features in the machine learning model training and testing process. Based on the results of this processing, each student was classified into one of three learning style categories, namely visual, auditory, and kinesthetic, thereby supporting the application of a supervised learning approach in this study.

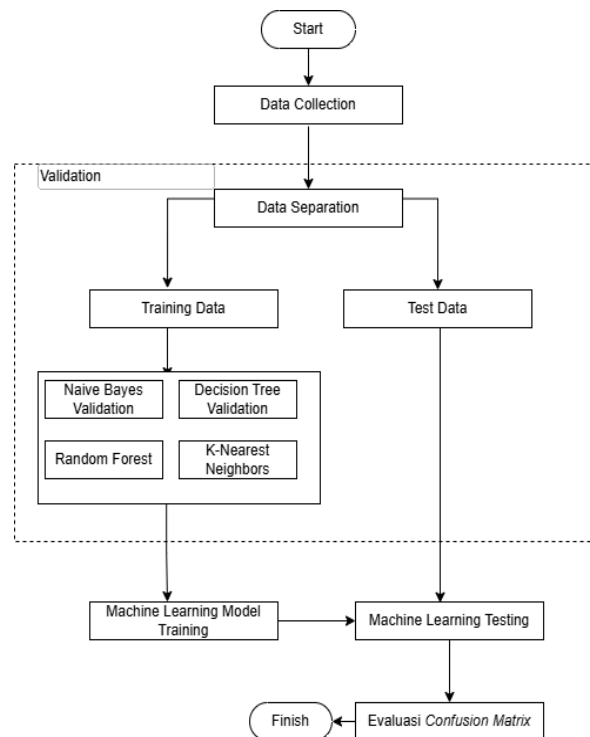


Figure 1. Research stages

Figure 1 shows the flow of the research stages in developing a machine learning-based student learning style classification model. The process begins with data collection, followed by data visualization and separation of data into training and test data. The training data is used to train several machine learning algorithms, namely NB, DT, RF, and KNN. Later, the trained models are tested using test data, and classification performance is evaluated using a confusion matrix to measure accuracy and other evaluation metrics.

2.1 Data Collection

Data collection is the first step in starting the research process to gather all the information needed for learning style classification. At this stage, raw data are collected and serve as the main foundation for the next process, namely data visualization and data separation. The collected data usually include variables that represent student learning patterns, which are later processed by algorithms such as NB, DT, RF, and KNN. The success of this stage is crucial because the quality

of the data obtained in determining the accuracy of the model in recognizing patterns and feature relationships at the training stage until the final evaluation using a confusion matrix.

The learning style assessment consists of five important aspects, namely Self-Character [8], Interests and Talents [9], Behaviorism [8], Emotions [10], and Cognition [11], as shown in Table 1. The Self-Character and Interests and Talents aspects describe students' tendencies and interests in certain learning methods, while the Behaviorism and Emotions aspects reflect students' behavior patterns and affective conditions during learning. The Cognitive aspect represents the students understand and process information in completing learning tasks. Students' responses to each statement are coded and processed into categorical data used as features in the machine learning model training and testing process. Based on the results of this processing, each student is classified into three learning style categories, namely visual, auditory, and kinesthetic, thus supporting the application of the supervised learning approach in this study.

Table 1 Instrument aspect

Code	Question	Code	Question
Self-Character		Behavioral	
S1	I tend to focus more on colors and visuals when learning new material	S17	I find it easier to understand IT material when it is presented through diagrams, screenshots, or other visual displays.
S2	I prefer listening to lectures or explanations rather than reading books or materials.	S18	I prefer to learn by reading materials or taking notes rather than just listening to explanations.
S3	I find it easier to remember information after listening to verbal explanations.	S19	I feel comfortable when the material is presented with logical and detailed explanations, so that it can be analyzed further.
S4	I feel that learning is more effective when I can discuss or listen to others talk about the material.	Emotions	
S5	I often review material I have just learned by speaking aloud or listening to audio recordings.	S20	I find it easier to understand concepts when there are visual presentations (e.g., slides, mind maps, videos).
S6	I prefer listening to explanations or tutorials about technology (such as podcasts or online lectures) rather than reading technical material.	S21	I prefer discussing or attending seminars related to technology rather than just reading textbooks or documentation.
S7	I tend to remember the steps or procedures for using software or applications after listening to audio instructions or tutorials.	S22	I prefer to learn by doing or trying something practically.
Interests and Talents		S23	I find it easier to understand a topic if I can be directly involved in physical activities or experiments.
S8	I find it easier to remember information when it is presented in the form of images, diagrams, or graphs.	S24	I tend to focus better if I can move around or engage in physical activity while learning.
S9	I often take notes or make sketches to help me understand the material being taught.	S25	I often feel more comfortable learning if I use props or physical media to explore the material.
S10	I find it easier to understand programming concepts or technological systems when listening to verbal instructions or discussions from an expert or instructor.	Cognitive	
S11	I prefer to learn about technology by directly operating hardware or trying out code/programs myself rather than just reading instructions or watching videos.	S26	I prefer to follow video tutorials or visual presentations (such as screencasts or slides) to learn new software or applications.
S12	I find it easier to understand new software if I try using it directly, rather than just listening to or watching tutorials.	S27	I understand concepts in software development or interface design more quickly if I can see visual examples or existing designs.
S13	I am more focused and quicker to understand IT concepts, such as application development or troubleshooting, when I experiment directly with the technology.	S28	I tend to remember information better when I see illustrations or diagrams that depict technological concepts.
S14	I tend to be more motivated to learn IT if I can interact directly with tools or hardware (e.g., assembling a computer or editing source code).	S29	I prefer to solve problems using a systematic, step-by-step approach rather than following instructions that are not very structured.
S15	I feel more motivated to learn when I know the purpose or outcome I want to achieve from the learning process.	S30	I prefer to learn through hands-on experience or using simulation software to understand difficult concepts.
S16	I tend to learn more efficiently when there are clear targets or deadlines for the material being studied.		

2.2 Data Separation

The data separation process is an important step in developing machine learning models, especially in research aimed at identifying students' learning styles. This step aims to ensure that

the developed model is able to generalize well to data that it has never encountered before. Data separation is done by dividing the dataset into two main parts, namely training data and testing data. To optimize model validation, data separation is performed through a cross-validation scheme with a multi-split strategy approach.

In this study, model validation was conducted using k-fold cross-validation to ensure reliable and unbiased performance evaluation. Two validation scenarios were applied, namely 5-fold and 10-fold cross-validation. In the 5-fold scheme, the dataset was divided into five equal subsets, where four subsets were used for training and one subset for testing in each iteration. Similarly, in the 10-fold scheme, nine subsets were used for training and one subset for testing.

The proportions commonly referred to as 80:20 and 90:10 in the previous version represent the approximate training–testing ratios resulting from the 5-fold and 10-fold configurations, respectively, rather than independent hold-out validation schemes. Therefore, all reported results in this study are based on k-fold cross-validation, ensuring methodological consistency and fair comparison among algorithms.

Two different data division scenarios are applied, namely a 5-fold (80:20) and 10-fold (90:10) ratio, to determine the composition of the training and testing data. This method is applied systematically to minimize the risk of bias in the test results and ensure that each data partition has a representative distribution. Through iterations on both splits, researchers can evaluate the stability of model accuracy on various scales of training data volumes, thereby obtaining more accurate and reliable performance parameters.

Training data is a subset used to train machine learning models. In this study, training data is used to teach models about the characteristics and patterns found in students' learning styles. Test data is a subset that is not used in the training process, allowing it to provide a more accurate picture of the model's performance on new data.

2.3 Data Training and Testing

2.3.1 Naive Bayes (NB)

The NB algorithm is one of the most popular machine learning algorithms for classification, based on the application of Bayes' Theorem with the "naive" assumption that each feature or variable is independent (does not influence each other) [12]. Naive Bayes aims to estimate probabilities simply and focuses on using known probabilities to calculate other probabilities [13].

Bayes' Theorem has a formula in Equation 1.

$$P(H|E) = \frac{P(E|H) \times P(H)}{P(E)} \quad (1)$$

$P(H|E)$ Posterior Probability (The probability of learning style H occurring after seeing the evidence of questionnaire answer E).

$P(H)$ Likelihood (Probability of questionnaire response E occurring in students with learning style H).

$P(H)$ Prior (Initial probability of learning style H among all students).

$P(E)$ Evidence (Probability of questionnaire response E appearing in general).

2.3.2 Decision Tree (DT)

The DT algorithm is a classification method in machine learning that represents decision rules in the form of a tree structure [14], [15]. This algorithm divides the dataset into smaller subsets based on certain criteria, until leaf nodes are formed that represent classes or prediction labels. This algorithm represents the decision-making process in a visual form resembling an inverted tree, starting from a root node, followed by internal nodes that represent tests on certain attributes, ending at leaf nodes which are the final classification results [16].

The main concept used in this process is Entropy, which is a metric for measuring the degree of uncertainty, disorder, or impurity in a data set. The higher the entropy value, the more heterogeneous or irregular the class distribution in the data set [17].

Entropy and is formulated in Equation 2.

$$H(X) = -\sum_{i=1}^n p(x_i) \log_b p(x_i) \quad (2)$$

After the entropy value before data separation is known, the next step is to calculate the Information Gain. Information Gain is defined as the difference between the entropy value before separation (at the parent node) and the total weighted average entropy value after the data is separated based on a specific attribute (at the child node). Information Gain has the formula shown in Equation 3.

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{S_i}{S} Entropy(S_i) \quad (3)$$

where $p(x_i)$ is the proportion of class i , and S_i is the subset for attribute value i .

2.3.3 Random Forest (RF)

The RF algorithm is an ensemble algorithm that builds multiple decision trees to improve prediction accuracy [18]. Its main formulation includes:

1. Decision Tree Construction: Each tree is trained using a randomly selected subset of data with replacement, known as the bootstrap sampling method. Mathematically, the bootstrap data set D_b is formulated in Equation 4.

$$D_b = \{(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)\} \quad (4)$$

where x_i represents the feature vector of the i -th sample and y_i denotes its corresponding class label.

2. Ensemble Prediction: The final prediction result is determined through a majority voting mechanism for classification tasks. The final prediction $\hat{y}$ is formulated in Equation 5.

$$\hat{y} = \arg \max_{c \in C} \sum_{t=1}^T I(h_t(x) = c) \quad (5)$$

where C is the set of possible classes, T is the total number of trees, $I(\cdot)$ is the indicator function, which equals 1 if the condition is true and 0 otherwise.

3. Gini Impurity is a function in Random Forest that is calculated for each split at a tree node. The Gini Impurity at node t is formulated in Equation 6.

$$Gini(t) = 1 - \sum_{i=1}^C p_i^2 \quad (6)$$

where C is the number of classes, p_i is the proportion of samples belonging to class i at node t .

2.3.4 K-Nearest Neighbors (KNN)

The KNN algorithm is a data classification method used to estimate the likelihood of a data point belonging to one group or another based on which group is the closest indicator data. Research shows that the use of various metrics can improve the effectiveness of KNN in various domains [19]–[21]. KNN consists of two steps: the first step is training, which stores each training pattern, and the second step is classification. Each time a pattern is classified, KNN must examine all training patterns to determine the K nearest patterns [22]. To find the nearest distance between:

1. Euclidean Distance

The Euclidean distance is one method for measuring the distance between two points in two, three, or more dimensions [23] and is formulated in Equation 7.

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (7)$$

where $x = (x_1, x_2, \dots, x_n)$ is the feature vector of the test sample, $y = (y_1, y_2, \dots, y_n)$ is the feature vector of the training sample, x_i and y_i represent the values of the i -th feature, n is the total number of features.

2. Manhattan Distance

Manhattan distance is a method for measuring distance by calculating the absolute difference between the coordinates of a pair of objects [23] and is formulated in Equation 8.

$$d(x, y) = \sum_{i=1}^n |x_i - y_i| \quad (8)$$

where x_i and y_i denote the values of the i -th feature of samples x and y , n is the dimensionality of the feature space.

2.4 Model Testing

The system testing process begins with the use of test data that has been separated from the training data through a data separation stage to ensure the objectivity of the evaluation. In the machine learning testing stage, models that have undergone validation and training using the Naive Bayes, Decision Tree, Random Forest, and KNN algorithms are given test data to produce predictions of student learning styles. The accuracy of these predictions is then tested in the confusion matrix evaluation stage by comparing the model's prediction labels with the original labels in the test data. Through this evaluation, the system's performance can be measured systematically before reaching the final stage, to ensure that the model is able to recognize learning style patterns accurately and reliably in real-world situations. A confusion matrix is a performance evaluation tool used to measure the performance of classification algorithms on a set of test data whose true values are known [24]. This matrix presents a comparison between the actual (target) values and the model's predicted values in the form of an $n \times n$ table, where n is the number of classes or categories predicted [25].

Table 2 represents a confusion matrix to evaluate the performance of a learning style classification model consisting of three classes, namely visual (V), auditory (A), and kinesthetic (K). This matrix compares the actual values with the predicted values generated by the system. The columns in the matrix show the model's predictions, while the rows show the actual conditions. Visual predictions for visual data are indicated as true visual (TV), auditory predictions for auditory data as true auditory (TA), and kinesthetic predictions for kinesthetic data as true kinesthetic (TK). Meanwhile, classification errors are expressed as false visual (FV), false auditory (FA), and false kinesthetic (FK), according to the prediction and actual value pairs.

Table 2 Confusion matrix

		Predicted		
		Visual (V)	Auditory (A)	Kinesthetic (K)
Actual	Visual (V)	TV	FV	FV
	Auditory (A)	FA	TA	FA
	Kinesthetic (K)	FK	FK	TK

Based on the matrix in Table 2, model performance was evaluated using accuracy, recall, and precision metrics. Accuracy describes the model's level of accuracy in classifying all data and is formulated in Equation 9.

$$Accuracy = \frac{TV+TA+TK}{TotalData} \quad (9)$$

Recall is used to measure the model's ability to correctly recognize data in each class. Recall for each class is formulated in Equations 10-12.

$$Recall(V) = \frac{TV}{TV+FV} \quad (10)$$

$$Recall(A) = \frac{TA}{TA+FA} \quad (11)$$

$$Recall(K) = \frac{TK}{TK+FK} \quad (12)$$

Precision is used to measure the accuracy of model predictions for each class, i.e., the proportion of correct predictions out of all predictions in that class. Precision for each class is formulated in Equations 13-15.

$$Presisi(V) = \frac{TV}{TV+FA+FK} \quad (13)$$

$$Presisi(A) = \frac{TA}{FV+TA+FK} \quad (14)$$

$$Presisi(K) = \frac{TK}{FV+FA+TK} \quad (15)$$

Using this confusion matrix, the performance of the visual, auditory, and kinesthetic learning style classification model can be comprehensively analyzed through the overall accuracy as well as the identification ability and prediction accuracy in each class (recall and precision).

3. RESULTS AND DISCUSSION

The performance of the algorithm was assessed based on accuracy, precision, recall, and F1-score to determine the best model. In addition, the interpretation of the research results and their influence on the context of learning style classification were also discussed.

3.1 Data Collection

The data used in this study came from questionnaires filled out by students. This dataset contained data collected from students at Muhammadiyah Al Mujahadin Junior High School and Muhammadiyah 08 Junior High School in Batu City. The dataset consisted of 30 feature columns and 1 label column, as shown in Table 3.

This study was conducted in accordance with standard ethical research practices. Prior to data collection, formal permission was obtained from the administrations of the participating junior high schools. All students participated voluntarily and were informed about the purpose of the study. Informed consent was obtained through the school authorities and class supervisors. To protect participants' privacy, all questionnaire responses were anonymized by removing personal identifiers such as names, student identification numbers, and contact information. The collected data were used exclusively for academic research purposes and were stored in password-protected digital repositories accessible only to the research team. No individual respondent can be identified from the published results.

After completing the learning style questionnaire, each student was asked to identify and declare the learning style that best represented their personal learning preference, namely visual, auditory, or kinesthetic. This self-assessment process was conducted under the guidance of teachers to ensure that students clearly understood the meaning of each learning style category. The declared learning style was then used as the class label for supervised machine learning training and testing. This labeling approach reflects students' perceived learning characteristics and is commonly adopted in educational studies that rely on self-reported learning behavior.

Table 3 Key features of the dataset

No	S1	S2	...	S30	Label
1	T	N	...	TS	Visual
	S				
2	N	TS	...	TS	Auditory
....	...	...	...	...	...
1,170	N	STS		TS	Auditory

In Table 3, there are S1 to S30, which means questionnaire instruments 1 to 30. The initial processing of this data was carried out through programming using the PHP programming language by creating our own library. Table 4 presents the distribution of learning style labels in the dataset used in this study. The dataset consists of a total of 1,170 student records, which are categorized into three learning style classes: visual, auditory, and kinesthetic. Based on the Table 4, 363 students are classified as visual learners, 440 students as auditory learners, and 367 students as kinesthetic learners. This distribution indicates that the auditory learning style has the highest number of samples, followed by kinesthetic and visual.

Table 4 Number of labels

No	Label	Number
1	Visual	363
2	Auditory	440
3	Kinesthetic	367
	Total	1,170

3.2. Classification Model Evaluation Results

Presents the evaluation results of various classification models used in the study. The evaluation was conducted to assess the performance of each algorithm in classifying learning style data. Each evaluation explains in detail the performance of each model based on evaluation metrics such as accuracy, precision, recall, and F1-score. The testing was carried out through two data division scenarios (k-fold cross validation), namely split data 1 with 10 folds and split data 2 with 5 folds.

Based on the data in Table 5 on the precision values of learning style classification, it can be seen that the four algorithms show very competitive performance with values above 79% for all categories. The NB and DT algorithms produced identical values in each test, where the highest precision value was achieved in the Kinesthetic category at 99.51% with a data split ratio of 90:10. Meanwhile, the Random Forest algorithm showed a consistent increase in precision in all categories when using an 80:20 split ratio compared to 90:10. On the other hand, KNN provided the best results for the Auditory category with a value of 96.15% at a ratio of 80:20, despite having the lowest precision value in the visual category at 79.53% when compared to the other algorithms. Overall, the kinesthetic category tended to have the highest precision rate compared to the visual and auditory categories in almost all test scenarios.

Table 5 Precision values for learning style classification

Algorithm	Split	Visual	Auditory	Kinesthetic
Naive Bayes	90:10	81.05	99.51	92.84
	80:20	81.02	99.50	92.64
Decision Tree	90:10	60.60	79.21	78.44
	80:20	55.94	76.79	76.56
Random Forest	90:10	80.86	91.78	89.02
	80:20	83.42	93.55	90.18
KNN	90:10	79.53	91.63	95.08
	80:20	79.65	90.92	96.15

Table 6 presents the Recall values, showing that the NB and DT algorithms demonstrate identical and superior performance compared to other algorithms, especially in the visual category with the highest value reaching 95.33% on a 90:10 data split ratio. In general, the kinesthetic category obtained relatively stable recall values ranging from 88% to 93% across all algorithms, while the auditory category tended to have lower values, with the lowest check on RF (82.47%). The effect of the data split ratio between 90:10 and 80:20 did not show a drastic change in recall results, but the KNN algorithm showed fairly good consistency in the range of 86% to 89% for all learning style categories.

Table 6 Recall values for learning style classification

Algorithm	Split	Visual	Auditory	Kinesthetic
Naive Bayes	90:10	95.33	84.21	93.64
	80:20	94.87	84.43	93.66
Decision Tree	90:10	65.52	74.24	80.32
	80:20	63.08	67.70	77.69
Random Forest	90:10	88.25	82.29	93.02
	80:20	87.32	88.73	91.87
KNN	90:10	88.55	87.84	88.63
	80:20	89.46	86.68	88.62

Table 7 shows that the Kinesthetic learning style consistently achieves the highest F1-Score in almost all tests, with the highest score reaching 93.21 in the NB and DT algorithms with a 90:10 split. The NB and DT algorithms produced identical performance in both data split scenarios, while RF and KNN showed more dynamic fluctuations in results depending on the data ratio used. Although the lowest score was in the visual category, all algorithms performed very well with F1-Scores remaining above 83. After going through a structured model development process, from data collection to the testing stage, the results of the system accuracy evaluation are presented in Table 8.

Table 7 F1-Score values for learning style classification

Algorithm	Split	Visual	Auditory	Kinesthetic
Naive Bayes	90:10	87.56	91.18	93.21
	80:20	87.37	91.33	93.13
Decision Tree	90:10	62.63	76.23	79.15
	80:20	59.02	71.83	76.93
Random Forest	90:10	84.14	86.47	90.57
	80:20	85.30	91.06	91.00
KNN	90:10	83.73	89.62	91.71
	80:20	84.08	88.64	92.21

Table 8 Accuracy values for learning style classification

Algorithm	Split	Accuracy
Naive Bayes	90:10	90.60
	80:20	90.60
Decision Tree	90:10	72.48
	80:20	69.23
Random Forest	90:10	87.09
	80:20	89.32
KNN	90:10	88.37
	80:20	88.46

The NB algorithm shows very stable performance with an accuracy value reaching 90.6% consistently in both data division methods. This indicates that NB has excellent generalization capabilities in recognizing student learning style data patterns without being affected by the number of folds in the validation. This performance far exceeds that of the DT algorithm, which is recorded as the model with the lowest accuracy, namely 72.48% at 90:10 and a decrease to 69.23% at 80:20.

The superior performance of the Naïve Bayes algorithm can be explained by its theoretical suitability for handling categorical questionnaire data. Naïve Bayes applies a probabilistic approach with an independence assumption between features, allowing each questionnaire response to contribute independently to the classification decision. Since the dataset in this study consists of 30 categorical indicators representing students' behavioral and cognitive learning characteristics, this probabilistic framework enables the model to effectively capture the relationship between responses and learning style categories. In addition, Naïve Bayes is known for its robustness in handling high-dimensional data with relatively simple model structures. This property allows the algorithm to generalize well on the dataset used in this study, resulting in stable performance across both cross-validation scenarios.

The attached Figure 2 presents a candlestick visualization of classification accuracy for all evaluated algorithms under two validation scenarios, namely 90:10 and 80:20 data splits. This visualization provides a clear and compact representation of the distribution, variability, and stability of model performance. In Figure 2(a) and Figure 2(b), each candlestick illustrates the minimum, maximum, and average accuracy values obtained during the cross-validation process for Naïve Bayes (NB), Decision Tree (DT), Random Forest (RF), and K-Nearest Neighbors (KNN). This allows to observe not only the average performance but also the fluctuation and consistency of each algorithm. The results show that Naïve Bayes and KNN exhibit relatively stable accuracy ranges across both validation scenarios, while Decision Tree demonstrates lower and more variable performance. Random Forest achieves competitive results but shows moderate variability. These trends indicate that Naïve Bayes provides the most consistent generalization capability among the evaluated models. Therefore, this candlestick visualization serves as a performance comparison plot and trend analysis tool, enabling intuitive interpretation of algorithm stability and accuracy behavior. Combined with the quantitative results in Tables 5–8, this figure strengthens the presentation of experimental findings and supports comprehensive comparative analysis.

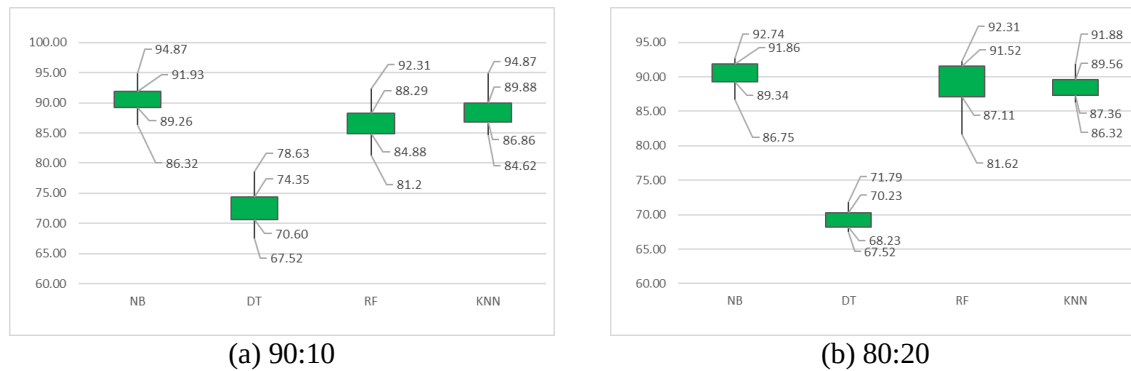


Figure 2. Candlestick of accuracy values for learning style classification

Compared with previous studies [2], [14], [15] which mainly focused on single validation scenarios or smaller datasets, this research provides a more extensive empirical evaluation using a larger sample size and multiple validation schemes. The experimental results demonstrate that the Naïve Bayes algorithm consistently outperforms other methods in terms of accuracy and stability. The use of two different cross-validation scenarios reveals that Naïve Bayes maintains stable performance regardless of data partitioning, indicating strong generalization capability. This finding highlights the suitability of probabilistic approaches for processing categorical questionnaire data with relatively clear behavioral patterns.

Furthermore, this research also provides a more comprehensive evaluation framework by using a larger dataset and multiple cross-validation scenarios. The results demonstrate that Naïve Bayes not only achieves high accuracy but also maintains stable performance across different validation settings. This strengthens the academic positioning of the study in providing empirical evidence for selecting suitable machine learning algorithms in questionnaire-based learning style classification.

4. CONCLUSION

This study has successfully conducted a comparative analysis of several machine learning algorithms in classifying student learning styles based on questionnaire data. The four algorithms tested, namely Naïve Bayes (NB), Decision Tree (DT), Random Forest (RF), and K-Nearest Neighbors (KNN), were evaluated using the k-fold cross-validation method with 5-fold and 10-fold scenarios and measured using the accuracy, precision, recall, and F1-score metrics. The results of the experiment show that the NB algorithm has the most optimal and stable performance compared to other algorithms, with the highest accuracy value of 90.60% consistent in both validation scenarios. These findings indicate that the NB probabilistic approach is very effective in handling learning style questionnaire data that is categorical and has relatively clear patterns.

This study contributes to the field of educational data mining by providing a systematic and empirically validated comparison of classical machine learning algorithms for learning style classification. Although no new algorithm is proposed, the strength of this research lies in its comprehensive evaluation framework, large-scale dataset, and dual cross-validation strategy. The results confirm that Naïve Bayes offers superior stability and accuracy in real-world educational data. These findings can serve as a reliable foundation for the development of adaptive and personalized learning systems. One limitation of this study is that the learning style labels were based on students' self-reported preferences, which may introduce subjective bias. Future studies may incorporate additional validation methods to obtain more objective labeling. Future studies are encouraged to integrate the proposed classification model into learning management platforms and explore hybrid or deep learning approaches to further enhance performance.

The implications of this study indicate that machine learning-based learning style classification models can be used as a basis for developing adaptive and personalized digital learning systems. For further research, it is recommended to develop a learning material recommendation system based on learning style classification results and integrate it into the online learning platform so that it can be accessed more widely and flexibly.

REFERENCES

- [1] D. Supit, M. Melianti, E. M. M. Lasut, and N. J. Tumbel, "Gaya Belajar Visual, Auditori, Kinestetik terhadap Hasil Belajar Siswa," *J. Educ.*, vol. 5, no. 3, pp. 6994–7003, Feb. 2023, doi: 10.31004/joe.v5i3.1487.
- [2] S. Ramadandi and J. Jahring, "Student Learning Style Classification Using Naïve Bayes Classifier Method," *J. Teknol. dan Inf.*, vol. 10, no. 2, pp. 170–179, 2020, doi: 10.34010/jati.v10i2.3096.
- [3] W. WAAM and P. HKS, "Identifying the Learning Style of Students Using Machine Learning Techniques: An Approach of Felder Silverman Learning Style Model (FSLSM)," *Asian J. Res. Comput. Sci.*, vol. 17, no. 3, pp. 15–37, Jan. 2024, doi: 10.9734/ajrcos/2024/v17i3422.
- [4] K. Daniel, M. M. Msafiri, F. Antony, and X. Wan, "Motivate Students for Better Academic Achievement: A Systematic Review of Blended Innovative Teaching and Its Impact on Learning," *Comput. Appl. Eng. Educ.*, vol. 32, no. 4, 2024, doi: 10.1002/cae.22733.
- [5] L. Rahmawati and S. Gumindari, "Identifikasi Gaya Belajar (Visual, Auditorial Dan Kinestetik) Mahasiswa Tadris Bahasa Inggris Kelas 3F IAIN Syekh Nurjati Cirebon," *Pedagog. J. Pendidik.*, vol. 16, no. 1, pp. 54–61, 2021, doi: 10.33084/pedagogik.v16i1.1876.
- [6] M. A. Hassan, U. Habiba, F. Majeed, and M. Shoaib, "Adaptive gamification in e-learning based on students' learning styles," *Interact. Learn. Environ.*, vol. 29, no. 4, pp. 545–565, May 2021, doi: 10.1080/10494820.2019.1588745.
- [7] A. A. Nugraha and U. Budiyanto, "Adaptive E-Learning System Berbasis Vark Learning Style dengan Klasifikasi Materi Pembelajaran Menggunakan K-NN (K-Nearest Neighbor)," *Technomedia J.*, vol. 7, no. 2, pp. 248–261, Sep. 2022, doi: 10.33050/tmj.v7i2.1900.
- [8] I. Martiana, "Peningkatan Hasil Belajar Siswa Dalam Mengikuti Pelajaran Sosiologi Melalui Strategi Pembelajaran Berdiferensiasi," *J. Locus Penelit. Dan Pengabd.*, vol. 2, no. 6, pp. 500–506, 2023, doi: 10.58344/locus.v2i6.1194.
- [9] E. J. Ngaga, E. Sulastri, F. Bukangdonu, I. Alip, K. W. Wae, and H. Hikmatiar, "Efektivitas Ekstrakurikuler Dan Dampak Pada Prestasi," *Bul. Edukasi Indones.*, vol. 2, no. 01, pp. 24–32, 2023, doi: 10.56741/bei.v2i01.254.
- [10] A. Komoder, L. Venaruzzo, and C. Young, "Up-Skilling Learners Through Story-Telling: Introducing the Learning Wave," *Ascilite Publ.*, 2023, doi: 10.14742/apubs.2023.468.
- [11] M. Simatauw, "Pentingnya Pendidikan Agama Kristen Dalam Menerapkan Teori Bloom Di Sekolah," *Ichtus J. Teol. Dan Pendidik. Kristiani*, vol. 6, no. 2, pp. 69–77, 2025, doi: 10.63830/91nrh307.
- [12] H. A. Setiawan and H. Yuliansyah, "Aspect-Based Sentiment Analysis of User Reviews on the Game 'Honkai: Star Rail' Using Naïve Bayes Classifier," *SISTEMASI*, vol. 13, no. 5, p. 1956, Sep. 2024, doi: 10.32520/stmsi.v13i5.4343.
- [13] A. Salsabila, M. Nasution, and I. Irmayanti, "Implementation of the Naïve Bayes Algorithm to Predict New Student Admissions," *Build. Informatics, Technol. Sci.*, vol. 6, no. 1, pp. 421–429, 2024, doi: 10.47065/bits.v6i1.5363.
- [14] M. Garonga and R. Tanduk, "Comparison of Naive Bayes, Decision Tree, and Random Forest Algorithms in Classifying Learning Styles of Universitas Kristen Indonesia Toraja Students," *J. Tek. Inform.*, vol. 4, no. 6, pp. 1507–1514, Dec. 2023, doi: 10.52436/1.jutif.2023.4.6.1020.
- [15] P. Ramírez-Correa, J. Alfaro-Pérez, and M. Gallardo, "Identifying Engineering Undergraduates' Learning Style Profiles Using Machine Learning Techniques," *Appl. Sci.*, vol. 11, no. 22, p. 10505, Nov. 2021, doi: 10.3390/app112210505.
- [16] T. T. Huynh-Cam, L. S. Chen, and H. Le, "Using decision trees and random forest algorithms to predict and determine factors contributing to first-year university students' learning performance," *Algorithms*, vol. 14, no. 11, 2021, doi: 10.3390/a14110318.
- [17] T. Sutikno, A. Fadlil, Sunardi, I. Riadi, and A. Yudhana, "A review of Naive Bayes and decision tree methods for predicting particle size distribution in pharmaceutical manufacturing," *J. Intell. Syst. Eng. Appl. Data Sci.*, vol. 1, no. 1, pp. 8–14, 2024.

- [18] I. Arfiani and H. Yuliansyah, "A Hybrid Model of Graph Attention Networks and Random Forests for Link Prediction in Co-Authorship Networks," *Int. J. Adv. Data Inf. Syst.*, vol. 6, no. 2, pp. 276–289, 2025, doi: 10.59395/ijadis.v6i2.1382.
- [19] M. David and M. A. Firdaus, "Defect Rate Prediction in Manufacturing Process Using K-Nearest Neighbor Algorithm," *Int. J. Informatics Econ. Manag. Sci.*, vol. 3, no. 2, p. 161, 2024, doi: 10.52362/ijiems.v3i2.1599.
- [20] S. Nayak, M. Bhat, N. V. S. Reddy, and B. A. Rao, "Study of Distance Metrics on K - Nearest Neighbor Algorithm for Star Categorization," *J. Phys. Conf. Ser.*, vol. 2161, no. 1, p. 12004, 2022, doi: 10.1088/1742-6596/2161/1/012004.
- [21] S. Uddin, I. Haque, H. Lu, M. A. Moni, and E. Gide, "Comparative Performance Analysis of K-Nearest Neighbour (KNN) Algorithm and Its Different Variants for Disease Prediction," *Sci. Rep.*, vol. 12, no. 1, 2022, doi: 10.1038/s41598-022-10358-x.
- [22] D. Wijaya and A. R. Widiarti, "Batik classification using KNN algorithm and GLCM features extraction," *E3S Web Conf.*, vol. 475, p. 02012, Jan. 2024, doi: 10.1051/e3sconf/202447502012.
- [23] M. Nishom, "Perbandingan Akurasi Euclidean Distance, Minkowski Distance, dan Manhattan Distance pada Algoritma K-Means Clustering berbasis Chi-Square," *J. Inform. J. Pengemb. IT*, vol. 4, no. 1, pp. 20–24, Jan. 2019, doi: 10.30591/jpit.v4i1.1253.
- [24] M. Hossin and M. . Sulaiman, "A Review on Evaluation Metrics for Data Classification Evaluations," *Int. J. Data Min. Knowl. Manag. Process*, vol. 5, no. 2, pp. 01–11, Mar. 2015, doi: 10.5121/ijdkp.2015.5201.
- [25] H. Yuliansyah, R. A. P. Imaniati, A. Wirasto, and M. Wibowo, "Predicting Students Graduate on Time Using C4.5 Algorithm," *J. Inf. Syst. Eng. Bus. Intell.*, vol. 7, no. 1, p. 67, Apr. 2021, doi: 10.20473/jisebi.7.1.67-73.