International Journal of Information Technology, Research and Applications (IJITRA)

Rani Kurnia Putri, Muhammad Athoillah, Prayogo, Annisa Dwisulistyaningtyas (2026). Support Vector Machine Based Classification of Students' Mathematics Learning Preferences Using Educational Data Mining, 5(1), 32-44.

ISSN: 2583-5343

DOI:10.59461/ijitra.v5i1.228

The online version of this article can be found at:
https://www.ijitra.com/index.php/ijitra/issue/archive

Published by:
PRISMA Publications

International Journal of Information Technology, Research and Applications (IJITRA) is a journal that publishes articles which contribute new theoretical results in all the areas of Computer Science, Communication Network and Information Technology. Research paper and articles on Big Data, Machine Learning, IOT, Blockchain, Network Security, Optical Integrated Circuits, and Artificial Intelligence are in prime position.

PRISMA
PUBLICATIONS
https://www.prismapublications.com/

Support Vector Machine Based Classification of Students' Mathematics Learning Preferences Using Educational Data Mining

Rani Kurnia Putri[1], Muhammad Athoillah[2], Prayogo[3], Annisa Dwisulistyaningtyas[4]
[1,4]Department of Mathematics Education, Universitas PGRI Adi Buana, Surabaya, 60234, Indonesia
[2]Department of Statistics, Universitas PGRI Adi Buana, Surabaya, 60234, Indonesia
[3]Department of Primary Education, Universitas PGRI Adi Buana, Surabaya, 60234, Indonesia

Article Info

Article history:
Received Feb 02, 2026
Revised Mar 30, 2026
Accepted Mar 30, 2026

Keywords:
Kernel methods
Feature extraction
Learning analytics
Pattern recognition
Predictive modeling

ABSTRACT

Educational institutions increasingly seek data-driven approaches to understand students' learning preferences, particularly in mathematics education where differences in conceptual understanding across domains may influence learning outcomes. This study aims to develop a machine learning-based model to identify students' mathematics learning preferences across Algebra, Geometry, and Statistics using questionnaire data. The research was conducted at Universitas PGRI Adi Buana Surabaya involving undergraduate students in mathematics education. A quantitative approach was employed using questionnaires as the primary instrument to capture students' learning preference characteristics. The dataset was analyzed using the Support Vector Machine algorithm with a hold-out validation technique, where the data were divided into 75% training data and 25% testing data. Model performance was evaluated through multiple experimental trials using accuracy, precision, recall, and F1-score. The results show that the model achieved strong and stable performance across ten trials, obtaining an average accuracy of 97.05%, precision of 94.88%, recall of 98.32%, and F1-score of 96.39%.

Corresponding Author:

Rani Kurnia Putri
Department of Mathematics Education,
Universitas PGRI Adi Buana, Surabaya, 60234, Indonesia
Email: rani@unipasby.ac.id

1. INTRODUCTION

The rapid development of digital technologies and artificial intelligence has significantly transformed many sectors, including education. In recent years, educational systems worldwide have increasingly adopted data-driven approaches to understand students' learning behaviors and improve instructional strategies[1], [2], [3]. Among various academic subjects, mathematics remains one of the most critical disciplines due to its fundamental role in science, technology, engineering, and economic development. International educational assessments such as the Programme for International Student Assessment (PISA) consistently highlight the importance of mathematics proficiency as a key indicator of students' readiness for the modern workforce[4], [5]. However, global reports indicate that many students continue to struggle with mathematics learning, often demonstrating varying levels of engagement, motivation, and achievement across different mathematical domains. Public perception toward mathematics learning has also been widely discussed in educational research. Mathematics is often viewed as a challenging and abstract subject, leading to decreased interest and negative attitudes among students. Studies have shown that students' perceptions of mathematics significantly influence their motivation and learning outcomes[6], [7]. Moreover, mathematics itself is not a homogeneous discipline; rather, it consists of multiple domains such as algebra, geometry, and statistics, each requiring distinct cognitive processes and problem-solving strategies. While some students may demonstrate strong interest in algebraic reasoning, others may find greater engagement in spatial reasoning or data interpretation.

Global educational statistics further emphasize the need for innovative approaches to understanding students' learning preferences in mathematics. Large-scale educational data collected through digital learning platforms, surveys, and institutional databases offer opportunities for advanced data analysis techniques to uncover hidden patterns in student learning behaviors. Educational researchers increasingly recognize that analyzing these data using intelligent computational methods can provide valuable insights for improving teaching strategies and curriculum design[8], [9], [10]. Therefore, integrating artificial intelligence techniques into educational research has become an important direction for advancing data-driven decision-making in education. In this context, educational data mining has emerged as a promising field that applies data analysis and machine learning techniques to educational datasets. By analyzing student-related data, educational data mining can identify learning patterns, predict academic outcomes, and support personalized learning environments. Understanding students' learning preferences in mathematics through intelligent data analysis is particularly important, as it may help educators design more adaptive and engaging instructional approaches. Consequently, exploring machine learning methods to analyze and classify students' mathematics learning preferences represents a valuable contribution to both educational research and intelligent learning systems[3], [11].

Educational data mining is an interdisciplinary research area that focuses on developing methods to explore data originating from educational environments. The primary goal of educational data mining is to extract meaningful information from large educational datasets in order to improve learning processes and educational outcomes. Techniques commonly used in this field include classification, clustering, prediction, and pattern discovery. Among these techniques, classification plays a particularly important role in identifying patterns in student data and categorizing learners based on specific characteristics[1], [12], [13]. Machine learning algorithms are widely used in educational data mining to build predictive and classification models. Machine learning refers to computational methods that allow systems to learn patterns from data and make decisions without explicit programming. These algorithms are capable of analyzing complex datasets and identifying relationships between variables that may not be easily observable through traditional statistical analysis[14], [15], [16]. In educational contexts, machine learning has been applied to tasks such as predicting student performance, detecting learning difficulties, identifying dropout risks, and analyzing learning behaviors[1], [3], [17].

One of the most widely used machine learning algorithms for classification tasks is the Support Vector Machine (SVM). SVM is a supervised learning algorithm designed to find an optimal hyperplane that separates data points into different classes. The algorithm is known for its effectiveness in handling high-dimensional data and its strong generalization capability. Due to these advantages, SVM has been successfully applied in various domains, including text classification, image recognition, medical diagnosis, and educational data analysis[18], [19]. Previous studies have demonstrated the potential of machine learning techniques in educational research. For example, researchers have used machine learning algorithms to predict students' academic performance, classify learning styles, and identify factors influencing student engagement[3], [12], [20]. In mathematics education specifically, data-driven approaches have been applied to analyze problem-solving strategies, predict learning outcomes, and detect misconceptions. These studies highlight the importance of integrating computational methods into educational research to better understand complex learning processes[21], [22]. Despite these advances, relatively few studies have focused on classifying students' preferences across different domains of mathematics learning. Understanding these preferences is important because students may demonstrate varying levels of interest and engagement depending on the type of mathematical content being studied. Therefore, applying machine learning techniques such as SVM to classify students' mathematics learning preferences may provide valuable insights for educators and curriculum designers.

Although mathematics plays a central role in modern education, many students still experience difficulties and lack of motivation in learning this subject. International assessment results continue to show disparities in mathematics achievement across countries and educational systems[23], [24], [25]. One important factor contributing to these challenges is the diversity of students' learning preferences and interests. Students do not necessarily engage with all mathematical domains in the same way; some may prefer algebraic reasoning, while others may find geometry or statistical reasoning more appealing. However, traditional classroom practices often treat mathematics as a uniform subject without considering variations in students' domain-specific interests. Instructional strategies are frequently designed using a generalized approach that assumes all students respond similarly to different mathematical topics. As a result, teaching methods may fail to address individual differences in learning preferences, which can lead to decreased motivation and reduced learning effectiveness. Another challenge lies in the limited use of advanced data analytics in understanding students' learning behaviors[10], [26].

Although educational institutions collect large amounts of student-related data, these data are often underutilized for learning analysis and decision-making. Without proper analytical methods, valuable insights regarding students' learning tendencies remain hidden within these datasets. Consequently, educators may lack the information needed to design more personalized learning experiences. If these issues are not addressed, several negative consequences may arise. Students who consistently encounter difficulties in certain mathematical domains may develop negative attitudes toward mathematics as a whole. This situation may ultimately reduce students' interest in pursuing STEM-related fields, which are increasingly important in the global knowledge economy[20], [27]. Therefore, identifying students' mathematics learning preferences through data-driven approaches is essential for supporting more adaptive and personalized learning environments. Machine learning methods offer a promising solution to this problem. By analyzing educational datasets, machine learning algorithms can identify patterns that reveal students' learning preferences and engagement tendencies. Among various algorithms, Support Vector Machine has demonstrated strong performance in classification tasks involving complex datasets[28], [29], [30]. Thus, applying SVM to classify students' mathematics learning preferences represents a valuable approach to addressing the challenges described above.

Based on the issues discussed above, this study aims to develop a machine learning–based classification model to identify students' mathematics learning preferences using educational data mining techniques. The study focuses on analyzing students' responses related to their learning experiences, interests, and perceptions toward different mathematical domains, including algebra, geometry, and statistics. By utilizing the Support Vector Machine algorithm, this research seeks to construct a reliable classification model capable of detecting patterns in students' learning preferences from educational datasets. In order to achieve this aim, the study first analyzes educational data collected from secondary school students to understand the characteristics of their learning tendencies. Subsequently, the research develops a classification model using the Support Vector Machine approach to categorize students according to their preferred mathematics domains. Finally, the performance of the proposed model is evaluated using standard machine learning evaluation metrics such as accuracy, precision, recall, and F1-score to determine the effectiveness of the classification process. Through these steps, the study is expected to contribute to the development of intelligent learning analytics and provide valuable insights for designing more adaptive and personalized mathematics learning environments.

2. METHOD

2.1 Research Design

This study employed a quantitative research design using a machine learning approach within the framework of educational data mining. The purpose of the study was to develop a classification model capable of identifying students' mathematics learning preferences based on data collected from educational questionnaires. The research adopted a data-driven approach in which students' responses were treated as a dataset for training and evaluating a classification model. The overall research procedure consisted of several stages, including data collection, data preprocessing, model development, and model evaluation (Figure 1). First, educational data were collected from secondary school students through a structured questionnaire designed to measure students' learning interests and perceptions toward different mathematics domains. Second, the collected data were processed and prepared for machine learning analysis. Third, a classification model was developed using the Support Vector Machine algorithm. Finally, the performance of the model was evaluated using standard classification metrics to determine its effectiveness in predicting students' mathematics learning preferences.

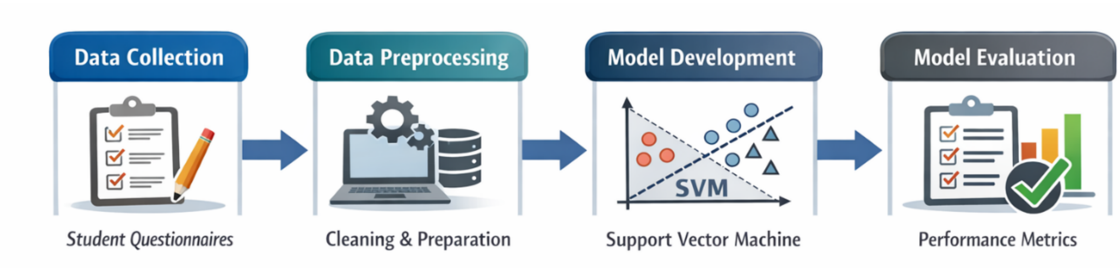

Figure 1. Research Design Diagram

2.2 Data Collection

The dataset used in this study was obtained from secondary school students who participated in a survey regarding their mathematics learning experiences and preferences. The questionnaire consisted of multiple indicators related to students' learning interest, perceived difficulty, motivation, and engagement in mathematics learning. These indicators were designed to capture students' attitudes toward different mathematical domains. A total of 449 student responses were collected and used as the dataset for this research. Each response contained 47 attributes representing students' answers to the questionnaire items designed to capture their learning tendencies and preferences when studying different areas of mathematics. These attributes describe various aspects of students' learning behaviors, such as their approaches to solving mathematical problems, their preference for symbolic or visual representations, and their ability to interpret numerical and contextual information. Each attribute corresponds to a specific questionnaire item and was transformed into numerical values so that the dataset could be processed using computational analysis techniques.

The target variable in the dataset represents the mathematics domain that is most preferred by each student. Based on the responses provided in the questionnaire, students' dominant learning preferences were categorized into three mathematics domains:

- ✓ Algebra
- ✓ Geometry
- ✓ Statistics

These domains were selected because they represent fundamental areas of mathematics learning that involve different cognitive processes and problem-solving strategies. Algebra is generally associated with symbolic manipulation, equation solving, and pattern recognition. Geometry focuses more on spatial reasoning, visualization, and the interpretation of geometric shapes and diagrams. Meanwhile, Statistics involves data interpretation, probability reasoning, and the analysis of numerical information within real-world contexts[31], [32].

The distribution of the dataset indicates that students' preferences vary across these three domains, reflecting the diversity of learning interests and cognitive strengths among students in mathematics education. Such variation is important for understanding how students approach mathematical problems and which types of mathematical concepts they feel more comfortable working with. From a computational perspective, the presence of multiple preference categories makes the dataset suitable for developing a multiclass classification model, where the objective is to predict the most dominant mathematics learning domain based on students' questionnaire responses.

2.3 Data Preprocessing

Before applying machine learning algorithms, several preprocessing steps were conducted to ensure the quality and usability of the dataset. Data preprocessing is an important step in educational data mining, as raw educational data often contain inconsistencies or require transformation before analysis. The preprocessing stage included data cleaning, attribute verification, and preparation of the dataset for machine learning analysis. Data cleaning involved checking for incomplete responses and removing any inconsistent or invalid entries. Attribute verification was performed to ensure that all variables were properly formatted and suitable for classification analysis. After cleaning and verification, the dataset was organized into feature variables and a target variable. The feature variables consisted of students' responses to the questionnaire items, while the target variable represented the preferred mathematics domain of each student. This structured dataset was then used as input for the machine learning classification model.

2.4 Classification Model

To classify students' mathematics learning preferences, this study employed the Support Vector Machine (SVM) algorithm. SVM is a supervised machine learning technique widely used for classification tasks due to its strong theoretical foundation and high generalization capability. The main principle of SVM is to determine an optimal hyperplane that separates data points belonging to different classes while maximizing the margin between the hyperplane and the closest data points from each class, known as support vectors[33], [34].

In this study, SVM was applied to perform multiclass classification in order to categorize students according to their preferred mathematics learning domain. The classification model was trained using students' responses obtained from a structured questionnaire. Each response forms a feature vector representing students' perceptions, learning interests, and experiences in mathematics learning. The target variable corresponds to the mathematics domain preferred by each student, which includes algebra, geometry, and statistics. Mathematically, the objective of SVM (Eq 1) is to determine the optimal separating hyperplane defined as:

$$w.x + b = 0 \tag{1}$$

where w represents the weight vector, x represents the input feature vector, and $b$represents the bias term. To maximize the separation margin while minimizing classification errors, the SVM optimization problem (Eq 2) can be formulated as follows:

$$\min_{w,b,\xi} \frac{1}{2} ||w||^2 + C \sum_{i=1}^{n} \xi_i \tag{2}$$

subject to the constraint

$$y_i(w.x_i + b) \geq 1 - \xi_{i,} \quad \xi_{i,} \geq 0$$

where

x_i represents the input data point,

y_i represents the class label,

$\xi_{i,}$ represents the slack variable allowing misclassification, and

C is the penalty parameter that controls the trade-off between maximizing the margin and minimizing classification errors.

Since many real-world datasets contain nonlinear relationships, SVM uses kernel functions to transform the original feature space into a higher-dimensional space where a linear separation can be achieved. In this study, the Gaussian kernel, also known as the radial basis function (RBF) kernel (Eq 3), was used. The Gaussian kernel is defined as:

$$K(x_i, x_j) = \exp\left(-\frac{\left\|x_i - x_j\right\|^2}{2\sigma^2}\right) \quad (3)$$

where x_i and , x_j represent two feature vectors and σ represents the kernel width parameter controlling the spread of the kernel function[19], [33], [35].

Because the dataset used in this research contains more than two classes, a multiclass classification strategy was required. The model implemented the one-versus-one (OvO) approach, in which a separate classifier is constructed for each pair of classes. For a dataset with three classes, the OvO method generates three binary classifiers (Figure 2). During prediction, each classifier votes for a class label, and the class receiving the highest number of votes is selected as the final prediction. Prior to model training, feature standardization was applied to normalize the dataset[19], [36]. Standardization transforms each feature to have a mean of zero and a standard deviation of one. This step ensures that all variables contribute equally to the learning process and prevents features with larger numerical scales from dominating the classification process.

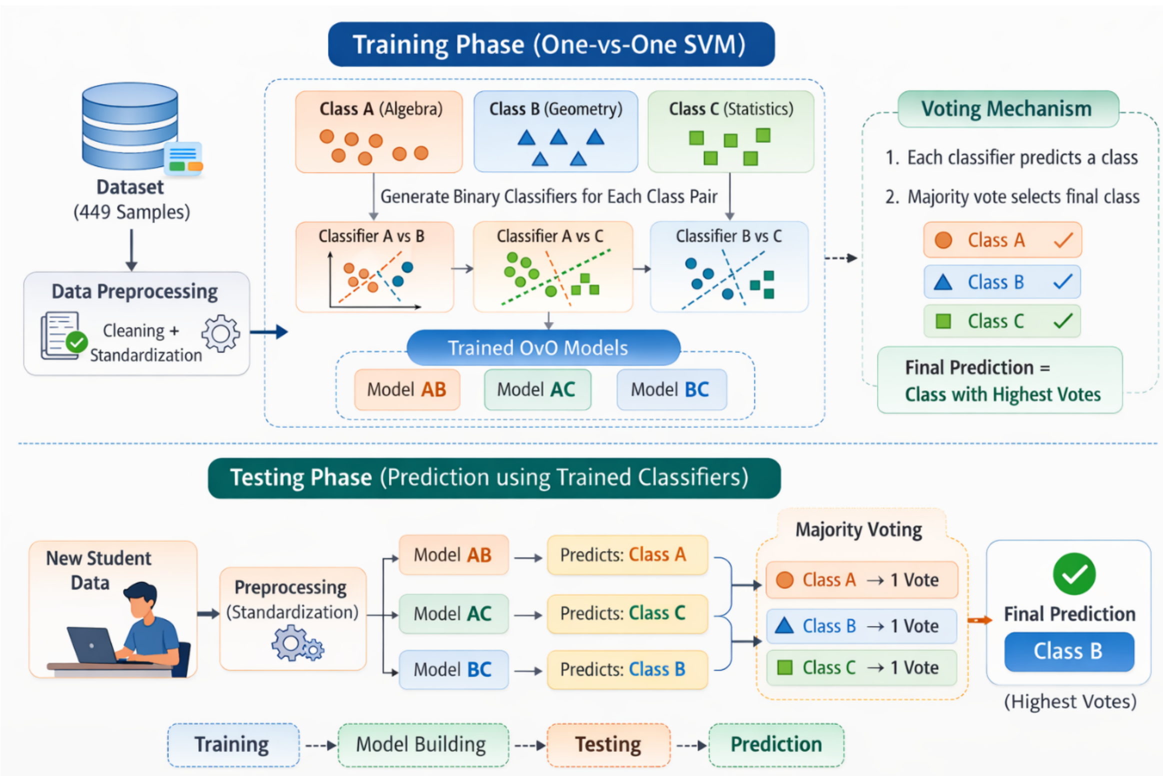

Figure 2. One-vs-One SVM classification workflow

The SVM model used in this study was configured using specific hyperparameter settings obtained during the experimental process. These parameters influence how the model learns patterns from the data and affect its classification performance.

Table 1. SVM Hyperparameter Configuration

Parameter	Value	Description
Kernel Function	Gaussian (RBF)	Kernel used to model nonlinear relationships
Box Constraint (C)	1	Controls trade-off between margin maximization and classification error
Kernel Scale	1.7	Determines width of the Gaussian kernel
Kernel Scale Mode	Manual	Kernel scale manually specified
Multiclass Method	One-vs-One	Strategy for handling multiple classes
Data Standardization	Enabled	Features normalized before training

Table 1 showed the configuration used in this framework. Using the above configuration, the SVM model was trained to identify patterns within the educational dataset and classify students according to their preferred mathematics learning domains. The trained model was subsequently evaluated using hold-out validation to assess its predictive performance on unseen data.

2.5 Model Evaluation

The performance of the classification model in this study was evaluated using the hold-out validation approach. Hold-out validation is a commonly used method in machine learning for assessing model performance by dividing the dataset into two independent subsets: a training set and a testing set. The training dataset is used to train the classification model, while the testing dataset is used to evaluate the model's predictive capability on unseen data. This approach allows researchers to assess how well the model generalizes to new data outside the training process[37], [38]. In this study, the dataset consisting of 449 student responses was randomly partitioned into two subsets. The majority of the data were used as the training set to construct the classification model using the Support Vector Machine algorithm, while the remaining portion was used as the testing set to evaluate the performance of the trained model. The separation between training and testing data ensures that the evaluation process reflects the model's ability to make predictions on previously unseen instances. To obtain reliable performance results, the hold-out validation process was repeated across multiple experimental runs. Each run involved randomly dividing the dataset into training and testing subsets, followed by model training and testing. This repeated evaluation helps reduce the influence of random variation in data splitting and provides a more stable estimate of the model's performance.

The effectiveness of the classification model was assessed using several standard performance metrics commonly applied in machine learning research, including accuracy, precision, recall, and F1-score. Accuracy measures the proportion of correctly classified instances relative to the total number of observations in the testing dataset. Precision evaluates the reliability of the model's predictions by measuring the proportion of correctly predicted instances among all predicted instances for a given class. Recall measures the model's ability to correctly identify instances belonging to a particular class. Meanwhile, the F1-score represents the harmonic mean of precision and recall, providing a balanced measure of classification performance[39], [40]. These evaluation metrics provide a comprehensive assessment of the effectiveness of the proposed classification model in identifying students' mathematics learning preferences.

3. RESULTS AND DISCUSSION

This section presents the findings obtained from the implementation of the Support Vector Machine (SVM) model for classifying students' mathematics learning preferences. The experimental evaluation was conducted through ten independent trials using hold-out validation, and the results are summarized based on the study objectives.

3.1 Classification Performance

To evaluate the effectiveness of the classification model, the experiment was conducted through ten independent trials using the hold-out validation approach. Each trial involved randomly splitting the dataset into training data (75%) and testing data (25%), after which the classification performance of the Support Vector Machine model was measured using four evaluation metrics: accuracy, precision, recall, and F1-score. The results obtained from all experimental trials are presented in Table 2.

Table 2. Performance Results of the SVM Model Across Ten Experimental Trials

Batch	Accuracy	Precision	Recall	F1-Score
1	97,32%	95,59%	98,44%	96,90%
2	96,43%	94,44%	97,98%	96,05%
3	97,32%	94,74%	98,46%	96,36%
4	94,64%	90,59%	97,01%	93,13%
5	97,32%	95,42%	98,46%	96,81%
6	96,43%	93,89%	97,98%	95,69%
7	98,21%	97,17%	98,96%	98,02%
8	96,43%	93,66%	97,98%	95,55%
9	97,32%	95,00%	98,44%	96,50%
10	99,11%	98,33%	99,47%	98,88%
Average	**97,05%**	**94,88%**	**98,32%**	**96,39%**

The results demonstrate that the classification model consistently achieved high performance across all experimental trials. The average accuracy obtained from the ten trials was 97.05%, indicating that the model correctly classified the majority of student learning preference instances. The precision value averaged 94.88%, suggesting that most predicted learning preference categories corresponded to the actual class labels. The recall metric obtained the highest average value of 98.32%, which indicates that the model was highly effective in identifying the correct classes with very few missed classifications. A high recall value is particularly important in classification tasks because it demonstrates that the model successfully detects most of the relevant instances within the dataset. Meanwhile, the average F1-score of 96.39% indicates a balanced performance between precision and recall, confirming that the classification model performs reliably across different evaluation perspectives.

An examination of the results across the ten trials also reveals the stability of the classification model. The accuracy values ranged between 94.64% and 99.11%, showing only a small variation despite the random partitioning of training and testing data. The highest classification performance was achieved in Batch 10, where the model reached 99.11% accuracy, 98.33% precision, 99.47% recall, and 98.88% F1-score. These values indicate that the model was able to identify nearly all student learning preference instances correctly in that trial. In contrast, the lowest performance occurred in Batch 4, where the model obtained 94.64% accuracy and 93.13% F1-score. Although this represents the lowest result among the trials, the performance level remains relatively high and still exceeds 90%, demonstrating that the model maintains reliable predictive capability even under less optimal data partitions. Another important observation is that the recall values remain consistently high across all trials, generally exceeding 97%. This suggests that the classification model rarely fails to identify the correct learning preference categories. The slight variations observed in precision values may indicate that some predicted classes overlap with other categories, which is common in educational datasets where student learning preferences may share similar characteristics.

Overall, the results indicate that the SVM-based classification model provides strong predictive performance and demonstrates consistent stability across multiple experimental trials. The high accuracy, precision, recall, and F1-score values confirm that the proposed model is effective in identifying patterns in students' responses and accurately classifying their mathematics learning preferences.

3.2 Class-Level Classification Performance

In addition to evaluating the overall performance of the classification model, further analysis was conducted to examine the classification performance for each mathematics learning preference category. The dataset in this study contains three learning preference domains: Algebra, Geometry, and Statistics. Evaluating the model at the class level allows a deeper understanding of how effectively the classification model identifies each learning preference category. Table 3 presents the average precision, recall, and F1-score for each mathematics learning domain obtained from the ten experimental trials.

Table 3. Average Classification Performance per Domain

Learning Domain	Precision	Recall	F1-Score
Algebra	97.39%	100.00%	97.39%
Geometry	87.26%	100.00%	93.10%
Statistics	100.00%	94.95%	97.40%

The results show that the classification model performs effectively across all three mathematics learning domains. The Algebra class achieved a precision of 97.39% and a perfect recall value of 100%, indicating that the model successfully identified all Algebra-related learning preference instances in the dataset. This suggests that the patterns associated with Algebra preferences are clearly distinguishable within the student response data. Similarly, the Statistics learning domain demonstrates strong classification performance. The precision value reached 100%, indicating that all predictions classified as Statistics were correct. The recall value of 94.95% indicates that the majority of Statistics preference instances were successfully detected, with only a small number of instances misclassified. The resulting F1-score of 97.40% confirms the balanced classification performance for this domain. The Geometry class shows slightly lower precision compared to the other domains, with a value of 87.26%, although its recall remains perfect at 100%. This means that while all Geometry instances were correctly identified, some instances predicted as Geometry may actually belong to other classes. This phenomenon may occur because certain characteristics of Geometry learning preferences share similarities with those of other mathematics domains, causing the model to occasionally assign overlapping predictions. Despite this slight variation, the overall class-level results indicate that the classification model is capable of effectively distinguishing between different mathematics learning preferences. The high recall values across all domains demonstrate that the model rarely fails to detect the correct class, while the high F1-scores confirm that the classification results remain balanced and reliable. These findings further confirm the effectiveness of the Support Vector Machine algorithm in modeling educational questionnaire data and identifying patterns related to students' mathematics learning preferences.

To provide a clearer understanding of the classification performance for each mathematics learning domain, graphical visualizations were generated for Algebra, Geometry, and Statistics based on the results obtained from the ten experimental trials (Figure 3). Each graph presents the performance metrics across all batch, allowing the trend and stability of the classification model to be observed for each domain.

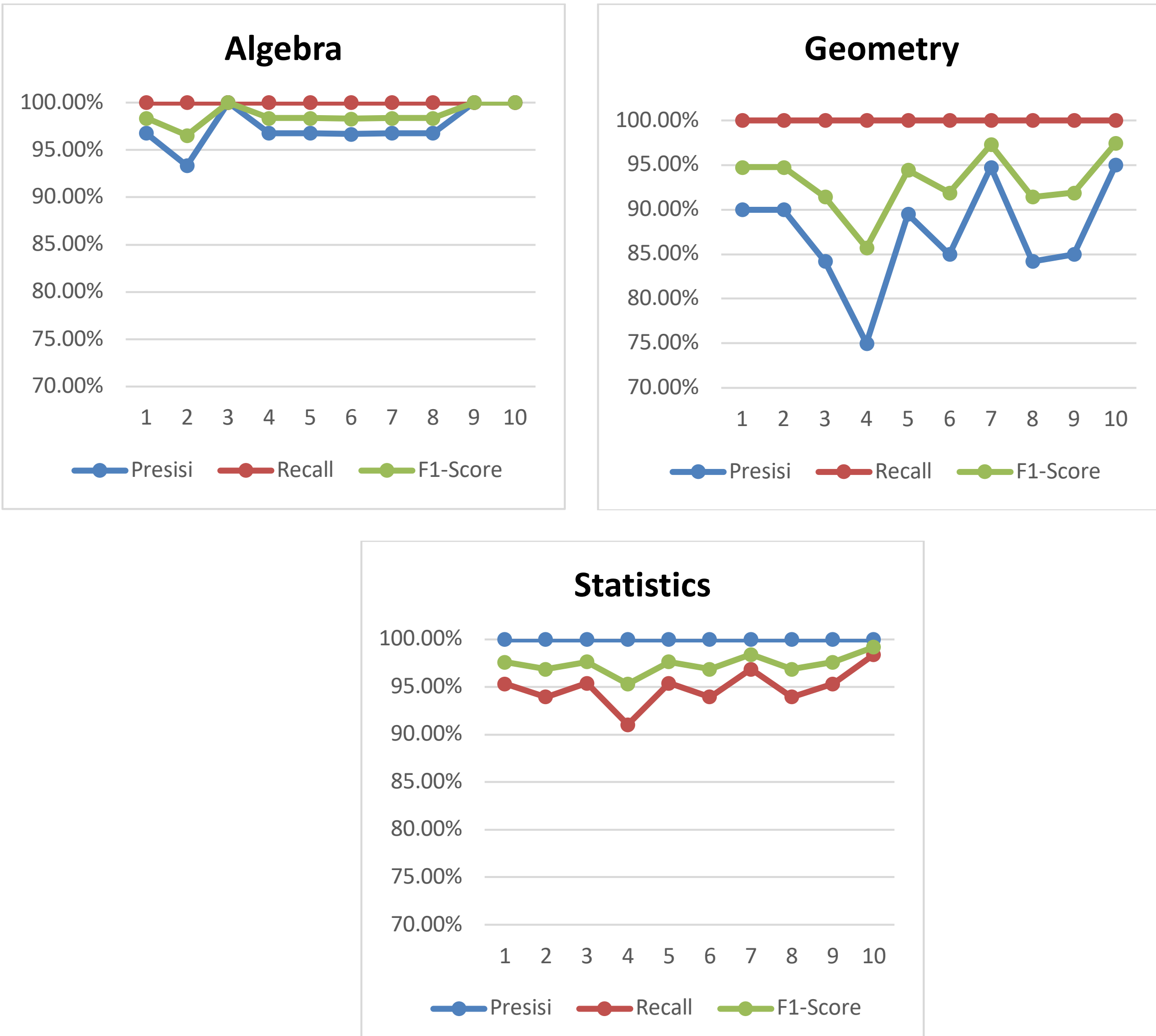

Figure 3. Classification Performance by Domain

For the Algebra domain, the graphical results show consistently high classification performance across all batches. The precision values remain stable throughout the experiments, reaching nearly perfect values in most trials. Meanwhile, recall values remain consistently high, indicating that the classification model successfully identified almost all Algebra-related instances. As a result, the F1-score values for Algebra remain close to 1.00 across the ten batches, demonstrating that the model performs reliably in identifying learning preferences associated with Algebra. The stability observed in the graph suggests that Algebra-related patterns within the dataset are clearly distinguishable by the Support Vector Machine classification model.

In contrast, the Geometry domain exhibits slightly greater variation across the ten batches. While recall values remain consistently high, indicating that the model successfully detects most Geometry instances, the precision values fluctuate more noticeably compared to the Algebra domain. This variation suggests that some predictions classified as Geometry may overlap with other learning preference domains. Such behavior may occur because the characteristics associated with Geometry learning preferences share similarities with those found in other mathematical topics. Despite this variation, the F1-score values remain relatively high across all batches, indicating that the classification model still performs effectively in identifying Geometry-related learning preferences.

In addition, for the Statistics domain, the graphical results demonstrate strong and stable classification performance similar to the Algebra domain. Precision values remain consistently high across all batches, often reaching perfect values, indicating that predictions for the Statistics class are highly accurate. Recall values also remain high throughout the experiments, showing that the majority of Statistics instances are successfully detected by the model. Consequently, the F1-score values remain consistently high across the ten batches, reflecting balanced classification performance. The stability observed in the Statistics graph suggests that the classification model effectively captures the distinguishing characteristics of this learning preference domain.

Overall, the graphical visualization of classification performance across the three domains highlights the robustness of the classification model. While minor variations occur across certain batches, particularly in the Geometry domain, the results consistently demonstrate high performance across all learning preference categories. These findings further confirm that the classification approach is capable of effectively identifying patterns within students' questionnaire responses and accurately classifying their mathematics learning preferences.

3.3 Discussion

The results obtained in this study demonstrate that the classification model achieved consistently high performance in identifying students' mathematics learning preferences based on questionnaire data. Across the ten experimental trials, the model produced an average accuracy of 97.05%, accompanied by high precision, recall, and F1-score values. These results indicate that the classification model is capable of effectively capturing patterns within students' responses and accurately categorizing their learning preferences into the appropriate mathematics domains. The relatively small variation observed across the experimental trials further confirms the stability of the model. Even when the dataset was randomly partitioned into training and testing subsets, the classification results remained consistently strong, suggesting that the model successfully learned meaningful patterns within the dataset rather than relying on a specific data distribution.

The strong classification performance observed in this study can largely be attributed to the characteristics of the Support Vector Machine algorithm. SVM is widely recognized for its ability to identify optimal decision boundaries in high-dimensional datasets. In the context of this research, each questionnaire response represents a feature describing a student's learning preference tendencies. When these responses are combined, they form a multidimensional feature space that reflects students' learning characteristics. The SVM algorithm constructs an optimal separating hyperplane that maximizes the margin between different classes, allowing the model to effectively distinguish between different learning preference categories. This capability is particularly useful when analyzing educational datasets where relationships between variables may be complex and multidimensional. Another factor contributing to the effectiveness of the model is the structured nature of questionnaire-based educational data. Students who prefer a particular mathematics domain often exhibit consistent response patterns across related questions. These patterns create distinguishable clusters in the dataset that can be effectively separated by the classification model. As a result, the model is able to achieve high recall values across the experimental trials, indicating that it successfully identifies the majority of relevant learning preference instances. The balanced precision and recall values reflected in the high F1-scores further confirm that the classification model performs reliably across different evaluation metrics. Further analysis of the results also reveals differences in classification performance across the three mathematics learning domains: Algebra, Geometry, and Statistics. The Algebra domain demonstrates the most stable classification performance among the three domains. Precision values remain consistently high across all experimental trials, while recall values approach perfect levels in most cases. These results indicate that the model is highly effective in identifying Algebra-related learning preferences and rarely misclassifies instances belonging to this domain. The strong performance observed in Algebra suggests that the learning characteristics associated with this domain are clearly represented in the dataset, making it easier for the classification model to identify patterns related to Algebra learning preferences.

The Geometry domain, on the other hand, exhibits slightly greater variation in classification performance compared to Algebra. While recall values remain consistently high across the experimental trials, precision values fluctuate more noticeably. This indicates that although the model successfully detects most Geometry-related instances, some predictions labeled as Geometry may actually belong to other domains. Such variations may occur because the learning characteristics associated with Geometry often overlap with those of other mathematical domains. Skills such as spatial reasoning, conceptual visualization, and logical thinking may also play roles in other areas of mathematics, which can lead to occasional classification overlap. Nevertheless, the F1-score values for Geometry remain relatively high across all batches, indicating that the model still performs effectively in identifying Geometry-related learning preferences. In contrast, the Statistics domain demonstrates strong classification performance similar to the Algebra domain. Precision values frequently reach perfect or near-perfect levels across the experimental trials, indicating that predictions assigned to the Statistics class are highly reliable. The recall values also remain consistently high, suggesting that the classification model successfully identifies most Statistics-related learning preference instances. Consequently, the F1-score values remain high throughout the experimental trials, confirming balanced performance between precision and recall. The stability observed in the Statistics domain suggests that the distinguishing characteristics associated with Statistics learning preferences are effectively captured by the classification model. Overall, the domain-level analysis indicates that the classification model is capable of effectively distinguishing between different mathematics learning preferences. While minor variations occur across certain domains, particularly in Geometry, the classification model consistently achieves high performance across all three learning preference categories. These findings highlight the potential of machine learning approaches for analyzing educational datasets and identifying patterns in students' learning behaviors.

From an educational perspective, the findings of this study demonstrate the potential benefits of applying machine learning techniques to analyze students' learning preferences. Identifying students' dominant mathematics learning domains may help educators design more personalized learning strategies that align with students' strengths and interests. For example, instructional approaches may be adapted to emphasize particular mathematical concepts that match students' preferred learning patterns. Additionally, automated classification of learning preferences may assist educators in identifying students who may require additional support or targeted interventions in specific mathematical areas. By integrating machine learning techniques into educational data analysis, researchers and educators can gain deeper insights into student learning behaviors and develop more adaptive learning environments that support diverse learning needs.

4. CONCLUSION

This study aimed to develop a classification model capable of identifying students' mathematics learning preferences based on questionnaire responses. The findings indicate that the proposed classification approach successfully achieved the objectives of the study. By applying the Support Vector Machine algorithm, the model demonstrated consistently high performance across multiple experimental trials, achieving an average accuracy of 97.05%, along with high precision, recall, and F1-score values. These results confirm that the classification model is capable of effectively identifying patterns in students' responses and accurately categorizing learning preferences into the three mathematics domains: Algebra, Geometry, and Statistics. The domain-level analysis further revealed that the model performed most consistently in the Algebra and Statistics domains, while minor variations were observed in the Geometry domain due to potential overlaps in learning characteristics. The findings of this study highlight the effectiveness of machine learning techniques for analyzing questionnaire-based educational data. From a practical perspective, the results suggest that such classification models can support educators in understanding students' dominant learning preferences and potentially assist in developing more personalized instructional strategies in mathematics education. The main contribution of this study lies in demonstrating the feasibility of applying a machine learning classification approach to identify students' mathematics learning domains using questionnaire data. In addition, the novelty of this research lies in integrating educational preference analysis with a Support Vector Machine–based classification framework to automatically detect patterns in students' learning tendencies across different mathematical domains.

Despite the promising results obtained in this study, several limitations should be acknowledged. First, the dataset used in this research was relatively limited in size and derived from a specific educational context, which may affect the generalizability of the findings to broader student populations. Second, the study focused only on three mathematics learning domains—Algebra, Geometry, and Statistics—while other mathematical areas or interdisciplinary learning preferences were not considered. Third, this research utilized a single classification algorithm, and the performance comparison with other machine learning models was not extensively explored. Future research may address these limitations by incorporating larger and more diverse datasets from different educational institutions to improve the robustness and generalizability of the classification model. Additionally, future studies may explore the integration of other machine learning approaches such as ensemble methods or deep learning techniques to compare classification performance and further improve prediction accuracy. Expanding the scope of analysis to include additional mathematical domains or incorporating behavioral learning data, such as students' interaction patterns in digital learning environments, may also provide deeper insights into students' learning preferences and contribute to the development of more adaptive educational systems.

FUNDING INFORMATION

The authors gratefully acknowledge the financial support provided by the Adi Buana Innovative Research Program at Universitas PGRI Adi Buana Surabaya through contract number 120.1.41/Kontrak/LPPM/XI/2025.

DATASET AVAILABILITY STATEMENT

The data supporting the results of this study can be obtained from the corresponding author upon reasonable request. Due to privacy considerations related to the research participants, the dataset is not publicly available and access is subject to certain restrictions.

CONFLICT OF INTEREST STATEMENT

All authors declare that they have no conflicts of interest.

REFERENCES

[1] D. Kučak, V. Juričić, and G. Đambić, "Machine Learning In Education-A Survey Of Current Research Trends," *Annals of DAAAM & Proceedings*, vol. 29, 2018.

[2] C. Korkmaz and A.-P. Correia, "A review of research on machine learning in educational technology," *EMI. Educ. Media Int.*, vol. 56, no. 3, pp. 250–267, 2019.

[3] K. Ahmad *et al.*, "Data-driven artificial intelligence in education: A comprehensive review," *IEEE Transactions on Learning Technologies*, vol. 17, pp. 12–31, 2023, doi: 10.1109/TLT.2023.3314610.

[4] O. Publishing, *PISA 2018 results (volume V): Effective policies, successful schools*. Organisation for Economic Co-operation and Development OECD, 2020.

[5] OECD, "PISA 2022 Results (Volume I and II) - Country Notes: Indonesia," 2022, Accessed: Feb. 22, 2026. [Online]. Available: https://www.oecd.org/en/publications/pisa-2022-results-volume-i-and-ii-country-notes_ed6fbcc5-en/indonesia_c2e1ae0e-en.html

[6] A. Renninger, S. Hidi, and M. Nieswandt, "Interest in mathematics and science learning," 2015.

[7] L. Verschaffel, S. Schukajlow, J. Star, and W. Van Dooren, "Word problems in mathematics education: a survey," *ZDM*, vol. 52, no. 1, pp. 1–16, Jan. 2026, doi: 10.1007/s11858-020-01130-4.

[8] K. Rybinski and E. Kopciuszewska, "Will artificial intelligence revolutionise the student evaluation of teaching? A big data study of 1.6 million student reviews," *Assess. Eval. High. Educ.*, vol. 46, no. 7, pp. 1127–1139, 2021.

[9] K. McKnight, K. O'Malley, R. Ruzic, M. K. Horsley, J. J. Franey, and K. Bassett, "Teaching in a Digital Age: How Educators Use Technology to Improve Student Learning," *Journal of Research on Technology in Education*, vol. 48, no. 3, pp. 194–211, Jul. 2016, doi: 10.1080/15391523.2016.1175856.

[10] K. Pechenkina, "Artificial intelligence for good? Challenges and possibilities of AI in higher education from a data justice perspective," *Higher Education for good: Teaching and learning futures (# HE4Good). Cambridge, UK: Open Book Publishers*, 2023.

[11] L. Lin, D. Zhou, J. Wang, and Y. Wang, "A systematic review of big data driven education evaluation," *Sage Open*, vol. 14, no. 2, p. 21582440241242180, 2024, doi: 10.1177/21582440241242180.

[12] N. A. Jalil, H. J. Hwang, and N. M. Dawi, "Machines learning trends, perspectives and prospects in education sector," in *Proceedings of the 3rd International Conference on Education and Multimedia Technology*, 2019, pp. 201–205.

[13] L. Zhou, S. Pan, J. Wang, and A. V Vasilakos, "Machine learning on big data: Opportunities and challenges," *Neurocomputing*, vol. 237, pp. 350–361, 2017.

[14] H. Taherdoost, "Machine learning algorithms: features and applications," in *Encyclopedia of Data Science and Machine Learning*, IGI Global Scientific Publishing, 2023, pp. 938–960.

[15] H. Li, L. Lin, and H. Zeng, *Machine learning methods*. Springer, 2004.

[16] R. K. Putri and M. Athoillah, "Artificial intelligence and machine learning in digital transformation: Exploring the role of AI and ML in reshaping businesses and information systems," in *Advances in Digital Transformation-Rise of Ultra-Smart Fully Automated Cyberspace*, IntechOpen, 2024.

[17] W. Forero-Corba and F. N. Bennasar, "Techniques and applications of Machine Learning and Artificial Intelli gence in education: a systematic review," *RIED-Revista Iberoamericana de Educación a Distancia*, vol. 27, no. 1, 2024, doi: 10.5944/ried.27.1.37491.

[18] J. Cervantes, F. Garcia-Lamont, L. Rodríguez-Mazahua, and A. Lopez, "A comprehensive survey on support vector machine classification: Applications, challenges and trends," *Neurocomputing*, vol. 408, pp. 189–215, 2020.

[19] D. A. Pisner and D. M. Schnyer, "Support vector machine," in *Machine learning*, Elsevier, 2020, pp. 101–121.

[20] O. F. Marzuki *et al.*, "Innovating Education: A Comprehensive Review of STEM Education Approaches," *International Journal of Academic Research in Progressive Education and Development*, vol. 13, Jan. 2024, doi: 10.6007/IJARPED/v13-i1/20490.

[21] M. Mavrikis and M. Margeti, *Review of mathematics education in the age of artificial intelligence*. Taylor &Francis, 2025.

[22] R. Hidayat, M. Z. bin Mohamed, N. N. binti Suhaizi, N. binti M. Sabri, M. K. H. bin Mahmud, and S. N. binti Baharuddin, "Artificial intelligence in mathematics education: A systematic literat ure review," *International Electronic Journal of Mathematics Education*, vol. 17, no. 3, p. em0694, Mar. 2026, doi: 10.29333/iejme/12132.

[23] OECD, "PISA 2022 Assessment and Analytical Framework," *PISA*, Aug. 2023, doi: 10.1787/dfe0bf9c-en.

[24] A. Fritz, V. G. Haase, and P. Rasanen, *International handbook of mathematical learning difficulties*. Springer, 2019.

[25] S. Schukajlow, K. Rakoczy, and R. Pekrun, "Emotions and motivation in mathematics education: Where we are today and where we need to go," *ZDM–Mathematics Education*, vol. 55, no. 2, pp. 249–267, 2023.

[26] L. Chen and S. Xiao, "Perceptions, challenges and coping strategies of science teachers in t eaching socioscientific issues: A systematic review," *Educ. Res. Rev.*, vol. 32, p. 100377, Jan. 2026, doi: 10.1016/j.edurev.2020.100377.

[27] K. Maaß, V. Geiger, M. Romero Ariza, and M. Goos, "The Role of Mathematics in interdisciplinary STEM education," *ZDM*, vol. 51, pp. 1–16, Oct. 2019, doi: 10.1007/s11858-019-01100-5.

[28] S. O. Abdulsalam, R. A. Ganiyu, E. O. Omidiora, S. O. Olabiyisi, and M. O. Arowolo, "Performance Evaluation of Support Vector Machine Kernel Functions on Students' Educational Data Set," in *Social Media and Crowdsourcing*, Auerbach Publications, 2023, pp. 125–149.

[29] F. Utaminingrum, I. K. Somawirata, S. Mayena, A. Septiarini, and T. K. Shih, "Analysis of Kernel Performance in Support Vector Machine Using Seven Features Extraction for Obstacle Detection," *Int. J. Control Autom. Syst.*, vol. 21, no. 1, pp. 281–291, Dec. 2025, doi: 10.1007/s12555-021-0702-z.

[30] M. Athoillah, I. Irawan, M., and M. Imah, Elly, "Study Comparison of SVM-, K-NN- and Backpropagation-Based Classifier for Image Retrieval," *Jurnal Ilmu Komputer dan Informasi (Journal of Computer Science and Information)*, 2015.

[31] S. Lerman, *Encyclopedia of mathematics education*. Springer, 2020.

[32] M. Van den Heuvel-Panhuizen and P. Drijvers, *Realistic Mathematics Education*. Springer International Publishing, 2026. doi: 10.1007/978-3-030-15789-0_170.

[33] C. Campbell and Y. Ying, *Learning with support vector machines*. Springer Nature, 2022.

[34] M. Athoillah, E. Purnaningrum, and R. K. Putri, "Modified Multi-Kernel Support Vector Machine for Mask Detection," *CommIT (Communication and Information Technology) Journal*, vol. 16, no. 2, pp. 159–166, 2022.

[35] T. Wang, L. Zhang, and W. Hu, "Bridging deep and multiple kernel learning: A review," *Information Fusion*, vol. 67, pp. 3–13, 2021.

[36] Y. Guo, Z. Zhang, and F. Tang, "Feature selection with kernelized multi-class support vector machine," *Pattern Recognit.*, vol. 117, p. 107988, 2021, doi: 10.1016/j.patcog.2021.107988.

[37] Y. Xu and R. Goodacre, "On splitting training and validation set: a comparative study of cross-validation, bootstrap and systematic sampling for estimating the generalization performance of supervised learning," *J. Anal. Test.*, vol. 2, no. 3, pp. 249–262, 2018.

[38] K. McAlinn and K. Takanashi, "Optimal Hold-Out Size in Cross-Validation," *arXiv preprint arXiv:2511.12698*, 2025, doi: 10.48550/arXiv.2511.12698.

[39] Ž. Vujović, "Classification model evaluation metrics," *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 6, pp. 599–606, 2021, doi: 10.14569/IJACSA.2021.0120670.

[40] G. Naidu, T. Zuva, and E. M. Sibanda, "A review of evaluation metrics in machine learning algorithms," in *Computer science on-line conference*, Springer, 2023, pp. 15–25. doi: 10.1007/978-3-031-35314-7_2.

BIOGRAPHIES OF AUTHORS

Rani Kurnia Putri earned her B.Sc. degree in Mathematics from Universitas Brawijaya and later obtained her M.Sc. degree in Mathematics from Institut Teknologi Sepuluh Nopember. She currently serves as a Lecturer (Lektor) in the Mathematics Education Study Program at Universitas PGRI Adi Buana Surabaya. Her research focuses on mathematics education and artificial intelligence, particularly in areas such as mathematical problem solving, machine learning, Support Vector Machines, multi-kernel learning, and deep learning applications for tasks including hoax detection and image recognition. She has published scholarly articles in nationally accredited journals as well as international journals indexed in major databases, including Scopus. She can be contacted via email at rani@unipasby.ac.id

Muhammad Athoillah obtained his B.Sc. and M.Sc. degrees in Mathematics from Institut Teknologi Sepuluh Nopember in 2013 and 2015, respectively. Since 2017, he has been serving as a Lecturer in the Statistics Study Program, Faculty of Science and Technology at Universitas PGRI Adi Buana Surabaya. His research interests encompass machine learning, data mining, pattern recognition, image processing, and text processing, with particular emphasis on Support Vector Machines and deep learning techniques. He has published research articles in nationally accredited and internationally indexed journals, including those indexed in Scopus and Web of Science, and has also authored several academic books. He can be contacted via email at athoillah@unipasby.ac.id.

Prayogo is a scholar and researcher who serves as a Lecturer in the Master's Program in Elementary Education at Universitas PGRI Adi Buana Surabaya. His academic expertise includes Elementary Mathematics Education, Computational Modeling, and Research Methodology, where he contributes to the development of pedagogical competencies among future educators. His research interests encompass mathematical communication skills, mathematical modeling and computation, and problem-solving structures, as well as the integration of ethnomathematics in educational practices. In addition, his work extends to areas of educational psychology and educational management, particularly focusing on student resilience, self-efficacy, and differentiated learning strategies. He can be contacted through his official university email at prayogo@unipasby.ac.id.

Annisa Dwi Sulistyaningtyas received her B.Sc. and M.Sc. degrees in Mathematics from Institut Teknologi Sepuluh Nopember. She currently serves as a Lecturer in the Mathematics Education Study Program at Universitas PGRI Adi Buana Surabaya. Her research interests include mathematical modeling, mathematics education, and the integration of technology in mathematics learning, particularly through the use of digital learning media and artificial intelligence–assisted instructional design. Her scholarly work also explores mathematical communication, problem-based learning, and the development of innovative learning resources such as GeoGebra-assisted materials and educational games for mathematics learning. She has published research articles in national and international academic forums, including conference proceedings and peer-reviewed journals in the fields of mathematics education and applied mathematics. She can be contacted via her official university email at: annisadwistyas@unipasby.ac.id