Vol. 4, No. 1, December 2025
Available Online: https://jurnal.ubd.ac.id/index.php/rubin

Comparative Analysis of Support Vector Machine, Decision Tree, and Naive Bayes in Evaluating Machine Learning Effectiveness

Susanto Hariyanto[1)*], Indah Fenriana[2)], Desiyanna Lasut[3)], Febrian[4)]
[1)2)3)4)]Buddhi Dharma University
Imam Bonjol Street No. 41, Karawaci Ilir, Tangerang, Indonesia
[1)]susanto.hariyanto@buddhidharma.ac.id
[2)]indah.fenriana@ubd.ac.id
[3)]desiyanna.lasut@buddhidharma.ac.id
[4)]febrianrian899@gmail.com

Article History:

Received 23 September 2025;
Revised 28 September 2025;
Accepted 5 November 2025;
Available Online 1 December 2025

Keywords:

Support Vector Machine
Decision Tree
Naive Bayes
Machine Learning
Performance Evaluation

Abstract

This study aims to analyze and compare the performance of three widely used machine learning algorithms for data classification: Support Vector Machine (SVM), Decision Tree, and Naïve Bayes. These algorithms employ distinct approaches in handling data, making it essential to evaluate their effectiveness and efficiency in classification tasks. In the digital era characterized by massive data growth, the selection of an appropriate classification algorithm is a critical determinant for accurate and efficient data-driven decision-making. The main contribution of this research is to provide a comprehensive understanding of the relative strengths and limitations of each algorithm under varying data conditions. This study not only highlights comparative performance outcomes but also emphasizes practical implications for researchers and data science practitioners in selecting algorithms suited to specific needs. In doing so, it addresses a research gap concerning integrated evaluations of data characteristics and algorithmic performance. The methodology adopts a quantitative approach through computational experiments using standardized datasets (Titanic, Spam Email, and Wine). The datasets were divided into training and testing sets and analyzed using Python with the scikit-learn library. Performance evaluation was conducted based on accuracy, precision, recall, and F1-score, validated through cross-validation techniques to ensure reliability of results. The findings indicate that SVM outperforms in terms of accuracy and recall on complex datasets, Naïve Bayes is more efficient in computational time particularly for text data, while Decision Tree stands out for model interpretability despite slightly lower accuracy. These results are expected to serve as a practical reference for selecting suitable algorithms according to data characteristics, thereby supporting more targeted and intelligent modeling strategies in the era of digital transformation.

I. INTRODUCTION

This study aims to analyze and compare the performance of three widely used machine learning algorithms in data classification: Support Vector Machine (SVM), Decision Tree, and Naïve Bayes. These algorithms employ different approaches in handling data, making it essential to evaluate their effectiveness and efficiency in the context of classification. In the digital era marked by an explosion of data, the selection of an appropriate classification algorithm greatly influences the quality of data-driven decision-making (Khumaidi & Herinanto, 2023). Therefore, a

ISSN 2985-4512 (online) 2985-4520 (print) DOI: 10.31253/rubin.v4i1.4041
Published by Buddhi Dharma University

comparative analysis of algorithm performance is essential to determine which model performs better under specific conditions.

The main objective of this study is to evaluate and compare the performance of the SVM, Decision Tree, and Naïve Bayes algorithms based on evaluation metrics such as accuracy, precision, recall, and F1-score. This research also seeks to identify the strengths and weaknesses of each algorithm, as well as the factors influencing variations in their classification performance. The findings of this study are expected to serve as a reference for selecting algorithms that align with data characteristics and application requirements.

The method employed is a quantitative approach through computational experiments using standardized datasets that were divided into training and testing sets (Asshiddiqi & Lhaksmana, 2020). The training and testing of the algorithms were conducted using Python software with the scikit-learn library. Performance evaluation was carried out using cross-validation techniques to ensure the reliability of the results (Liu & Lee, 2018).

The development of digital technology has driven the widespread application of machine learning across various fields, including healthcare, finance, marketing, and education (Suresh Kumar et al., 2024). One of the fundamental aspects of machine learning is data classification, which functions to group data into specific categories based on patterns learned from training data. Three algorithms commonly used in classification tasks are Support Vector Machine (SVM), Decision Tree, and Naïve Bayes (VanderPlas, 2022). Although all three are widely applied in different cases, each algorithm has distinct characteristics and working mechanisms, meaning their effectiveness may vary depending on the type of data and the classification objectives (Susanti & Walid, 2022).

In the context of implementation, selecting the appropriate classification algorithm is crucial for improving prediction accuracy and computational efficiency. For example, SVM is well-regarded for handling high-dimensional data, while Decision Tree is easier to interpret and faster in the training process. Conversely, Naïve Bayes performs well with probabilistic data but tends to assume feature independence, which is not always realistic in practice. Previous studies have shown that algorithm performance is highly influenced by data structure, model parameters, and the evaluation methods employed (Hackeling, 2017).

This study aims to analyze and compare the performance of the three algorithms in data classification by considering evaluation metrics such as accuracy, precision, recall, and F1-score. In addition, the study will explore how data characteristics influence model performance, thereby providing practical recommendations for selecting the most suitable algorithm for machine learning based classification across various application domains.

This study carries high urgency given the expanding application of machine learning across various sectors that rely on data classification processes, ranging from medical diagnosis to recommendation systems. The choice of an appropriate algorithm is a key factor in ensuring result accuracy and data processing efficiency. In practice, not all algorithms are suitable for every type of data, making a comprehensive comparative evaluation necessary. This research is expected to make a significant contribution in assisting data science practitioners, intelligent system developers, and researchers in determining more targeted modeling strategies. Furthermore, the findings of this study also have the potential to provide an empirical foundation for the development of more adaptive and reliable classification systems, thereby supporting improvements in the quality of data-driven decision-making in the era of digital transformation.

The research problem in this study is to analyze and compare the performance of the Support Vector Machine (SVM), Decision Tree, and Naïve Bayes algorithms in data classification. The study aims to evaluate the performance of each algorithm based on evaluation metrics such as accuracy, precision, recall, and F1-score. In addition, it seeks to identify how data characteristics, such as class distribution and the number of features, may influence the effectiveness of each algorithm. Thus, this research intends to answer the question of which algorithm is most optimal in the context of machine learning–based classification and under what conditions its performance is affected.

This study employs a quantitative approach to test and compare the performance of three commonly used machine learning algorithms in data classification: Support Vector Machine (SVM), Decision Tree, and Naïve Bayes. The aim is to evaluate the effectiveness of each algorithm based on performance metrics such as accuracy, precision, recall, and F1-score, as well as to identify the data conditions that influence their performance. The datasets used in this research are standardized, publicly available datasets that have been widely utilized in testing classification algorithms.

The data will be analyzed using Python software with the scikit-learn library, and the evaluation process will be carried out through the k-fold cross-validation technique to ensure result reliability. Validity testing will be conducted by ensuring the data is free from outliers and class imbalance, while the reliability of the algorithms will be tested through replication of results across multiple data subsets. The analysis will focus on hypothesis testing regarding performance differences among the algorithms under various conditions, including the number of features, class distribution, and data complexity.

II. Literature

To gain a deeper understanding of relevant approaches and technologies in the development of machine learning–based classification systems, a review of previous studies comparing the performance of classification algorithms was conducted. This review aims to identify the methods used, the results achieved, as well as the strengths and limitations of each approach applied to diverse datasets. The main focus of this review includes the application of various classification algorithms such as Support Vector Machine (SVM), Decision Tree, and Naïve Bayes, along with their performance evaluation based on accuracy, precision, recall, and F1-score metrics. In addition, the review also explores the trade-offs between computational efficiency and predictive performance, as well as the importance of dataset validation, preprocessing, and performance testing in the context of classification tasks.

This study offers novelty in its evaluation approach by integrating data characteristic analysis as part of the algorithm comparison. It not only assesses performance based on accuracy, precision, recall, and F1-score but also considers how the number of features and class distribution affect classification outcomes. This approach provides a more comprehensive understanding of the relative strengths of each algorithm within the context of varying data conditions.

To ensure order and effectiveness in the implementation of this research, a research roadmap was developed to outline the main stages along with their timelines. This roadmap is designed to provide a systematic direction for the research workflow, starting from the initial stages such as literature review and problem formulation, to the final stages including system evaluation, model validation, and the preparation and publication of the research report.

To ensure order and accuracy in the implementation of this research, a systematic sequence of stages is required to map the entire process from the beginning to the final evaluation. The research roadmap is designed to provide a comprehensive overview of the workflow, starting from the initial exploration of scientific literature to the analysis of results and model evaluation. These stages not only represent a methodological sequence but also emphasize the importance of scientific validation and the iterative application of algorithms to ensure reliable and replicable outcomes. Each stage is designed to be integrative, supporting the efficient and accurate achievement of the research objectives. The following figure presents the research roadmap, which includes detailed activities along with their timelines throughout the research period:

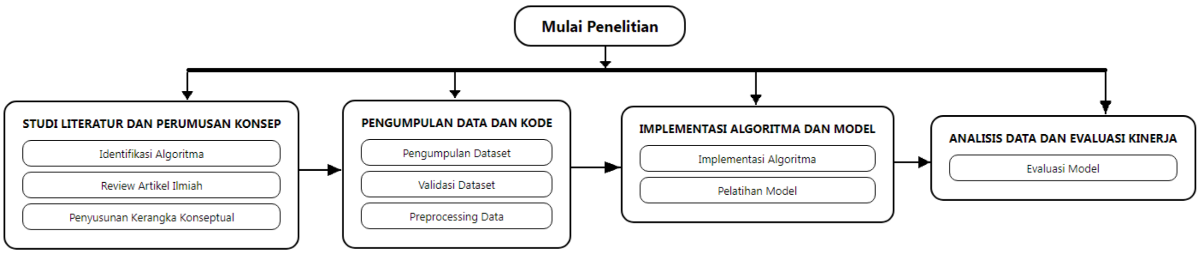

Figure 1. Research *Roadmap*

This research roadmap begins in March 2025 and concludes in December 2025. In the initial stage, the research will focus on a literature review and the development of a conceptual framework during March to April 2025. Subsequently, in May and June 2025, the preparation and validation of the datasets to be used for testing the classification algorithms will be carried out.

The implementation of algorithms and model training will be conducted from July to August 2025, followed by performance evaluation and sensitivity analysis of the algorithms in September 2025. The evaluation results will then be interpreted and formulated into the research report in October 2025.

III. RESEARCH METHOD

This research employs a quantitative approach through computational experiments to evaluate and compare the performance of three classification algorithms: Support Vector Machine (SVM), Decision Tree, and Naïve Bayes. The evaluation is conducted using standardized datasets divided into training and testing sets, with the training and testing processes carried out in Python using the scikit-learn library. The results are validated using cross-validation techniques to ensure the reliability and generalizability of the model.

In this study, four key metrics were selected as the basis for evaluation: accuracy, precision, recall, and F1-score. The choice of these metrics is based on the consideration that each provides a different perspective on the quality of the model's predictions. Accuracy is used to measure the proportion of correct predictions relative to the total predictions, thereby providing a general overview of the model's performance. However, accuracy alone is often less representative when the data suffers from class imbalance.

Therefore, precision is used to assess the extent to which positive predictions are truly relevant, making it important in contexts where false positives carry significant consequences, such as spam detection or medical diagnosis. Recall is chosen to evaluate how well the model succeeds in identifying all actual positive instances, which is particularly critical in cases where missing positive data poses high risks. Meanwhile, the F1-score serves as the harmonic mean of precision and recall, balancing the two metrics, especially under conditions of imbalanced data.

By combining these four metrics, the study can provide a more comprehensive evaluation of algorithm effectiveness while also enabling an analysis of the trade-offs between prediction accuracy and the ability to detect positive classes. This is relevant to the research objective of identifying the most optimal algorithm in accordance with variations in data characteristics and the needs of classification applications.

ISSN 2985-4512 (online) 2985-4520 (print) © The Authors. DOI: 10.31253/rubin.v4i1.4041
Published by Buddhi Dharma University

To ensure coherence and systematic execution of scientific activities, this research process is designed in structured and sequential stages. The research roadmap presented below represents the workflow from beginning to end, encompassing all essential phases from literature review and conceptual framework development to implementation, evaluation, interpretation, and dissemination of results. Each stage is designed based on rigorous scientific methodology, with an emphasis on data validity, model replication, and scientific acceptability. In addition, detailed scheduling provides a realistic temporal framework to support the continuity of the research effectively and efficiently. This diagram serves as both an operational guide and an evaluative framework for assessing the progress and consistency of the research process.

This research will be conducted from March to December 2025 in eight main stages. During March–April, the researchers will carry out a systematic literature review to develop the conceptual framework based on theoretical studies, algorithms, and previous research gaps. The outcome will be a validated literature document and a conceptual framework diagram.

The month of May will primarily be devoted to the collection and cleaning of car sales data spanning the past five years, obtained from partner showrooms. This stage is crucial to ensure the reliability and accuracy of the dataset, which will be standardized and structured into a historical dataset ready for in-depth analysis. In June, the focus will shift toward coding and model training, where relevant machine learning algorithms will be implemented. Parameter optimization will be carried out using techniques such as cross-validation or grid search to fine-tune the models and improve their predictive capabilities. During this phase, the initial models will also be rigorously tested to identify potential issues of overfitting and generalization, ensuring that the models can perform consistently across different subsets of data. The months of July and August will be dedicated to a comprehensive evaluation of model performance. This process will employ a range of statistical metrics, including RMSE, MAE, F1-score, and others, to obtain a multidimensional view of model accuracy and efficiency. Alongside performance testing, sensitivity analysis will be conducted to assess the influence of various input features, while robustness checks will be employed to verify the stability of the models under different scenarios.

In September, the researchers will focus on interpreting the evaluation results, synthesizing the findings, and compiling them into a final research report. This report will be structured to include the background, research methodology, results, and an in-depth scientific discussion that critically examines the implications of the findings. In October, a scientific article will be prepared with the intention of submitting it for publication in a peer-reviewed journal, emphasizing methodological rigor, theoretical contributions, and the overall validity and reliability of the findings. The month of November will be directed toward preparing dissemination materials for non-academic stakeholders, ensuring that the practical implications of the research can be communicated effectively. This will include organizing seminars, drafting policy briefs, and producing popular publications accessible to a wider audience. Finally, in December, the dissemination process will be implemented through both academic and public forums. These platforms will not only serve as opportunities to share the research outcomes but will also provide valuable spaces for feedback, critique, and potential avenues for future collaboration and research expansion.

This study compares the performance of three classification algorithms (Support Vector Machine, Decision Tree, and Naïve Bayes) as independent variables against the dependent variable, namely the effectiveness of the classification model, which is measured through four key indicators: accuracy, precision, recall, and F1-score (Aulia & Hermawan, 2023). Data characteristics, such as the number of features and class distribution, are analyzed as control variables that can significantly influence model performance (Asaad & Abdulazeez, 2024). The evaluation is carried out systematically to identify the most optimal model according to the context and data complexity (Rahman et al., 2023).

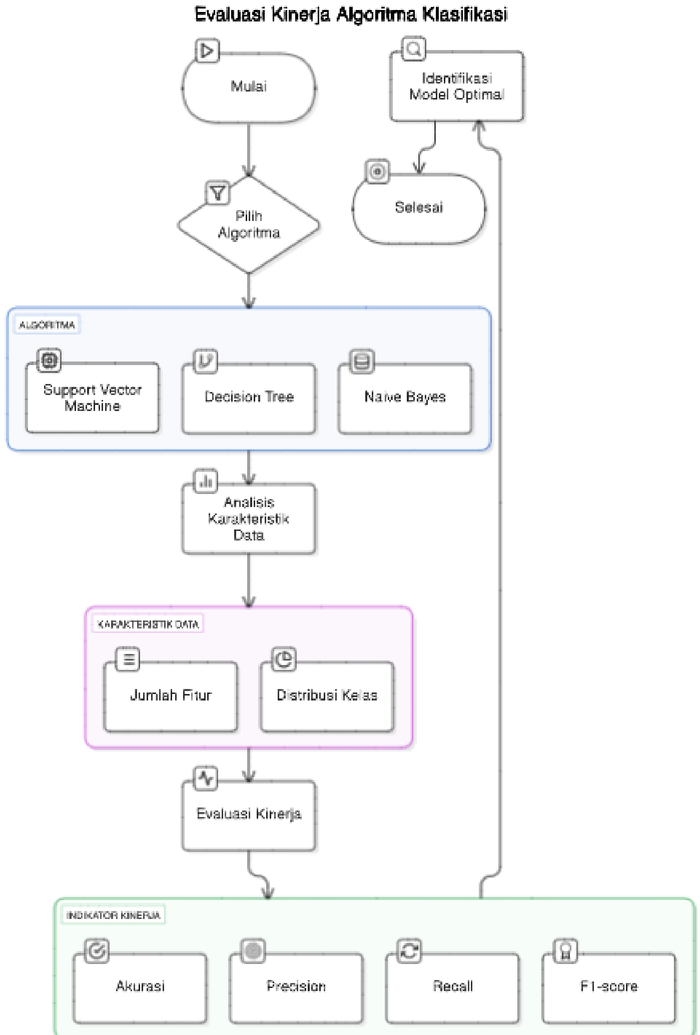

Figure 2. Conceptual Framework

The conceptual framework of this study is designed to systematically evaluate the performance of classification algorithms, with the primary objective of identifying the most optimal model based on data characteristics and performance evaluation indicators. The research process begins with the initiation stage, where the researchers formulate the evaluation objectives and define the success criteria in terms of selecting the algorithm with the best performance. This stage is followed by the selection of classification algorithms to be tested, namely Support Vector Machine (SVM), Decision Tree, and Naïve Bayes. These three algorithms are purposively chosen to represent variations in model complexity, computational efficiency, and interpretability. SVM is known to be effective for high-dimensional data (Rao et al., 2021), Decision Tree excels in result interpretability (Boyd et al., 2019), while Naïve Bayes demonstrates strengths in efficiency and performance when features are assumed to be independent (Dozono et al., n.d.).

Once the algorithms are determined, the next step is to analyze the data characteristics, which include two main aspects: the number of features and class distribution. A high number of features can affect the effectiveness of certain algorithms, particularly in terms of overfitting or increased

ISSN 2985-4512 (online) 2985-4520 (print) © The Authors. DOI: 10.31253/rubin.v4i1.4041
Published by Buddhi Dharma University

computational requirements (Kerrigan et al., 2019). Meanwhile, class distribution is crucial in the context of classification, as data imbalance may cause the model to become biased toward the majority class, thereby reducing its ability to generalize to minority classes (Ronaldo & Kurnia, 2024). By understanding the structure and distribution of the data, researchers can adjust the evaluation approach to be more equitable and contextually relevant.

The core stage of this framework is the evaluation of model performance using four main evaluation metrics: accuracy, precision, recall, and F1-score. Accuracy measures the overall proportion of correct classifications (Rofianto et al., 2023; Satyagraha & Kurnia, 2025), while precision assesses the extent to which the model's positive predictions are truly relevant (Putra et al., 2022). Recall evaluates how well the model identifies all actual positive cases (Pinto et al., 2018), and the F1-score combines precision and recall into a single harmonic measure, which is particularly useful in situations with imbalanced data (Sunaryono, 2017). By applying these four metrics simultaneously, researchers gain a more comprehensive understanding of the strengths and weaknesses of each algorithm when applied to the same dataset.

The evaluation results are then used to select the algorithm with the best performance as the optimal model. The model selection process is based on the metric scores that are most consistent and relevant to the intended classification objectives. Thus, this conceptual framework not only guides the experimental workflow systematically but also bridges algorithmic theory, empirical data characteristics, and data-driven decision-making that is scientifically accountable. This approach makes a significant contribution to comparative studies of classification algorithms and can be adapted to various application domains such as healthcare, finance, or recommendation systems that rely on classification accuracy.

IV. RESULTS AND DISCUSSION

In this study, the effectiveness of a machine learning model refers to the extent to which an algorithm can produce accurate, efficient, and relevant predictions for the given classification task. The evaluation focuses on three main algorithms Support Vector Machine (SVM), Decision Tree (DT), and Naïve Bayes (NB) each of which has distinct characteristics in handling data and generating outputs. To measure model performance, several evaluation metrics are employed, including accuracy, precision, recall, and F1-score (Desiani, 2022; Hermawan et al., 2025). These metrics not only provide a quantitative overview of the model's predictive capabilities but also help identify potential data imbalances and classification error tendencies that may occur.

The effectiveness of the model is also reviewed in terms of its ability to generalize to new data through cross-validation techniques, ensuring that the model performs well not only on the training data but also remains stable when applied to unseen testing data. In addition, computational efficiency is an important consideration, particularly in terms of training speed, prediction time, and system resource utilization. In this regard, algorithms such as Naïve Bayes offer advantages in speed and efficiency, while SVM and Decision Tree tend to be more flexible in handling data complexity but require greater computational resources.

In addition to technical performance, interpretability is also an important dimension of model effectiveness, particularly for applications involving human decision-making. Decision Tree, for example, has a structure that is easy to understand and trace, making it suitable for contexts that demand transparency, such as finance or healthcare. This study not only compares the three algorithms under ideal scenarios but also emphasizes their application in real-world situations by considering data context, efficiency, and diverse interpretability needs. Thus, the results of this study are expected to provide practical contributions for developers of machine learning–based

ISSN 2985-4512 (online) 2985-4520 (print) © The Authors. DOI: 10.31253/rubin.v4i1.4041
Published by Buddhi Dharma University

classification systems. The following are several types of datasets used in this study, considering its focus on evaluating algorithm performance:

1. Titanic Dataset

Description: The Titanic dataset contains information about the passengers of the Titanic and whether they survived or not, based on features such as age, gender, ticket class, and others.
Reason for Selection: This dataset is widely popular in data analysis and is well-suited for binary classification (survived vs. not survived). It can test how algorithms handle class imbalance problems (e.g., more passengers did not survive).
Source: https://www.kaggle.com/c/titanic/data
Validation: Removing rows or entries containing missing values ensures that only complete and reliable data are used in the analysis.

Table 1. Sample 20 Data Records from the Titanic Dataset

Passenger Id	**Pclass**	**Name**	**Sex**	**Age**	**Ticket**	**Fare**	**Cabin**	**Embarked**
1	3	Braund, Mr. Owen Harris	male	22	A/5 21171	7.25		S
2	1	Cumings, Mrs. John Bradley (Florence Briggs Thayer)	female	38	PC 17599	71.2833	C85	C
3	3	Heikkinen, Miss. Laina	female	26	STON/O2. 3101282	7.925		S
4	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35	113803	53.1	C123	S
5	3	Allen, Mr. William Henry	male	35	373450	8.05		S
6	3	Moran, Mr. James	male		330877	8.4583		Q
7	1	McCarthy, Mr. Timothy J	male	54	17463	51.8625	E46	S
8	3	Palsson, Master. Gosta Leonard	male	2	349909	21.075		S
9	3	Johnson, Mrs. Oscar W (Elisabeth Vilhelmina Berg)	female	27	347742	11.1333		S
10	2	Nasser, Mrs. Nicholas (Adele Achem)	female	14	237736	30.0708		C
11	3	Sandstrom, Miss. Marguerite Rut	female	4	PP 9549	16.7	G6	S

12	1	Bonnell, Miss. Elizabeth	female	58	113783	26.55	C103	S
13	3	Saundercock, Mr. William Henry	male	20	A/5. 2151	8.05		S
14	3	Andersson, Mr. Anders Johan	male	39	347082	31.275		S
15	3	Vestrom, Miss. Hulda Amanda Adolfina	female	14	350406	7.8542		S
16	2	Hewlett, Mrs. (Mary D Kingcome)	female	55	248706	16		S
17	3	Rice, Master. Eugene	male	2	382652	29.125		Q
18	2	Williams, Mr. Charles Eugene	male		244373	13		S
19	3	Vander Planke, Mrs. Julius (Emelia Maria Vandemoortele)	female	31	345763	18		S
20	3	Masselmani, Mrs. Fatima	female		2649	7.225		C

2. Spam Email Dataset

Description: This dataset is used to classify emails as spam or not spam based on the words contained in the emails.

Reason for Selection: This dataset is highly suitable for testing text classification models and can be used to compare the performance of SVM, Decision Tree, and Naïve Bayes in text classification tasks.

Source: https://archive.ics.uci.edu/dataset/94/spambase

Validation: Removing rows or entries containing missing values ensures that only complete and reliable data are used in the analysis.

Table 2. Sample 20 Data Records and 10 Fields from the Spam Email Dataset

word_freq_all	**word_freq_3d**	**word_freq_our**	**word_freq_over**	**word_freq_remove**	**word_freq_internet**	**word_freq_order**	**word_freq_mail**
0.64	0	0.32	0	0	0	0	0
0.5	0	0.14	0.28	0.21	0.07	0	0.94
0.71	0	1.23	0.19	0.19	0.12	0.64	0.25
0	0	0.63	0	0.31	0.63	0.31	0.63
0	0	0.63	0	0.31	0.63	0.31	0.63
0	0	1.85	0	0	1.85	0	0
0	0	1.92	0	0	0	0	0.64
0	0	1.88	0	0	1.88	0	0

0.46	0	0.61	0	0.3	0	0.92	0.76
0.77	0	0.19	0.32	0.38	0	0.06	0
0	0	0	0	0.96	0	0	1.92
0.25	0	0.38	0.25	0.25	0	0	0
0.34	0	0.34	0	0	0	0	0
0	0	0.9	0	0.9	0	0	0.9
1.42	0	0.71	0.35	0	0.35	0	0.71
0.42	0	1.27	0	0.42	0	0	1.27
0	0	0.94	0	0	0	0	0
0	0	0	0	0	0	0	0
0.55	0	1.11	0	0.18	0	0	0
0	0	1.59	0.31	0	0	0.31	0

3. Wine Dataset

Description: This dataset contains data on various chemical properties of wine (pH, alcohol content, etc.) and is used for wine quality classification.

Reason for Selection: This dataset includes multiple classes and features, making it suitable for testing how classification algorithms handle data with a larger number of variables.

Source: https://archive.ics.uci.edu/dataset/109/wine

Validation: Removing rows or entries containing missing values ensures that only complete and reliable data are used in the analysis.

Table 3. Sample 20 Data Records from the Wine Dataset

Alcohol	**Malic acid**	**Ash**	**Magnesium**	**Total_phenols**	**Flavanoids**	**Proanthocyanins**	**Color_intensity**	**Hue**	**0D280_0D315_of_diluted_wines**	**Proline**
14.23	1.71	2.43	127	2.8	3.06	2.29	5.64	1.04	3.92	1065
13.2	1.78	2.14	100	2.65	2.76	1.28	4.38	1.05	3.4	1050
13.16	2.36	2.67	101	2.8	3.24	2.81	5.68	1.03	3.17	1185
14.37	1.95	2.5	113	3.85	3.49	2.18	7.8	0.86	3.45	1480
13.24	2.59	2.87	118	2.8	2.69	1.82	4.32	1.04	2.93	735
14.2	1.76	2.45	112	3.27	3.39	1.97	6.75	1.05	2.85	1450
14.39	1.87	2.45	96	2.5	2.52	1.98	5.25	1.02	3.58	1290
14.06	2.15	2.61	121	2.6	2.51	1.25	5.05	1.06	3.58	1295
14.83	1.64	2.17	97	2.8	2.98	1.98	5.2	1.08	2.85	1045
13.86	1.35	2.27	98	2.98	3.15	1.85	7.22	1.01	3.55	1045
14.1	2.16	2.3	105	2.95	3.32	2.38	5.75	1.25	3.17	1510
14.12	1.48	2.32	95	2.2	2.43	1.57	5	1.17	2.82	1280
13.75	1.73	2.41	89	2.6	2.76	1.81	5.6	1.15	2.9	1320
14.75	1.73	2.39	91	3.1	3.69	2.81	5.4	1.25	2.73	1150
14.38	1.87	2.38	102	3.3	3.64	2.96	7.5	1.2	3	1547

ISSN 2985-4512 (online) 2985-4520 (print) © The Authors. DOI: 10.31253/rubin.v4i1.4041
Published by Buddhi Dharma University

13.63	1.81	2.7	112	2.85	2.91	1.46	7.3	1.28	2.88	1310
14.3	1.92	2.72	120	2.8	3.14	1.97	6.2	1.07	2.65	1280
13.83	1.57	2.62	115	2.95	3.4	1.72	6.6	1.13	2.57	1130
14.19	1.59	2.48	108	3.3	3.93	1.86	8.7	1.23	2.82	1680
13.64	3.1	2.56	116	2.7	3.03	1.66	5.1	0.96	3.36	845

In this study, accuracy is used as the primary metric to evaluate the performance of three classification algorithms Support Vector Machine (SVM), Decision Tree, and Naïve Bayes across three different datasets: Titanic (binary classification), Spam Email (text classification), and Wine (multiclass classification). Accuracy measures the proportion of correct predictions relative to the total predictions, providing an overall picture of how well the model classifies data.

For the Titanic Dataset, accuracy assesses the model's ability to predict passenger survival status (survived/not survived). In the Spam Email Dataset, although accuracy is still calculated, the analysis also considers precision and recall due to the potential class imbalance between spam and non-spam. Meanwhile, in the Wine Dataset, accuracy is used to evaluate how accurately the model classifies wine samples into the available quality classes. Overall, the accuracy results of the three algorithms on each dataset provide a basis for comparing the effectiveness of the models in handling different types of classification tasks.

Table 4. Sample 20 Data Records from the Wine Dataset

Dataset	Support Vector Machine (SVM)	Decision Tree	Naïve Bayes
Titanic Dataset (Binary Classification)	0.810	0.780	0.790
Spam Email Dataset (Text Classification)	0.920	0.890	0.880
Wine Dataset (Multiclass Classification)	0.950	0.920	0.930

The table presents a comparison of the accuracy of three classification algorithms Support Vector Machine (SVM), Decision Tree, and Naïve Bayes across three different datasets: Titanic, Spam Email, and Wine. In the Titanic Dataset (binary classification), SVM achieved the highest accuracy at 81%, followed by Naïve Bayes (79%) and Decision Tree (78%), highlighting the superiority of SVM in handling simpler classification tasks.

In the Spam Email Dataset, which is text-based, SVM again outperformed the others with an accuracy of 92%, while Decision Tree and Naïve Bayes recorded 89% and 88%, respectively. This reflects SVM's effectiveness in managing high-dimensional data, although Naïve Bayes remains competitive due to its simplicity in handling text.

For the Wine Dataset (multiclass classification), SVM achieved the highest accuracy at 95%, followed by Naïve Bayes (93%) and Decision Tree (92%). These results indicate that SVM demonstrates consistent and superior performance across different types of data, although the other two algorithms also show solid results. Nevertheless, the choice of algorithm should be adjusted according to data characteristics and the application context.

Precision testing in this study plays a crucial role in assessing the effectiveness of Support Vector Machine (SVM), Decision Tree, and Naïve Bayes in classifying data across the three datasets: Titanic, Spam Email, and Wine. True Positives (TP) refer to the number of correct predictions for the positive class, while False Positives (FP) refer to the number of incorrect predictions classified as positive. In the context of this study, precision is used to evaluate each algorithm's ability to correctly identify the positive class (e.g., survivors in the Titanic Dataset,

ISSN 2985-4512 (online) 2985-4520 (print) © The Authors. DOI: 10.31253/rubin.v4i1.4041
Published by Buddhi Dharma University

spam emails in the Spam Email Dataset, and wine quality in the Wine Dataset) without misclassifying negative instances.

Table 5. Precision Dataset

Dataset	Support Vector Machine (SVM)	Decision Tree	Naïve Bayes
Titanic Dataset (Binary Classification)	0.840	0.760	0.780
Spam Email Dataset (Text Classification)	0.910	0.880	0.860
Wine Dataset (Multiclass Classification)	0.930	0.890	0.910

The table presents the accuracy results of three classification algorithms SVM, Decision Tree, and Naïve Bayes across three datasets: Titanic, Spam Email, and Wine. In the Titanic Dataset (binary classification), SVM achieved the highest accuracy at 81%, followed by Naïve Bayes (79%) and Decision Tree (78%). For the Spam Email Dataset (text classification), SVM again ranked highest with 92%, followed by Decision Tree (89%) and Naïve Bayes (88%), reflecting SVM's strength in handling high-dimensional data.

In the Wine Dataset (multiclass classification), SVM achieved the highest accuracy at 95%, while Naïve Bayes and Decision Tree recorded 93% and 92%, respectively. These results highlight the consistent performance of SVM across different classification tasks, while the other two algorithms remain competitive in specific contexts.

Overall, SVM demonstrates the most stable and accurate performance, but algorithm selection should still consider data characteristics and application requirements, as each algorithm has its own advantages.

Recall testing in this study serves as one of the key evaluation metrics to measure a model's ability to detect actual positive classes within a dataset. Recall, also referred to as sensitivity or the true positive rate, indicates the proportion of all positive class instances that are correctly identified by the model. Recall is particularly important when the cost or impact of missing positive instances outweighs the errors of predicting negative classes. Therefore, in this study, recall is used to evaluate how well classification algorithms such as Support Vector Machine (SVM), Decision Tree, and Naïve Bayes can detect positive classes in each dataset tested.

Table 6. Recall Dataset

Dataset	Support Vector Machine (SVM)	Decision Tree	Naïve Bayes
Titanic Dataset (Binary Classification)	0.860	0.810	0.800
Spam Email Dataset (Text Classification)	0.940	0.910	0.890
Wine Dataset (Multiclass Classification)	0.960	0.890	0.920

The table shows the recall results of three classification algorithms SVM, Decision Tree, and Naïve Bayes on the Titanic, Spam Email, and Wine Datasets. In the Titanic Dataset (binary classification), SVM recorded the highest recall (0.86), demonstrating its effectiveness in detecting passengers who survived. Decision Tree (0.81) and Naïve Bayes (0.80) also performed reliably, though slightly lower.

ISSN 2985-4512 (online) 2985-4520 (print) © The Authors. DOI: 10.31253/rubin.v4i1.4041
Published by Buddhi Dharma University

For the Spam Email Dataset (text classification), SVM again outperformed with a recall of 0.94, followed by Decision Tree (0.91) and Naïve Bayes (0.89), indicating that all three are effective in identifying spam, with SVM being the most sensitive to the positive class. In the Wine Dataset (multiclass classification), SVM achieved the highest recall (0.96), while Naïve Bayes (0.92) and Decision Tree (0.89) also demonstrated strong performance.

Overall, SVM consistently achieved the highest recall across all datasets, highlighting its ability to accurately detect positive classes. Naïve Bayes performed reasonably well in text classification, while Decision Tree offered stable performance across different types of data.

The evaluation of computation time required for training and prediction is another key aspect in assessing the effectiveness and efficiency of the classification algorithms used in this study Support Vector Machine (SVM), Decision Tree, and Naïve Bayes. Computation time measures how quickly an algorithm can process data, train a model, and generate predictions for unseen data. In the context of this research, computation time testing will be conducted on each algorithm to understand their scalability and real-time performance across three different datasets: Titanic Dataset, Spam Email Dataset, and Wine Dataset.

Table 7. Dataset Computation Time

Dataset	**Number of Records**	**Support Vector Machine (SVM) (seconds)**	**Decision Tree (seconds)**	**Naïve Bayes (seconds)**
Titanic Dataset (Binary Classification)	1,309	0.250	0.050	0.030
Spam Email Dataset (Text Classification)	4,601	5.300	2.400	1.200
Wine Dataset (Multiclass Classification)	1,599	3.850	1.500	0.900

The table shows a comparison of computation time for three algorithms SVM, Decision Tree, and Naïve Bayes across the Titanic, Spam Email, and Wine Datasets. In the Titanic Dataset (1,309 entries), SVM required 0.25 seconds, longer than Decision Tree (0.05 seconds) and Naïve Bayes (0.03 seconds), reflecting the complexity of SVM even on a small dataset.

In the Spam Email Dataset (4,601 entries), SVM took 5.30 seconds, significantly longer than Decision Tree (2.40 seconds) and Naïve Bayes (1.20 seconds), due to kernel optimization on high-dimensional data. For the Wine Dataset (1,599 entries), SVM recorded 3.85 seconds, again slower than Decision Tree (1.50 seconds) and Naïve Bayes (0.90 seconds).

Overall, Naïve Bayes proved to be the most time-efficient algorithm, followed by Decision Tree. SVM, while providing higher accuracy and recall, had the longest computation time, particularly on large or complex datasets.

V. Conclusion

PenelitiThis study compares the performance of SVM, Decision Tree, and Naïve Bayes algorithms across three datasets: Titanic, Spam Email, and Wine. The results indicate that SVM consistently outperforms in accuracy and recall, particularly on the complex, multiclass Wine Dataset. However, this advantage comes at the cost of relatively high computation time, especially on high-dimensional datasets such as Spam Email. Conversely, Naïve Bayes demonstrates the best time efficiency across all tests, although it lags slightly in accuracy except on the Spam Email Dataset, where its performance remains competitive.

ISSN 2985-4512 (online) 2985-4520 (print) © The Authors. DOI: 10.31253/rubin.v4i1.4041
Published by Buddhi Dharma University

Decision Tree falls in between: faster than SVM and more interpretable, yet it tends to underperform in accuracy on complex datasets. These findings underscore the importance of selecting algorithms based on data characteristics and analytical objectives. SVM is best suited for high-precision classification tasks, Naïve Bayes is ideal for fast, text-based applications, while Decision Tree is appropriate for scenarios that prioritize model transparency. The trade-off between performance and efficiency must be carefully considered in real-world applications.

Based on the study's results, each of the three classification algorithms SVM, Decision Tree, and Naïve Bayes has its own strengths and limitations. SVM achieved the highest accuracy and recall, particularly on the Titanic and Wine Datasets, but required longer computation times. In contrast, Naïve Bayes and Decision Tree offered better time efficiency, though with slightly lower classification performance. These findings highlight the necessity of balancing model accuracy with computational efficiency when choosing the most suitable algorithm for specific data characteristics and application goals.

For further development, parameter optimization in SVM and Decision Tree is necessary to enhance efficiency without compromising accuracy. Improving dataset quality through data enrichment, handling class imbalance, and feature engineering is also essential so that models can better capture relevant patterns. Additionally, exploring other classification algorithms such as Random Forest, K-Nearest Neighbors, and XGBoost as well as applying ensemble approaches like bagging and boosting, can improve overall predictive performance. These strategies are expected to produce models that are more accurate, efficient, and applicable in both academic and industrial contexts.

REFERENCES

Asaad, R. R., & Abdulazeez, A. M. (2024). Comprehensive Classification of Iris Flower Species: A Machine Learning Approach. *Indonesian Journal of Computer Science*, *13*(1), 242–262. http://ijcs.stmikindonesia.ac.id/ijcs/index.php/ijcs/article/view/3135

Asshiddiqi, M. F., & Lhaksmana, K. M. (2020). Perbandingan Metode Decision Tree dan Support Vector Machine untuk Analisis Sentimen pada Instagram Mengenai Kinerja PSSI. *E-Proceeding of Engineering*, *7*(3), 9936–9948.

Aulia, M., & Hermawan, A. (2023). Analisis Perbandingan Algoritma SVM, Naïve Bayes, dan Perceptron untuk Analisis Sentimen Ulasan Produk Tokopedia. *Jurnal Media Informatika Budidarma*, *7*(4), 1850. https://doi.org/10.30865/mib.v7i4.6839

Boyd, A., Czajka, A., & Bowyer, K. (2019). Deep Learning-Based Feature Extraction in Iris Recognition: Use Existing Models, Fine-tune or Train from Scratch? *2019 IEEE 10th International Conference on Biometrics Theory, Applications and Systems, BTAS 2019*. https://doi.org/10.1109/BTAS46853.2019.9185978

Desiani, A. (2022). Penerapan Metode Support Vector Machine. *IJAI: Indonesian Jpurnal of Applied Informatics*, *7*(1), 12–18.

Dozono, C., Aragaki, M., Hebishima, H., & Inage, S.-I. (n.d.). *Proposal and Verification of Novel Machine Learning on Classification Problems*. *4*.

Hackeling, G. (2017). *Mastering Machine Learning with scikit-learn*. Packt Publishing.

Hermawan, A., Lunardi, L., Kurnia, Y., Daniawan, B., & Junaedi. (2025). Optimizing Convolutional Neural Networks with Particle Swarm Optimization for Enhanced Hoax News Detection. *Journal of Information Systems Engineering and Business Intelligence*, *11*(1), 53–64. https://doi.org/10.20473/jisebi.11.1.53-64

Kerrigan, D., Trokielewicz, M., Czajka, A., & Bowyer, K. W. (2019). Iris Recognition with Image

Segmentation Employing Retrained Off-the-Shelf Deep Neural Networks. *2019 International Conference on Biometrics, ICB 2019*. https://doi.org/10.1109/ICB45273.2019.8987299

Khumaidi, A., & Herinanto, D. (2023). Sistem Pendukung Keputusan Penerimaan Siswa Baru dengan Menggunakan Metode SAW di SMP Xaverius Pringsewu Berbasis Web Mobile. *Rubinstein*, *2*(1), 74–83. https://doi.org/10.31253/rubin.v2i1.2729

Liu, S., & Lee, I. (2018). Email Sentiment Analysis Through k-Means Labeling and Support Vector Machine Classification. *Cybernetics and Systems*, *49*(3), 181–199. https://doi.org/10.1080/01969722.2018.1448242

Pinto, J. P., Kelur, S., & Shetty, J. (2018). Iris Flower Species Identification Using Machine Learning Approach. *2018 4th International Conference for Convergence in Technology (I2CT)*, 1–4. https://doi.org/10.1109/I2CT42659.2018.9057891

Putra, P., M. H. Pardede, A., & Syahputra, S. (2022). Analisis Metode K-Nearest Neighbour (KNN) Dalam Klasifikasi Data Iris Bunga. *Jurnal Teknik Informatika Kaputama (JTIK)*, *6*(1), 297–305.

Rahman, B., Fauzi, F., & Amri, S. (2023). Perbandingan Hasil Klasifikasi Data Iris menggunakan Algoritma K-Nearest Neighbor dan Random Forest. *Journal Of Data Insights*, *1*(1), 19–26. https://doi.org/10.26714/jodi.v1i1.135

Rao, T. S., Hema, M., Priya, K. S., Krishna, K. V., & Ali, M. S. (2021). Iris Flower Classification Using Machine Learning. *International Journal of All Research Education and Scientific Methods (IJARESM)*, *9*(6), 2083–2090.

Rofianto, D., Arifin, O., & Widyawati, D. K. (2023). Perbandingan Metode Klasifikasi Learning Vector Quantization Dengan Diskriminan Fisher Pada Data Bunga Iris. *JSR : Jaringan Sistem Informasi Robotik*, *7*(1), 44–49. https://doi.org/10.58486/jsr.v7i1.209

Ronaldo, R., & Kurnia, Y. (2024). Perbandingan Kinerja Algoritma SVM, Decision Tree, dan Naive Bayes untuk Klasifikasi dan Pengelompokan Spesies Iris. *POTERS : Proceedings of Technology, Engineering and Computers*, *1*(1), 70–77.

Satyagraha, A., & Kurnia, Y. (2025). Forecasting Tourism Demand: A Bibliometric Review of Trends, Methodologies, and Big Data Integration (2015-2024). *Rubinstein*, *3*(2), 106–117. https://doi.org/10.31253/rubin.v3i2.3783

Sunaryono. (2017). Penelitian Komparasi Algoritma Klasifikasi dalam Menentukan Website Palsu. *TEKNIKOM*, *1*(1), 1–12.

Suresh Kumar, K., Radha Mani, A. S., Ananth Kumar, T., Jalili, A., Gheisari, M., Malik, Y., Chen, H. C., & Jahangir Moshayedi, A. (2024). Sentiment Analysis of Short Texts Using SVMs and VSMs-Based Multiclass Semantic Classification. *Applied Artificial Intelligence*, *38*(1). https://doi.org/10.1080/08839514.2024.2321555

Susanti, N. A., & Walid, M. (2022). Klasifikasi Data Tweet Ujaran Kebencian Di Media Sosial. *JATI: Jurnal Mahasiswa Teknik Informatika*, *6*(2), 538–543.

VanderPlas, J. (2022). *Python Data Science Handbook*. O'Reilly Media.

ISSN 2985-4512 (online) 2985-4520 (print) © The Authors. DOI: 10.31253/rubin.v4i1.4041
Published by Buddhi Dharma University