

Comparison Of Decision Tree Algorithms And Support Vector Machine (SVM) In Depression Classification In Students

M. Khoirul Risqi, Ifnu Wisma Dwi Prasty, Muhammad Jauhar Vikri

Universitas Nahdatul Ulama Sunan Giri, Indonesia

Email: risqisgb@gmail.com, ifnuprasty@unugiri.ac.id, vikri@unugiri.ac.id

ABSTRACT

Mental health in adolescents, especially students, is an important concern in the world of education. Early detection of symptoms of depression in students can help preventive efforts in handling them. This study aims to compare the performance of two classification algorithms, namely Decision Tree and Support Vector Machine (SVM) in detecting the level of depression in students based on data obtained from the Kaggle platform. The dataset used consisted of 502 student data with 10 features that caused depression and 1 target class. The research stage includes data preprocessing, which includes data cleaning, categorical value encoding, and normalization with the Min-Max Scaling method. The model was developed using the 5-Fold Cross Validation method to evaluate the classification performance of each algorithm. Model evaluation was carried out using precision, recall, and accuracy metrics. The test results showed that the SVM algorithm had better performance with a precision value of 93.63%, recall of 95.21%, accuracy of 94.22%, and F1-score of 94.68%. Meanwhile, Decision Tree obtained a precision of 81.77%, a recall of 84.90%, an accuracy of 82.86%, and an F1-score of 83.64%. Based on these results, it can be concluded that the Support Vector Machine is superior in classifying depression in students compared to Decision Tree.

KEYWORDS Depression, Student, Decision Tree, Support Vector Machine, K-Fold Cross Validation, Classification.



This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International

INTRODUCTION

Depressive disorder is one of the most common mental health problems and has a significant impact on the quality of life of sufferers. Depression is characterized by sadness, loss of interest or pleasure, feelings of guilt or low self-esteem, sleep disturbances or appetite, feelings of tiredness, and poor concentration [1]. According to research by Ahuvia et al., (2023), adolescents who experience depression tend to attribute their condition to various factors, especially dysfunctional family relationships (52%) and academic stress (42%). Other factors such as childhood trauma (11%), social media use (12%), and biogenetic causes such as chemical imbalances in the brain (19%) are also frequently mentioned [2].

Risqi (2025). Comparison of Decision Tree Algorithms and Support Vector Machine (SVM) In Depression Classification In Students. Journal Eduvest.

How to cite: 5 (4): 4557-4567
E-ISSN: 2775-3727

Globally, the total number of people living with depression stands at 322 million, with nearly half of these living in the Southeast Asia Region and the Western Pacific Region [1]. In Indonesia itself, a 2022 adolescent mental health survey showed that 5.5% of adolescents aged 10-17 years had mental disorders, with 1% of them experiencing depression [3]. This shows that depression in adolescents is a problem that needs more attention, especially in prevention efforts and early intervention to improve the mental well-being of the younger generation.

There are many ways that can be done to detect depression, one of which is by using data mining techniques. Data mining is a knowledge processing technique based on big data, data that is commonly used is taken from databases, data warehouses, the web and others to be processed into interesting information [4]. In practice, data mining is often used to analyze sales patterns, forecast production results, and various other applications. However, in this study, the application of data mining is focused on disease classification. Various techniques in data mining, such as association, clustering, classification, and regression, are used to process and extract information from large data. Among these methods, classification is one of the most commonly used techniques in data mining, especially in the process of identifying and predicting a condition based on the data patterns that have been analyzed.

Classification is one of the data mining algorithms that has the concept of grouping data into certain criteria by reading pre-existing data. The concept of classification algorithms is to predict the categorical class labels of a data to group it into one of the predefined classes [5]. In the classification process, there are various algorithms that can be used, such as Decision Tree and Support Vector Machine (SVM). Decision Tree is known as an efficient method of classification and prediction by building a model based on the structure of the decision tree. Meanwhile, the Support Vector Machine works by comparing a number of standard parameters with discrete values to determine the data categories, resulting in a classification with a high degree of accuracy [6].

In a study conducted by Arifuddin et al., (2024), two algorithms were compared, namely Decision Tree and SVM, to determine which algorithm is most effective in predicting heart disease. SVM outperforms Decision Tree in terms of accuracy, precision, recall, and F1-score. In a study conducted by Nurnawati et al., (2023), a model was developed to predict the predicate of lecturers based on the activities carried out. This study compares two algorithms, namely Decision Tree and Naïve Bayes, using the CRISP-DM data mining method which includes business understanding, data understanding, data preparation, modeling, evaluation, and development. The performance testing of the training data was carried out using K Fold Cross Validation. The test results showed that the Decision Tree algorithm performed better with an accuracy of 94.70%, a precision of 93.24%, and a recall of 96.33%. Meanwhile, the Naïve Bayes algorithm achieved an accuracy of 92.95%, a precision of 90.08%, and a recall of 96.33%. These findings indicate that Decision Tree-based models are more effective in determining lecturer performance.

Follow-up research by Helmi et al., (2021), compared the Support Vector Machine (SVM) and Naïve Bayes algorithms to determine the method with the best

accuracy. The study used microarray data of 80 individuals, each with 2,408 genetic expressions. Of these, 60 individuals were categorized as cancer patients, while 20 individuals were included in the normal group. The results of the analysis showed that the SVM algorithm achieved an accuracy of 90%, while Naïve Bayes obtained an accuracy of 75%, confirming the SVM's superiority in the classification of microarray data.

Based on previous research, the Decision Tree and SVM algorithms both show a high level of accuracy in the classification process. Therefore, these two algorithms are considered to be the right option in classifying depression. Thus, this study is expected to provide insight into the comparison of the accuracy level between the Decision Tree algorithm and the Support Vector Machine (SVM) in depression classification.

Although existing literature has extensively compared the effectiveness of various machine learning algorithms, such as Decision Tree and Support Vector Machine (SVM), in different classification tasks, there is a lack of studies specifically focused on comparing these algorithms for depression classification in students. Most studies have either focused on a single algorithm or generalized findings across different datasets without considering the specific characteristics of depression-related data. Moreover, while SVM has been shown to outperform other algorithms in various medical and psychological classifications, the comparative performance analysis of Decision Tree and SVM in detecting depression in students remains underexplored. This research addresses this gap by evaluating and comparing these algorithms specifically for classifying depression in students based on a specialized dataset.

The novelty of this study lies in its focus on comparing the performance of two widely used machine learning algorithms, Decision Tree and Support Vector Machine (SVM), specifically in the context of classifying depression in students. While both algorithms have been applied to a variety of classification tasks, this study is unique in its application of these models to detect depression, utilizing a publicly available dataset of student data that includes various factors influencing depression. The study's use of 5-fold cross-validation for model evaluation and comparison on multiple performance metrics, such as precision, recall, accuracy, and F1-score, provides new insights into the suitability of each algorithm for mental health-related classifications in educational settings.

The primary objective of this study is to compare the performance of Decision Tree and Support Vector Machine (SVM) algorithms in classifying depression levels in students. By analyzing the effectiveness of both algorithms on a dataset that includes various factors associated with depression, the study seeks to identify which algorithm is more accurate and reliable in detecting depression among students. The findings aim to contribute to the development of machine learning-based tools for early depression detection and preventive mental health measures in educational environments.

This research offers both theoretical and practical benefits. Theoretically, it contributes to the field of machine learning by providing a detailed comparison of the Decision Tree and SVM algorithms in the specific context of mental health classification. The study offers valuable insights into the strengths and weaknesses

of these algorithms for detecting depression, which can inform future research on machine learning applications in psychology and healthcare. Practically, the research can serve as a reference for educational institutions and mental health practitioners in developing early detection systems for depression, enabling timely intervention and support for students experiencing mental health challenges.

RESEARCH METHOD

The research method provides a comprehensive overview of the stages carried out in the research process to achieve the goals that have been set. This process begins with data collection to the evaluation stage to ensure the accuracy and relevance of the research results. The series of stages of this research can be seen in Figure 1.

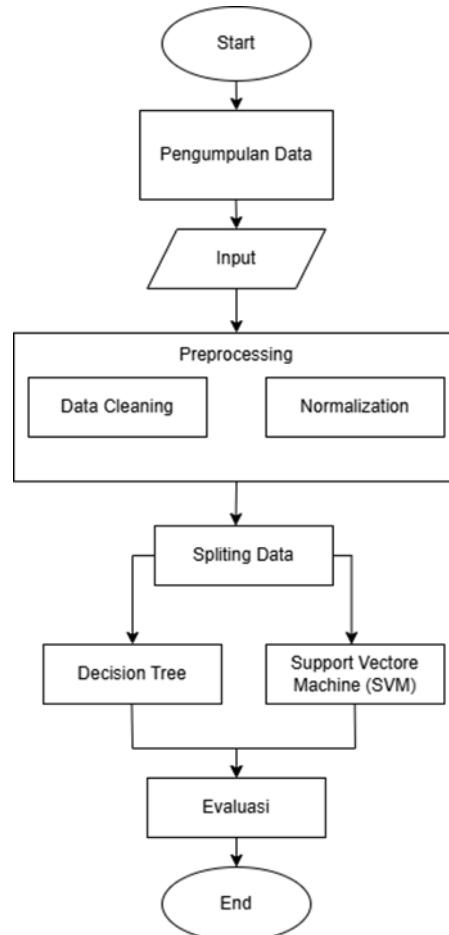


Figure 1 Research flow

Data Collection

The data collection process is carried out by searching for suitable datasets for classification purposes. In this study, the dataset used comes from Kaggle with the title Depression Student Dataset, which can be accessed via the following link: <https://www.kaggle.com/datasets/ikynahidwin/depression-student-dataset>.

This dataset consists of 502 data with 10 attributes that are factors that cause depression in students and 1 class for depression classification. The 10 attributes

analyzed in this dataset include Gender, Age, Academic Pressure, Study Satisfaction, Sleep Duration, Dietary Habits, Have you ever had suicidal thoughts?, Study Hours, Financial Stress, Family History of Mental Illness, and Depression class as classification labels. The data is shown in Table 1.

Table 1 Dataset attributes

No	Attribution	Description
1	Age	Respondents' age in years
2	Academic Pressure	The level of academic stress felt by the student
3	Study Satisfaction	The level of student satisfaction with the learning process
4	Sleep Duration	Average sleep duration per day in hours
5	Dietary Habits	Students' dietary habits
6	Have you ever had suicidal thoughts?	Have or have not experienced suicidal thoughts
7	Study Hours	Average student learning hours in a day
8	Financial Stress	Stress levels caused by financial conditions
9	Family History of Mental Illness	Family history of mental disorders

Table 2 Dataset Classes

No	Class	Description
1	Depression	indicates the class of Depression (no = not depressed, yes = depressed)

Preprocessing

Preprocessing is a crucial stage in the data mining process. The data used in the analysis is often not in the ideal condition to be processed immediately. Sometimes, the data contains various problems that can affect the accuracy of the analysis results, such as missing data, redundancy, the existence of outliers, or data formats that are not compatible with the system. To overcome these obstacles, a preprocessing process is needed. This step aims to clean and adjust the data so that it is better prepared for use in the classification and further analysis process [10]. Here are some of the stages of data preprocessing carried out in this study:

1. Data Cleaning or known as data cleaning aims to improve the quality of data to be more accurate and reliable in analysis. In this study, the cleanup stage includes the elimination of duplicate data, the identification and removal of data that contains anomalies or anomalies, and the elimination of attributes that are considered irrelevant to the study [11].
2. Normalization is an important technique in data processing that aims to align the values in the dataset so that they are on a uniform scale. This process is often referred to as feature scaling [12]. One of the commonly used methods for data normalization is Min-Max Normalization. This technique transforms the value of each feature by subtracting the minimum value of that feature and then

dividing it by the value range (the difference between the maximum and minimum values), so that all data values are in the range of 0 to 1. The formula used in Min-Max Normalization is:

$$X_{new} = \frac{X_{old} - X_{min}}{X_{max} - X_{min}}$$

Table 3: Depression of students after processing

Gender	Age	Academic Pressure	Study Satisfaction	Sleep Duration	Diet Habits	Have you ever had suicidal thoughts?	Study Hours	Financial Stress	Familly History	Depression
1.0 25	0.6 25	0.25 75	0.75 125	0.7 0.0	0.5 0.0	1.0 0.0	0.75 1.0	0.25 1.0	1.0 0.0	0.0
1.0 25	0.6 375	0.75 0.4	1.0 0.0	0.300000000 0.300000000	1.0 0.0	1.0 0.0	0.583333333 0.833333333	0.0 0.75	1.0 0.0	0.0
1.0 375	0.3 125	0.0 0.25	0.75 0.0	1.0 0.0	0.0 0.0	1.0 1.0	0.583333333 0.666666666	0.25 1.0	1.0 0.0	0.0
...
1.0 75	0.3 125	0.25 0.5	0.0 1.0	0.0 0.300000000	0.0 1.0	1.0 0.0	0.666666666 0.083333333	1.0 1.0	0.0 1.0	1.0
1.0 375	0.9 0.0	0.75 1.0	0.75 0.5	1.0 1.0	1.0 0.0	0.0 0.0	0.666666666 0.666666666	0.0 0.25	1.0 1.0	0.0
1.0	0.0	1.0	0.5	1.0	0.0	0.0	0.5	0.25	1.0	1.0

Splitting Data

Splitting Data is a dataset divided into two parts, namely training data and test data, with a certain proportion. For example, 80% of the dataset is used to train the model, while the remaining 20% is used to test the model's performance in making predictions [13].

Based on research conducted by Nurnawati, E. K., et al. (2023), the model evaluation process in this study uses K-Fold Cross Validation, where the data is divided into k subsets (folds). In each iteration, one subset is used as test data, while the other k-1 subset is used for training. This process is repeated k times, so that each data gets the opportunity to become a one-time test data and training data in other iterations.

In this study, the 5-Fold Cross Validation method was applied (K = 5) to evaluate the performance of the model with Decision Tree and Support Vector Machine (SVM) algorithms. This scheme was chosen to ensure more accurate evaluation results and reduce bias in machine learning.

Classification

Classification is a method used to determine the category or label of a data instance based on the patterns that have been studied. This technique groups the data into predefined classes, allowing for more accurate predictions. The classification falls under the type of supervised learning because the model is trained using data that already has a known class label [14].

Decision Tree

A Decision Tree is a tree-shaped model that resembles a flowchart, where each internal node represents a test against a specific feature, each branch shows the results of that test, and the leaf node (terminal node) contains a class label as the final output. If the target variable has a finite number of values, this model is called a classification tree. In this structure, each leaf depicts a class category, while a branch indicates a combination of feature paths that lead to a final decision regarding the predicted class [15].

Support Vector Machine (SVM)

Support Vector Machine (SVM) is a classification method that can be used to separate data in a linear and non-linear manner. The process of determining the optimal hyperplane is carried out by looking for the maximum margin, which is the farthest distance between the data points of each class against the hyperplane [16]. This algorithm is able to overcome problems with non-linear patterns because it utilizes the kernel concept, which allows data to be mapped to higher-dimensional spaces to find more optimal separators [17].

Google Colaboratory

Google Colab, or commonly called Google Colab, is an interactive document that can be run directly in the browser and allows users to write, store, and share programming code through integration with Google Drive [18]. Google Colab is in great demand by various people because it provides a variety of useful features, both for beginners and professionals, especially in the fields of data science, machine learning, and data processing in general.

Evaluation

The evaluation is carried out after the algorithm testing stage is completed. The purpose of this evaluation is to ensure that the model is constructed to truly represent the data according to the modeling design, as well as to assess and compare the performance of the two algorithms used in the study [6].

In this study, the Confusion Matrix was used to assess the accuracy of the results of the two algorithms. The confusion matrix is a table used to show the number of correct and incorrect predictions from the test data in the classification process [19].

There are four main metrics used to measure the performance of a classification system, namely Precision, Recall, Accuracy, and F1-Score. These four metrics provide a comprehensive picture of how well the model predicts data accurately and consistently.

$$Precision = \frac{TP}{FP + TP}$$

$$Recall = \frac{TP}{FN + TP}$$

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN}$$

$$F1\ Score = 2X \frac{Precision \times Recall}{Precision + Recall}$$

RESULT AND DISCUSSION

Based on the results of tests that have been carried out using both algorithms, namely Decision Tree and Support Vector Machine (SVM), it is shown that each algorithm provides different accuracy values in classifying the level of depression in students. The dataset used in this study comes from public sources available on the Kaggle platform, with a total of 502 entries consisting of 10 features and 1 target class, namely Depression. In the preprocessing stage, the process of data cleaning, category value handling, and normalization is carried out using the Min-Max Scaling method to ensure that the data is on a uniform scale before being used in the model training process.

After going through the data preprocessing stage, the next step in this study is to model the Decision Tree and Support Vector Machine (SVM) algorithms. The modeling process is carried out using the 5-Fold Cross Validation approach and supported by the Google Colab platform, which allows efficiency in data processing and supports optimal visualization of model evaluation results.

The classification model is built to detect potential depression in students, and their performance is evaluated using the Precision, Recall, and Accuracy metrics. The test results show that the SVM algorithm provides higher performance than Decision Tree. The evaluation values of the model are shown in the following table:

Table 4. Results of Decision Tree and SVM Model Evaluation in Detecting Depression in Students

Fold	Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
1	Decision Tree	85.15	86.00	84.31	85.15
2	Decision Tree	83.17	79.31	90.20	84.40
3	Decision Tree	82.00	80.77	84.00	82.35
4	Decision Tree	83.00	81.13	86.00	83.50
5	Decision Tree	81.00	81.63	80.00	80.81
1	SVM	98.02	96.23	100.00	98.08

2	SVM	93.07	89.29	98.04	93.46
3	SVM	94.00	100.00	88.00	93.62
4	SVM	91.00	90.20	92.00	91.09
5	SVM	95.00	92.45	98.00	95.15

Based on the results of the evaluation using the 5-Fold Cross Validation method, the performance results of two classification algorithms, namely Decision Tree and Support Vector Machine (SVM) in detecting the level of depression in students, were obtained.

In the 1st Fold, the Decision Tree algorithm produces an accuracy of 85.15%, precision of 86.00%, recall of 84.31%, and an F1-Score of 85.15%. Meanwhile, SVM showed excellent performance with an accuracy of 98.02%, accuracy of 96.23%, recall of 100%, and an F1-Score of 98.08%. This shows that in the first fold, SVM is able to recognize all positive data very well (100% recall).

In the 2nd Fold, Decision Tree recorded an accuracy of 83.17%, precision of 79.31%, recall of 90.20%, and an F1-Score of 84.40%. On the other hand, SVM also showed high performance with an accuracy of 93.07%, accuracy of 89.29%, recall of 98.04%, and an F1-Score of 93.46%. This shows that the SVM is not only accurate, but also consistent in correctly recognizing positive data.

In the 3rd Fold, Decision Tree obtained an accuracy of 82.00%, accuracy of 80.77%, recall of 84.00%, and an F1-Score of 82.35%. On the other hand, SVM gets 94.00% accuracy, 100% perfect precision, 88.00% recall, and 93.62% F1-Score. Although SVM recall decreased slightly, the accuracy remained very high.

In the 4th Fold, the results for the Decision Tree were 83.00% accuracy, 81.13% accuracy, 86.00% recall, and 83.50% F1-Score. Meanwhile, SVM recorded an accuracy of 91.00%, precision of 90.20%, recall of 92.00%, and F1-Score of 91.09%, indicating good performance stability.

In the 5th Fold, Decision Tree obtained an accuracy of 81.00%, accuracy of 81.63%, recall of 80.00%, and an F1-Score of 80.81%. Meanwhile, the SVM shows an accuracy of 95.00%, precision of 92.45%, recall of 98.00%, and an F1-Score of 95.15%, which again shows the superior performance of this model.

Overall, based on the average of all five folds, the SVM algorithm showed a more consistent and superior performance compared to Decision Tree in all evaluation metrics, namely accuracy, precision, recall, and F1-score. This shows that SVM is more effectively used for the detection of depression levels in the dataset used in this study.

CONCLUSION

This study compares the performance of Decision Tree and Support Vector Machine (SVM) algorithms in classifying depression levels in students using a dataset obtained from Kaggle, consisting of 502 student data points with 10 features related to depression. The data underwent preprocessing, including cleaning, categorical value transformation, and normalization via Min-Max Scaling. The models were evaluated using 5-Fold Cross Validation to reduce bias and enhance

objectivity. The results indicated that SVM outperformed Decision Tree with an average precision of 93.63%, recall of 95.21%, accuracy of 94.22%, and F1-score of 94.68%, whereas Decision Tree yielded lower performance with a precision of 81.77%, recall of 84.90%, accuracy of 82.86%, and F1-score of 83.64%. Therefore, it can be concluded that SVM is more effective and accurate in detecting depression in students based on the given dataset. These results provide valuable insights for the development of machine learning-based early depression detection systems, which could assist in promotive and preventive efforts in adolescent mental health. For future research, it is recommended to explore the integration of other algorithms, such as Random Forest or Neural Networks, to further improve classification accuracy, as well as to apply these models to larger, more diverse datasets for better generalizability. Additionally, incorporating real-time data from social media or wearable health devices could enhance the accuracy and timeliness of depression detection systems.

REFERENCES

1. World Health Organization. (2017). *Depression and Other Common Mental Disorders: Global Health Estimates*. Retrieved from <https://www.who.int/publications/i/item/depression-global-health-estimates>
2. Ahuvia, I. L., Chen, S., Gordon, L. H., Fox, K. R., & Schleider, J. L. (2023). *A mixed-methods investigation of adolescents' beliefs about the causes of depression*. Stony Brook University & University of Denver.
3. Center for Reproductive Health. (2022). *Indonesia – National Adolescent Mental Health Survey (I-NAMHS) Report*. Gadjah Mada University. Accessed from <https://qcmhr.org/outputs/reports/12-i-namhs-report-bahasa-indonesia>
4. Sofyan, F. M. A., Voutama, A., & Umaidah, Y. (2023). Application of C4.5 Algorithm for Lung Disease Prediction Using RapidMiner. *JATI (Journal of Informatics Engineering Students)*, 7(2), 1409.
5. Barata, M. A., Noersasongko, E., Purwanto., & Soeleman, M. A. (2023). Improvement of C4.5 Algorithm Accuracy with Chi-Square Method on Classification of Pure Tea Using Electronic Nose. *RESTI JOURNAL (Systems Engineering and Information Technology)*, 7(2), 226-235.
6. Laili, E. F., Alawi, Z., Rohmah, R., & Barata, M. A. (2025). Comparison of Decision Tree and Support Vector Machine (SVM) algorithms in the classification of heart attacks. *Journal of Information Systems and Informatics (Simika)*, 8(1), 67. <https://doi.org/10.47080/simika.v8i1.3683>
7. Arifuddin, A., Buana, G. S., Vinarti, R. A., & Djunaidy, A. (2024). Performance comparison of Decision Tree and Support Vector Machine algorithms for heart failure prediction. *Procedia Computer Science*, 234, 628–636. <https://doi.org/10.1016/j.procs.2024.03.048>
8. Nurnawati, E. K., Sholeh, M., Ariyana, R. Y., & Almuntaha, E. (2023). Comparison of Decision Tree and Naïve Bayes algorithms in classification models to determine lecturer performance using K Fold Cross Validation. *Jurnal TAM (Technology Acceptance Model)*, 14(2), 152–157.

9. Helmi, M. I. Y., Anggraeni, D., & Hadi, A. F. (2021). *Diagnosis of lung cancer patients using Support Vector Machine and Naïve Bayes*. Statistics, 21(1), 1–4.
10. Dwifebri, M., Irwan, M., Imam, A., & Adiwijaya. (2021). RESTI JOURNAL Comparison of Support Vector Machine and Modified Balanced Random. 1(10), 393–399.
11. Prabaswara, R., Lemantara, J., & Jusak, J. (2023). Classification of secondary school destination for inclusive students using Decision Tree Algorithm. *Journal of RESTI (Systems Engineering and Information Technology)*, 7(5), 1009–1018.
12. Saputra, I., & Kristiyanti, D. A. (2022). *Machine Learning for Beginners*. Bandung: Informatics.
13. Eina, M. F., Chrisnanto, Y. H., & Melina, M. (2024). Klasifikasi Telemarketing Menggunakan Naïve Bayes Classification Dan Wrapper Sequential Feature Selection. *INTECOMS: Journal of Information Technology and Computer Science*, 7(4), 1189–1198. <https://doi.org/10.31539/intecoms.v7i4.10846>
14. Rahayu, D. S., Afifah, J., & Intan, S. (2023). Classification of Diabetes Mellitus Using C4 . 5 Algorithm, Support Vector Machine (SVM) and Linear Regression Classification of Diabetes Mellitus Using C4 Algorithm. 5, Support Vector Machine (SVM) and Linear Regression. SENTIMAS: National Seminar on Research and Community Service, 1(1 SE-), 56–63.
15. Lagzi, M. D., Sajadi, S. M., & Taghizadeh-Yazdi, M. (2024). A hybrid stochastic data envelopment analysis and decision tree for performance prediction in retail industry. *Journal of Retailing and Consumer Services*, 80, 103908. <https://doi.org/10.1016/j.jretconser.2024.103908>
16. Adi, S., & Wintarti, A. (2022). Comparison of Support Vector Machine (SVM), K-Nearest Neighbors (KNN), and Random Forest (RF) methods for the prediction of heart failure. *MATHunesa: Scientific Journal of Mathematics*, 10(2), 258–268. <https://doi.org/10.26740/mathunesa.v10n2.p258-268>
17. Sari, P. K., & Suryono, R. R. (2024). Comparison of Support Vector Machine and Random Forest algorithms for Metaverse sentiment analysis. *MNEMONIC Journal*, 7(1), 31–40.
18. Maulana, G., & Dana, R. D. (2024). Prediction of corn production in West Java using the linear regression algorithm method using Google Collab. *JATI (Student Journal of Informatics Engineering)*, 8(1), 827–834.
19. Normawati, D., & Prayogi, S. A. (2021). Implementation of the Naïve Bayes Classifier and Confusion Matrix on text-based sentiment analysis on Twitter. *Journal of Computer Science & Informatics (J-SAKTI)*, 5(2), 697–711.