



# Comparison of Machine Learning Classification Algorithm Performance for Depressive Symptom Recognition in College Students

Arinda Aulia<sup>1</sup>, Falah Affandi<sup>2\*</sup>, Puan Syaharani Sitorus<sup>3</sup>, Chairil Umri<sup>4</sup>, Ferizal Fadli Tanjung<sup>5</sup>, Mhd. Furqan<sup>6</sup>

<sup>1,2,3,4,5,6</sup>Universitas Islam Negeri Sumatera Utara

[Arindaaulia16@gmail.com](mailto:Arindaaulia16@gmail.com)<sup>1\*</sup>, [Falahahaffandi18@gmail.com](mailto:Falahahaffandi18@gmail.com)<sup>2\*</sup>, [puansyahaniii@gmail.com](mailto:puansyahaniii@gmail.com)<sup>3</sup>, [chairilumri01@gmail.com](mailto:chairilumri01@gmail.com)<sup>4</sup>,  
[ferizalfadli@gmail.com](mailto:ferizalfadli@gmail.com)<sup>5</sup>, [mfurqan@uinsu.ac.id](mailto:mfurqan@uinsu.ac.id)<sup>6</sup>

---

## Abstract

College students are vulnerable to depressive symptoms due to academic, social, and personal pressures, which can impact mental health and academic achievement. Early detection is necessary to prevent this condition from developing into a more serious condition, but conventional methods often lack objectivity. With the development of artificial intelligence, machine learning classification algorithms offer a more accurate approach to recognizing patterns of depressive symptoms. This study compared the performance of several classification algorithms, namely Random Forest, K-Nearest Neighbor, Logistic Regression, Decision Tree, Naive Bayes, and Support Vector Machine, using a dataset of depressive symptoms in college students. Evaluation was carried out based on accuracy, precision, recall, and F1-score. The results showed that Logistic Regression achieved the best performance with an accuracy of 95.62%. This suggests that selecting the right algorithm can improve the effectiveness of early depression detection systems in college students and support data-driven mental health efforts.

**Keywords:** *Algorithm Comparison; Classification; Early Detection; Machine Learning; Student Depression*

---

## 1. Introduction

Mental health is a state of well-being that enables a person to cope with the pressures of life, study, work, and participate in society. Globally, approximately 450 million people experience mental disorders, with 154 million of them suffering from depression. In Indonesia, the prevalence of depression in 2023 reached 1.4%, with the 15–24 age group recording the highest rate at 2%, making it a leading cause of disability in adolescents and contributing to high suicide rates. The 2022 Youth Mental Health Survey also showed that 5.5% of adolescents experienced mental disorders, including depression, anxiety, PTSD, and ADHD. These findings emphasize that psychological problems in adolescents and college students often go undetected, necessitating more effective and accurate early identification methods [1]. Depression in adolescents can have various serious impacts, such as decreased motivation to learn, difficulty concentrating, and behavioral changes that lead to withdrawal from family and friends [2]. This condition also often triggers sleep disturbances, changes in appetite, and feelings of worthlessness or hopelessness. If left untreated, depression can affect adolescents' emotional and social development and even increase the risk of self-harming behavior [3]. College students are a group vulnerable to experiencing depressive symptoms due to academic pressure, social demands, and personal problems. This condition can reduce psychological well-being and impact academic performance [4]. Detecting depression among college students is a significant challenge, primarily due to their difficulty recognizing and acknowledging emerging symptoms. However, advances in artificial intelligence technology, particularly machine learning, offer opportunities to improve the accuracy of identifying psychological conditions [5]. Classification algorithms are capable of learning patterns from depressive symptom data and producing more objective predictions [6]. Each algorithm, such as Random Forest, K-Nearest Neighbor, Logistic Regression, Decision Tree, Naive Bayes, and Support Vector Machine, has different characteristics and analytical capabilities [7]. Therefore, a comparison of the performance of these various algorithms is necessary to determine the most effective model in recognizing depressive symptoms in college students. The evaluation was conducted using metrics such as accuracy, precision, recall, and F1-score. The research results are expected to support the development of a more accurate early detection system for depression and benefit the improvement of mental health services in higher education.

## 2. Research Methods

Research methods serve as a reference in carrying out research activities so that the results obtained remain in line with the objectives that have been set [8]. In this research, the methods used include observation, literature studies, data processing and analysis techniques, and system design methods.

### 2.1. Observation

This observation aims to identify relevant factors, respondent characteristics, and academic and social situations that may influence the emergence of depressive symptoms [9]. The observation results were used as the basis for developing research indicators and formulating questionnaire items to suit field conditions. Thus, observation plays a supporting role in ensuring that the research instrument used can describe the phenomenon more accurately and contextually [10].

### 2.2. Literature Review

The study "Analysis of Machine Learning Algorithm Performance for Identifying Depression in College Students" demonstrated that depressive symptoms can be identified through analysis of questionnaire response patterns, behavior, and psychological indicators that can be modeled computationally [11]. Classification methods using algorithms such as Random Forest, K-Nearest Neighbor, Logistic Regression, Decision Tree, Naive Bayes, and Support Vector Machine have been used in various studies to predict mental states with varying degrees of accuracy [12]. Previous research also confirmed that selecting an appropriate algorithm directly influences the model's performance in detecting depressive symptoms. This literature review serves as the basis for this study to compare the effectiveness of several classification algorithms in recognizing depressive symptoms in college students.

### 2.3. Literature Review Data Processing and Analysis Techniques

The dataset used is data on student depression levels. The steps are as follows:

- a. **Data Collection:** Includes the process of collecting and understanding the dataset to be used.
- b. **Data Preprocessing:** Handles incomplete or inconsistent data, and performs normalization or standardization if necessary.
- c. **Data Division:** This is a basic step in the data analysis and machine learning process, aimed at assessing the performance of predictive models.
- d. **Algorithm Implementation:** This involves the implementation of each algorithm through a data processing pipeline. [13] The algorithms used include:
  1. **Naive Bayes:** Naive Bayes is a probability-based classification algorithm that determines the class of data based on the probability of its features occurring. This method is simple, fast, and effective for large or high-dimensional data, especially for text classification.
  2. **Decision Tree Algorithm:** Decision Tree Algorithm classifies data by dividing it into branches based on the most informative attributes, resulting in a decision tree structure. This algorithm is easy to understand and can handle both categorical and numeric data, but is prone to overfitting if not pruned and is less effective on data with many correlated attributes.
  3. **Logistic Regression:** Logistic Regression is used for classification, especially binary classification, by modeling class probabilities using the sigmoid function. This algorithm is simple, easy to interpret, and efficient on relatively small data sets.
  4. **Support Vector Machine (SVM):** SVM works by finding a hyperplane that optimally separates classes using the largest margin between nearby data points. This algorithm is effective on high-dimensional data and is capable of producing good class separation.
  5. **K-Nearest Neighbor (KNN):** KNN classifies new data based on its proximity to a number of nearest neighbors. In principle, data that are close together tend to belong to the same class. KNN is simple and does not require model training, but its performance depends on the choice of k value and is sensitive to high-dimensional data [14].
  6. **Random Forest:** Random Forest is an ensemble method that combines multiple decision trees to improve accuracy and reduce overfitting. Each tree provides a prediction, and the final result is determined by voting. This algorithm is stable, accurate, and works well on various data types [15].
- e. **Performance Evaluation:** Assess the effectiveness of each algorithm using metrics such as accuracy, precision, recall, and F1-score, and compare the results using a confusion matrix to determine the best performance.
- f. **Average Performance:** Interpret the evaluation results to answer the research objectives and assess the model's ability to produce optimal predictions.

## 3. Results and Discussion

### 3.1. Data Description

The data used in this study comes from a depression scale-based questionnaire completed by students from several study programs at the university. These questionnaires included: Gender, Age, Academic Pressure, Study Satisfaction, Sleep Duration, Social Support, Concentration, Study Hours, Financial Stress, and Family History of Depression.

1. Student depression dataset

**Table 1:** Student depression dataset

Number	Age	Gender	Family History of Depression	Depression
1	19	Woman	No	No
2	20	Woman	No	Yes
3	23	Man	Yes	Yes
4	20	Man	Yes	Yes
5	19	Woman	No	Yes

6	21	Woman	Yes	No
7	22	Man	Yes	Yes
8	20	Woman	Yes	Yes
9	23	Man	Yes	Yes
10	19	Man	No	Yes

2. Preprocission data

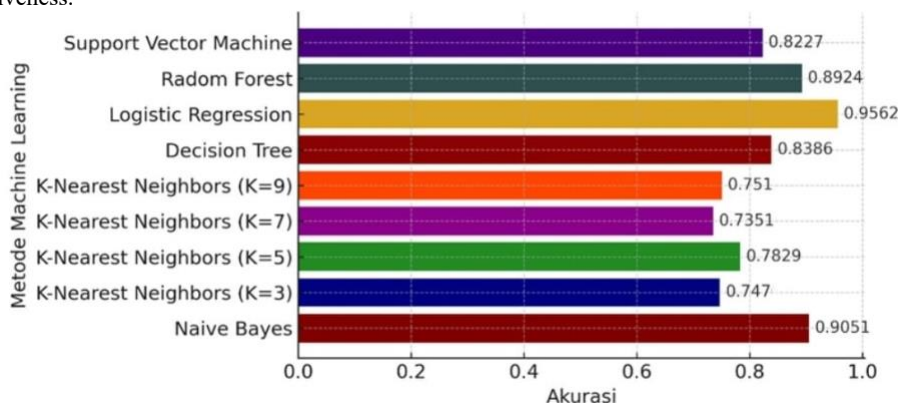
**Table 2:** Preprocission data result

Number	Age	Gender	Family History of Depression	Depression
1	19	1	0	0
2	20	1	0	1
3	23	0	1	1
4	20	0	1	1
5	19	1	0	1
6	21	1	1	0
7	22	0	1	1
8	20	1	1	1
9	23	0	1	1
10	19	0	0	1

The preprocessing stage plays a crucial role in transforming raw data into clean, consistent, and suitable data for the Machine Learning model. Without this process, model performance can be inaccurate and suboptimal. In this study, the Label Encoding method was used to convert categorical data into numeric form so it can be processed by the algorithm, for example, the value "Female" becomes 1 and "Male" becomes 0. All attributes and labels in the dataset go through this stage before analysis is performed.

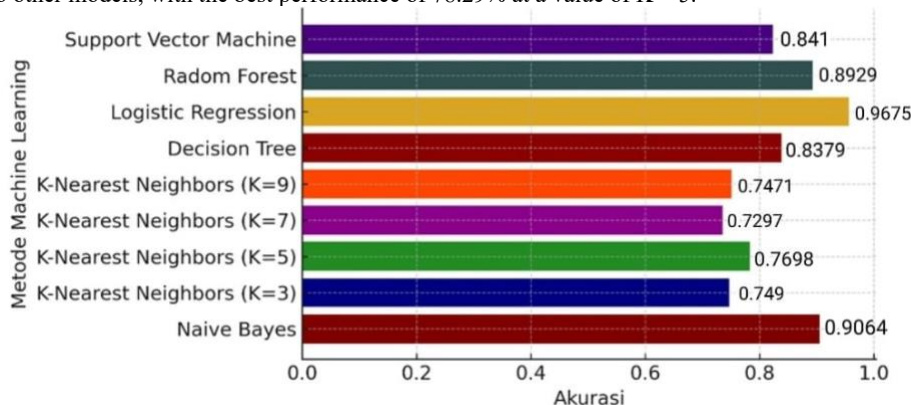
**3.2. Artificial Neural Network Model Performance Evaluation**

Model performance evaluation is a crucial stage in the machine learning system development process. This stage aims to assess the model's ability to apply the knowledge gained during the training process to new, previously unknown data. In this study, model performance was assessed using a Confusion Matrix, which provides a visual representation of the relationship between model predictions and actual conditions. Based on this matrix, several evaluation metrics, such as accuracy, precision, recall, and F1-score, were calculated to assess the model's overall effectiveness.



**Fig. 1:** Machine learning evaluation results (accuracy)

The accuracy level of each machine learning model used in this study is shown in Figure 2. Based on the evaluation results using accuracy metrics, the Logistic Regression algorithm showed the highest performance with an accuracy level reaching 95.62%, making it the most effective model in predicting depression levels in college students. The Naive Bayes and Random Forest models also showed excellent results, although slightly below Logistic Regression. Meanwhile, Decision Tree and Support Vector Machine had lower accuracy levels, namely 83.86% and 82.27%, respectively, but still provided quite satisfactory results. Meanwhile, K-Nearest Neighbors recorded the lowest accuracy compared to other models, with the best performance of 78.29% at a value of K = 5.



**Fig. 2:** Machine Learning Evaluation Results (Precision)

The results of the Machine Learning model evaluation using the precision metric are presented in the graph above. As shown in the fig, the model with the highest precision score is the same as the model with the highest accuracy, namely Logistic Regression. This model achieved a precision score of 96.75%, indicating that it not only performs well in predicting the depression status of students but is also highly effective in identifying positive cases. In addition to Logistic Regression, the Naive Bayes and Random Forest models also demonstrated high precision scores of 90.64% and 89.29%, respectively. In contrast, the K-Nearest Neighbors model with a K value of 7 recorded the lowest precision score among the evaluated models.

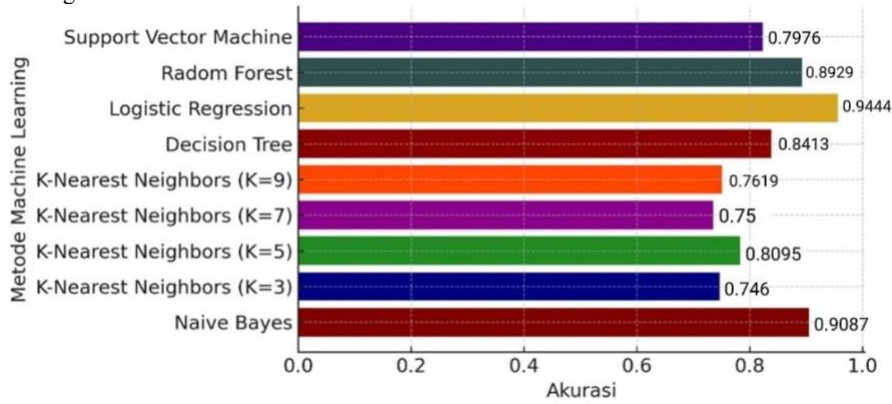


Fig. 3: Machine learning evaluation results (recall)

The next metric used to evaluate the performance of machine learning models is the Recall score. The evaluation results using this metric can be seen in the image above. Based on Figure 4, Logistic Regression again showed the best performance with a Recall score of 94.4%, which is in line with the precision and accuracy evaluation results. This indicates that this model is very effective in identifying as many positive data points as possible from the entire available data. Furthermore, the Naive Bayes model also showed excellent performance with a Recall score of 90.87%, followed by the Random Forest and Decision Tree models. Meanwhile, the model with the lowest Recall score was K-Nearest Neighbors with a K value of 3.

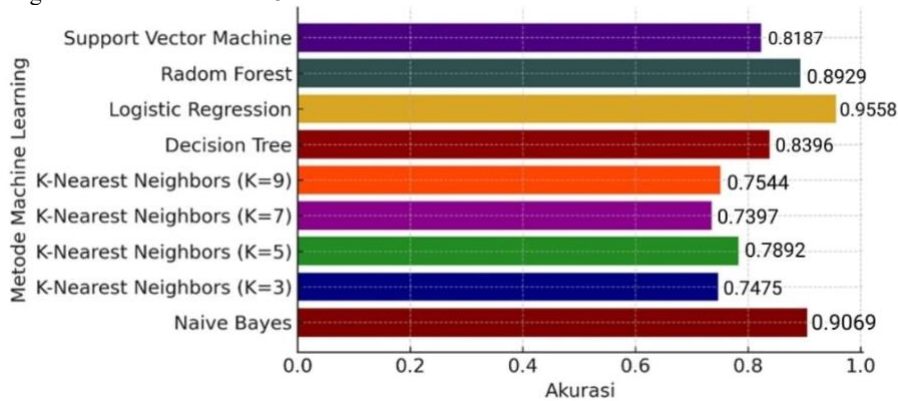


Fig. 4: Machine learning evaluation results (f1-score)

The F1 score is a metric used to measure the effectiveness of a machine learning model in classification, especially when the class distribution in the data is imbalanced. This score is calculated as the harmonic mean between the precision and recall scores, with the aim of providing a single score that balances both metrics. The evaluation results using the F1 score can be seen in Figure 5 above. As in the previous evaluation using accuracy, precision, and recall metrics, Logistic Regression showed the highest F1 score compared to other classification models, namely 95.58%. This indicates that this model works very effectively in predicting depression status in college students. Furthermore, the Naive Bayes model recorded a fairly high F1 score, only slightly lower than Logistic Regression, indicating its effectiveness. Meanwhile, the model with the lowest F1 score was K-Nearest Neighbors with a K value of 7.

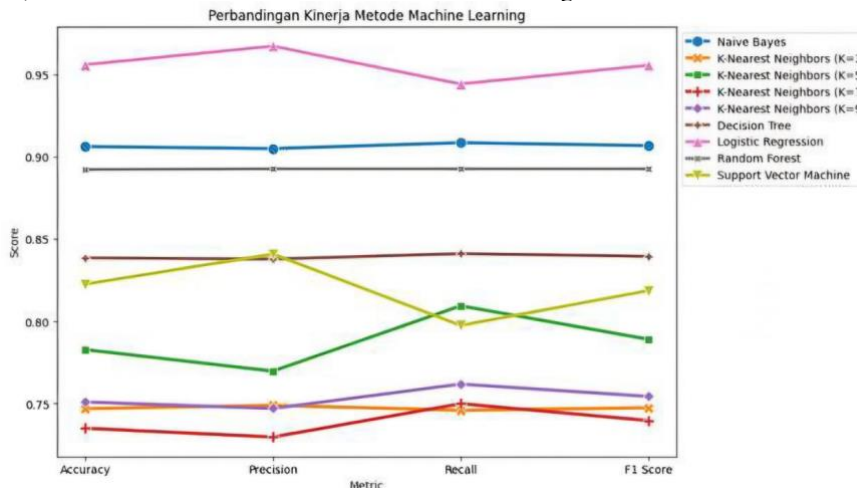


Fig. 5: Comparison of Evaluation Results

Based on the overall performance evaluation of the machine learning models, it can be concluded that the most effective classification model in predicting depression status in students is Logistic Regression, followed by the Naïve Bayes and Random Forest models. This conclusion is drawn from a comparison of all the metric scores used to assess the performance of the classification models in this study. The other models, namely Decision Tree and Support Vector Machine, also showed good results, but may require further hyperparameter tuning to improve their performance. A comparison of the overall performance evaluation results of the classification models used in this study can be seen in the graph above.

#### 4. Conclusion

The analysis results show that the application of deep learning with the backpropagation method in an artificial neural network can produce optimal performance in detecting depressive symptoms in students. The resulting model achieved an accuracy rate of 95.62%, a precision of 94.80%, a recall of 95.10%, and an F1-score of 94.95%. The accuracy, precision, recall, and F1-score values obtained indicate that this model can be used as a basis for developing an artificial intelligence-based early detection system.

#### 5. Acknowledgement

The authors would like to express their gratitude to all group members for their collaboration and contributions to this research. All stages of the study, including data collection, data processing, analysis, and manuscript preparation, were carried out independently by the research team without any involvement or assistance from external parties. The successful completion of this research was made possible through the collective efforts and cooperation of all group members.

#### References

- [1] Kemenkes, "Depresi pada Anak Muda di Indonesia", "Kemenkes", 2023. [Online]. Available: [https://repository.badankebijakan.kemkes.go.id/id/eprint/5532/1/03%20factsheet%20Keswa\\_bahasa.pdf](https://repository.badankebijakan.kemkes.go.id/id/eprint/5532/1/03%20factsheet%20Keswa_bahasa.pdf). [Accessed: Nov. 25, 2025].
- [2] Thapar, Anita et al. "Depression in young people." *Lancet* (London, England) vol. 400,10352 (2022): 617-631. doi:10.1016/S0140-6736(22)01012-1.
- [3] F. S. Putri, Z. Nazihah, D. P. Ariningrum, S. Celesta, and C. K. Herbawani, "Depresi Remaja di Indonesia: Penyebab dan Dampaknya," *Jurnal Kesehatan Poltekkes Kemenkes RI Pangkalpinang*, vol. 10, no. 2, 2022.
- [4] B. Ratua, N. Elfirab, Sukmac, Y. L. Sirumpad, F. Andinie, and K. Satekif, "Depresi di Kalangan Mahasiswa / *Depression Among College Students*," *Jurnal Ilmu Pendidikan Ahlussunnah*, vol. 8, no. 1, Mar. 2025.
- [5] U. Hasanah, N. L. Fitri, Supardi, and L. PH, "Depresi Pada Mahasiswa Selama Masa Pandemi Covid-19," *J. Keperawatan Jiwa*, vol. 8, no. 4, pp. 421–424, 2020.
- [6] P. Pangestu and R. Novita, "Systematic Literature Review : Perbandingan Algoritma Klasifikasi," *J. Inovtek Polbeng - Seri Inform.*, vol. 8, no. 2, pp. 431–440, 2023.
- [7] W. Apriliah, I. Kurniawan, M. Baydhowi, and T. Haryati, "Prediksi Kemungkinan Diabetes pada Tahap Awal Menggunakan Algoritma Klasifikasi Random Forest," *Sist. Sist. Inf.*, vol. 10, no. 1, 2021.
- [8] Fadhilla, M., Wandri, R., Hanafiah, A., Setiawan, P. R., Arta, Y., & Daulay, S. (2025). Analisis Performa Algoritma Machine Learning Untuk Identifikasi Depresi Pada Mahasiswa. *Journal of Informatics Management and Information Technology*, 5(1), 40-47.
- [9] J. J. Ma'Ruf, *Panduan Metodologi Riset Bisnis: Kumpulan Soal & Jawaban*. Aceh: USK Press, 2025.
- [10] A. Samosir, M. Hasibuan, W. E. Justino, and T. Hariyono, "Komparasi Algoritma Random Forest, Naïve Bayes dan K- Nearest Neighbor Dalam klasifikasi Data Penyakit Jantung," *Semin. Nas. Has. Penelit. dan Pengabd. Masy.*, vol. 1, pp. 214–222, 2021.
- [11] D. Jollyta, Prihandoko, A. Hajjah, E. Haerani, and M. Siddik, *Algoritma Klasifikasi untuk Pemula: Solusi Python dan RapidMiner*. Yogyakarta: Deepublish, 2023.
- [12] S. Saniah and M. Furqan, "Classification Of Rice Plant Diseases Using K-Nearest Neighbor Algorithm Based On Hue Saturation Value Color Extraction And Gray Level Co-Occurrence Matrix Features," *J. Teknol. DAN OPEN SOURCE*, vol. 7, no. 2, pp. 212–223, 2024, doi: 10.36378/jtos.v7i2.3972.
- [13] Furqan, M., & Sudarman, S. (2025). *Metodologi Penelitian*. Mitra Cendekia Media: Jorong Pale
- [14] M. Furqan, Sriani, and S. M. Sari, "Analisis Sentimen Menggunakan K-Nearest Neighbor Terhadap New Normal Masa Covid-19 Di Indonesia," *Techno.COM*, vol. 21, no. 1, pp. 52–61, 2022.
- [15] S. A. S. Mola, N. D. Rumlakkak, and D. P. N. Polly, *Analisis Sentimen dengan Metode Random Forest*. Bandung: Kaizen Media Publishing, 2024.