JoCoSiR
Journal of Computer Science Research
ISSN 2986-2337 (Online)

Comparison of Naïve Bayes, K-Nearest Neighbors, and Decision Tree Methods for Classifying Heart Disease Risk Factors

Ahmad Jihad Al Fayed [a], Surya Darma [b], Zailani Sinabariba [c], Surya Maruli P Pardede [d]

[a,c,d] *Magister Teknologi Informasi, Universitas Pembangunan Panca Budi, Medan, Indonesia*
[b] *Film and Televisi, Universitas Potensi Utama, Medan, Indonesia*

Email: [a]Jihadahmad000@gmail.com, [b]suryadarma090693@gmail.com, [c] zailanisinabariba88@gmail.com, [d]sp.trainer.belajar.id@gmail.com

ARTICLE INFO

Keywords:
Classification, Heart Disease, Naïve Bayes, K-Nearest Neighbors, Decision Tree

IEEE style in citing this article:
A. J. Al Fayed, S. Darma, Z. Sinabariba, and S. M. P. Pardede, " Comparison of Naïve Bayes, K-Nearest Neighbors, and Decision Tree Methods for Classifying Heart Disease Risk Factors," *JoCoSiR: Jurnal Ilmiah Teknologi Sistem Informasi*, vol. 3, no. 3, pp. 81-88, 2025.

ABSTRACT

Heart disease is the leading cause of death and poses a major challenge to global health systems. The classification of heart disease risk factors is crucial for preventing serious indications, but the challenge is that detection of this disease is often hampered because the classification process is not yet sufficiently accurate. This study aims to develop a heart disease risk classification model using a machine learning approach on a 2025 dataset consisting of 6025 patient data with 14 features. After going through the data collection stage and determining the attributes for comparing the performance of machine learning algorithms (Naive Bayes, K-Nearest Neighbors, and Decision Tree), it was found that the Decision Tree algorithm provided the best performance with an accuracy of 86%, followed by the K -Nearest Neighbors algorithm with an accuracy of 78% and the Naive Bayes algorithm with an accuracy of 76%.

1. Introduction

Heart disease is the leading cause of death, which has become a major global health challenge. According to a report by the World Health Organization (WHO), more than 17.9 million deaths per year are caused by cardiovascular disease, and more than 80% of these deaths are caused by heart attacks and strokes. Meanwhile, 31% account for total global deaths. [1][2]

Meanwhile, in Indonesia, the Ministry of Health (2022) reported that the prevalence of heart disease is increasing every year, especially among people of productive age. These cases indicate that heart disease is not only a medical problem, but also has a significant impact on public health, which will affect productivity and the national economy. [3]

Unhealthy habits are a major factor in heart disease, including lack of exercise, a diet high in fat and salt, smoking, alcohol consumption, and stress. Other risk factors for heart disease include obesity, diabetes, heredity, and high cholesterol. [4][5]

Rapid technological advances to date have continued to undergo changes over time, encouraging researchers to conduct research with the aim of developing tools to assist in diagnosing diseases, particularly heart disease, through the application of machine learning [6]. Machine learning has become very popular in recent years in analyzing medical data due to its ability to identify patterns in large amounts of data (big data). This study uses several types of algorithms to classify the risk of heart disease in individuals using a medical record dataset. The algorithms used are the Naïve Bayes Algorithm, the K-Nearest Neighbors (KNN) Algorithm, and the Decision Tree as a comparison of the classification process.

The Naïve Bayes method is based on Bayes' theorem with a probabilistic approach. The assumption of the Naïve Bayes algorithm is that the attributes used in the dataset are independent of each other. The advantages of the Naïve Bayes algorithm are that it can handle large amounts of data and has high computational speed. Han, Kamber, and Pei [7] explain that Naïve Bayes is very effective in text and medical classification with stable results even when there is an increase in data. However, the weakness of Naïve Bayes is that the assumption of independence between attributes in medical data is unrealistic.

Meanwhile, the K-Nearest Neighbors (KNN) algorithm classifies new data by considering the k closest and most similar neighbors using the Euclidean Distance metric. The advantage of KNN is that it is easy to implement and does not require training data. Tan, Steinbach, and Kumar [8] explain that the main weaknesses of KNN are its sensitivity to unbalanced data and the number of k selected. In addition, adding data requires more computation time, because predictions require distance calculations against training data.

ISSN 2502-3357 (Online)

A significant difference can be seen in Decision Trees, namely the use of a rule-based approach in presenting the decision-making process in a decision tree structure. The nodes are attributes, and the branches contain the results of the process. The advantage of Decision Trees is that they are easy to understand. Quinland [9] explains that Decision Tree algorithms can handle categorical and numeric data well. However, Decision Trees have the disadvantage of tending to overfit when the tree structure is too complex, if there is a lot of noise in the training data.

This study aims to compare the performance of the Naïve Bayes, K-Nearest Neighbors (KNN), and Decision Tree algorithms in classifying risk factors for heart disease. The test was conducted by measuring the accuracy, precision, recall, and F1 score of each algorithm. It is hoped that this study will identify the most optimal algorithm for improving the accuracy of early detection of heart disease risk, thereby assisting medical professionals in diagnosing heart disease.

2. State of the Art

Predicting the risk of heart disease using machine learning has become a growing focus of research due to its ability to support early screening and clinical decision-making. Several classical algorithmic methods such as Naïve Bayes (NB), K-Nearest Neighbors (KNN), and Decision Tree (DT) are still widely used due to their ease of implementation, high interpretability (especially in Decision Tree), and computational efficiency (Naïve Bayes). However, based on several research results, there is no single algorithm that is consistently accurate across all types of datasets. Model performance is highly dependent on data quality, preprocessing, feature selection techniques, and validation strategies. [10]

2.1. Naïve Bayes Algorithm

Naïve Bayes is a probabilistic algorithm based on Bayes' theorem. This algorithm assumes that features are independent, so it can be used with large data sets [11].

Bayes' Theorem Formula:

$$P(y|X) = \frac{P(X|y)P(y)}{P(X)} \quad (1)$$

Conditions:
y = class variable (heart attack: yes/no)
X = risk factor for heart attack If feature X is numerical data, then the conditional probability

P(X|y) can be calculated using the Gaussian distribution:

$$P(X|y) = \frac{1}{\sqrt{2\pi\sigma^{2}}} e^{\frac{-(x-\mu)^{2}}{2\sigma^{2}}} \quad (2)$$

Conditions:
μ = mean of data in a particular class
σ^2 = variance of data in a particular class

2.2. K-Nearest Neighbors Algorithm

The K-Nearest Neighbors (K-NN) algorithm is a supervised learning method and belongs to the instance-based learning group [12]. The lazy learning technique in the KNN method is commonly used in data prediction or classification [13]. The initial stage of the KNN method process flow is to determine the K training data objects that are closest to the test data objects [14].
The Euclidean distance formula is as follows:

$$d_i = \sqrt{\Sigma_{n=1}^{p}(x_{2i} - x_{1i})} \quad (3)$$

Conditions:
d_i = Distance
x_i = Training Dataset
y_i = Testing Dataset
i = Variable Dataset
n = Dimension Dataset

The stages of implementation using the K-Nearest Neighbor (KNN) method are described in the following explanation:

1. Initial stage of determining the value of k
2. Determine the distance between the testing dataset and the training data
3. Next, sort the distances and determine the objects closest to k
4. Determine the closest category
5. Next, determine the majority of the K closest neighbors
6. The final stage is the prediction result with the test data point

2.3. Decision Tree Algorithm

A decision tree is a decision tree algorithm used for structured decision making. Decision trees are widely used by researchers for data classification [15]. This algorithm selects the best attribute as the root of the tree

DOI: https://doi.org/10.65126/jocosir.v3i3 Methods

ISSN 2502-3357 (Online)

based on Information Gain or Gini Index [16]. In line with this explanation, decision trees have inputs that are tested based on sample data used for decision trees that will be tested for accuracy [17]. The Decision Tree algorithm has been extensively tested and classified by researchers [17][18].
With Gian's Formula :

$$IG(S, A) = Entropy(S) - \Sigma \frac{|S_v|}{|S|} Entropy(S_v) \quad (4)$$

Conditions:
S = Initial dataset
A = Attributes tested
Sv = Data subset based on attribute A

With the Entropy Formula:

$$Entropy(S) = -\Sigma P_i Log_2 P_i \quad (5)$$

If the Gini Index is used, then the calculation is as follows:

$$Gini(S) = 1 - \Sigma P_i^2 \quad (6)$$

2.4. Previous Research

Related research conducted by Dikayanto, Agil Langga, and Nilogiri Agung entitled "Comparison of Naïve Bayes Algorithm with K -Nearest Neighbor Algorithm for Heart Disease Prediction," tested the Naive Bayes algorithm, which produced 10-fold cross-validation by testing the fourth data set, achieving an accuracy of 90.00% and a precision of 86.67%. Then, the K-Nearest Neighbors algorithm resulted in an accuracy of 7 neighbors with 10-fold cross-validation testing by testing the 4th data folder with an accuracy of 80.00% and a precision of 90.00% [19]. The similarity between this study and the author's study is that both use the Naïve Bayes algorithm and the K-Nearest Neighbor algorithm as problem-solving methods, while the difference is that the author's study adds a Decision Tree algorithm as an additional problem-solving method, and the related study only uses two algorithms.

Then, another study conducted by Muthohhar and Prihanto compared the Decision Tree, Naïve Bayes, and Random Forest Classifier algorithms with cases of heart disease classification. Testing results with average data obtained a value of 0.844 with random search using the Decision Tree Algorithm and grid search 0.84. Naïve Bayes was tested without any difference between evaluations using random search and grid search, only 0.85. Random Forest was also tested in this study with a classifier value of 0.852 and grid search 0.868. The classification results showed that the Random Forest Classifier Algorithm was the best at classifying heart disease [20]. The similarity with the author's research was in the use of the Decision Tree Algorithm and Naïve Bayes for heart disease classification, while the difference was that the author's research added K-Nearest Neighbors, whereas the related research used Random Forest.

Then, research by Tuloli, et al compared the performance of the C4.5, Naïve Bayes, and K-Nearest Neighbors (KNN) algorithms for predicting heart disease using a clinical dataset. The test results showed that the C4.5 algorithm was superior to the other algorithms, with an accuracy rate of 81.07%, while Naïve Bayes had an accuracy of 79.10% and KNN had an accuracy of 75.68%, indicating that C4.5 is effective in detecting the risk of heart disease [21]. From this data, it can be seen that machine learning is effective in classifying cases. Researchers developed a comparison of Naïve Bayes, K-Nearest Neighbors, and Decision Tree to analyze data accuracy with the latest, larger dataset (big data). The research conducted by previous researchers appears to be significantly different from the research conducted by the author.

3. Research Methodology

This study uses three methods of classification, namely the Naïve Bayes method, K-Nearest Neighbors (KNN), and Decision Tree in analyzing and classifying big data. The initial stage involves identifying problems and formulating objectives so that the research remains within a focused context.

Next, data collection was carried out related to the classification of the Naïve Bayes, K-Nearest Neighbors (KNN), and Decision Tree methods using relevant and credible sources. Next, data cleaning was performed on the obtained data to clean up inefficient, duplicated, and invalid data. Then, data preprocessing was performed, including feature selection, which has a significant influence on the classification results.

The K-Nearest Neighbors (KNN) method is implemented by determining the optimal K value, which aims to measure data proximity. Then, the Naïve Bayes method is implemented to calculate probabilities according to attributes and classify the highest probability values. Then, evaluate the model's performance by measuring the accuracy and credibility of the method used. The evaluation stage is carried out using accuracy, precision, and recall metrics to measure the method's ability to classify accurately and precisely.

After testing these three methods, the final step was to compare the Naïve Bayes, K-Nearest Neighbors (KNN), and Decision Tree methods to find patterns, relationships, and new information that could contribute to the classification model in this study.

DOI: https://doi.org/10.65126/jocosir.v3i3 Methods

ISSN 2502-3357 (Online)

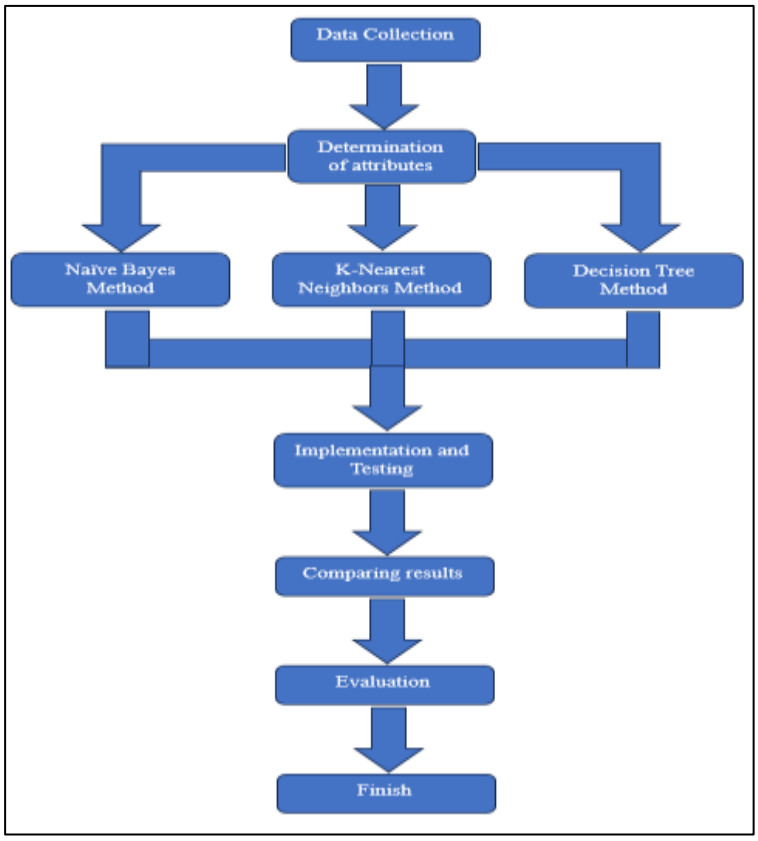

Figure 1. Research Methods and Stages [22]

4. Results and Discussion

In this study, researchers used machine learning classification models with the Naive Bayes, K-Nearest Neighbors (KNN), and Decision Tree methods using Python. The Heart Disease dataset used in this study was obtained from the website kaggle.com under the title "Heart Disease Prediction Dataset." This data contains 6025 patient data, where the dataset is divided into test data and training data with a ratio of 80% test data and 20% training data, compiled with 14 attributes or clinical features with the classification of patients who have heart disease or not.

4.1 Dataset Collection

In this study, the data used was obtained from a publicly available heart disease dataset from Kaggle containing 6025 data points, accessible via the link https://www.kaggle.com/datasets/iamcaano/heart-disease-prediction-dataset. This dataset contains various health parameters related to heart disease risk, such as age, gender, blood pressure, cholesterol levels, and several other relevant variables. Table 1 is the dataset collection for testing in this study :

Table 1. Test Collection Dataset Comparing the Naïve Bayes, K-Nearest Neighbors, and Decision Tree Methods for Classifying Heart Disease Risk Factors

DOI: https://doi.org/10.65126/jocosir.v3i3 Methods

ISSN 2502-3357 (Online)

No	age	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak	slope	ca	thal	target
1	52	1	0	125	212	0	1	168	0	1	2	2	3	0
2	53	1	0	140	203	1	0	155	1	31	0	0	3	0
3	70	1	0	145	174	0	1	125	1	26	0	0	3	0
4	61	1	0	148	203	0	1	161	0	0	2	1	3	0
5	62	0	0	138	294	1	1	106	0	19	1	3	2	0
6	58	0	0	100	248	0	0	122	0	1	1	0	2	1
7	58	1	0	114	318	0	2	140	0	44	0	3	1	0
8	55	1	0	160	289	0	0	145	1	8	1	1	3	0
9	46	1	0	120	249	0	0	144	0	8	2	0	3	0
10	54	1	0	122	286	0	0	116	1	32	1	2	2	0
6018	630	10	30	1100	2020	0	0	1460	10	19	20	0	20	10
6019	420	10	10	1330	1950	0	10	1320	0	0	10	10	30	10
6020	500	10	20	1180	1450	0	0	1250	0	9	0	30	0	10
6021	540	0	20	1230	2820	0	0	990	10	22	10	20	20	10
6022	440	10	10	1300	2190	0	0	1880	0	0	20	0	20	10
6023	540	0	0	1800	3250	0	20	1180	10	36	10	40	20	0
6024	570	10	20	1520	1320	10	10	1770	0	1	20	30	20	10
6025	520	10	0	1120	2300	0	10	1600	0	0	20	10	20	0

(Source: https://www.kaggle.com/datasets/iamcaano/heart-disease-prediction-dataset. Accessed on October 22, 2025, at 10:21 p.m. WIB)

4.2. Testing The Naïve Bayes Algorithm

Based on testing conducted on a heart disease dataset using the Naïve Bayes Algorithm in Google Colab (Python), with a test data proportion of 80% and training data proportion of 20%, performance evaluation results showed an accuracy of 76%, precision of 75%, recall of 76%, and an F1-Score of 76%. A visualization of the calculations is available in Figure 2 :

```
==== Naive Bayes ====
Accuracy   : 0.7635
Precision   : 0.7554
Recall     : 0.7635
F1-Score   : 0.7584

Classification Report:
              precision    recall  f1-score   support

         0.0       0.59      0.52      0.55       340
         1.0       0.82      0.86      0.84       865

    accuracy                           0.76      1205
   macro avg       0.71      0.69      0.70      1205
weighted avg       0.76      0.76      0.76      1205
```

Figure 2. Results of Naïve Bayes Algorithm Testing

4.3. Testing the K-Nearest Neighbors Algorithm

The results of testing the heart disease dataset with a comparison of 80% test data and 20% training data in the K-Nearest Neighbors (KNN) algorithm -Nearest Neighbors (KNN) algorithm in Python using Google Colab yielded an accuracy score of 78%, a precision score of 81%, a recall score of 78%, and an F1-Score of 79%. The results of the calculations can be seen in Figure 3 below :

DOI: https://doi.org/10.65126/jocosir.v3i3 Methods

ISSN 2502-3357 (Online)

```
==== K-Nearest Neighbors ====
Accuracy   : 0.7801
Precision  : 0.8078
Recall     : 0.7801
F1-Score   : 0.7878

Classification Report:
              precision    recall  f1-score   support

         0.0       0.58      0.77      0.66       340
         1.0       0.90      0.78      0.84       865

    accuracy                           0.78      1205
   macro avg       0.74      0.78      0.75      1205
weighted avg       0.81      0.78      0.79      1205
```

Figure 3. Results of K-Nearest Neighbors (KNN) Algorithm Testing

4.4. Testing the Decision Tree Algorithm

The results of testing the heart disease dataset with a comparison of 80% test data and 20% training data in the Decision Tree Algorithm in Python using Google Colab Research yielded an accuracy score of 86%, a precision score of 87%, a recall score of 86%, and an F1-Score of 86%. The calculation results can be seen in Figure 4 below :

```
==== Decision Tree ====
Accuracy   : 0.8589
Precision  : 0.8671
Recall     : 0.8589
F1-Score   : 0.8616

Classification Report:
              precision    recall  f1-score   support

         0.0       0.72      0.82      0.77       340
         1.0       0.93      0.87      0.90       865

    accuracy                           0.86      1205
   macro avg       0.82      0.85      0.83      1205
weighted avg       0.87      0.86      0.86      1205
```

Figure 4. Results of K-Nearest Neighbors (KNN) Algorithm Testing

After testing, the researchers then compared the test results from several algorithms, using the Accuracy, Precision, Recall, and F1-Score metrics. This was done to identify the most effective and appropriate algorithm for the heart disease data set, as presented in Table 2.

Table 2. Comparison Table of Results Comparing the Naïve Bayes, K-Nearest Neighbors, and Decision Tree Methods for Classifying Heart Disease Risk Factors

=== Comparison Table of Results ===					
	Model	Accuracy	Precision	Recall	F1-Score
0	Naïve Bayes	0.763485	0.755393	0.763485	0.758371
1	K-Nearest Neighbors	0.780083	0.807777	0.780083	0.787784
2	Decision Tree	0.858921	0.867138	0.858921	0.861559

DOI: https://doi.org/10.65126/jocosir.v3i3 Methods

ISSN 2502-3357 (Online)

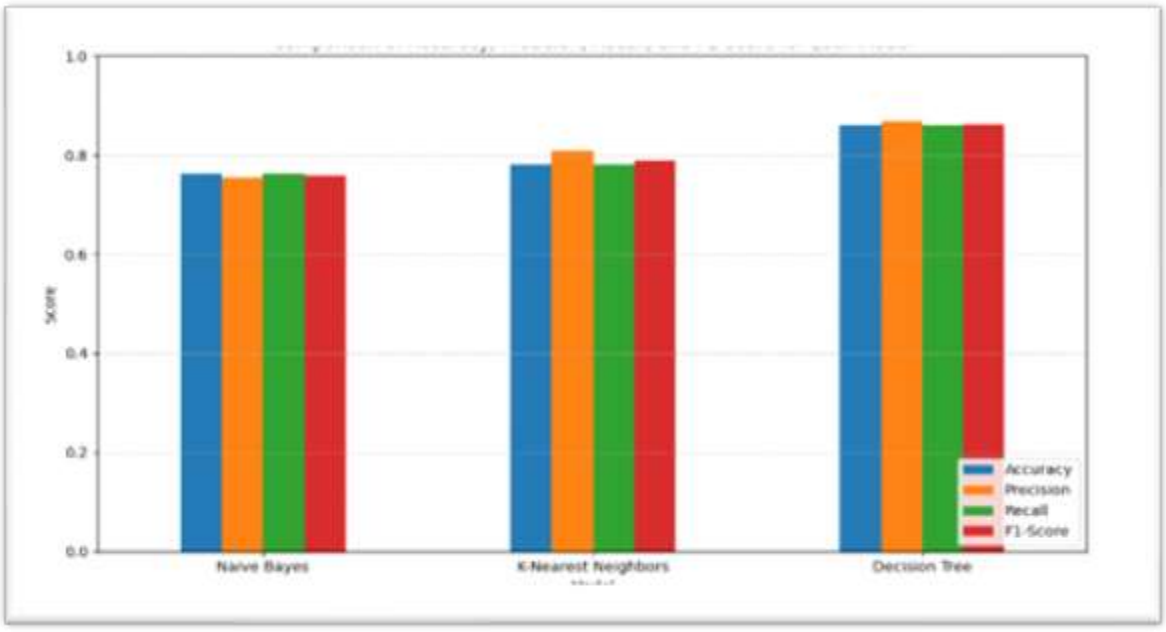

Figure 5. Comparison of Accuracy, Precision, Recall, and F1 Score for Each Model

Table 2 and Figure 5 show that the Naïve Bayes method has an accuracy value of 76%, precision value of 75%, recall value of 76%, and f1-score value of 76%, while the K -Nearest Neighbors method has results of 78% accuracy, 81% precision, 78% recall, and 79% f1-score, and the Decision Tree method has results of 86% accuracy, 87% precision, 86% recall, and 86% f1-score.

5. Conclusions

The results of this study conclude that the application of the Decision Tree method to the heart disease dataset yielded the best performance on the 20% training data, with an accuracy of 86%, precision of 87%, recall of 86%, and an F1-score of 86%. The application of the K-Nearest Neighbors method to the heart disease dataset obtained values on 20% training data, namely accuracy of 78%, precision of 81%, recall of 78%, and f1-score of 79%. The application of the Naïve Bayes method to the heart disease dataset obtained values on 20% training data, namely accuracy of 76%, precision of 75%, recall of 76%, and f1-score of 76%. Based on the results of the method comparison described above, the Decision Tree method has more effective and accurate performance in detecting patients with a high risk of heart disease compared to the K-Nearest Neighbors and Naïve Bayes methods in terms of accuracy, precision, recall, and F1-score.

6. References

[1] W. H. O. (WHO), "Cardiovascular diseases," World Health Organization (WHO). [Online]. Available: https://www.who.int/health-topics/cardiovascular%0Adiseases#tab=tab_1

[2] M. Anita, I. G. D. Yulianti, and S. V. Pasaribu, "Klasifikasi Faktor Risiko Penyakit Jantung Menggunakan Machine Learning," *HOAQ (High Educ. Organ. Arch. Qual. J. Teknol. Inf.*, vol. 16, no. 1, pp. 68–78, 2025.

[3] Kementerian Kesehatan Republik Indonesia. "Profil Kesehatan Indonesia Tahun 2022." Jakarta: Kemenkes RI, 2022.

[4] Winda Sinthya Naomi, Intje Picauly, and Sarci Magdalena Toy, "FAKTOR RISIKO KEJADIAN PENYAKIT JANTUNG KORONER(Studi Kasus di RSUD Prof. Dr. W. Z. Johannes Kupang)," *Media Kesehat. Masy.*, vol. 3, no. 1, pp. 99–107, 2021.

[5] S. N. N. Arif, A. M. Siregar, S. Faisal, and A. R. Juwita, "Klasifikasi Penyakit Serangan Jantung Menggunakan Metode Machine Learning K-Nearest Neighbors (KNN) dan Support Vector Machine (SVM)," *J. Media Inform. Budidarma*, vol. 8, no. 3, p. 1617, 2024.

[6] S. C. Dewi, C. E. Putra, and A. G. Nugraheni, "Implementasi Metode K-Nearest Neighbors (KNN) dan Naïve Bayes untuk Klasifikasi Penyakit Jantung," *Technol. Informatics Insight J.*, vol. 3, no. 2, pp. 76–94, 2024.

[7] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, 3rd ed., Elsevier, 2012.

[8] P.-N. Tan, M. Steinbach, and V. Kumar, *Introduction to Data Mining*, 2nd ed., Pearson, 2019.

[9] J. R. Quinlan, *C4.5: Programs for Machine Learning*, Morgan Kaufmann Publishers, 1993.

[10] M. Hasan and A. M. Abdulazeez, "A Review of Heart Disease Classification Based on Machine Learning Algorithms," *Indonesian Journal of Computer Science*, vol. 13, no. 2, Apr. 2024.

[11] K. Kartianom, A. Arpandi, G. K. Kassymova, and O. Ndayizeye, "Prediction model of teacher candidate student graduation status: Decision Tree C4.5, Naive Bayes, and k-NN," *Ekspose J. Penelit. Huk. dan Pendidik.*, vol. 21, no. 2, pp. 1419–1427, 2022.

DOI: https://doi.org/10.65126/jocosir.v3i3 Methods

ISSN 2502-3357 (Online)

[12] M. Kurniawan Khamdani, N. Hidayat, and R. K. Dewi, "Implementasi Metode K-Nearest Neighbor Untuk Mendiagnosis Penyakit Tanaman Bawang Merah," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 1, pp. 11–16, 2021.

[13] D. Cahyanti, A. Rahmayani, and S. A. Husniar, "Analisis performa metode Knn pada Dataset pasien pengidap Kanker Payudara," *Indones. J. Data Sci.*, vol. 1, no. 2, pp. 39–43, 2020.

[14] A. Yudhana, S. Sunardi, and A. J. S. Hartanta, "Algoritma K-Nn Dengan Euclidean Distance Untuk Prediksi Hasil Penggergajian Kayu Sengon," *Transmisi*, vol. 22, no. 4, pp. 123–129, 2020.

[15] F. S. Pattiiha and H. Hendry, "Perbandingan Metode K-NN, Naïve Bayes, Decision Tree untuk Analisis Sentimen Tweet Twitter Terkait Opini Terhadap PT PAL Indonesia," *JURIKOM (Jurnal Ris. Komputer)*, vol. 9, no. 2, p. 506, 2022.

[16] J. R. Quinlan, *C4.5: Programs for Machine Learning*, San Mateo, CA: Morgan Kaufmann Publishers, 1996.

[17] F. A. Sianturi, "Analisa Decision Tree Dalam Pengolahan Data Siswa," *MEANS (Media Informasi Analisa dan Sistem)*, vol. 3, no. 2, pp. 166–172, 2018.

[18] S. Saifullah, M. Zarlis, Z. Zakaria, and R. W. Sembiring, "Analisa Terhadap Perbandingan Algoritma Decision Tree Dengan Algoritma Random Tree Untuk Pre-Processing Data," *Jurnal Sains Komputer & Informatika (J-SAKTI)*, vol. 1, no. 2, pp. 180–186, Sept. 2017

[19] D. A. Langga and Dkk, "Perbandingan Algoritma Naive Bayes Dengan Algoritma K-Nearest Neighbor Untuk Prediksi Penyakit Jantung," *J. Chem. Inf. Model.*, vol. 53, no. 9, pp. 1689–1699, 2019.

[20] J. D. Muthohhar and A. Prihanto, "Analisis Perbandingan Algoritma Klasifikasi untuk Penyakit Jantung," *J. Informatics Comput. Sci.*, vol. 04, pp. 298–304, 2023.

[21] M. S. Tuloli, T. S. Kinanti, and L. N. Amali, "Perbandingan Algoritma C4.5, Naive Bayes, dan K-Nearest Neighbors untuk Prediksi Penyakit Jantung," *Jambura J. Informatics*, vol. 1, no. 1, pp. 11–21, 2025.

[22] S. C. Dewi, C. E. Putra, and A. G. Nugraheni, "Implementasi Metode K-Nearest Neighbors (KNN) dan Naive Bayes untuk Klasifikasi Penyakit Jantung," *Technology and Informatics Insight Journal*, vol. 3, no. 2, pp. 76–94, 2024.

DOI: https://doi.org/10.65126/jocosir.v3i3 Methods