

Comparison of Decision Tree C4.5 and Single Layer Perceptron (SLP) Implementation for Stroke Classification

Indah Marsya Fitadea¹, Sulistiani², Anita Desiani^{3*}, Yuli Andriani⁴, Endang Sri Kresnawati⁵, Siti Rusdiana Puspa Dewi⁶

^{1,2,3,4,5}Mathematics Department Sriwijaya University; Jl. Raya Palembang – Prabumulih Km. 32, Indralaya Indah, Kec. Indralaya, Kabupaten Ogan Ilir; (+62)(711)580-069

⁶Department of Dentistry Sriwijaya University; Jl. Raya Palembang – Prabumulih Km. 32, Indralaya Indah, Kec. Indralaya, Kabupaten Ogan Ilir; (+62)(711)580-069

Received: 14-08-2025
Accepted: 11-08-2025

Keywords:

Stroke;
Data mining;
SLP; Decision tree C4.5;
Artificial Intelligence.

Correspondent Email:
anita_desiani@unsri.ac.id

Abstract. Stroke is a condition that occurs when blood flow to the brain is blocked due to a blockage of an artery, a buildup of blood, or a rupture of a blood vessel. The cause can be abnormalities like the blood or weak blood vessel walls. Stroke ranks as the second leading cause of death and the third leading cause of disability worldwide. To prevent a more severe impact, early detection is very important as many small factors are often not realized to be the cause of stroke. Early detection is done by utilizing information technology such as machine learning. Machine learning was used in Decision Tree Classifier 4.5 (C4.5) and Single Layer Perceptron (SLP). The results of research conducted with the percentage split method show that the C4.5 algorithm produces an accuracy value of 71%, while the SLP algorithm produces a higher accuracy of 76%. In addition, the precision and recall values in class 0 (negative) are 71% and 70%, respectively, while class 1 (positive) produces values of 74% and 72%. Then, precision and recall in the SLP algorithm in class 0 (negative) resulted in values of 76% and 71% while in class 1 (positive) resulted in values of 76% and 81%. From both algorithms, it can be concluded that the SLP algorithm provides more accurate and effective prediction results for stroke disease detection.

1. INTRODUCTION

A stroke is a medical condition that occurs when the blood supply to the brain is disrupted. This condition can be caused by a blockage in an artery or a ruptured blood vessel [1]. Blockages are often caused by clots or fatty deposits in the arteries, while ruptured blood vessels can occur due to weakened vessel walls or blood disorders. As a result, the brain is deprived of oxygen and nutrients, which can cause brain tissue damage. The most common symptom of stroke is frequent weakness or numbness in the hands, feet, or face [2]. The World Health Organization (WHO) reports that there are 13.7 million cases of stroke each year, and about 5.5 million of those cases result in

death. The high number of stroke cases is due to delays in diagnosing the disease. Early detection is essential to help prevent more severe strokes. Early detection that has been proven effective involves the use of machine learning. Machine learning is a data mining method that allows computers to recognize patterns and make diagnoses without manually entering program code continuously. Several methods that have been developed in machine learning are Decision Tree Classifier 4.5 and Single Layer Perceptron (SLP).

The C4.5 algorithm is a Decision Tree algorithm that works by calculating the entropy and gain values of each attribute variable used. Other decision tree algorithms such as ID3 can

only be used on categorical data, whereas C4.5 is capable of handling both categorical and numerical attributes. In addition, the C4.5 algorithm has several advantages, including being easy to implement with low computation time, reducing the use of unnecessary features, and identifying attributes that have high and low influence [3]. In the medical field, doctors are greatly assisted by algorithms in making important decisions for specific pathology reports [4]. Several studies have used the C4.5 decision tree algorithm in classification, namely Maulana Sidiq *et al.* [5], who implemented the C4.5 algorithm in the classification of stroke, producing a fairly good accuracy of 91%, but precision and recall values of only 51.6% and 53.88%. Gusriani Fitri *et al.* [6] in the classification of stroke using the C4.5 algorithm produced an accuracy value of 90.22%, but the precision and recall values were still below 60%, namely precision of 57.135% and recall of 55.19%. Additionally, Afifuddin and Hakim [7] used the C4.5 algorithm in the classification of diabetes mellitus with an accuracy of 96%, precision of 96.5%, and recall of 95.5%. Although the C4.5 algorithm is easy to apply, it has limitations in reading large amounts of data because it will cause overly complex branching, making the use of the algorithm suboptimal [3]. The C4.5 algorithm is also prone to overfitting, a condition where the program learns the data in too much detail, which will result in lower accuracy values [8]. Another method that can be used in stroke classification is the SLP algorithm.

Single Layer Perceptron (SLP) is an artificial neural network that only has one layer, namely input and output. Some of the advantages of SLP are its ability to acquire knowledge even though the data used is uncertain, its ability to generalize and extract patterned data, and its ability to create knowledge patterns through learning or self-adjustment [9]. Several previous studies have used the SLP algorithm in the field of health, such as Dewi [10], who showed an accuracy value of 99% for the classification of Alzheimer's disease. [11] on the classification of hepatitis-c disease produced an accuracy value of 95%. In addition, Annisa Nurba Iffah'da and Anita Desiani [12] conducted research to predict primary biliary cirrhosis and produced an accuracy of 63%. From these

accuracy results, it can be concluded that the SLP algorithm has an accuracy value above 60%. One of the shortcomings of the SLP algorithm is that it is not effective for handling data problems with unbalanced classes, where one class has a much larger amount of data than the other classes. This can cause the model to be too inclined to predict the dominant class, resulting in low accuracy for the minority class. In this case, the SLP algorithm cannot be used for tasks that require data with complex classifications or many classes. In addition, SLP also has limitations in grouping data with a high level of accuracy.

Based on the advantages and disadvantages of both algorithms, this study will conduct tests to compare the C4.5 decision tree algorithm and SLP using the percentage split method. In the percentage split method, the dataset is divided into two parts, namely training data and testing data, with 30% of the dataset used for testing data and 70% used as training data. This study aims to determine the algorithm with the best performance in stroke classification based on the precision, recall, and accuracy values of both algorithms. The results of this study are also expected to provide insight into selecting the best algorithm for more accurate and efficient stroke classification, thereby helping to reduce the incidence of stroke. In addition, this study can be an effective solution in preventing and managing risk factors that contribute to stroke.

2. LITERATURE REVIEW

2.1. Stroke

A stroke is a condition in which the blood supply to part of the brain is disrupted due to a blockage or rupture of a blood vessel, so that brain tissue does not receive enough oxygen, resulting in damage or death of brain cells [13]. The signs of stroke appear suddenly, including weakness of the facial muscles, arms, or legs on one side of the body, difficulty speaking or understanding language, blurred or lost vision in one or both eyes, and sudden dizziness accompanied by loss of balance [14].

2.2. Classification

Classification is a technique for recognizing a characteristic of data that will then be grouped into classes [15]. Classification

in data mining is the process of predicting or classifying data into a specific group or category based on existing attributes [16]. Classification is used in various fields, ranging from disease prediction, document identification, to creditworthiness determination. This technique helps provide a clearer picture of the data, thereby facilitating the analysis process.

2.3. Decision Tree C4.5

Decision Tree is one of the most commonly used supervised learning classification methods. Decision Tree combines branching techniques to show every possible decision [17]. Decision Tree C4.5 is a development of the ID3 algorithm created by Ross Quinlin [5]. The rules generated from the classification process using the Decision Tree C4.5 algorithm can be used to predict attribute values in new data [18].

2.4 Single Layer Perceptron (SLP)

Single Layer Perceptron (SLP) is an artificial neural network model consisting of a single layer of neurons that receive input from data through weights and biases (thresholds) [19]. SLP has five main components, namely input values, input layers, weighting, output layers, and output values [9].

3. RESEARCH METHOD

This study will use several methods, namely data collection, data preprocessing, data splitting, analysis of results, and model comparison. The flow of stages to be carried out in this study is shown in Figure 1.

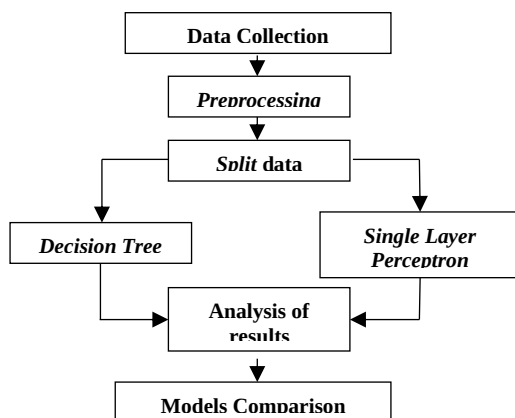


Figure 1. Research Flow

3.1. Data Description

The data in this study will use a dataset from the Kaggle website titled “Stroke Prediction Dataset” in Excel CSV format [20]. The dataset consists of 5,110 rows and 12 attributes. The rows contain information about the symptoms experienced by stroke patients and predictions of whether they have stroke or not. The attributes used in this dataset can be seen in Table 1.

Table 1. Data Attributes and Descriptions

No	Attribute	Data Type	Number of Null Data
1	Id	Categories	0
2	Gender	Categories	0
3	Age	Numeric	0
4	Hypertension	Numeric	0
5	Heart Disease	Numeric	0
6	Ever Married	Categories	0
7	Work Type	Categories	0
8	Residence Type	Categories	0
9	Avg Glucose	Numeric	0
10	BMI (Body Mass Index)	Numeric	201
11	Smoking Status	Categories	18528
12	Stroke	Categories	0

3.2. Preprocessing Data

Before data is processed, data preprocessing is essential in data mining. Data preprocessing is the process of processing raw data in preparation for further processing. One preprocessing technique is data cleaning. Data cleaning is a process of removing noise and variable and irrelevant data [21]. In this study, attributes that have no influence will be removed, such as the Id and Gender attributes,

leaving 10 attributes. In addition, empty data or null values will also be removed. The purpose of preprocessing in data mining is to convert data into a form or format that can facilitate and optimize the process according to user needs.

3.3. Split Data

Percentage split is a technique of dividing data into two parts, namely training data and testing data, with a certain percentage [10]. Percentage split can be used to evaluate an algorithm that can predict data percentages. In this study, the data percentage used is 30:70, with 30% of the dataset used as testing data and 70% used as training data selected at random. Thus, this method can help optimize the accuracy of the prediction model and reduce errors.

3.4. Data Training

3.4.1. Decision Tree C4.5

The C4.5 algorithm generates a simple decision tree by converting complex data into easy-to-understand rules so that the class type of new records can be predicted [18]. The stages in classification using the C4.5 algorithm are:

- Selecting attributes as the root of the decision tree.
- The attribute is selected as the root based on the highest Information Gain (IG). Gain is a quantity that measures how well an attribute classify the data [8]. Equation 1 is used to obtain the gain value for each attribute.
- Each value is made into a branch.
- Select the attribute that will become the branch.
- Repeat the process until all cases in node N branch obtain a class.

$$Gain(S, A) = Entropy(S) - \left(\sum_{i=1}^n \frac{|S_i|}{|S|} Entropy(S_i) \right) \quad (1)$$

Explanation :

S : Set of cases

A : Attribute

n : Number of partitions of attribute A

|S_i| : Number of cases in partition i

The method for calculating entropy is shown in Equation 2.

$$Entropy(P) = \sum_{i=1}^n -p_i \log_2 p_i \quad (2)$$

P : Set of cases

n : number of partitions P

p_i : number of cases in partition i

3.4.2. Algorithm Single Layer Perceptron (SLP)

A Single Layer Perceptron (SLP) is a simple Artificial Neural Network (ANN) model consisting of only one layer with interconnected weights. SLP directly processes inputs and generates outputs without having to go through a hidden layer. The main components of SLP are input values, input layer, weighting, output layer, and output values [9]. The architecture of the SLP algorithm can be seen in Figure 2.

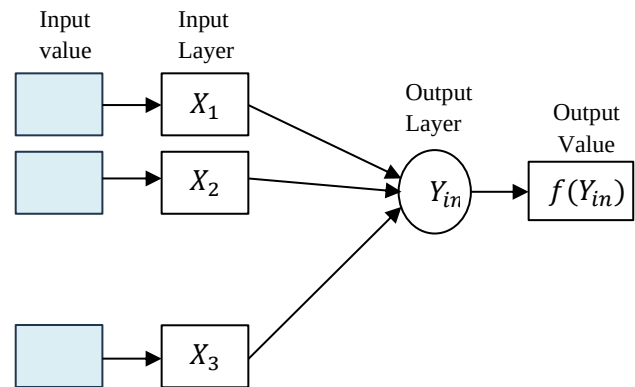


Figure 2. Single Layer Perceptron Architecture

To calculate the SLP algorithm, you can use the following steps:

- Determine the learning rate (α) value, which is greater than zero.
- Set the threshold (θ) equal to zero.
- Initialize the weight and bias values. This step involves setting the initial values for the weight and bias. There are two approaches to initialization: zero initialization and random initialization.
- Set the activation for the input vector.

$$x = s^{(q)} \quad (3)$$

Explanation :

x : Activation vector

s^(q) : Activation function

e. Calculating the output unit response.

$$y_{in} = \sum x_q + w_q + b \quad (4)$$

Explanation :

y_{in} : Input vector output unit
 w : Weight
 b : Bias

f. Calculating output values and error values.

$$y = f(y_{in}) = \frac{1}{1+e^{-y_{in}}} \quad (5)$$

$$t = c - y \quad (6)$$

Explanation :

y : output value
 $f(y_{in})$: Sigmoid activation function
 e : Constant
 t : error value
 c : Target Value

g. If in step number 5 the classification is according to the target, namely $y \neq t$ or the value of $t = \theta$, then the condition stops and will be true. If the value of $t = \theta$, then the condition will be false, so the weight and bias are corrected using the formula:

$$w_{new} = w_{old} + \alpha \cdot t^{(q)} \cdot x \quad (7)$$

$$w_{new} = b + \alpha \cdot t^{(q)} \quad (8)$$

h. Repeat steps 4 to 7 using epoch limits to process the training data, where the dataset has input unit values $q = 1, 2, 3, \dots Q$. This process will continue until the model produces class values that match the target.

3.5. Confusion Matrix

A confusion matrix is a table that serves as a testing method to calculate classification performance results by obtaining Accuracy, Precision, and Recall [15]. The Confusion Matrix table can be seen in Table 2. As can be seen in Table 2, True Negative (TN) is data that is correctly predicted as negative, and False Negative (FN) is data that should be positive but is predicted as negative. Meanwhile, True Positive (TP) is data that is correctly predicted as positive, and False Positive (FP) is data that should be negative but is predicted as positive.

Table 2. Confusion Matrix

		Prediction	
		Negative	Positive
Actual	Negative	True Negative (TN)	False Positive (FP)
	Positive	False Negative (FN)	True Positive (TP)

3.5.1. Accuracy

Error rate, also known as accuracy. Accuracy is a number that indicates how often a model makes correct or incorrect predictions from a data set. Generally, accuracy is calculated using test data that is independent from the training data to provide an objective evaluation of the model's capabilities. The formula for calculating accuracy can use equation (9) as follows [10].

$$Accuracy = \frac{TP+TN}{TP+FN+FP+TN} \times 100\% \quad (9)$$

3.5.2. Precision

Precision is a matrix used to measure a system's capacity to generate important data. Precision is the ratio between True Positive (TP) predictions and all predicted positive results. In data mining, the formula for calculating precision can use the following equation (10) [15].

$$Precision = \frac{TP}{TP+FP} \times 100\% \quad (10)$$

3.5.3. Recall

Recall, also known as sensitivity, is the ratio between the number of True Positive (TP) predictions combined with the total number of True Positives (TP), or an estimate of how accurately the original positive predictions are identified as positive. Recall relates to the ability of a test to identify True Positive (TP) results against all data that is actually positive [15]. Recall can be calculated using equation (11).

$$Recall = \frac{TP}{TP+FN} \times 100\% \quad (11)$$

4. RESULTS AND DISCUSSION

4.1. Decision Tree C4.5

The 419 data points will then be divided into two using a 30:70 percentage split technique. The result of the data division is 126 training data points and 293 testing data points, which is 40% more. From the available data, a decision tree is then formed using the C4.5 algorithm to produce a decision flow. The results of the attribute calculation as the root based on the highest gain show that if the patient meets the root decision attribute requirements, it can be predicted that the patient has a stroke, but if the opposite is true, then the next decision branch needs to be looked at until a conclusion is reached as to whether the patient's characteristics are positive or negative for stroke. The success of the C4.5 algorithm in classifying stroke using the percentage split technique can be seen in Table 3.

Table 3. Confusion Matrix C4.5

		Prediction	
		Negative	Positive
Actual	Negative	42	17
	Positive	19	48

Based on Table 3, there are a total of 126 data points that were successfully predicted using the percentage split method. Of the 126 data points, 42 were correctly predicted as Negative (0) and 17 were incorrectly predicted as Positive (1), which should have been classified as Negative (0). In the Positive (1) class, there were 47 data points that were predicted correctly and 19 data points that were incorrectly predicted, which should have been in the Positive class but were predicted to be in the Negative (0) class. With the results from the confusion matrix, the precision, recall, and accuracy values of each class were then calculated, the results of which can be seen in Table 4.

Table 4 shows the performance results of the C4.5 algorithm in the testing stage using the stroke dataset. The precision for the Negative (0) class was 71% and for the Positive (1) class was 74%, the Recall for the Negative

(0) class was 70% and for the Positive (1) class was 72%, and the accuracy value obtained was 71%.

Table 4. Precision, Recall, and Accuracy Values C4.5

Kelas	Presisi	Recall	Accuracy
Negative (0)	71%	70%	71%
Positive (1)	74%	72%	

4.2. Single Layer Perceptron

In this study, the perceptron model on the stroke dataset will be trained and validated 100 times. The training process uses a learning rate (α) value of 0.3, a threshold value of 0, and all weights and biases are initialized with a value of 0. The algorithm will repeatedly calculate the input and output values until the output value obtained is as expected. If the output value matches the expected value, the process stops. However, if the output value does not match, the weights and biases will be adjusted. During the training process, the model's performance will be evaluated at each epoch using the percentage split method. The training results will be compared at each epoch to ensure that the model is approaching optimal performance, as shown in Figure 3.

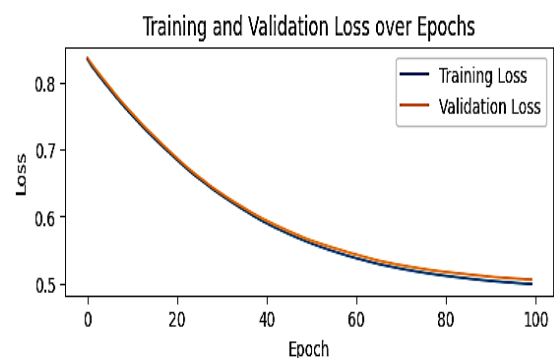


Figure 3. Epoch Training Results Using the Percentage Split Method

Figure 3 shows that the epoch value of 100 training and validation data is decreasing to 0.5. This indicates that the training and validation results of each epoch in the data are optimal, and the iteration is stopped and continued with the testing data. The testing data will produce precision, recall, and accuracy values as shown in Table 5.

Table 5. Precision, Recall, and Accuracy Values SLP

Class	Presisi	Recall	Accuracy
Negative (0)	76%	71%	76%
Positive (1)	76%	81%	

Table 5 shows that the results of the percentage split are a precision value of 76% for the negative class (0) and 76% for the positive class (1). The recall value for the negative class (0) is 71% and for the positive class (1) is 81%. The accuracy value obtained for the percentage split method on the stroke dataset is 76%.

4.3. Comparison of C4.5 and SLP Algorithm Results

Based on the results of both algorithms tested using the percentage split method, this study shows that the SLP algorithm provides better results with an accuracy of 76%. A comparison of precision, recall, and accuracy values can be seen in Figure 4.

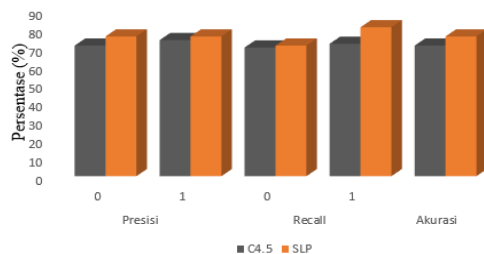


Figure 4. Comparison Diagram of Precision, Recall, and Accuracy of the C4.5 and SLP Algorithms

In Figure 4, it can be seen that the precision value in class 0 (negative), the recall value in class 1 (positive), and the accuracy of the SLP algorithm are greater than those of the C4.5 algorithm. On the other hand, the precision value in class 1 (positive) of both algorithms shows the same value, and the recall value in class 1 (positive) shows that the C4.5 algorithm value is greater than the SLP algorithm. The precision of the C4.5 algorithm in class 0 (negative) produces a value of 71% and in class 1 (positive) of 74%, while the precision of the SLP algorithm in class 0 (negative) produces a value of 76% and in class 1 (positive) of 76%.

Meanwhile, the recall of the C4.5 algorithm in class 0 (negative) produced a value of 70% and in class 1 (positive) 72%, while the precision of the SLP algorithm in class 0 (negative) produced a value of 71% and in class 1 (positive) 81%. The SLP algorithm produced an accuracy value of 76%, higher than the C4.5 algorithm, which only produced an accuracy value of 71%. This shows that SLP has superior performance, not only in terms of accuracy but also in terms of precision and recall. Therefore, it can be concluded that the SLP algorithm is more accurate and effective in classifying stroke datasets.

5. CONCLUSION

Based on the results of research on the classification of stroke using the C4.5 and Single Layer Perceptron (SLP) algorithms, it was found that the C4.5 algorithm achieved an accuracy of 71%, while SLP achieved an accuracy of 76%. Both algorithms can be considered quite good at detecting stroke, but SLP shows superior performance in terms of accuracy, precision, and recall. In the SLP algorithm, the range of accuracy, precision, and recall values is between 74% and 100%, while in the C4.5 Decision Tree algorithm, the range is between 70% and 100%. Thus, it can be concluded that SLP has better performance than the C4.5 algorithm in detecting stroke.

ACKNOWLEDGMENTS

The author would like to thank Dr. Anita Desiani S.Si., M.Kom and the Computation Laboratory I for their guidance, support, and facilities provided so that this research could be completed successfully. Thanks to all the lecturers of the Mathematics Department, Faculty of Mathematics and Natural Sciences (FMIPA) for all the support given to the author. The author also thanks the managers and review team of the Electrician Journal for providing a platform for the publication of this research.

REFERENCES

- [1] N. Y. Paramitha, A. Nuryaman, A. Faisol, E. Setiawan, and D. E. Nurvazly, "Klasifikasi Penyakit Stroke Menggunakan Metode Naïve Bayes," *J. Siger Mat.*, vol. 04, no. 01, pp. 11–16, 2023.
- [2] A. Puspitawuri, E. Santoso, and C. Dewi,

- "Diagnosis Tingkat Risiko Penyakit Stroke Menggunakan Metode K-Nearest Neighbor dan Naïve Bayes," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 3, no. 4, pp. 3319–3324, 2019.
- [3] A. A. Mortara, M. Permatasari, A. Desiani, Y. Andriani, and M. Arhami, "Perbandingan Algoritma C4.5 dan Adaptive Boosting dalam Klasifikasi Penyakit Alzheimer Comparison of C4.5 and Adaptive Boosting Algorithms in Alzheimer's Disease Classification," vol. 13, no. September, 2023, doi: 10.34010/jati.v13i2.
- [4] R. E. Pambudi, Sriyanto, and Firmansyah, "Klasifikasi Penyakit Stroke Menggunakan Algoritma Decision Tree C4.5," *J. Tenknika*, vol. 16, no. 02, pp. 221–226, 2022.
- [5] C. M. Sidiq, A. Faqih, and G. Dwilestari, "Algoritma Decision Tree C4.5 Digunakan Untuk Mengklasifikasikan Data Stroke," *JATI (Jurnal Mhs. Tek. Inform.*, vol. 8, no. 2, pp. 1869–1874, 2024, doi: 10.36040/jati.v8i2.8388.
- [6] N. Gusrialni Fitri, S. Adilya, and F. Azizi, "Comparison of the Naive Bayes Classification System and C4.5 for the Diagnosis of Stroke Perbandingan Sistem Klasifikasi Naive Bayes dan C4.5 Untuk Diagnosa Penyakit stroke," *SENTIMAS Semin. Nas. Penelit. dan Pengabd. Masy.*, pp. 49–55, 2023.
- [7] A. Afifuddin and L. Hakim, "Deteksi Penyakit Diabetes Mellitus Menggunakan Algoritma Decision Tree Model Arsitektur C4.5," *J. Krisnadana*, vol. 3, no. 1, pp. 25–33, 2023, doi: 10.58982/krisnadana.v3i1.470.
- [8] V. Junita and F. A. Bachtiar, "Klasifikasi Aktivitas Manusia menggunakan Algoritme Decision Tree C4.5 dan Information Gain untuk Seleksi Fitur," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 3, no. 10, pp. 9426–9433, 2020.
- [9] T. M. Abdillah, "Rancang Bangun Sistem Pengklasifikasi Kecepatan Maksimum Kereta Api Pada Jalur Klakah-Pasirian Menggunakan Metode Single Layer Perceptron," in *Digital Repository Univ Jember*, 2017.
- [10] N. R. Dewi, A. Desiani, F. Salamah, and Y. Andriani, "Algoritma K-Nearest Neighbor (K-NN) dan Single Layer Perceptron (SLP) untuk Klasifikasi Penyakit Alzheimer," vol. 9, no. September, 2023.
- [11] A. Desiani, S. I. Maiyanti, B. Suprihatin, R. Kurniawan, and A. A. Nabila, "Implementasi Algoritma Adaptive Boosting (ADABOOST) Dan Single Layer Perceptron (SLP) Pada Klasifikasi Penyakit Hepatitis-c," *Sci. Comput. Sci. Informatics J.*, vol. 6, no. 2, pp. 41–54, 2023.
- [12] A. N. Iffah'da and A. Desiani, "Implementasi Algoritma K-Nearest Neighbor (K-NN) dan Single Layer Perceptron (SLP) Dalam Prediksi Penyakit Sirosis Biliari Primer," *J. Ilm. Inform.*, vol. 7, no. 1, pp. 65–74, 2022, doi: 10.35316/jimi.v7i1.65-74.
- [13] M. Ali *et al.*, "Peningkatan Kesadaran Masyarakat Terhadap Pentingnya Pencegahan Dini Stroke," *J. Pengabd. Masy. Fisioter. dan Kesehat. Indones.*, vol. 2, no. 01, pp. 65–71, 2023, doi: 10.59946/jpmfki.2023.199.
- [14] M. Ibnu Sina *et al.*, "Penyuluhan 'Kenali Gejala Stroke di Rumah Sakit Amin,'" *J. Abdimas Kedokt. dan Kesehat.*, vol. 3, no. 1, pp. 37–42, 2025.
- [15] F. M. Hana, "Klasifikasi Penderita Penyakit Diabetes Menggunakan Algoritma Decision Tree C4.5," *J. SISKOM-KB (Sistem Komput. dan Kecerdasan Buatan)*, vol. 4, no. 1, pp. 32–39, 2020, doi: 10.47970/siskom-kb.v4i1.173.
- [16] Ardi Ramdani, Christian Dwi Sofyan, Fauzi Ramdani, Muhamad Fauzi Arya Tama, and Muhammad Angga Rachmatsyah, "Algoritma Klasifikasi Data Mining Untuk Memprediksi Masyarakat Dalam Menerima Bantuan Sosial," *J. Ilm. Sist. Inf.*, vol. 1, no. 2, pp. 39–47, 2022, doi: 10.51903/juisi.v1i2.363.
- [17] Tiya Muthia, N. Nurrahma, and Yudi Eka Putra, "Perbandingan Akurasi Model Pembelajaran Mesin SVM, KNN, Decision Tree, dan Naive Bayes pada Klasifikasi Gangguan Kesehatan Mental," *Electr. J. Rekayasa dan Teknol. Elektro*, vol. 18, no. 3, pp. 363–368, 2024, doi: 10.23960/elc.v18n3.2758.
- [18] Muhamad, A. P. Windarto, and Suhada, "Penerapan Algoritma C4.5 pada Klasifikasi Potensi," vol. 3, pp. 1–8, 2019, doi: 10.30865/komik.v3i1.1688.
- [19] A. R. Putra *et al.*, "Penerapan K-Nearest Neighbors Pada Perceptron untuk Klasifikasi Dataset Kecil dengan Tiga Fitur," *OKTAL J. Ilmu Komput. dan Sci.*, vol. 3, no. 11, pp. 2866–2872, 2024, [Online]. Available: <https://journal.mediapublikasi.id/index.php/oktal>
- [20] Fedesoriano, "Stroke Prediction Dataset," kaggle.com. [Online]. Available: <https://www.kaggle.com/datasets/fedesoriano/stroke-prediction-dataset>
- [21] F. Akbar, H. W. Saputra, A. K. Maulaya, M. F. Hidayat, and Rahmaddeni,

“Implementation of Decision Tree Algorithm C4.5 and Support Vector Regression for Stroke Disease Prediction,” *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 2, no. 2, pp. 61–67, 2022, doi: 10.1088/1742-6596/1641/1/012025.