

Contents list available at [www.jurnal.unimed.ac.id](http://www.jurnal.unimed.ac.id)

**CESS**  
**(Journal of Computing Engineering, System and Science)**

journal homepage: <https://jurnal.unimed.ac.id/2012/index.php/cess>



**Data Mining Algoritma *Decision Tree Iterative Dechotomiser 3* (ID3) untuk  
Klasifikasi Penyakit Stroke**

***Data Mining Algorithm Decision Tree Iterative Dechotomiser 3 (ID3) for  
Classification of Stroke***

Zunaida Sitorus<sup>1</sup>, Adi Widarma<sup>2\*</sup>

<sup>1,2</sup> Program Studi Teknik Informatika, Universitas Asahan  
Jl. Jend. Ahmad Yani, Kisaran, Sumatera Utara 21216  
email: <sup>1</sup> [z\\_sitorus@yahoo.com](mailto:z_sitorus@yahoo.com), <sup>2</sup> [adiwidarma10@gmail.com](mailto:adiwidarma10@gmail.com)

**ABSTRAK**

Penyakit stroke atau *cerebrovascular* merupakan penyakit yang terjadi karena terputusnya suplai pasokan darah ke suatu bagian otak sehingga mengganggu sistem syaraf pusat. Penyakit ini sangat serius dan harus segera cepat ditangani karena dapat menyebabkan salah satu kematian sesuai data WHO (*World Health Organization*) akibat stroke terjadi 70% kematian dunia. Penanganan yang cepat dan tepat serta pengetahuan masyarakat akan penyakit stroke sangat dibutuhkan agar dapat segera diatasi. Perkembangan teknologi seperti *Machine Learning* sangat dibutuhkan karena pendekatan yang populer untuk mampu melakukan prediksi stroke dengan akurat. Algoritma *Machine Learning* yaitu Data Mining dengan metode *Decision Tree* akan diterapkan. Dalam penelitian ini, kerangka kerja dilakukan yang bertujuan untuk menganalisis kinerja model klasifikasi metode *Decision Tree* menggunakan ID3 dalam bidang prediksi penyakit stroke. Dataset *public* yang bersumber dari kaggle dengan jumlah record sebanyak 5110 dipilih dan diterapkan untuk membangun model klasifikasi dan menguji kinerjanya serta pengujian model akan dilakukan menggunakan aplikasi RapidMiner. Uji performance untuk evaluasi model data mining dengan *Confusion Matrix* digunakan sebagai indikator akurasi dalam kerangka untuk mengevaluasi kinerja klasifikasi. Perbandingan nilai evaluasi model data mining dengan membagi data menjadi data *training* dan data *testing* dan menghasilkan nilai *accuracy* dengan proporsi 90:10 sebesar 94,72%, 80:20 sebesar 95,21%, 70:30 sebesar 95,04% dan 60:40 sebesar 94,81%. Hasilnya menunjukkan bahwa proporsi data 80:20 memiliki nilai akurasi paling besar dibandingkan dengan proporsi data yang lainnya.

**Kata Kunci:** *Data Mining, Klasifikasi, Decision Tree ID3, Penyakit Stroke.*

## ABSTRACT

Stroke or cerebrovascular disease is a disease that occurs due to the interruption of the blood supply to a part of the brain that disrupts the central nervous system. This disease is very serious and must be treated immediately because it can cause one of the deaths according to WHO (World Health Organization) data due to stroke occurring 70% of world deaths. Quick and precise treatment as well as public knowledge of stroke is urgently needed so that it can be resolved immediately. Technological developments such as Machine Learning are urgently needed because of the popular approach to being able to predict strokes accurately. Machine Learning Algorithm, namely Data Mining with the Decision Tree method will be applied. In this study, a framework was carried out which aimed to analyze the performance of the Decision Tree method classification model using ID3 in the field of stroke prediction. A public dataset sourced from kaggle with a total of 5110 records is selected and applied to build a classification model and test its performance and model testing will be carried out using the RapidMiner application. Evaluation of the Confusion Matrix data mining model is used as an indicator of accuracy in a framework for evaluating classifier performance. Comparison of the evaluation value of the data mining model by dividing the data into training data and testing data. The results show that the accuracy value with the proportion of 90:10 is 94.72%, 80:20 is 95.21%, 70:30 is 95.04% and 60:40 is 94.81%.

**Keywords:** *Data Mining, Classification, ID3 Decision Tree, Stroke Disease.*

---

## 1. PENDAHULUAN

Stroke adalah adalah suatu kondisi yang terjadi ketika suplai darah ke suatu bagian otak tiba-tiba terganggu, akibat matinya sebagian sel otak akibat terhentinya aliran darah akibat sumbatan atau pecahnya pembuluh darah di otak. Di jaringan otak, kurangnya aliran darah memicu serangkaian reaksi biokimia yang dapat merusak atau membunuh sel saraf di otak. Kematian jaringan otak dapat menyebabkan hilangnya fungsi yang dikendalikan oleh jaringan ini. Terhentinya peredaran darah menghentikan suplai oksigen dan nutrisi ke otak, sehingga sebagian otak tidak dapat berfungsi dengan baik.

Salah satu penyakit yang paling umum dan penyebab kematian utama di Indonesia adalah stroke. Di Indonesia, stroke merupakan salah satu penyakit yang paling umum dan penyebab kematian utama, diikuti oleh diabetes dan hipertensi. Menurut data yang dihimpun Badan Penelitian dan Pengembangan Kementerian Kesehatan RI, jumlah korban stroke setiap tahunnya terus bertambah. Gejala stroke dilaporkan oleh 8,3 orang dari setiap 1000 penduduk Indonesia pada tahun 2007. Setelah itu, naik menjadi 12,1, atau lebih tepatnya diperkirakan 2.137.941 orang, pada tahun 2013. Di tahun-tahun mendatang, diperkirakan angka kejadian stroke saat ini korban kira-kira 25 sampai 30 per 1.000 orang akan meningkat.

Masyarakat yang kekurangan pengetahuan tidak menyadari potensi tanda-tanda peringatan yang sudah ada. Penderita stroke dengan penanganan yang cepat dan efektif lebih mungkin untuk pulih dari kondisinya dibandingkan mereka yang menerima perawatan terlambat, sehingga dapat menyebabkan kecacatan permanen, kerusakan otak lanjutan, atau bahkan kematian. Karena itu, penyakit stroke memerlukan penggunaan teknologi yang disebut machine learning yang dalam hal ini data mining digunakan sebagai teknik pengolahan data. Data Mining merupakan bidang yang menggabungkan ilmu komputer dan disiplin ilmu lainnya, adalah proses penggalian pengetahuan dari basis data menggunakan teknik seperti

kecerdasan buatan (*artificial intelligence*), pembelajaran mesin (*machine learning*), dan statistik untuk menemukan pola dan informasi yang menarik[1].

Klasifikasi adalah salah satu peran di mana data mining dapat digunakan. Klasifikasi adalah teknik langsung untuk menemukan kelas atau model data, yang selanjutnya digunakan sebagai strategi untuk prediksi masalah. Dalam penelitian ini algoritma *Decision Tree* (ID3) yang merupakan algoritma untuk membuat pohon keputusan digunakan sebagai metode klasifikasi penyakit stroke.

Penelitian tentang klasifikasi data mining untuk penyakit stroke sudah banyak dilakukan dengan berbagai metode seperti klasifikasi penyakit stroke dengan metode Naive Bayes[2][3], metode *Support Vector Machine* untuk klasifikasi penyakit stroke[4][5], Klasifikasi Penyakit Stroke Menggunakan Algoritma *Decision Tree* C.45[6], klasifikasi penyakit stroke menggunakan algoritma [7][8], klasifikasi Penyakit Stroke Menggunakan Metode K-Nearest Neighbor[9].

Penelitian sebelumnya yang pernah dilakukan yaitu menganalisis perbandingan algoritma C4.5 dan CART untuk mendapatkan hasil klasifikasi yang terbaik dan disimpulkan bahwa algoritma C4.5 merupakan metode klasifikasi terbaik untuk penyakit stroke ketika dibandingkan dengan algoritma CART[1]. Pada penelitian yang dilakukan oleh Randi Estian Pambudi (2022) yaitu klasifikasi Penyakit Stroke Menggunakan Algoritma *Decision Tree* C.45 dimana algoritma *Decision Tree* C4.5 menghasilkan tingkat akurasi yang sangat baik pada penyakit stroke. Penelitian yang dilakukan oleh [10] menggunakan perbandingan model klasifikasi *Random Forest*, *Naïve Bayes* dan *KNN* dimana dataset kurang banyak dan disarankan menggunakan jumlah dataset yang lebih banyak lagi serta menggunakan metode klasifikasi yang lain untuk melihat setiap akurasi. Dari beberapa penelitian sebelumnya dan hasil yang didapatkan maka, pada penelitian ini penulis melakukan klasifikasi penyakit stroke menggunakan metode *Decision Tree* ID3.

## 2. TINJAUAN TEORI

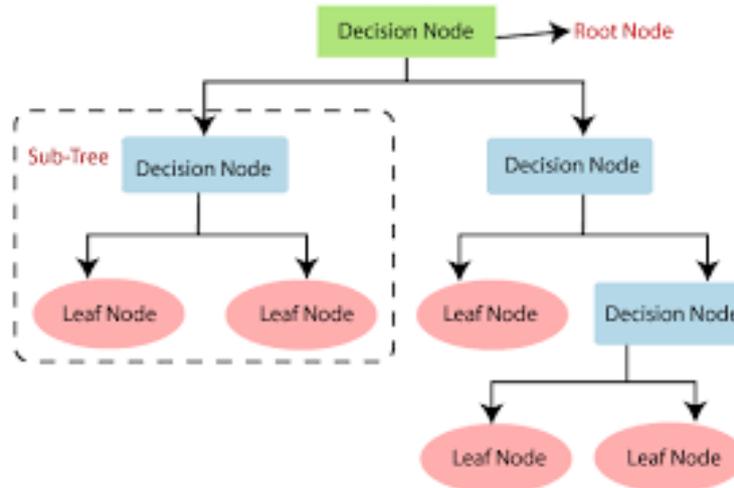
### 2.1. Data Mining

Data mining adalah proses penggalian informasi untuk mengidentifikasi pola, tren dan data yang berguna dalam proses bisnis guna dijadikan sebagai referensi dalam langkah pengambilan keputusan berbasis data dari kumpulan data yang besar. Data Mining atau sering dikenal penambangan data merupakan salah satu teknik yang berguna dalam membantu pengusaha, peneliti dan individu lainnya untuk mengekstrak informasi berharga dan kumpulan data yang sangat besar[11]. Beberapa peran data mining yaitu Estimasi, Forecasting, Klasifikasi, Clustering, Asosiasi.

### 2.2. Algoritma *Decision Tree* ID3

Metode yang populer digunakan untuk klasifikasi dengan metode data mining adalah *decision tree*. Pohon keputusan atau *decision tree* adalah bagan alir struktur yang menyerupai Pohon (*tree*), dimana setiap node internal menandakan pengujian pada atribut, setiap cabang mewakili hasil pengujian, dan node distribusi kelas atau kelas direpresentasikan oleh daun[6]. Pohon keputusan (*decision tree*) adalah metode dalam data mining dan pembelajaran mesin yang digunakan untuk membuat model prediktif berdasarkan serangkaian keputusan atau kondisi yang saling terkait. Pohon keputusan adalah representasi grafis dari algoritma yang menggambarkan serangkaian keputusan dan konsekuensinya. Pohon keputusan memiliki

proses dengan membentuk data (tabel) menjadi bentuk model pohon, kemudian mengubah pohon menjadi sebuah rule, dan merubah rule menjadi lebih sederhana.



**Gambar 1.** Decision Tree (Pohon Keputusan)

Salah satu metode decision tree adalah algoritma ID3 yang melakukan pencarian secara menyeluruh pada semua kemungkinan pohon keputusan. Algoritma ID3 berusaha membangun decision tree (pohon keputusan) secara top-down (dari atas ke bawah). Tahapan Algoritma Decision Tree (ID3) yaitu:

1. Menyiapkan dataset
2. Memilih atribut sebagai akar dan menghitung nilai Entropy dengan rumus:

$$Entropy(S) = \sum_{i=1}^n p_i * \log_2 p_i \quad (1)$$

Keterangan:

- S = himpunan (dataset) Kasus
- n = jumlah partisi S
- $p_i$  = proporsi dari  $S_i$  terhadap S

3. Menghitung nilai information gain dari setiap variabel

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{S} * Entropy(S_i) \quad (2)$$

Keterangan:

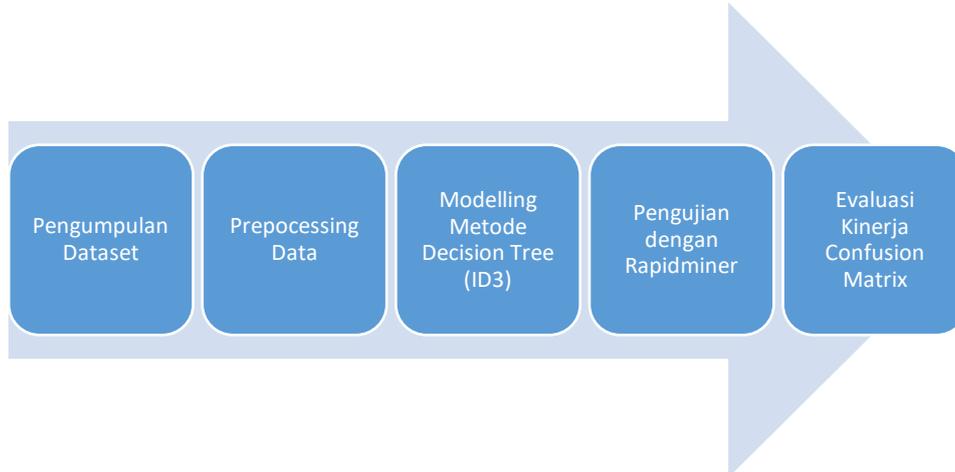
- S = Ruang data (sampel)
- A = Atribut
- $|S_i|$  = jumlah sampel data
- $|S|$  = jumlah seluruh sampel data

Entropy ( $S_i$ ) = untuk sampel-sampel yang memiliki nilai i

4. Membuat node cabang untuk tiap-tiap nilai
5. Ulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yg sama hingga membentuk pohon keputusan.

### 3. METODE PENELITIAN

Penelitian yang dilakukan ini terdiri dari beberapa tahapan penelitian, dimana tahapan penelitian merupakan langkah-langkah yang akan dilakukan yang digambarkan pada diagram alir pada gambar 2.



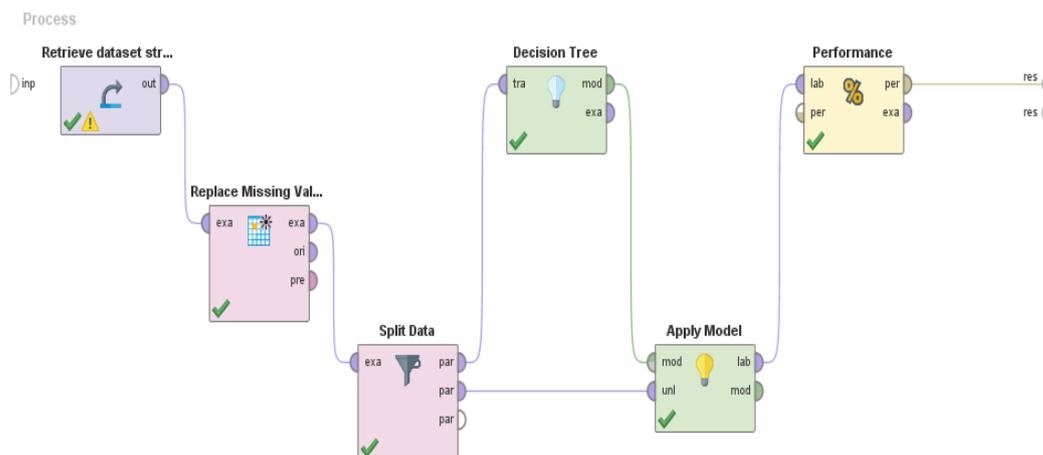
**Gambar 2.** Tahapan Penelitian

Adapun tahapan penelitian yang akan dilakukan dijelaskan sebagai berikut:

1. Dataset adalah himpunan data yang terdiri dari kumpulan data yang akan dilakukan proses yang nantinya akan digunakan untuk proses penggalian data.
2. *Preprocessing* data atau preparation data dilakukan untuk membersihkan data yang kosong (*missing value*) atau *noisy* data yang berupa isi data yang tidak konsisten (*inconsistency*)
3. Metode data mining dipilih sesuai karakter data dengan tujuan untuk menghasilkan pengetahuan.
4. Pengujian dilakukan untuk menguji hasil pengetahuan yang selanjutnya akan dihasilkan pemodelan yang tepat.
5. Evaluasi dilakukan untuk mengetahui tingkat keakuratan dari model yang didapatkan.

#### 4. HASIL DAN PEMBAHASAN

Pada penelitian ini akan dilakukan analisis metode decision tree menggunakan Algoritma ID3 untuk klasifikasi penyakit stroke. *Design* model penelitian menggunakan aplikasi RapidMiner untuk membantu peneliti dalam menganalisisnya. Adapun *design* model yang dilakukan menggunakan RapidMiner dapat dilihat pada gambar 3.



**Gambar 3.** Design model penelitian dengan aplikasi RapidMiner

#### 4.1. Dataset

Data merupakan hal yang utama dalam data mining. Pada penelitian ini menggunakan dataset public yang bersumber dari situs Kaggle. Dataset yang diambil terdiri dari 11 atribut dan 5110 record data[12]. Dataset dapat dilihat pada tabel 1 berikut:

**Tabel 1.** Dataset Penyakit Stroke

No	Gender	Age	Hypertension	Heart disease	Ever married	Work type	Residence type	Avg Glucose level	bmi	Smoking status	Stroke
1	Male	67	0	1	Yes	Private	Urban	228.69	36.6	formerly smoked	1
2	Female	61	0	0	Yes	Self-employed	Rural	202.21	N/A	never smoked	1
....	....	....	....	....	....	....	....	....	....	....	....
5109	Male	51	0	0	Yes	Private	Rural	166.29	25.6	formerly smoked	0
5110	Female	44	0	0	Yes	Govt_job	Urban	85.28	26.2	Unknown	0

#### 4.2. Preprocessing Data

Dalam melakukan proses mining data kita akan menjumpai beberapa masalah khususnya terkait kualitas data. Data yang memiliki kualitas yang buruk atau rendah akan mempengaruhi pada hasil mining. *Preprocessing* data dilakukan dengan menggunakan *software* RapidMiner untuk mengetahui nilai atribut yang *missing value*, *noise* atau *duplicate* data. Hasil menggunakan RapidMiner ditunjukkan pada gambar 4.

Name	Type	Missing	Statistics	Filter (11 / 11 attributes)
smoking_status	Binominal	2333	Least formerly smoked (885) Most never smoked (1892)	Values never smoked (1892), formerly smoked (885)
bmi	Integer	201	Min 10 Max 97	Average 28.449
gender	Binominal	1	Least Male (2115) Most Female (2994)	Values Female (2994), Male (2115)
stroke	Binominal	0	Least 1 (249) Most 0 (4861)	Values 0 (4861), 1 (249)
age	Real	0	Min 0.080 Max 82	Average 43.227
hypertension	Integer	0	Min 0 Max 1	Average 0.097
heart_disease	Integer	0	Min 0 Max 1	Average 0.054
ever_married	Binominal	0	Least No (1757) Most Yes (3353)	Values Yes (3353), No (1757)
work_type	Polynomial	0	Least Never_worked (22) Most Private (2925)	Values Private (2925), Self-employed (22)

**Gambar 4.** Proses *Preprocessing* dengan RapidMiner.

Dari hasil menggunakan RapidMiner terdapat tiga atribut yang ada *missing value* yaitu atribut *smoking\_status*, *bmi* dan *gender*, kemudian nilai atribut yang *missing value* kita *replace* dengan operator *Replace Missing Value* dengan pilih *average* dari nilai atribut *smoking\_status*,

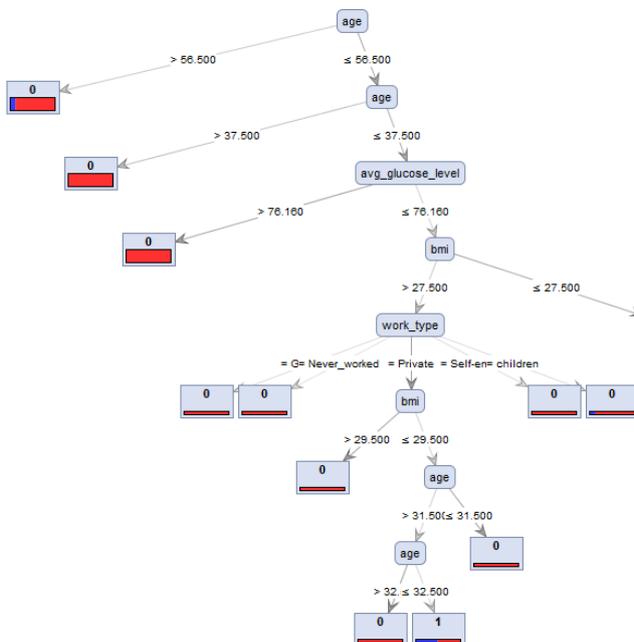
bmi dan *gender*. Setelah *Replace Missing Value* tidak ada lagi atribut yang *missing* seperti ditunjukkan pada gambar 5.

Name	Type	Missing	Statistics	Filter (11 / 11 attributes)
stroke	Binominal	0	Least 1 (249) Most 0 (4861)	Values 0 (4861), 1 (249)
gender	Binominal	0	Least Male (2115) Most Female (2995)	Values Female (2995), Male (2115)
bmi	Integer	0	Min 10 Max 97	Average 28.431
smoking_status	Binominal	0	Least formerly smoked (885) Most never smoked (4225)	Values never smoked (4225), formerly smoked (885)
age	Real	0	Min 0.080 Max 82	Average 43.227
hypertension	Integer	0	Min 0 Max 1	Average 0.097
heart_disease	Integer	0	Min 0 Max 1	Average 0.054
ever_married	Binominal	0	Least No (1757) Most Yes (3353)	Values Yes (3353), No (1757)
work_type	Polynomial	0	Least Never_worked (22) Most Private (2925)	Values Private (2925), Self-employed (22)

Gambar 5. Hasil setelah *Preprocessing Data*

### 4.3. Modelling

Pada penelitian ini digunakan metode Decision Tree algoritma ID3 (*Iterative Dichotomizer 3*) untuk melakukan klasifikasi penyakit stroke. Algoritma ID3 bekerja dengan memilih atribut pengukuran Information Gain. Klasifikasi penyakit stroke menggunakan algoritma ID3 dengan aplikasi RapidMiner menghasilkan pola pohon keputusan seperti pada gambar 6 dan model tree seperti pada gambar 7.



Gambar 6. Pola Pohon Keputusan

### Tree

```

age > 56.500: 0 {1=207, 0=1412}
age ≤ 56.500
| age > 37.500: 0 {1=39, 0=1425}
| age ≤ 37.500
| | avg_glucose_level > 76.160: 0 {1=0, 0=1496}
| | avg_glucose_level ≤ 76.160
| | | bmi > 27.500
| | | | work_type = Govt_job: 0 {1=0, 0=18}
| | | | work_type = Never_worked: 0 {1=0, 0=2}
| | | | work_type = Private
| | | | | bmi > 29.500: 0 {1=0, 0=96}
| | | | | bmi ≤ 29.500
| | | | | | age > 31.500
| | | | | | | age > 32.500: 0 {1=0, 0=7}
| | | | | | | age ≤ 32.500: 1 {1=1, 0=1}
| | | | | | | age ≤ 31.500: 0 {1=0, 0=25}
| | | | | | | work_type = Self-employed: 0 {1=0, 0=12}
| | | | | | | work_type = children: 0 {1=2, 0=12}
| | | | | | | | bmi ≤ 27.500: 0 {1=0, 0=355}
    
```

Gambar 7. Model Tree yang dihasilkan

#### 4.4. Pengujian

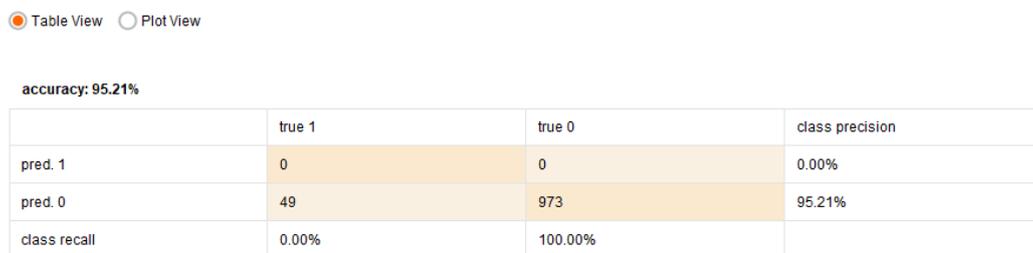
Pada tahap ini dilakukan pembagian data menjadi data *training* dan data *testing* dengan proporsi pembagian 90:10, 80:20, 70:30 dan 60:40. Proses pembagian data menggunakan aplikasi RapidMiner dengan memakai operator split data. Hasil pembagian jumlah data training dan data testing dapat dilihat pada tabel 2.

**Tabel 2.** Jumlah data training dan data testing

	Proporsi Data			
	90:10	80:20	70:30	60:40
Data Training	4599	4088	3577	3066
Data Testing	511	1022	1533	2044

#### 4.5. Evaluasi Model dan Kinerja Metode

Pada tahap ini dilakukan proses evaluasi model dan kinerja metode menggunakan *Confusion Matrix* untuk melihat akurasi. Proses *Confusion Matrix* dilakukan dengan aplikasi RapidMiner dengan memakai operator *Performance*. Salah satu hasil yang ditampilkan untuk melihat akurasi dengan aplikasi RapidMiner pada proporsi 80:20 seperti pada gambar 8 berikut.



**Gambar 8.** Hasil *Performance Confusion Matrix* untuk proporsi data 80:20

Untuk hasil akurasi dari setiap proporsi menggunakan aplikasi RapidMiner setiap pembagian data dapat dilihat pada tabel 3.

**Tabel 3.** Hasil *Confusion Matrix* untuk melihat akurasi dari setiap proporsi data

Proporsi Data	Accuracy
90:10	94.72%
80:20	95.21%
70:30	95.04%
60:40	94.81%

Dari tabel 3 diatas menunjukkan bahwa pembagian data dengan proporsi 80:20 dengan jumlah data *training* 4088 record dan jumlah data *testing* 1022 record menghasilkan nilai akurasi yang paling besar yaitu 95.21% dibandingkan dengan proporsi data lainnya.

## 5. KESIMPULAN

Penelitian yang dilakukan ini merupakan implementasi algoritma *Decision Tree* (ID3) untuk klasifikasi penyakit stroke. Dataset yang digunakan dari *public* dataset yang bersumber dari situs Kaggle dengan jumlah 5110 record. Design model yang dirancang menggunakan aplikasi RapidMiner untuk membantu proses analisisnya. Dataset dibagi menjadi data *training* dan data *testing*. Untuk menguji *performance* dilakukan evaluasi kinerja model yang dihasilkan menggunakan *Confusion Matrix* untuk melihat akurasi dari setiap proporsi dari data *training* dan data *testing*. Berdasarkan hasil analisis dan pembahasan yang dilakukan pada penelitian ini diperoleh kesimpulan bahwa nilai akurasi dari setiap proporsi data yaitu proporsi data 90:10 sebesar 94.72%, proporsi data 80:20 sebesar 95.21%, proporsi data 70:30 sebesar 95.04% dan proporsi data 60:40 sebesar 94.81%. Dari hasil akurasi tersebut bahwasanya proporsi data 80:20 memiliki nilai akurasi paling besar.

## REFERENSI

- [1] Suryani, D. Rahmadani, A. A. Muzafar, A. Hamid, R. Annisa, and Mustakim, "Comparative Analysis of C4. 5 and CART Algorithms for Classification of Stroke," in *SENTIMAS: Seminar Nasional Penelitian dan Pengabdian Masyarakat*, 2022. [Online]. Available: <https://journal.irpi.or.id/index.php/sentimas/article/view/279>
- [2] N. Y. Paramitha, A. Nuryaman, A. Faisol, E. Setiawan, and D. E. Nurvazly, "Klasifikasi Penyakit Stroke Menggunakan Metode Naïve Bayes," *J. Siger Mat.*, vol. 4, no. 1, 2023, [Online]. Available: <https://jurnal.fmipa.unila.ac.id/JSM/article/view/9236>
- [3] N. Y. Paramitha, "Penerapan Metode Naïve Bayes Untuk Klasifikasi Penyakit Stroke," *digilib.unila.ac.id*, 2023. [Online]. Available: <http://digilib.unila.ac.id/id/eprint/71038>
- [4] L. ZAHRA, "Optimasi Metode Support Vector Machine Menggunakan Algoritma Genetika Untuk Klasifikasi Penyakit Stroke," *repository.ump.ac.id*, 2023. [Online]. Available: <https://repository.ump.ac.id/15184/>
- [5] K. R. Sulaeman, C. Setianingsih, and R. E. Saputra, "Analisis Algoritma Support Vector Machine Dalam Klasifikasi Penyakit Stroke," in *eProceedings of Engineering*, 2022, vol. 9, no. 3. [Online]. Available: <https://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/article/view/17909>
- [6] R. E. Pambudi, Sriyanto, and Firmansyah, "Klasifikasi Penyakit Stroke Menggunakan Algoritma Decision Tree C. 45," *Teknika*, vol. 16, no. 2, 2022, [Online]. Available: <https://jurnal.polsri.ac.id/index.php/teknika/article/view/4914>
- [7] R. M. Nafisah, "Implementasi Metode Sampling Pada Klasifikasi Penyakit Stroke Menggunakan Algoritma Random Forest," *repository.um.ac.id*, 2022. [Online]. Available: <http://repository.um.ac.id/264944/>
- [8] N. A. Iskandar, I. Ernawati, and Y. Widiastiwi, "Klasifikasi Diagnosis Penyakit Stroke Dengan Menggunakan Metode Random Forest," in *Prosiding SEMANIKA*, 2022. [Online]. Available: <https://conference.upnvj.ac.id/index.php/senamika/article/view/2190>
- [9] R. HIDAYAT, *Klasifikasi Penyakit Stroke Menggunakan Metode K-Nearest Neighbor Studi Kasus: Puskesmas Karangbinangun*. eprints.umg.ac.id, 2021. [Online]. Available: <http://eprints.umg.ac.id/4746/>

- [10] F. A. H. Airi, T. Suprpti, and A. Bahtiar, "Komparasi Metode Klasifikasi Data Mining Untuk Prediksi Penyakit Stroke," *E-Link J. Tek. Elektro dan Inform.*, vol. 18, no. 1, 2023, [Online]. Available: <http://journal.umg.ac.id/index.php/e-link/article/view/5271>
- [11] A. Surahmat, N. Ahmad, H. Fitri, A. Widarma, and dkk, "Kecerdasan Buatan Dalam Data Mining," Pertama., A. Surahmat, Ed. Bandung: Widina Bhakti Persada, 2023.
- [12] Fedesoriano, "Stroke Prediction Dataset." <https://www.kaggle.com/datasets/fedoriano/stroke-prediction-dataset> (accessed Jun. 16, 2023).