



# Implementasi Algoritma CART dan Naïve Bayes Untuk Mendeteksi Penyakit Stroke

Anita Desiani<sup>1\*</sup>, Puspa Sari<sup>2</sup>, Indri Ramayanti<sup>3</sup>, Muhammad Arhami<sup>4</sup>

<sup>1,2</sup>Matematika, Universitas Sriwijaya,

Jl. Raya Palembang-Prabumulih Km. 32, Indralaya, Indonesia

<sup>3</sup>Departemen Parasitologi, Fakultas Kedokteran, Universitas Muhammadiyah Palembang

<sup>4</sup>Teknik Informatika, Politeknik Lhokseumawe Aceh

\*Email Penulis Koresponden: [anita\\_desiani@unsri.ac.id](mailto:anita_desiani@unsri.ac.id)

## **Abstrak (TNR 10 Bold Italic):**

Penyakit stroke merupakan penyakit pembuluh darah yang disebabkan oleh kurangnya sirkulasi oksigen dan darah ke otak sehingga menyebabkan kerusakan pada jaringan otak. Stroke dapat menyebarkan perubahan pada fungsi dan fisiologi anatomi yang letaknya jauh dari kerusakan. Penyakit stroke merupakan penyakit penyebab utama kematian setelah penyakit jantung sehingga diperlukan suatu system untuk mendeteksi penyakit stroke sebagai bentuk pencegahan dini agar tidak terserang penyakit stroke. Sistem deteksi menggunakan kecerdasan buatan dengan teknik klasifikasi. Klasifikasi telah digunakan oleh para peneliti untuk mendeteksi penyakit dengan hasil yang memuaskan. Pada penelitian ini menggunakan teknik klasifikasi menggunakan algoritma CART (*Classification and Regression Tree*) dan algoritma Naïve Bayes. Penelitian ini membagi data menggunakan persentase split sebesar 90% data latih dan sisanya berupa data uji dengan dataset dari Kaggle. Berdasarkan penggunaan kedua algoritma dalam mendeteksi penyakit stroke, algoritma CART menghasilkan akurasi sebesar 95.57% sedangkan naïve bayes memiliki akurasi sebesar 85%. Pada sisi presisi model rata-rata yang dihasilkan oleh algoritma CART yaitu 88% sedangkan naïve bayes memiliki presisi 66%. Recall rata-rata yang dihasilkan oleh algoritma CART yaitu 61% sedangkan naïve bayes memiliki recall sebesar 58%. Dari perbandingan antara akurasi, presisi, dan recall yang dihasilkan algoritma CART dan Naïve bayes dapat disimpulkan jika algoritma CART sangat baik dalam mendeteksi penyakit stroke secara dini.

*This is an open access article under the [CC BY-NC](https://creativecommons.org/licenses/by-nc/4.0/) license*



## **Kata Kunci:**

Stroke;  
CART;  
Naïve Bayes;  
Klasifikasi;

## **Riwayat Artikel:**

Diserahkan 26 April, 2023  
Direvisi 17 Nopember, 2025  
Diterima 19 Februari, 2026

## **DOI:**

10.22441/incomtech.v16i1.20209

## 1. PENDAHULUAN

Stroke (gangguan serebrovascular) merupakan penyakit penyebab utama kematian setelah penyakit jantung [1]. Stroke merupakan gangguan otak yang disebabkan karena berkurangnya aliran darah dan oksigen yang menyebabkan kematian pada jaringan otak [2]. Stroke dapat menyebabkan perubahan pada fungsi dan fisiologi daerah anatomi lainnya walaupun berada jauh dari letak kerusakan [3]. Berdasarkan data WHO (1996) bahwa penyakit pada peredaran darah menyebabkan kematian sekitar 30% setiap tahun atau 15 juta kematian dengan 4,5 juta kematian di antaranya disebabkan oleh stroke, selain itu stroke juga menjadi penyebab dua per tiga kematian di negara non industri [4].

Berdasarkan data Riskesdas (2007) bahwa prevalensi stroke yaitu delapan perseribu penduduk di Indonesia dengan prevalensi 10.9 per mil kemudian dengan prevalensi pengidap penyakit stroke di Sulawesi Selatan yaitu 10.6 % per mil menempati provinsi tertinggi kasus stroke [5] kemudian prevalensi di kota Bogor sekitar 1,1 % atau 11 penduduk per seribu [6]. Kasus penyakit stroke di Indonesia diperkirakan 500.000 setiap tahun dengan sekitar 2.5% atau 12.500 orang meninggal dunia dan lainnya menderita cacat ringan dan juga cacat berat [6]. Dengan memperhatikan tingginya kasus penyakit stroke di Indonesia maka diperlukan suatu sistem yang digunakan untuk mendeteksi pola penyakit dengan sistem yang digunakan adalah data mining.

Menurut [2] data mining adalah suatu kegiatan pengumpulan dan penggunaan data yang digunakan untuk menentukan pola hubungan sebagai pelajaran di masa depan. Data mining memiliki beberapa teknik antara lain klasifikasi. Teknik klasifikasi adalah suatu pekerjaan yang menilai objek dan memasukkannya ke dalam kelas dengan membangun model yang berdasarkan data latih untuk mengklasifikasikan data yang baru [7]. Algoritma klasifikasi telah diterapkan untuk mendiagnosa penyakit dan hasilnya sangat menjanjikan untuk mempercepat serta menyerderhanakan proses diagnosis [8]. Salah satu teknik klasifikasi adalah dengan menggunakan algoritma CART dan Naïve bayes. Penelitian klasifikasi penyakit stroke pernah dilakukan oleh Ohoud dkk [9] dengan menggunakan metode C4.5, Jrip, dan MLP mendapat akurasi 95% namun tidak menghitung evaluasi kinerja lainnya. Penelitian lainnya dilakukan Indarto dkk [10] dengan menggunakan metode C4.5, Jrip, dan MLP dengan masing-masing evaluasi kinerja berupa akurasi, presisi, *recall*, dan F1-Score dibawah 89% dalam mendeteksi stroke hemorragic. Penelitian mendeteksi penyakit stroke dilakukan oleh Gangavarapu dkk [11] dengan menggunakan Naïve Bayes menghasilkan akurasi sebesar 82% tanpa menghitung kinerja lainnya. Penelitian oleh Xuemeng dkk [12] menggunakan *decision tree* dan *random forest* menghitung *recall* masing-masing sebesar 99,94% dan 97.33% namun tidak menghitung evaluasi kinerja lainnya seperti akurasi dan sebagainya.

Algoritma CART adalah suatu algoritma yang menggunakan sampel pembelajaran dari data-data dengan kelas yang telah ditentukan untuk membuat pohon keputusan yang dapat digunakan dalam eksplorasi dan prediksi data [13]. Algoritma CART pernah digunakan dalam mendiagnosa penyakit diabetes menghasilkan akurasi sebesar 78% dan hepatitis dengan akurasi sebesar 83% [14]. CART dapat menangani variabel numerik dan kategorik serta mengidentifikasi variabel yang paling signifikan dengan mudah menangani outlier. Namun, CART memiliki kelemahan berupa pohon keputusan yang tidak stabil dengan pembagi

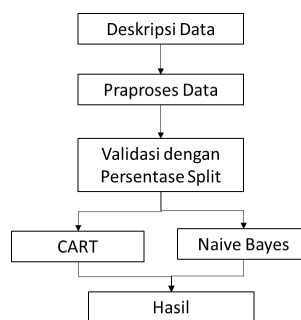
satu variabel yang tidak signifikan [15]. Oleh karena itu, diperlukan algoritma yang dapat menutupi kekurangan dari CART yaitu algoritma naïve bayes.

Naïve Bayes merupakan algoritma klasifikasi sederhana yang menghitung dengan probabilitas terhadap frekuensi dan kombinasi nilai pada kumpulan data tertentu dengan menggunakan Teorema Bayes [16]. Naïve bayes membutuhkan sedikit pelatihan untuk memperkirakan parameter yang diperlukan dalam klasifikasi dengan menggunakan model probabilitas serta dapat menggabungkan pengetahuan dengan data yang diamati [17]. Naïve bayes pernah digunakan untuk mengklasifikasikan penyakit liver dengan akurasi 61% dan presisi 55.58% [18].

Berdasarkan kelebihan dan kekurangan dari masing-masing algoritma CART dan Naïve Bayes dalam mengklasifikasi. Pada penelitian ini akan mendeteksi penyakit stroke menggunakan algoritma CART dan Naïve Bayes agar dapat membantu dalam mempercepat deteksi penyakit stroke dengan hasil yang akurat.

## 2. METODE

Dalam penelitian ini akan digunakan beberapa metode berupa deskripsi data, praproses data, validasi, implementasi algoritma, dan evaluasi hasil. Alur metode tertera di gambar 1.



Gambar 1. Alur Metode Penelitian.

### 2.1 Deskripsi Data

Data yang digunakan diperoleh dari situs Kaggle mengenai prediksi penyakit stroke (<https://www.kaggle.com/datasets/fedesoriano/stroke-prediction-dataset>) dengan format csv. Data terkait penyakit stroke terdiri 5.092 data yang terdiri dari 12 atribut. Rincian data disajikan pada Tabel 1.

Tabel 1. Deskripsi data

Atribut	Type	Data kosong
Id	Numerik	0
Gender	Kategorik	0
Usia	Numerik	0
Hipertensi	Kategorik	0
Sakit jantung	Kategorik	0
Status menikah	Kategorik	0
Pekerjaan	Kategorik	0
Domisili	Kategorik	0
Rata-rata glukosa	Numerik	0
Bmi	Numerik	184
Rokok	Kategorik	1536
Stroke	Kategorik	0

## 2.2 Praproses Data

Pada praproses data, data awal memiliki 9 atribut dengan 3 atribut ditiadakan yaitu id, status menikah, dan pekerjaan sehingga digunakan 9 atribut yaitu gender, usia, hipertensi, sakit jantung, tempat tinggal, rata-rata glukosa, bmi, rokok, dan stroke. Pada praproses data karena dilakukan proses cleaning yang merupakan pemeriksaan data yang tidak konsisten dan memperbaiki kesalahan pada data [19]. Terdapat data hilang pada rokok sehingga data tersebut dihapus menghasilkan sisa data sebanyak 3.556 data dengan pengidap stroke sebanyak 197 data dan bukan pengidap penyakit stroke sebanyak 3.359 data. Selain itu dilakukan transformasi data menjadi data label pada atribut usia, rata-rata glukosa, dan BMI.

Atribut usia yang dilabelkan berdasarkan [20] yang dimana penggolongan usia dapat digolongkan menjadi muda jika usia kurang dari 40 tahun, baya jika usia di antara 41 sampai 60 tahun, dan tua jika usia lebih dari 60 tahun. Pada atribut rata-rata glukosa darah data digolongkan menurut ADA yaitu normal apabila kurang dari 114mg/dL, prediabetes apabila di antara 114mg/dL sampai 140 mg/dL, dan diabetes apabila lebih besar dari 140 mg/dL. Pada atribut BMI data digolongkan berdasarkan [21] yaitu kurang apabila BMI lebih kecil dari 18,5, normal apabila di antara 18,5-22,9 dan obesitas apabila lebih dari 23. Hasil praproses data pada setiap atribut dapat dilihat pada Tabel 2.

Tabel 2. Hasil Praproses Data

Nama Atribut	Partisi Nilai
Usia	0 : Muda, 1: baya, 2 : tua
Gender	0: laki-laki, 1: Perempuan
Hipertensi	0: Tidak, 1: ya
Sakit jantung	0 : Tidak, 1: ya
Tempat tinggal	0 : kota, 1: ya
Rata-rata glukosa	0: Normal ; 1: Pre-diabetes; 2: Diabetes
Bmi	0 : Normal; 1: Kurang.; 2: Overweight
Rokok	0 : Tidak pernah; 1: jarang; 2: Perokok
Stroke	0 : Tidak; 1: Ya

Sistem validasi dilakukan dengan persentase split. *Split* validasi adalah teknik validasi yang membagi data secara acak menjadi dua bagian yaitu data latih dan data uji dengan pengujiannya berdasarkan persentase yang ditentukan. Pada *system* klasifikasi penyakit stroke menggunakan 90% data latih dan 10% data uji.

## 2.3 Algoritma CART

Algoritma CART (*Classification and Regression Tree*) tergolong dalam model tidak parametrik yang mengandalkan partisi biner rekursif dari data dengan simpul pohon akar dan node untuk mendirikan pohon regresi [22]. Pada algoritma CART apabila variabel kontinu maka menghasilkan pohon regresi sedangkan apabila variabel kategorik maka menghasilkan pohon klasifikasi [23].

Menurut [24] langkah-langkah algoritma CART :

1. Menyiapkan dataset
2. Menghitung nilai index gini dari semua atribut. Nilai index gini dapat dihitung dengan persamaan 1.

$$IndexGini = 1 - \sum_{i=1}^k p_i^2 \quad (1)$$

3. Menghitung nilai gini gain. Gini Gain dapat dihitung dengan persamaan 2.  

$$Gini\ Gain = Gini(A, S) - \sum_{i=1}^n \frac{|S_i|}{|S|} \times Gini(S_i) \quad (2)$$
4. Membuat akar dan cabang dari nilai Gini Gain tertinggi
5. Ulangi Langkah 2 sampai 4 hingga semua variabel terpartisi.

## 2.4 Algoritma Naïve Bayes

Klasifikasi Naïve bayes merupakan klasifikasi yang menggunakan aturan naïve bayes. Langkah-langkah proses implementasi naïve bayes adalah sebagai berikut [25]:

1. Siapkan dataset latih
2. Hitung jumlah dan probabilitas setiap atribut. Dalam menghitung probabilitas dapat digunakan rumus [26] seperti persamaan 3:

$$P(R|S) = \frac{P(R)P(S|R)}{P(S)} \quad (3)$$

Keterangan :

R : Data yang belum diketahui

S : Hipotesis pada data R

P(R|S) : Peluang R dengan syarat S

P(R) : Peluang dari nilai R

P(S|R) : Peluang S dengan syarat R

P(S) : Peluang S

3. Hitung probabilitas jika data kategorik atau rata-rata jika data numerik.
4. Hitung data uji, jika data bertipe numerik dapat dicari dengan distribut gaussian setiap atribut
5. Probabilitas akhir dapat ditentukan dengan memasukkan dan menghitung semua data untuk klasifikasi pada kelas yang sama kemudian hasilnya dikalikan setiap probabilitas kelas.

## 2.5 Evaluasi

Hasil validasi disajikan dalam *confusion matrix*. *Confusion matrix* adalah alat yang digunakan dalam mengevaluasi model klasifikasi dengan mengelompokkan jumlah prediksi objek benar atau salah. Matriks terdiri dari prediksi yang akan dibandingkan dengan kelas asli berisi informasi nilai aktual dan prediksi dari klasifikasi [27].

Tabel 3. *Confusion* Matriks

Kelas	Prediksi Positif	Prediksi Negatif
Aktual Positif	TP	FN
Aktual Negatif	FP	TN

TP (*True Positive*) dan TN (*True Negative*) merupakan kondisi dimana hasil prediksi sama dengan keadaan sebenarnya. FP (*False Positif*) dan FN (*False Negative*) merupakan kondisi dimana hasil prediksi tidak sama dengan keadaan sebenarnya [28]. Dari confusion matriks dapat ditentukan nilai akurasi, presisi, dan recall sebagai evaluasi model.

Akurasi merupakan perbandingan dari jumlah nilai prediksi *true* dari semua data. Perhitungan akurasi dapat menggunakan Persamaan (4) [29].

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (4)$$

Presisi [29] merupakan ukuran ketepatan dari hasil model yang dinyatakan pada persamaan 5 yaitu :

$$presisi = \frac{TP}{TP+FP} \quad (5)$$

Recall [29] merupakan ukuran kelengkapan dari sebuah model yang dapat dinyatakan dengan menggunakan persamaan 6 :

$$recall = \frac{TP}{TP+FN} \quad (6)$$

### 3. HASIL DAN PEMBAHASAN

#### 3.1 Implementasi CART

Implementasi algoritma CART pada penelitian ini diawali dengan perhitungan indeks gini dan gini gain untuk menilai atribut yang memiliki kemampuan pemisahan kelas paling optimal. Hasil perhitungan indeks gini setiap elemen menghasilkan nilai gini gain sebagai dasar pemilihan atribut pemecah. Nilai indeks gini dan gini gain dari setiap atribut ditampilkan pada Tabel 4.

Tabel 4. Nilai *indeks gini* dan *gini gain* pada setiap atribut

Atribut	Indeks Gini	Gini Gain
Gender	0,17—0,19	0,00004
Usia	0,02—0,35	0,01611
Hipertensi	0,14—0,36	0,00534
Sakit Jantung	0,16—0,39	0,00448
Tempat Tinggal	0,17—0,18	0,00001
Rata-rata Glukosa	0,14—0,30	0,00389
Rokok	0,15—0,23	0,00089
BMI	0,13—0,28	0,00068

Berdasarkan Tabel 4, atribut usia memperoleh nilai gini gain sebesar 0,01611, yang merupakan nilai tertinggi dibandingkan atribut lain sehingga ditetapkan sebagai node akar dalam pembentukan pohon keputusan CART. Atribut usia terdiri atas tiga kategori, masing-masing dengan nilai indeks gini muda = 0,0182, baya = 0,1283, dan tua = 0,3510. Nilai indeks gini kategori tua merupakan yang tertinggi sehingga kategori tersebut diprioritaskan untuk proses pemecahan lanjutan, diikuti oleh kategori baya, sedangkan kategori muda memiliki nilai indeks gini paling rendah karena proporsi terbesar kasus berada pada kelas tidak stroke, sehingga kategori tersebut diinterpretasikan sebagai kondisi dengan kecenderungan tidak mengalami stroke. Proses pemisahan pada node berikutnya dilanjutkan secara rekursif berdasarkan perhitungan nilai indeks gini dan gini gain hingga tidak ditemukan pemisahan yang lebih informatif dan node dinyatakan mencapai kondisi akhir (terminal node). Tahapan pemisahan ini menghasilkan struktur pohon keputusan yang memuat hubungan antar atribut beserta arah keputusan yang terbentuk. Adapun bentuk aturan klasifikasi yang diperoleh dari hasil pembentukan pohon CART adalah sebagai berikut.

1. *if* usia=muda *then* stroke=tidak.
2. *if* usia=baya *and* rata-rata glukosa=prediabetes *then* stroke=ya.
3. *if* usia=baya *and* rata-rata glukosa=normal *then* stroke=tidak

4. *if usia=baya and rata-rata glukosa=diabetes and hipertensi=tidak then stroke=tidak*
5. *if usia=baya and rata-rata glukosa=diabetes and hipertensi=ya and bmi=normal then stroke =tidak*
6. *if usia=baya and rata-rata glukosa=diabetes and hipertensi=ya and bmi=kurang then stroke =tidak*
7. *if usia=baya and rata-rata glukosa=diabetes and hipertensi=ya and bmi=overweight and rokok=jarang then stroke=tidak*
8. *if usia=baya and rata-rata glukosa=diabetes and hipertensi=ya and bmi=overweight and rokok=tidak pernah then stroke=tidak*
9. *if usia=baya and rata-rata glukosa=diabetes and hipertensi=ya and bmi=overweight and rokok=perokok and sakit\_jantung=tidak then stroke=tidak*
10. *if usia=baya and rata-rata glukosa=diabetes and hipertensi=ya and mi=overweight and rokok=perokok and sakit\_jantung=ya and tempat\_tinggal=desa then stroke=tidak*
11. *if usia=baya and rata-rata glukosa=diabetes and hipertensi=ya and mi=overweight and rokok=perokok and sakit\_jantung=ya and tempat\_tinggal=kota and gender=pria then stroke=tidak*
12. *if usia=baya and rata-rata glukosa=diabetes and hipertensi=ya and mi=overweight and rokok=perokok and sakit\_jantung=ya and tempat\_tinggal=kota and gender=wanita then stroke=ya*
13. *if usia=tua and hipertensi=tidak then stroke=tidak*
14. *if usia=tua and hipertensi=ya and bmi=normal then stroke=tidak*
15. *if usia=tua and hipertensi=ya and bmi=kurang then stroke=tidak*
16. *if usia=tua and hipertensi=ya and bmi=overweight and sakit\_jantung=tidak then stroke=tidak*
17. *if usia=tua then hipertensi=ya and bmi=overweight and sakit\_jantung=ya and rata-rata glukosa=prediabetes then stroke=tidak*
18. *if usia= tua then hipertensi=ya and bmi=overweight and sakit\_jantung=ya and rata-rata glukosa=normal then stroke=ya*
19. *if usia= tua then hipertensi=ya and bmi=overweight and sakit\_jantung=ya and rata-rata glukosa=diabetes and rokok=tidak then stroke=tidak*
20. *if usia= tua then hipertensi=ya and bmi=overweight and sakit\_jantung=ya and rata-rata glukosa=diabetes and rokok=perokok then stroke=ya*
21. *if usia= tua then hipertensi=ya and bmi=overweight and sakit\_jantung=ya and rata-rata glukosa=diabetes and rokok=jarang and gender=pria then stroke=tidak*
22. *if usia= tua then hipertensi=ya and bmi=overweight and sakit\_jantung=ya and rata-rata glukosa=diabetes and rokok=jarang and gender=wanita and tempat tinggal=desa then stroke=tidak*
23. *if usia= tua then hipertensi=ya and bmi=overweight and sakit\_jantung=ya and rata-rata glukosa=diabetes and rokok=jarang and gender=wanita and tempat tinggal=kota then stroke=ya*

Persentase *split* yang digunakan adalah 90% data latih dan 10% data uji. Hasil klasifikasi menggunakan algoritma CART dinyatakan dengan menggunakan matriks confusion yang dapat dilihat pada Tabel 5.

Tabel 5. Confusion Matriks CART

Kelas	Aktual		
	0	1	
Prediksi	0	337	1
	1	14	4

Berdasarkan Tabel 5 terdapat 337 kasus yang dikategorikan sebagai *true positif*, 1 *false negative*, 14 *false positif*, dan 4 *true negative* dengan kelas 0 bukan pengidap stroke dan kelas 1 adalah pengidap stroke. Dari hasil *confusion* matriks tersebut terdapat 337 data yang diprediksi secara benar sebagai kelompok 0 yaitu bukan penderita stroke, 4 data yang diprediksi secara benar sebagai kelompok 1 yaitu penderita stroke, 1 data penderita stroke dikenali bukan penderita stroke dan 14 data bukan penderita stroke dikenali sebagai penderita stroke.

Dari *confusion* matriks tersebut dapat dihitung nilai akurasi dari data yang berhasil diprediksi secara benar yaitu sebesar 95,57%. Hasil pengukuran recall untuk masing-masing kelompok adalah 100% untuk kelompok 0 (tidak mengidap penyakit stroke) dan 22% untuk pengidap penyakit stroke. Kemudian hasil pengukuran presisi untuk masing-masing kelompok adalah 96% untuk kelompok 0 dan 80% untuk kelompok 1.

### 3.2 Implementasi Naïve Bayes

Pada algoritma naïve bayes digunakan persentase split sebesar 90% data training dan 10% data testing. Hasil *Confusion* matriks pada klasifikasi menggunakan naïve bayes dapat dilihat pada Tabel 6.

Tabel 6. Confusion Matriks Naive bayes

Kelas	Aktual		
	0	1	
Prediksi	0	292	41
	1	13	10

Berdasarkan Tabel 6, hasil *confusion* matriksnya terdapat 292 *true positif*, 41 *false negative*, 13 *false positif*, dan 10 *true negative* dengan kelas 0 merupakan bukan pengidap stroke dan kelas 1 pengidap stroke. Berdasarkan hasil *confusion* matriks tersebut terdapat 292 data yang diprediksi secara benar sebagai kelompok 0 yaitu tidak mengidap penyakit stroke, 10 data yang diprediksi secara benar sebagai kelompok 1 yaitu pengidap penyakit stroke, 41 data penyakit stroke yang dikenali tidak mengidap penyakit stroke dan 13 data yang tidak mengidap penyakit stroke dikenali menderita penyakit stroke.

Kemudian diperoleh hasil akurasi sebesar 85% jauh lebih kecil daripada algoritma CART. Hasil recall untuk masing-masing kelas yaitu 88% untuk kelas 0 (Bukan penderita stroke) dan 43% untuk kelas 1 (penderita stroke). Hasil presisi untuk masing masing kelas yaitu 96% untuk kelas 0 (bukan penderita stroke) dan 20% untuk kelas 1 (penderita stroke).

### 3.3 Perbandingan CART dan Naïve Bayes

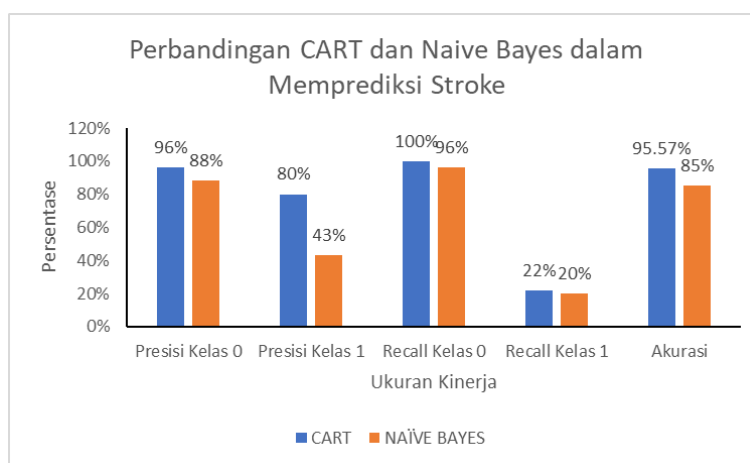
Hasil prediksi kedua algoritma CART dan Naïve bayes menghasilkan hasil prediksi yang baik dari segi akurasi, presisi, dan juga recall sehingga kedua algoritma ini memiliki kinerja yang baik dalam mendeteksi pengidap penyakit

stroke. Perbandingan hasil pemodelan algoritma CART dan naïve bayes ditunjukkan pada Tabel 7.

Tabel 7. Perbandingan hasil algoritma CART dan Naïve Bayes.

Metode	Kelas	Presisi	Recall	Akurasi
CART	0	96%	100%	95,57%
	1	80%	22%	
Naïve bayes	0	88%	96%	85%
	1	43%	20%	

Tabel 7 menyajikan perbandingan hasil kinerja algoritma CART dan Naïve Bayes berdasarkan metrik presisi, recall, dan akurasi. Secara visual, hasil tersebut juga divisualisasikan dalam bentuk grafik yang dapat dilihat pada Gambar 2 untuk memberikan gambaran perbandingan yang lebih jelas.



Gambar 2. Grafik perbandingan CART dan Naive bayes

Gambar 2 memperlihatkan bahwa kinerja algoritma CART menunjukkan performa yang lebih baik dibandingkan Naïve Bayes pada seluruh metrik evaluasi yang digunakan. Pada metrik presisi, CART menghasilkan nilai 96% pada kelas 0 dan 80% pada kelas 1, sedangkan Naïve Bayes masing-masing menghasilkan 88% dan 43%. Pada metrik recall, CART mencapai 100% pada kelas 0 dan 22% pada kelas 1, sementara Naïve Bayes memperoleh 96% untuk kelas 0 dan 20% untuk kelas 1. Selain itu, akurasi keseluruhan CART sebesar 95,57%, lebih tinggi 10,57% dibandingkan Naïve Bayes yang memperoleh akurasi 85%. Berdasarkan hasil tersebut, dapat disimpulkan bahwa algoritma CART memiliki kemampuan yang lebih baik dalam memprediksi kasus stroke pada dataset yang digunakan dibandingkan dengan Naïve Bayes, baik dari sisi presisi, recall, maupun akurasi.

### 3.4 Perbandingan dengan Metode Lain

Penelitian tentang mendeteksi penyakit stroke telah banyak dilakukan dengan menggunakan algoritma selain algoritma CART dan naïve bayes. Perlu dilakukan perbandingan nilai akurasi untuk mendapatkan metode yang sangat baik dalam mendeteksi penyakit stroke. Perbandingan akurasi dari algoritma CART dan naïve bayes dengan algoritma *decision tree* J48 dan C4.5 yang pernah dilakukan oleh [5] [30] dapat dilihat pada Tabel 8.

Tabel 8. Perbandingan dengan Metode lain untuk Mendeteksi stroke.

Algoritma	Presisi	Recall	Akurasi
Naïve Bayes	65.50%	58.00%	85.00%
CART	88.00%	61.00%	97.50%
C4.5[10]	88.60%	88.90%	88.90%
Jrip [10]	87.80%	88.10%	88.10%
MLP [10]	88.20%	88.40%	88.40%
Decision tree [12]	-	99.94%	-
Random Forest [12]	-	97.33%	-

Berdasarkan Tabel 8, Implementasi algoritma dengan klasifikasi CART dan Naïve bayes yang dilakukan pada penelitian ini dibandingkan dengan 5 algoritma lainnya yaitu C4.5, Jrip, MLP, Decision Tree, dan Random Forest. Algoritma C4.5 memiliki ukuran kinerja presisi terbesar dibandingkan algoritma lainnya, Algoritma Decision Tree menghasilkan ukuran kinerja Recall tertinggi dibandingkan algoritma lainnya, dan algoritma CART menghasilkan ukuran akurasi tertinggi dibandingkan algoritma lainnya. Dari hasil akurasi tersebut, CART cukup baik dalam mengklasifikasikan penyakit stroke.

#### 4. KESIMPULAN

Berdasarkan dataset penyakit stroke yang telah diperoleh di Kaggle dilakukan system klasifikasi dengan menggunakan algoritma CART dan algoritma Naïve bayes. Hasil implementasi algoritma CART dan algoritma naïve bayes menghasilkan nilai akurasi, presisi, dan recall yang berbeda. Hasil presisi, recall, dan akurasi dari algoritma CART lebih tinggi daripada hasil naïve bayes baik dari segi rata-rata maupun setiap kelasnya yang mencakup pengidap penyakit stroke ataupun tidak. Tidak hanya itu, perbandingan juga dilakukan terhadap akurasi dari algoritma lain yang telah digunakan dalam mendeteksi penyakit stroke dengan hasil akurasi algoritma CART lebih tinggi daripada algoritma lain. Dapat disimpulkan bahwa algoritma CART sangat baik dalam mendeteksi penyakit stroke.

#### REFERENSI

- [1] A. Towfighi and J. L. Saver, "Stroke declines from third to fourth leading cause of death in the United States: Historical perspective and challenges ahead," *Stroke*, vol. 42, no. 8, pp. 2351–2355, 2011, doi: 10.1161/STROKEAHA.111.621904.
- [2] I. Lishania, R. Gojantoro, and Y. N. Nasution, "Perbandingan Klasifikasi Metode Naive Bayes dan Metode Decision Tree Algoritma (J48) pada Pasien Penderita Penyakit Stroke di RSUD Abdul Wahab Sjahranie Samarinda," *J. Eksponensial*, vol. 10, no. 2, pp. 135–142, 2019, [Online]. Available: <http://jurnal.fmipa.unmul.ac.id/index.php/exponensial/article/view/571>
- [3] G. C. Araujo *et al.*, "Profiles of Executive Function Across Children with Distinct Brain Disorders: Traumatic Brain Injury, Stroke, and Brain Tumor," *J. Int. Neuropsychol. Soc.*, vol. 23, no. 7, pp. 529–538, 2017, doi: 10.1017/S1355617717000364.
- [4] C. D. A. Wolfe, "The impact of stroke," *British Medical Bulletin*, vol. 56, no. 2, pp. 275–286, 2000. doi: 10.1258/0007142001903120.
- [5] A. Byna and M. Basit, "Penerapan Metode Adaboost Untuk Mengoptimasi Prediksi Penyakit Stroke Dengan Algoritma Naïve Bayes," *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 9, no. 3, pp. 407–411, 2020, doi: 10.32736/sisfokom.v9i3.1023.

- [6] W. Riyadina and E. Rahajeng, "Determinan Penyakit Stroke," *Kesmas Natl. Public Heal. J.*, vol. 7, no. 7, p. 324, 2013, doi: 10.21109/kesmas.v7i7.31.
- [7] D. P. Utomo and M. Mesran, "Analisis Komparasi Metode Klasifikasi Data Mining dan Reduksi Atribut Pada Data Set Penyakit Jantung," *J. Media Inform. Budidarma*, vol. 4, no. 2, p. 437, 2020, doi: 10.30865/mib.v4i2.2080.
- [8] D. Jain and V. Singh, "Feature selection and classification systems for chronic disease prediction: A review," *Egypt. Informatics J.*, vol. 19, no. 3, pp. 179–189, 2018, doi: 10.1016/j.eij.2018.03.002.
- [9] O. Almadani and R. Alshammari, "Prediction of stroke using data mining classification techniques," *Int. J. Adv. Comput. Sci. Appl.*, vol. 9, no. 1, pp. 457–460, 2018, doi: 10.14569/IJACSA.2018.090163.
- [10] Indarto, E. Utami, and S. Raharjo, "Mortality Prediction Using Data Mining Classification Techniques in Patients with Hemorrhagic Stroke," *2020 8th Int. Conf. Cyber IT Serv. Manag. CITSM 2020*, 2020, doi: 10.1109/CITSM50537.2020.9268802.
- [11] G. Sailasya and G. L. A. Kumari, "Analyzing the Performance of Stroke Prediction using ML Classification Algorithms," *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 6, pp. 539–545, 2021, doi: 10.14569/IJACSA.2021.0120662.
- [12] X. Li, D. Bian, J. Yu, H. Mao, M. Li, and D. Zhao, "Using Machine Learning Models To Improve Stroke Risk Level Classification Methods of China National Stroke Screening," *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS*, vol. 2, pp. 1386–1390, 2019, doi: 10.1109/EMBC.2019.8857657.
- [13] S. Kalmegh, "Analysis of WEKA Data Mining Algorithm REPTree , Simple Cart and RandomTree for Classification of Indian News," *Int. J. Innov. Sci. Eng. Technol.*, vol. 2, no. 2, pp. 438–446, 2015.
- [14] M. Fatima and M. Pasha, "Survey of Machine Learning Algorithms for Disease Diagnostic," *J. Intell. Learn. Syst. Appl.*, vol. 09, no. 01, pp. 1–16, 2017, doi: 10.4236/jilsa.2017.91001.
- [15] P. G. Sonia Singh, "Comparative Study ID3,CART AND C4.5 Decision Tree Algorithm," *Int. J. Adv. Inf. Sci. Technol.*, vol. 27, no. 27, p. 98, 2014.
- [16] M. M. Saritas and A. Yasar, "Performance Analysis of ANN and Naive Bayes Classification Algorithm for Data Classification," *Int. J. Intell. Syst. Appl. Eng.*, vol. 7, no. 2, pp. 88–91, 2019, doi: 10.1039/b000000x.
- [17] L. Dey, S. Chakraborty, A. Biswas, B. Bose, and S. Tiwari, "Sentiment Analysis of Review Datasets Using Naïve Bayes and K-NN Classifier," *Int. J. Inf. Eng. Electron. Bus.*, vol. 8, no. 4, pp. 54–62, 2016, doi: 10.5815/ijieeb.2016.04.07.
- [18] D. Vijayarani, "Liver Disease Prediction using SVM and Naïve Bayes Algorithms," *Int. J. Sci. Eng. Technol. Res.*, vol. 4, no. 4, pp. 816–820, 2015.
- [19] K. Fatmawati and A. P. Windarto, "Data Mining: Penerapan Rapidminer Dengan K-Means Cluster Pada Daerah Terjangkit Demam Berdarah Dengue (Dbd) Berdasarkan Provinsi," *Comput. Eng. Sci. Syst. J.*, vol. 3, no. 2, p. 173, 2018, doi: 10.24114/cess.v3i2.9661.
- [20] D. Ratnasari, A. Mughni, E. Yudhanto, and S. Budijitno, *Perbedaan Derajat Diferensiasi Adenokarsinoma Kolorektal Pada Golongan Usia Muda, Baya, Dan Tua Di Rsup Dr.Kariadi Semarang*, vol. 1, no. 1. 2012.
- [21] T. Rismawan and D. S. Kusumadewi, "Aplikasi K-Means Untuk Pengelompokkan Mahasiswa Berdasarkan Nilai Body Mass Index (Bmi) & Ukuran Kerangka," *Semin. Nas. Apl. Teknol. Inf.*, vol. 21, no. 01, pp. 1907–5022, 2008.
- [22] S. Monalisa and F. Hadi, "Penerapan Algoritma CART Dalam Menentukan Jurusan Siswa di MAN 1 Inhil," *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 9, no. 3, pp. 387–394, 2020, doi: 10.32736/sisfokom.v9i3.932.
- [23] N.- Insan, M. Hadijati, and I. Irwansyah, "Perbandingan Metode Classification and Regression Trees (CART) dengan Naïve Bayes Classification (NBC) dalam Klasifikasi Status Gizi Balita di Kelurahan Pagesangan Barat," *Eig. Math. J.*, vol. 1, no. 2, p. 9, 2020, doi: 10.29303/emj.v1i2.68.
- [24] J. Suntoro, *Data Mining : Algoritma dan Implementasi dengan Pemrograman PHP*. Jakarta: Elex Media Komputindo, 2019.
- [25] M. Guntur, J. Santony, and Y. Yuhandri, "Prediksi Harga Emas dengan Menggunakan Metode Naïve Bayes dalam Investasi untuk Meminimalisasi Resiko," *J. RESTI (Rekayasa*

- Sist. dan Teknol. Informasi*), vol. 2, no. 1, pp. 354–360, 2018, doi: 10.29207/resti.v2i1.276.
- [26] Y. I. Kurniawan, “Perbandingan Algoritma Naive Bayes dan C.45 dalam Klasifikasi Data Mining,” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 5, no. 4, p. 455, 2018, doi: 10.25126/jtiik.201854803.
- [27] Dwi Untari, K. Hastuti, E. Y. Hidayat, Dwi Untari, N. Limão, and N. Y. L. Gaol, “Data Mining untuk Menganalisa Prediksi Mahasiswa Berpotensi Non-Aktif Menggunakan Metode Decision Tree C4.5,” *Fak. Ilmu Komput. Univ. Dian Nuswantoro*, vol. 2013, no. November, pp. 31–48, 2010.
- [28] J. Susilo, T. Pujiatna, and S. Firmasari, “Pembinaan Tata Bahasa dan Bentuk Surat-Menyurat Indonesia Berbasis Microsoft di Desa Mandala, Dukupuntang Kabupaten Cirebon,” *JPPM (Jurnal Pengabd. dan Pemberdaya. Masyarakat)*, vol. 4, no. 1, p. 173, 2020, doi: 10.30595/jppm.v0i0.5498.
- [29] M. R. A. Nasution and M. Hayaty, “Perbandingan Akurasi dan Waktu Proses Algoritma K-NN dan SVM dalam Analisis Sentimen Twitter,” *J. Inform.*, vol. 6, no. 2, pp. 226–235, 2019, doi: 10.31311/ji.v6i2.5129.
- [30] N. D. Saputri, “Komparasi Penerapanmetode Bagging Dan Adaboostpada Algoritma C4.5 Untuk Prediksi Penyakit Stroke,” 2021.