

PENERAPAN METODE *CLASSIFICATION AND REGRESSION TREE* (CART) DALAM KLASIFIKASI STROKE DI RS X

Ayu Pangestuti^{1*}, Rachmah Indawati², Diah Indriani³

Program Studi D3 Rekam Medis dan Informasi Kesehatan Fakultas Teknologi dan Manajemen Kesehatan Institut Ilmu Kesehatan Bhakti Wiyata Kediri¹, Program Studi S2 Kesehatan Masyarakat Fakultas Kesehatan Masyarakat Universitas Airlangga Surabaya²³

*Corresponding Author : ayu.pangestuti@iik.ac.id

ABSTRAK

Metode klasifikasi, baik parametrik maupun non-parametrik, sering digunakan dalam penelitian untuk mengelompokkan data secara sistematis. Namun metode parametrik memiliki keterbatasan seperti pemenuhan asumsi dan penyederhanaan interpretasi. Metode non-parametrik, meskipun lebih mudah diinterpretasikan. Prevalensi *stroke* di Indonesia meningkat, dengan data rekam medis pasien *stroke* menunjukkan pentingnya penerapan metode klasifikasi dalam memahami dan mengelola risiko *stroke* untuk penanganan yang lebih cepat dan tepat. Penelitian ini menggunakan data rekam medis pasien *stroke* yang dirawat inap di RS X dan mengaplikasikan metode CART untuk klasifikasi faktor risiko dengan jumlah 480 kasus *stroke* pertama kali dengan 14 variabel prediktor. Metode yang digunakan dalam analisis menggunakan *Classification and Regression Tree* (CART). Tahapan analisis CART meliputi pembentukan pohon klasifikasi maksimal, pemilahan variabel prediktor, perhitungan keberagaman data, pemilahan simpul berdasarkan indeks gini, dan pemangkasan pohon untuk mendapatkan pohon optimal. Setelah dilakukan analisis, maka dapat dihitung akurasi, *sensitivity*, *specivity*, APER, dan presisi dalam metode yang dihasilkan CART. Berdasarkan hasil analisis, dapat diketahui bahwa ada 8 variabel yang penting, yaitu GCS UGD, gejala klinis muntah, tekanan darah sistolik rawat inap, riwayat diabetes mellitus, tekanan darah diastolik rawat inap, GDA dan kolesterol total. Hasil prediksi untuk metode CART pada akurasi sebesar 81%, *sensitivity* 82%, *specivity* 75%, APER 19%, dan presisi 95%. Simpulan dari penelitian ini menunjukkan metode CART efektif dalam klasifikasi faktor risiko *stroke* dengan akurasi 81%. Delapan variabel utama telah diidentifikasi, menunjukkan pentingnya penerapan metode klasifikasi untuk memahami dan mengelola risiko *stroke* guna meningkatkan efektivitas penanganan pasien.

Kata kunci : *Classification and Regression Tree* (CART), *stroke* pertama kali

ABSTRACT

Classification methods, both parametric and non-parametric, are widely used in research to systematically categorize data. Non-parametric methods, while easier to interpret, also have their own challenges. This study utilizes medical records of stroke inpatients at Hospital X and applies the Classification and Regression Tree (CART) method to classify risk factors. The study includes 480 first-time stroke cases with 14 predictor variables. The analysis process using the CART method involves several steps, including constructing a maximum classification tree, selecting predictor variables, calculating data diversity, splitting nodes based on the Gini index, and pruning the tree to obtain an optimal model. After conducting the analysis, the accuracy, sensitivity, specificity, APER (Apparent Error Rate), and precision of the CART method were calculated. The results indicate that eight key variables play a significant role: Emergency Department GCS, clinical symptoms of vomiting, inpatient systolic blood pressure, history of diabetes mellitus, inpatient diastolic blood pressure, random blood glucose, and total cholesterol. The prediction results using the CART method show an accuracy of 81%, sensitivity of 82%, specificity of 75%, APER of 19%, and precision of 95%. In conclusion, this study demonstrates that the CART method is effective in classifying stroke risk factors with an accuracy of 81%. The identification of eight key variables highlights the importance of classification methods in understanding and managing stroke risk to enhance patient treatment effectiveness.

Keywords : *Classification and Regression Tree* (CART), first-time stroke

PENDAHULUAN

Metode klasifikasi sering menjadi topik penelitian dalam statistika karena mampu mengelompokkan data secara sistematis (Siahaan et al., 2016). Klasifikasi merupakan proses mengelompokkan objek berdasarkan kesamaan karakteristik atau ciri tertentu ke dalam berbagai kategori. Metode ini telah diterapkan di berbagai bidang seperti sosial, teknologi, pendidikan, dan kesehatan. Pengklasifikasian data dapat dilakukan menggunakan pendekatan parametrik maupun non-parametrik (Keputusan Dirjen Penguatan Riset dan Pengembangan Ristek Dikti et al., 2017). Dalam metode parametrik, regresi dan analisis diskriminan digunakan sebagai teknik utama. Sementara itu, dalam pendekatan non-parametrik, pohon klasifikasi sering kali diterapkan. Struktur pohon klasifikasi mudah dipahami dan diinterpretasikan. Pohon ini terdiri dari sejumlah simpul (node) yang saling terhubung melalui cabang dan berkembang dari root node ke arah bawah hingga mencapai *leaf* node. Hasil klasifikasi ditentukan oleh variabel yang paling berpengaruh pada *leaf* node, dengan faktor-faktor lain yang turut memengaruhi klasifikasi pada *leaf* node berikutnya (Blockeel et al., 2023).

Salah satu metode pohon klasifikasi non-parametrik yaitu *Classification and Regression Tree* (CART). Metode *Classification and Regression Tree* (CART) kini banyak digunakan dalam berbagai penelitian. Popularitasnya didorong oleh kemampuannya dalam menghasilkan estimasi dengan tingkat kesalahan rendah serta kemudahan dalam menginterpretasikan hasil analisis. Selain itu, perkembangan perangkat lunak turut berkontribusi terhadap meningkatnya penggunaan metode ini, terutama di kalangan praktisi yang tidak memiliki latar belakang statistik yang kuat (Bagus Sartono, 2010). Algoritma *Classification and Regression Tree* (CART) adalah metode yang sederhana namun memiliki kinerja yang optimal. Algoritma ini mampu melakukan proses klasifikasi dengan efisiensi tinggi dan waktu komputasi yang cepat (Wuryani et al., 2021)

Stroke adalah penyakit yang terjadi akibat gangguan pada sistem pembuluh darah di otak. Penyakit ini memiliki dampak besar dalam kehidupan sehari-hari, seperti penurunan daya ingat, keterbatasan dalam bergerak, gangguan penglihatan dan bicara, dan dalam beberapa kasus dapat berujung pada kematian. *Stroke* merupakan salah satu jenis penyakit tidak menular (PTM) yang dapat terjadi secara mendadak akibat terganggunya aliran darah ke otak selama lebih dari 24 jam. Baik di negara maju maupun berkembang, *stroke* menjadi permasalahan kesehatan yang signifikan (Dian et al, 2021). Faktor determinan yang mempengaruhi *stroke* adalah usia, jenis kelamin, hipertensi, kolesterol, serta diabetes mellitus. Risiko *stroke* meningkat seiring bertambahnya usia seseorang. Meskipun *stroke* dapat terjadi pada segala usia, lebih dari 70% kasus dialami oleh individu berusia di atas 65 tahun. Selain itu, pria memiliki kecenderungan lebih tinggi untuk mengalami *stroke* dibandingkan wanita. Hal ini disebabkan oleh tingginya prevalensi faktor risiko *stroke*, seperti hipertensi, pada pria (Rizaldy Taslim Pinzon et al., 2021).

Kasus *stroke* di Indonesia mengalami peningkatan yang signifikan, menjadikan negara ini sebagai salah satu dengan jumlah penderita *stroke* tertinggi di Asia. Berdasarkan data Riskesdas 2018, prevalensi *stroke* di Indonesia meningkat menjadi 10,9%, dibandingkan dengan 7,0% pada tahun 2013. Berdasarkan data dinas Kesehatan Jawa Timur tahun 2022 penyakit *stroke* masuk dalam 10 besar penyakit terbanyak, dengan jumlah 30854 jiwa. Kematian akibat *stroke* sulit diperkirakan karena gejala klinisnya tidak terduga dan berkembang sangat cepat. Untuk memprediksi penyakit *stroke*, maka metode klasifikasi dapat menjadi alternatif untuk melihat variabel *prediktor* yang mempengaruhi terjadinya *stroke*. Metode CART untuk klasifikasi penyakit hepatitis pada anak, diketahui algoritma CART memiliki tingkat akurasi yang baik, yakni 94% (Hendri et al, 2021). Tujuan pada penelitian ini adalah untuk mengetahui ketepatan klasifikasi CART dalam menentukan klasifikasi kejadian *stroke*.

METODE

Jenis penelitian ini adalah penelitian analitik observasional. Populasi dan sampel dalam penelitian ini adalah seluruh pasien rawat inap dengan kode diagnosis ICD-10 di *Chapter IX* (I60 – I69) di bangsal Wisnu/*Stroke Center* Rumah Sakit X pada tahun 2022-2023 yang memenuhi kriteria inklusi: pasien yang menderita *stroke* pertama kali, isi rekam medis lengkap, pasien dengan diagnosis utama dengan kode ICD-10 subbab I60-I63. Besar sampel yang digunakan adalah 480 yang memenuhi kriteria inklusi.

Variabel independen dalam penelitian ini adalah umur, jenis kelamin, gula darah acak (GDA), kolesterol total, tekanan sistolik UGD, tekanan diastolik UGD, tekanan sistolik rawat inap, tekanan diastolik rawat inap, riwayat diabetes mellitus, riwayat hipertensi, *Glasgow Coma Scale* (GCS) UGD, *Glasgow Coma Scale* (GCS) rawat inap, gejala klinis muntah, gejala klinis nyeri kepala, sedangkan variabel dependen adalah diagnose *stroke*. Data dalam penelitian ini diambil dari data sekunder yang diperoleh dari *medical record* pasien *stroke* di rumah sakit tahun 2022-2023. Instrumen dalam pengumpulan data penelitian berupa formulir pengumpulan data yaitu berupa lembar observasi yang terdiri dari variabel penelitian. Data yang telah dikumpulkan kemudian dilakukan proses *cleaning* menggunakan perangkat lunak *Microsoft excel* dan *STATA*, sehingga dihasilkan data yang siap untuk dianalisis. Penelitian ini menggunakan perangkat lunak *open-access*, yaitu RStudio, untuk analisis data.

HASIL

Data yang digunakan dalam penelitian ini adalah data dokumen rekam medis pasien *stroke* yang rawat inap pada ruang Wisnu atau *stroke center* di RS X. Berdasarkan data laporan penyakit pada tahun 2022-2023 jumlah pasien *stroke* adalah sebanyak 1165. Dari data laporan penyakit pasien *stroke* 2022-2023, didapatkan pasien yang dirawat pada unit *stroke center* adalah 617 pasien. Sedangkan yang sesuai dengan kriteria inklusi sebanyak 480 pasien. Pemilihan variabel dalam penelitian ini disesuaikan dengan kerangka konsep penelitian. Ada 14 variabel yang akan digunakan, yaitu umur, jenis kelamin, gula darah acak (GDA), kolesterol total, tekanan sistolik UGD, tekanan diastolik UGD, tekanan sistolik rawat inap, tekanan diastolik rawat inap, riwayat diabetes mellitus, riwayat hipertensi, *Glasgow Coma Scale* (GCS) UGD, *Glasgow Coma Scale* (GCS) rawat inap, gejala klinis muntah, dan gejala klinis nyeri kepala.

Identifikasi Indikator Kejadian *Stroke*

Tabel 1. Distribusi Frekuensi Pasien *Stroke* di RS X Tahun 2022-2023

Jenis <i>Stroke</i>	Frekuensi	Persentase (%)
<i>Infark</i>	354	73,7
<i>Haemorrhage</i>	126	26,3
Total	480	100

Berdasarkan tabel 1, dapat diketahui bahwa mayoritas pasien yang dirawat pada unit *stroke center* termasuk kedalam jenis *stroke* infark yakni sebanyak 354 pasien, artinya terdapat sebesar 73,7% pasien dengan jenis *stroke* infark. Sedangkan sebanyak 126 atau 26,3% pasien terdiagnosa *stroke* haemorrhage.

Berdasarkan tabel 2, menunjukkan bahwa sebanyak 39,6% pasien *stroke* di RS X berusia di bawah 60 tahun, sementara 60,4% berusia 60 tahun ke atas. Dari segi jenis kelamin, 50,8% pasien adalah laki-laki, dan 49,2% adalah perempuan. Sebagian besar pasien *stroke* (62,0%) memiliki gula darah acak (GDA) dalam kategori normal, sedangkan 21,7% berada di kategori pre-diabetes dan 16,3% dalam kategori diabetes. Sebanyak 65,6% pasien memiliki kolesterol

total normal, 21,7% borderline, dan 12,7% tinggi. Tekanan darah sistolik di UGD paling banyak ditemukan pada kategori grade 2 hipertensi (61,5%), diikuti grade 1 hipertensi (20,0%), kategori normal (9,7%), dan high normal (8,7%). Tekanan darah diastolik di UGD didominasi oleh grade 1 hipertensi (59,8%), kategori normal (29,6%), dan high normal (10,6%), tanpa kasus grade 2 hipertensi.

Tabel 2. Data Pasien Stroke di RS X Tahun 2022-2023

No	Variabel	Kategori	Frekuensi	Percentase (%)
1	Umur	Kurang dari 60	190	39,6
		60 tahun ke atas	290	60,4
2	Jenis Kelamin	Laki-laki	244	50,8
		Perempuan	236	49,2
3	GDA	Normal	298	62,0
		Pre-Diabetes	104	21,7
		Diabetes	78	16,3
4	Kolesterol Total	Normal	315	65,6
		Borderline	104	21,7
		Tinggi	61	12,7
5	Tekanan Sistolik UGD	Normal	47	9,7
		High-Normal	42	8,7
		Grade 1 Hipertensi	96	20,0
		Grade 2 Hipertensi	295	61,5
6	Tekanan Diastolik UGD	Normal	142	29,6
		High-Normal	51	10,6
		Grade 1 Hipertensi	287	59,8
		Grade 2 Hipertensi	0	0
7	Tekanan Sistolik Rawat Inap	Normal	53	11,0
		High-Normal	50	10,4
		Grade 1 Hipertensi	115	24,0
		Grade 2 Hipertensi	262	54,6
8	Tekanan Diastolik Rawat Inap	Normal	167	34,8
		High-Normal	78	16,2
		Grade 1 Hipertensi	90	18,8
		Grade 2 Hipertensi	145	30,2
9	Riwayat Diabetes Mellitus	Tidak	397	82,7
		Ya	83	17,3
10	Riwayat Hipertensi	Tidak	141	29,4
		Ya	339	70,6
11	GCS UGD	GCS >14	302	62,9
		GCS 9 – 14	93	19,4
		GCS 3-8	85	17,7
12	GCS Rawat Inap	GCS >14	317	66,0
		GCS 9 -14	86	17,9
		GCS 3 – 8	77	16,1
13	Gejala Klinis Muntah	Tidak	358	74,6
		Ya	122	25,4
14	Gejala Klinis Nyeri Kepala	Tidak	251	52,3
		Ya	229	47,7

Di ruang rawat inap, tekanan sistolik pasien terbanyak berada pada kategori grade 2 hipertensi (54,6%), grade 1 hipertensi (24,0%), normal (11,0%), dan high normal (16,2%). Tekanan diastolik rawat inap terbanyak pada kategori normal (34,8%), grade 2 hipertensi (30,2%), grade 1 hipertensi (18,8%), dan high normal (16,2%). Pasien dengan riwayat diabetes mellitus tercatat sebanyak 17,3%, sementara 70,6% pasien memiliki riwayat hipertensi.

Sebagian besar pasien *stroke* memiliki GCS >14 (66,0%), GCS 9–14 (17,9%), dan GCS 3–8 (16,0%). Gejala muntah terjadi pada 25,4% pasien, dan 47,7% pasien merasa nyeri kepala.

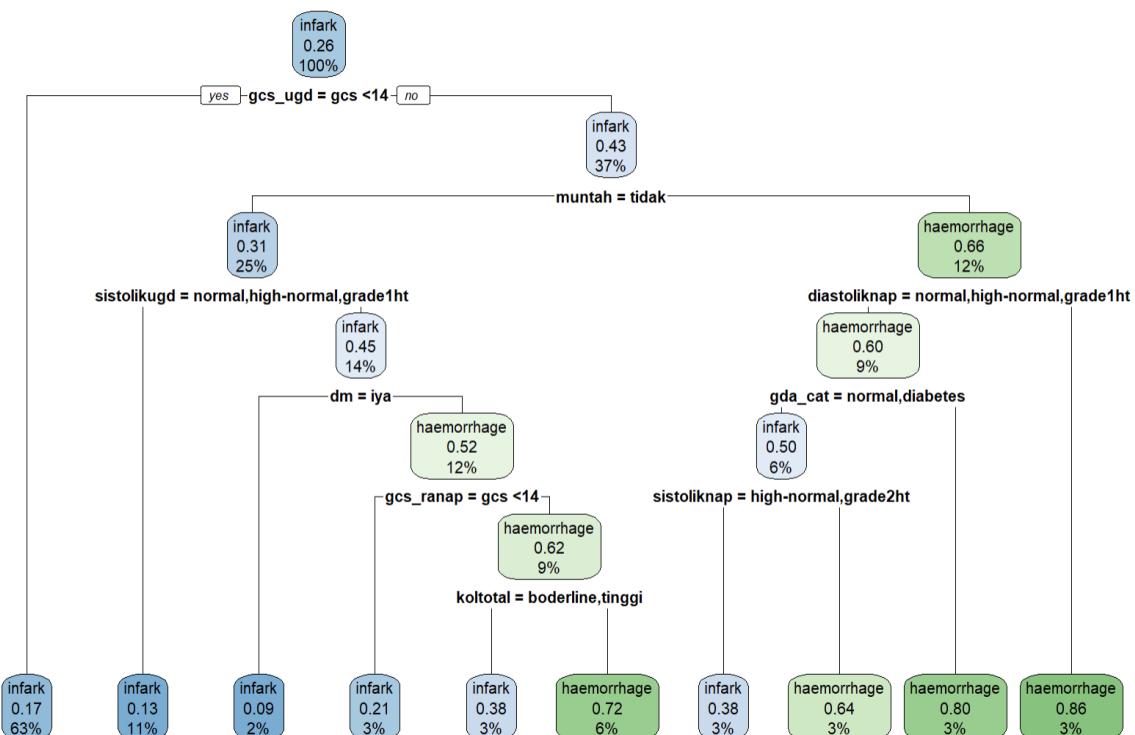
Analisis Klasifikasi Kejadian *Stroke* Menggunakan CART

Tahapan pembentukan pohon klasifikasi dengan metode CART secara umum dibagi menjadi lima tahap sebagai berikut: Pembentukan pohon klasifikasi maksimal CART diawali dengan mengimport data dan memeriksa keberadaan nilai hilang dan data duplikat. Kemudian memastikan bahwa data telah dalam format yang sesuai. Data frame yang digunakan sudah diperiksa untuk memastikan tidak ada yang *missing*.

```
summary(data)

##      jenis_stroke      sex      dm      ht      muntah
##  infark    :354  laki-laki:244  tidak:397  tidak:141  tidak:358
##  haemorrhage:126  perempuan:236  iya   : 83  iya   :339  iya   :122
##
##      nyeri_kepala      sistolikugd      diastolikgd      koltotal
##  tidak:251  normal       : 47  normal       :142  normal       :315
##  iya   :229  high-normal: 42  high-normal: 51  borderline:104
##          grade1ht     : 96  grade1ht     :287  tinggi     : 61
##          grade2ht     :295  grade2ht     : 0
##          gda_cat      :298  sistoliknap  :295  diastoliknap: 0
##  normal    :298  normal       : 53  normal       :167  Kurang dari 60 :190
##  pre-diabetes:104  high-normal: 50  high-normal: 78  60 tahun ke atas:290
##  diabetes   : 78  grade1ht     :115  grade1ht     : 90
##          grade2ht     :262  grade2ht     :145
##          gcs_ranap    :317  gcs_ugd      :302
##  gcs <14   :317  gcs <14     :302
##  gcs 9-14  : 86  gcs 9-14    : 93
##  gcs 3-8   : 77  gcs 3-8    : 85
##
```

Gambar 1. Data Frame yang Digunakan Dalam Analisis



Gambar 2. Pohon Klasifikasi CART Optimal untuk *Stroke* pada Pasien di RS X Tahun 2023 – 2024

Memilih variabel prediktor yang digunakan untuk membangun model CART. Variabel *prediktor* yang digunakan ada 14 variabel, yaitu umur, jenis kelamin, GDA, kolesterol total, tekanan sistolik UGD, tekanan diastolik UGD, tekanan sistolik rawat inap, tekanan diastolik rawat inap, riwayat diabetes mellitus, riwayat hipertensi, GCS UGD, GCS rawat inap, gejala klinis muntah dan gejala klinis nyeri kepala. Menghitung keragaman data digunakan dalam memilih variabel terbaik untuk melakukan split pada *node* pemisah. Pemilahan simpul berdasarkan indeks gini. Indeks gini merupakan *impurity measure* dapat membantu menemukan fungsi pemilah yang optimal. Nilai *impurity measure* (*importance*) tertinggi akan digunakan dalam tahap pemilihan pemisah. Pemangkasan yang bertujuan untuk mendapatkan pohon optimal (layak dan tidak melebar). Pohon klasifikasi CART ditunjukkan dalam gambar 2.

Gambar 2 menunjukkan hasil pemangkasan (*pruning*) untuk menghasilkan pohon klasifikasi optimal yang lebih sederhana. Proses *pruning* ini dilakukan dengan menggunakan *command* dari Rstudio. Dengan begitu dapat membentuk pohon optimal yang memiliki risiko sangat minimum mengalami *overfitting* (pohon yang lebar dan kompleks). Berdasarkan gambar 2, pohon keputusan dimulai dengan *node* akar yang mempertimbangkan nilai *Glasgow Coma Scale* di UGD (GCS UGD). Jika GCS UGD kurang dari 14, ada probabilitas 26% pasien mengalami infark. Sebaliknya, jika GCS IGD tidak kurang dari 14, langkah selanjutnya adalah mempertimbangkan apakah pasien mengalami muntah. Jika pasien tidak muntah, probabilitas infark meningkat menjadi 43%.

Jika pasien mengalami muntah, pohon keputusan memperhitungkan beberapa faktor tambahan. Pada cabang pertama, jika tekanan darah sistolik di UGD berada dalam kategori normal, tinggi-normal, atau hipertensi tingkat 1, probabilitas infark adalah 31%. Cabang ini kemudian membagi lebih lanjut berdasarkan apakah pasien memiliki diabetes melitus (DM). Jika pasien tidak memiliki diabetes, probabilitas infark adalah 17%. Jika pasien memiliki diabetes, *node* berikutnya mempertimbangkan apakah GCS rawat inap kurang dari 14. Jika ya, probabilitas perdarahan adalah 52%. Jika tidak, probabilitas infark adalah 3%.

Pada cabang kedua, jika tekanan darah diastolik di rawat inap berada dalam kategori normal, tinggi-normal, atau hipertensi tingkat 1, probabilitas perdarahan meningkat menjadi 66%. Cabang ini juga membagi lebih lanjut berdasarkan beberapa faktor lain, seperti tekanan darah sistolik di rawat inap dan kadar gula darah. Jika tekanan darah sistolik di rawat inap berada dalam kategori tinggi-normal atau hipertensi tingkat 2, probabilitas perdarahan meningkat lebih jauh menjadi 80%. Pohon keputusan ini membantu dalam menentukan indikator dan prediksi klinis dalam menentukan apakah kasus tersebut akan diklasifikasikan sebagai "infark" atau "haemorrhage".

Tabel 3. Confusion Matrix CART Optimal

Hasil Prediksi	Hasil Aktual	
	Stroke Infark	Stroke Haemorrhage
n	n	
Stroke Infark	336	72
Stroke Hemorrhage	18	54

Berikut ini adalah perhitungan akurasi, sensitivitas, spesifisitas, APER, dan presisi dari hasil klasifikasi:

$$1-\text{APER} = \frac{TP+TN}{TP+FP+TN+FN} = \frac{336+54}{336+18+54+72} = 0,81$$

$$\text{Sensitivity} = \frac{TP}{TP+FN} = \frac{336}{336+72} = 0,82$$

$$\text{Specificity} = \frac{TN}{TN+FP} = \frac{54}{54+18} = 0,75$$

$$\text{APER} = \frac{FN+FP}{TP+FP+TN+FN} = \frac{72+18}{336+18+54+72} = 0,19$$

$$\text{Presisi} = \frac{TP}{TP+FP} = \frac{336}{336+18} = 0,95$$

PEMBAHASAN

Metode *Classification and Regression Tree* (CART) memiliki beberapa keunggulan dibandingkan teknik klasifikasi lainnya. CART dapat digunakan pada kumpulan data berukuran besar, dengan salah satu penelitian menyebutkan bahwa jumlah sampel minimal adalah 100. Selain itu, metode ini juga mampu menangani banyak variabel bebas, hingga 11 variabel atau lebih (Reny Roswita Nazar, 2018). Penelitian metode CART ini dilakukan dengan sampel 480 pasien *stroke* dengan 14 variabel independen. Kelemahan metode CART adalah hanya bisa dibentuk binary split, tidak stabil dalam pohon klasifikasi karena CART sangat sensitif dengan data baru, CART sangat bergantung dengan besar sampel, dan tiap pemilahan bergantung pada nilai yang hanya berasal dari satu variabel penjelas. Keterbatasan lainnya adalah cenderung membentuk pohon pada banyak tingkat sehingga hasilnya tidak efisien (Suwardika et al., 2019).

Pohon klasifikasi terdiri dari sejumlah simpul (node) yang saling terhubung melalui cabang, bergerak dari root node ke bawah hingga mencapai *leaf* node. Hasil klasifikasi ditentukan oleh variabel yang paling berpengaruh pada *leaf* node, dengan faktor-faktor lain yang turut memengaruhi klasifikasi pada *leaf* node berikutnya (Blockeel et al., 2023). Tahapan pembentukan pohon klasifikasi dengan metode CART secara umum dapat dibagi menjadi lima tahapan, yaitu mengimpor data dan memeriksa keberadaan nilai hilang; pada pembentukan pohon klasifikasi maksimal yang terdiri atas tahapan pemilihan pemisah, tanpa penetapan simpul terminal, dan tahap penandaan label kelas; dilanjutkan dengan tahapan pemangkasan pohon klasifikasi untuk penentuan pohon klasifikasi optimal. Pemilihan dilakukan dengan menggunakan hasil nilai improvement yang berasal dari perhitungan indeks gini. Metode pemilahan indeks gini karena lebih mudah diterapkan dalam berbagai kasus dan perhitungan lebih sederhana serta cepat (Rahmi et al., 2017). Improvement adalah ukuran tingkat keheterogenan suatu kelas pada simpul tertentu dalam pohon klasifikasi, yang berperan dalam menentukan fungsi pemisah yang paling optimal.

Pembentukan pohon maksimal CART akan berhenti ketika simpul terminal terbentuk. Proses ini akan terjadi apabila jumlah kasus minimum pada suatu simpul terminal adalah 5 kasus (Siahaan et al., 2016). Pemangkasan adalah langkah terakhir dalam metode CART yang dilakukan dengan tujuan mendapat pohon CART yang optimal. Pohon optimal adalah pohon yang tidak berukuran besar dan tidak menyebabkan kompleksitas tinggi. Pohon optimal juga sering dikatakan pohon yang sederhana (Rahmadani et al., 2023). Hasil klasifikasi CART optimal memiliki 18 simpul (*node*). Simpul tersebut terdiri dari 1 simpul induk (*node 0*), 9 simpul keputusan (*node 1,2,3,5,7,11,15,30, dan 60*) dan 9 simpul terminal (*node 4,6,10,14,22,31,61,120 dan 121*).

Berdasarkan hasil analisis menggunakan *Classification and Regression Trees* (CART), dapat diketahui bahwa ada 13 variabel yang penting, yaitu GCS UGD, riwayat diabetes mellitus, tekanan darah sistolik UGD, tekanan darah sistolik rawat inap, tekanan darah diastolik UGD, gejala klinis muntah, jenis kelamin, tekanan darah diastolik rawat inap, usia, gejala klinis nyeri kepala, riwayat hipertensi, GDA, dan kolesterol total. Tingkat akurasi yang diperoleh pada penelitian ini adalah sebesar 80%. Hasil akurasi pada penelitian ini lebih baik dibandingkan dengan penelitian sebelumnya, dikarenakan adanya data *ct-scan* yang sudah diketahui oleh peneliti untuk memastikan diagnosis *stroke* di rumah sakit. Penelitian yang dilakukan oleh (Subarkah et al., 2021) mencatat tingkat akurasi sebesar 63,4%, yang lebih rendah dibandingkan dengan hasil yang dicapai dalam penelitian (Fadillah et al, 2022) (yang mencapai 74,73%). Hasil ini menunjukkan bahwa peningkatan dalam metode analisis atau penyesuaian pada variabel yang digunakan mungkin dapat meningkatkan performa prediksi

model. Penelitian ini sendiri mencatat tingkat akurasi sebesar 80% yang menunjukkan kemungkinan adanya penggunaan variabel (GCS UGD, riwayat diabetes mellitus, tekanan darah sistolik UGD, tekanan darah sistolik rawat inap, tekanan darah diastolik UGD, gejala klinis muntah, jenis kelamin, tekanan darah diastolik rawat inap, usia, gejala klinis nyeri kepala, GDA, dan kolesterol total) atau pemilihan parameter yang lebih optimal dalam membangun model CART. Namun, perlu dilakukan studi lebih lanjut untuk menguji hasil ini secara luas dan memastikan keandalannya dalam berbagai situasi serta berbagai kelompok populasi yang berbeda.

KESIMPULAN

Performansi akurasi ketepatan klasifikasi menggunakan Classification and Regression Tree (CART) dalam mengklasifikasikan faktor risiko (umur, jenis kelamin, GDA, kolesterol total, tekanan darah sistolik UGD, tekanan darah diastolik UGD, tekanan darah sistolik rawat inap, tekanan darah diastolik rawat inap, riwayat diabetes mellitus, riwayat hipertensi, GCS UGD, GCS rawat inap, gejala klinis muntah, gejala klinis dan nyeri kepala) yang berpengaruh terhadap klasifikasi *stroke* infark dan haemorrhage yaitu sebesar 81,0%.

UCAPAN TERIMAKASIH

Ucapan terimakasih peneliti sampaikan kepada jajaran pimpinan dan seluruh staff Rumah Sakit, partner peneliti, dan semua pihak yang telah membantu peneliti dalam menyelesaikan kegiatan penelitian dan juga dalam penyusunan jurnal.

DAFTAR PUSTAKA

- Angraini Simamora, F., Simamora, A. A., & Silvia. (2021). Pengaruh mirror therapy terhadap peningkatan kekuatan otot pada pasien *stroke* di Rumah Sakit Umum Daerah Kota Padangsidimpuan. *CHMK Nursing Scientific Journal*, 5(2).
- Blockeel, H., Devos, L., Frénay, B., Nanfack, G., & Nijssen, S. (2023). Decision trees: From efficient prediction to responsible AI. *Frontiers in Artificial Intelligence*, 6. Frontiers Media SA. <https://doi.org/10.3389/frai.2023.1124553>
- Fadillah Hermawan, A., Rakhmat Umbara, F., & Kasyidi, F. (2022). MIND (Multimedia Artificial Intelligent Networking Database) prediksi awal penyakit *stroke* berdasarkan rekam medis menggunakan metode algoritma CART (Classification and Regression Tree). *MIND Journal*, 7(2), 151–164. <https://doi.org/10.26760/mindjournal.v7i2.151-164>
- Jones, H. S., Makmun, A. S., & Muhchromin. (2021). Implementasi metode CART untuk klasifikasi diagnosis penyakit hepatitis pada anak. *Journal of Informatics, Information System, Software Engineering and Applications*, 3(2), 61–70. <https://doi.org/10.20895/INISTA.V3I2>
- Keputusan Dirjen Penguatan Riset dan Pengembangan Ristek Dikti, S., Chairunisa, R., & Astuti, W. (2017). Terakreditasi SINTA peringkat 2 perbandingan CART dan Random Forest untuk deteksi kanker berbasis klasifikasi data microarray. *Masa Berlaku Mulai*, 1(3), 805–812.
- Nazar, R. R. (2018). Penerapan metode CHIAD (Chi-Squared Automatic).
- Pinzon, R. T., Kedokteran, F., Rs, U., & Yogyakarta, B. (2021). *Rizaldy Taslim Pinzon*.
- Rahmadani, D., Muzafar, A. A., Hamid, A., & Annisa, R. (2023). Comparative analysis of C4.5 and CART algorithms for classification of *stroke*. *Seminar Nasional Penelitian dan Pengabdian Masyarakat (SENTIMAS)*. <https://journal.irpi.or.id/index.php/sentimas>

- Rahmi, I. H., Rahmy, H. A., Matematika, J., Andalas, U., Nutrisi, J., & Kesehatan Masyarakat, F. (2017). Telaah faktor-faktor yang mempengaruhi status gizi balita di Kota Padang berdasarkan berat badan per tinggi badan menggunakan metode CART. *Eksakta*, 18(2). <http://eksakta.ppj.unp.ac.id>
- Saraswati, D. R., & Penelitian dan Pengembangan Biomedis dan Teknologi Dasar Kesehatan, P. (2021). Transisi epidemiologi *stroke* sebagai penyebab kematian pada semua kelompok usia di Indonesia. *Seminar Nasional Riset Kedokteran (SENSORIK II)*.
- Sartono, B. U. D. (2010). 4895-Article Text-13304-1-10-20120420. *Statistika Dan Komputasi*, 15(1), 1–7.
- Siahaan, D., Wahyuningsih, S., Deny, F., & Amijaya, T. (2016). Aplikasi classification and regression tree (CART) dan regresi logistik ordinal dalam bidang pendidikan (Studi Kasus: Predikat Kelulusan Mahasiswa S1 Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Mulawarman). *Jurnal EKSPONENSIAL*, 7(1).
- Subarkah, P., Abdallah, M. M., Oktaviani, S., & Hidayah, N. (2021). Komparasi akurasi algoritme CART dan neural network untuk diagnosis penyakit diabetic retinopathy. *Cogito Smart Journal*, 7(1).
- Suwardika, G., Ketut, I., & Suniantara, P. (2019). Analisis random forest pada klasifikasi CART ketidaktepatan waktu kelulusan mahasiswa Universitas Terbuka. *BAREKENG: Jurnal Ilmu Matematika dan Terapan*, 13(3). <https://doi.org/10.30598/barekengvol13iss3pp179-186ar910>
- Wuryani, N., Agustiani, S., Komputer, I., & Mandiri, N. (2021). Random forest classifier untuk deteksi penderita COVID-19 berbasis citra CT scan. *Jurnal Teknik Komputer AMIK BSI*, 7(2). <https://doi.org/10.31294/jtk.v4i2>