



Klasifikasi Penyakit Stroke Iskemik Menggunakan Algoritma C4.5 di Rumah Sakit Citra Husada Jember

Wahyu Triana*, Angga Rahagiyanto, Ervina Rachmawati, Bakhtiyar Hadi Prakoso

Jurusan Kesehatan, Politeknik Negeri Jember, Indonesia

Abstrak: Stroke Iskemik adalah kondisi dimana terjadinya disfungsi otak akibat kerusakan pada jaringan otak. Pada pelaporan 10 besar penyakit rawat inap di Rumah Sakit Citra Husada tahun 2021 sampai dengan 2023 mengalami kenaikan, Selain itu Stroke Iskemik menempati urutan peringkat ke-1 pada 10 besar kematian. Penelitian ini bertujuan untuk melakukan klasifikasi penyakit Stroke Iskemik berdasarkan faktor risiko menggunakan Algoritma C4.5 di Rumah Sakit Citra Husada Jember yang dapat berperan sebagai upaya pencegahan penyakit stroke iskemik. Jenis penelitian ini adalah penelitian kuantitatif yang kemudian diolah menggunakan metode Algoritma C4.5 dikarenakan dapat melakukan klasifikasi atau pengelompokan pada dataset dengan bantuan tools RapidMiner. Pengambilan sampel menggunakan teknik simple random sampling dengan jumlah sampel sebanyak 366 dokumen rekam medis, dengan kriteria inklusi yaitu dokumen pasien stroke iskemik dan kriteria eksklusi yaitu dokumen pasien non stroke iskemik. Hasil penelitian menunjukkan variabel hipertensi menjadi variabel yang paling berpengaruh dikarenakan memiliki nilai 96,17%. Hasil confusion matrix menunjukkan tingkat nilai accuracy, precision dan recall tertinggi pada split 90:10 dengan nilai accuracy sebesar 97,22%, nilai precision sebesar 94,73% dan nilai recall yaitu 100%.

Kata Kunci: Algoritma C4.5, Confusion Matrix, Stroke Iskemik

DOI:

<https://doi.org/10.47134/mpk.v2i2.6095>

*Correspondence: Wahyu Triana

Email: wahyuria68@gmail.com

Received: 15-10-2024

Accepted: 30-12-2024

Published: 30-01-2025



Copyright: © 2025 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (BY SA) license (<http://creativecommons.org/licenses/by/4.0/>).

Abstract: Ischemic Stroke is a condition in which brain dysfunction occurs due to damage to brain tissue. In the reporting of the top 10 inpatient diseases at Citra Husada Hospital from 2021 to 2023, there was an increase. In addition, Ischemic Stroke is ranked 1st in the top 10 deaths. This study aims to classify Ischemic Stroke disease based on risk factors using the C4.5 Algorithm at Citra Husada Hospital Jember which can play a role as an effort to prevent ischemic stroke disease. This type of research is quantitative research which is then processed using the C4.5 Algorithm method because it can classify or group the dataset with the help of the RapidMiner tool. Sampling using a simple random sampling technique with a sample size of 366 medical record documents, with inclusion criteria, namely ischemic stroke patient documents and exclusion criteria, namely non-ischemic stroke patient documents. The results of the study showed that the hypertension variable was the most influential variable because it had a value of 96.17%. The confusion matrix results show the highest levels of accuracy, precision and recall values at split 90:10 with an accuracy value of 97.22%, a precision value of 94.73% and a recall value of 100%.

Keywords: C4.5 Algorithm, Confusion Matrix, Ischemic Stroke

Pendahuluan

Penyakit stroke ialah penyakit yang dapat menyerang dan menyebabkan kematian secara mendadak pada penderitanya tanpa mengenal usia, ras, dan jenis kelamin. Stroke Iskemik menurut WHO adalah tanda-tanda klinis yang dapat berkembang dengan cepat akibat gangguan fungsi otak foklat atau global karena adanya sumbatan atau pecahnya

pembuluh darah di otak dengan gejala yang berlangsung selama 24 jam atau lebih [1]. Menurut WHO stroke merupakan penyebab kecacatan nomor satu dan penyebab kematian ketiga di dunia setelah jantung dan kanker [2]. Berdasarkan data *World Stroke Organization* setiap tahunnya terdapat 13,7 juta kasus baru stroke dan sekitar 5,5 juta kematian terjadi akibat penyakit stroke. Sekitar 70% penyakit stroke dan 87% kematian dan disabilitas akibat stroke terjadi pada negara berpendapatan rendah dan menengah [3].

Berdasarkan laporan nasional riskesdas tahun 2018 penduduk di Indonesia dengan usia lebih dari 15 tahun yang terkena stroke sejumlah 713.783 penduduk, sedangkan pada daerah Jawa Timur jumlah penduduk yang terkena stroke sejumlah 75.490 penduduk. Pada tahun 2023 di Kabupaten Jember jumlah kasus baru penderita stroke mencapai 1.190 jiwa. Pada Rumah Sakit Citra Husada Jember terdapat peningkatan secara signifikan terjadi pada tahun 2022 ke tahun 2023. Jumlah kasus pada tahun 2022 sebesar 176, kemudian meningkat sebanyak 47 kasus pada tahun 2023 sehingga total kasus morbiditas menjadi 223.

Tingginya angka kasus kematian pasien stroke iskemik memerlukan langkah antisipasi dari pihak pemberi pelayanan kesehatan seperti di rumah sakit dengan memberikan pencegahan agar tidak terjadi ledakan. Pada Rumah Sakit Citra Husada Jember faktor risiko stroke iskemik dapat berperan sebagai upaya pencegahan terjadinya stroke iskemik. Dengan memahami faktor risiko stroke iskemik, Rumah Sakit dapat menerapkan program pencegahan yang efektif. Terdapat beberapa faktor risiko dari penyakit stroke ada faktor risiko yang tidak dapat diubah seperti usia, jenis kelamin dan sejenisnya. Selain itu, terdapat beberapa faktor kondisi kesehatan seperti hipertensi, penyakit jantung dan lain-lain yang berkontribusi terhadap risiko stroke [4].

Faktor risiko dari penyakit stroke dapat diketahui dari data rekam medis pasien di rumah sakit. Pemanfaatan data rekam medis pasien dapat dilakukan untuk menangani faktor risiko dari penyakit stroke iskemik dengan cara pemanfaatan teknologi menggunakan pemodelan *Machine Learning* yang dapat digunakan untuk mencari variabel mana yang paling berpengaruh dalam memprediksi penanganan penyakit. Pengklasifikasian penyakit stroke iskemik dapat memberikan beberapa manfaat diantaranya yaitu dapat mendeteksi risiko pasien terkena penyakit stroke dengan lebih efektif, prediksi kejadian risiko stroke berulang dan dapat menjadi pengembangan sistem deteksi dini untuk suatu penyakit. Terdapat beberapa cara yang dapat dilakukan untuk memprediksi atau mengklasifikasikan berbagai penyakit dengan membandingkan kinerja teknologi *data mining* [5].

Penerapan teknologi dalam bidang kesehatan sangat dibutuhkan untuk menyediakan alat yang mampu mendiagnosis penyakit dengan berbagai pertimbangan. Dengan demikian, teknik *data mining* dapat dimanfaatkan untuk memprediksi atau mengklasifikasikan suatu penyakit berdasarkan kumpulandata yang tersedia di rumah sakit atau layanan kesehatan lainnya [6]. *Data mining* ialah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam suatu *database*. *Data mining* termasuk dalam proses yang digunakan dalam sebuah teknik statistik, matematika, kecerdasan buatan, dan *machine learning* yang digunakan untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang berkaitan dengan database besar [7]. Ada berbagai metode untuk membangun pohon keputusan, salah satunya adalah menggunakan

algoritma C4.5. Algoritma C4.5 merupakan suatu algoritma yang dapat digunakan untuk melakukan klasifikasi atau pengelompokan pada dataset [8].

Berdasarkan uraian latar belakang diatas maka peneliti tertarik melakukan penelitian dengan judul “Klasifikasi Penyakit Stroke Iskemik Menggunakan Algoritma C4.5 di Rumah Sakit Citra Husada Jember”. Sehingga, penelitian ini memiliki tujuan untuk memperluas pengetahuan tentang penggunaan data mining dalam klasifikasi penyakit stroke iskemik. Dengan mengintegrasikan teknologi komputer dan informasi medis penelitian ini diharapkan dapat memberikan manfaat yang berarti dalam pencegahan, diagnosis dan pengembangan sistem deteksi dini penyakit stroke iskemik. Beberapa variabel yang digunakan untuk melakukan klasifikasi penyakit stroke iskemik di Rumah Sakit Citra Husada adalah usia, jenis kelamin, riwayat hipertensi, riwayat merokok, riwayat diabetes melitus, riwayat penyakit jantung, kolesterol, obesitas, riwayat mengkonsumsi alkohol dan riwayat keluarga terkena stroke.

Metodologi

Jenis Penelitian

Jenis Penelitian yang digunakan oleh peneliti yaitu penelitian kuantitatif yang dimana penelitian ini memiliki tujuan untuk mengungkap gejala dalam konteks holistik dengan mengumpulkan data dari lingkungan alam, dengan menggunakan peneliti sendiri sebagai instrumen kuncinya. Pada penelitian ini data yang diperoleh dari dokumen rekam medis akan diolah dengan *tools* RapidMiner dengan Algoritma C4.5 dikarenakan C4.5 merupakan salah satu solusi untuk pemecahan kasus pada teknik klasifikasi.

Populasi dan Sampel

Populasi dalam penelitian ini yaitu dokumen rekam medis pasien stroke iskemik rawat inap tahun 2022 dan 2023 yang berjumlah 4.123 dokumen. Sampel pada penelitian ini yaitu 366 dokumen rekam medis yang dibagi menjadi 2 kategori yaitu terbagi menjadi kriteria inklusi dan eksklusi. Kriteria inklusi dalam penelitian ini ialah dokumen rekam medis pasien dengan diagnosis stroke iskemik dan Kriteria eksklusi dalam penelitian ini adalah dokumen rekam medis pasien penyakit dalam non kardiovaskuler, diabetes melitus dan kolesterol, sampel tiap masing-masing kategori berjumlah 183.

Teknik dan Instrumen Penelitian

Teknik pengambilan data penelitian ini menggunakan observasi dokumen rekam medis, sedangkan instrumen penelitian yang digunakan adalah lembar *checklist*. Sebelum melakukan proses penelitian lembar *checklist* dalam penelitian ini dilakukan proses validitas kepada tenaga medis di Rumah Sakit Citra Husada Jember.

Variabel Penelitian

Variabel bebas yang digunakan dalam penelitian ini berjumlah 10 yaitu usia, jenis kelamin, riwayat hipertensi, riwayat diabetes melitus, riwayat penyakit jantung, riwayat merokok, riwayat kolestrol, riwayat penggunaan alkohol, riwayat keluarga terkena stroke dan obesitas.

Tabel 1. Atribut, Hasil Ukur dan Keterangan Kategori dalam Klasifikasi Penyakit Stroke Iskemik

Variabel Terikat		
Atribut	Hasil Ukur	Keterangan
Kejadian Stroke Iskemik	Stroke Iskemik	Diagnosa utama stroke iskemik
	Non Stroke Iskemik	Diagnosa utama bukan stroke iskemik
Variabel Bebas		
Atribut	Hasil Ukur	Keterangan
Usia	≥ 55 tahun	Berusia ≥ 55 tahun
	< 55 tahun	Berusia < 55 tahun
Jenis Kelamin	L	Berjenis kelamin laki-laki
	P	Berjenis kelamin perempuan
Riwayat Merokok	Ya	Merokok
	Tidak	Tidak merokok
Riwayat Hipertensi	Ya	Penderita hipertensi
	Tidak	Penderita bukan hipertensi
Riwayat Diabetes Melitus	Ya	Penderita diabetes melitus
	Tidak	Bukan penderita diabetes melitus
Riwayat Penyakit Jantung	Ya	Penderita penyakit jantung
	Tidak	Bukan penderita penyakit jantung
Riwayat Kolesterol	Ya	Penderita kolesterol
	Tidak	Bukan penderita kolesterol
Obesitas	Ya	Penderita obesitas
	Tidak	Bukan penderita obesitas
Riwayat Alkohol	Ya	Penderita mengkonsumsi alkohol
	Tidak	Penderita tidak mengkonsumsi alkohol
Riwayat Keluarga yang terkena stroke	Ya	Ada riwayat keluarga terkena stroke
	Tidak	Tidak ada riwayat keluarga terkena stroke

Pengolahan Data

Pengolahan data dilakukan pada penelitian ini menggunakan metode *Knowledge Discovery in Database* yaitu :

- Data selection* yaitu proses pengumpulan data stroke iskemik dan penyakit dalam non kardiovaskuler, diabetes melitus dan kolesterol yang didapatkan dari observasi dokumen rekam medis.
- Pre-processing* yaitu proses pengolahan data dengan mengisi data yang kosong (missing value), inkonsistensi data dan menghapus data yang terjadi duplikasi menggunakan bantuan *tools RapidMiner*.
- Transformation* yaitu proses dimana data diklasifikasikan berdasarkan kategori yang sesuai dan disimpan ke dalam format .xls, .csv, .arf dan lainnya untuk proses menggunakan *tools rapidminer*.
- Data Mining* yaitu proses data *mining* menggunakan algoritma C4.5
- Evaluation* yaitu proses evaluasi dan interpretasi menggunakan beberapa teknik yang terdapat pada *tools RapidMiner*.

Hasil dan Pembahasan

Tahapan Data Mining

Data Selection

Penelitian ini dimulai dengan melakukan observasi data rekam medis kemudian dilakukan pengumpulan data rekam medis pasien rawat inap stroke iskemik dan pasien rawat inap penyakit dalam non kardiovaskuler, diabetes melitus dan kolesterol tahun 2022 dan 2023 yang berjumlah 366 data rekam medis.

Pre-processing

Setelah data terkumpul, maka dilakukan tahapan *preprocessing* data. *Preprocessing* data melibatkan pembersihan data dengan menghapus informasi duplikat, mengidentifikasi ketidakkonsistenan data, dan memperbaiki kesalahan data. Langkah ini diperlukan sebelum penambangan data dapat dilakukan [9]. Pada tahap ini dilakukan penghapusan 2 variabel yaitu riwayat mengkonsumsi alkohol dikarenakan memiliki kondisi yang sama dan riwayat keluarga terkena stroke dikarenakan tidak ditemukan riwayat keluarga yang terisi pada dokumen rekam medis. Setelah itu dilakukan proses pengisian *missing value* menggunakan *tools* RapidMiner.

Transformation

Tahap selanjutnya yaitu *transformation*. *Transformation* adalah langkah di mana data diubah menjadi bentuk yang sesuai dengan kebutuhan [10]. Pada tahap ini dilakukan perubahan data atau proses inialisasi pada variabel usia.

Data Mining

Tahap selanjutnya yaitu proses data *mining*. Data *mining* adalah proses yang digunakan untuk mengidentifikasi pengetahuan potensial yang berguna dan bermanfaat dari database besar dengan menggunakan teknik statistik, matematika, kecerdasan buatan, dan machine learning [10]. Pada tahap ini dilakukan proses data *mining* dengan algoritma C4.5. menggunakan *tools* RapidMiner.

Evaluation

Tahapan terakhir dari proses data *mining* ialah *evaluation*. Pada tahap ini dilakukan untuk mengevaluasi hasil penelitian dari proses perhitungan algoritma C4.5 menggunakan *confusion matrix*, sehingga didapatkan nilai *accuracy*, *precision* dan *recall* untuk perbandingan rasio data.

Mengidentifikasi karakteristik penyakit stroke iskemik yang digunakan sebagai variabel untuk mengklasifikasikan penyakit stroke iskemik di Rumah Sakit Citra Husada Jember

Identifikasi Variabel Jenis Kelamin

Tabel 2. Distribusi Frekuensi Jenis Kelamin

Jenis Kelamin	Stroke Iskemik		Non Stroke Iskemik	
	Jumlah	%	Jumlah	%

Laki-Laki	81	44,26	87	47,54
Perempuan	102	55,74	96	52,46
Total	183	100	183	100

Hasil dari tabel 3.1 dapat diambil kesimpulan bahwa pasien stroke iskemik tertinggi pada pasien berjenis kelamin perempuan dengan presentase 55,74%. Hal ini dikarenakan mayoritas perempuan yang sudah memasuki masa menopause akan menyebabkan penurunan hormon estrogen. Hormon estrogen berperan dalam melindungi pembuluh darah dari aterosklerosis dengan cara meningkatkan HDL dan menurunkan LDL. Oleh karena itu, pada masa menopause, proteksi terhadap proses aterosklerosis tidak lagi ada [11].

Identifikasi Variabel Usia

Tabel 3. Distribusi Frekuensi Usia

Usia	Stroke Iskemik		Non Stroke Iskemik	
	Jumlah	%	Jumlah	%
≥ 55 tahun	117	63,93	60	32,79
< 55 tahun	66	36,07	123	67,21
Total	183	100	183	100

Hasil dari tabel 3.2 menunjukkan bahwa pasien stroke iskemik dengan kategori usia ≥ 55 tahun memiliki kasus sebanyak 63,93%. Hal tersebut terjadi karena seseorang berusia di atas 55 tahun memiliki peluang 3,23 kali lebih besar untuk menderita stroke dibandingkan dengan mereka yang berusia di bawah 55 tahun. Risiko kejadian stroke cenderung meningkat seiring dengan bertambahnya usia, dan peluang terjadinya stroke akan meningkat dua kali lipat setelah usia 55 tahun [4].

Identifikasi Variabel Riwayat Merokok

Tabel 4. Distribusi Frekuensi Riwayat Merokok

Riwayat Merokok	Stroke Iskemik		Non Stroke Iskemik	
	Jumlah	%	Jumlah	%
Iya	67	36,61	62	33,88
Tidak	116	63,39	121	66,12
Total	183	100	183	100

Hasil dari tabel 3.3 menunjukkan bahwa pasien stroke iskemik dengan riwayat merokok memiliki kasus sebanyak 36,61%. Dari 132 pasien yang memiliki riwayat merokok 67 diantaranya menderita penyakit stroke. Hal ini dikarenakan kebiasaan merokok merupakan kebiasaan negatif yang dapat menyebabkan penyakit, karena setiap batang rokok mengandung lebih dari 4000 zat, dengan 2000 di antaranya memiliki efek buruk bagi kesehatan tubuh, termasuk bahan radioaktif serta zat-zat yang digunakan dalam cat, pembersih lantai, kapur barus, gas beracun, dan lainnya [12].

Identifikasi Variabel Riwayat Hipertensi

Tabel 5. Distribusi Frekuensi Riwayat Hipertensi

Riwayat Hipertensi	Stroke Iskemik		Non Stroke Iskemik	
	Jumlah	%	Jumlah	%
Iya	176	96,17	53	28,96

Tidak	7	3,83	130	71,04
Total	183	100	183	100

Hasil dari tabel 3.4 menunjukkan bahwa pasien stroke iskemik dengan riwayat hipertensi memiliki kasus sebanyak 96,17%. Hal ini dikarenakan hipertensi merupakan penyebab utama terjadinya stroke. Sehingga tekanan darah yang tidak normal mengakibatkan kerusakan sel-sel endotel pembuluh darah, yang menimbulkan jejas pada rongga vaskuler sehingga menyebabkan penyempitan pembuluh darah [13].

Identifikasi Variabel Riwayat Diabetes Melitus

Tabel 6. Distribusi Frekuensi Riwayat Diabetes Melitus

Riwayat Diabetes Mellitus	Stroke Iskemik		Non Stroke Iskemik	
	Jumlah	%	Jumlah	%
Iya	82	44,81	11	6,01
Tidak	101	55,19	172	93,99
Total	183	100	183	100

Hasil dari tabel 3.5 dapat dilihat 93 dari seseorang yang memiliki riwayat diabetes melitus 82 diantaranya menderita penyakit stroke iskemik. Hal ini dikarenakan diabetes mellitus dapat meningkatkan risiko stroke karena kelebihan glukosa dalam darah dapat menyebabkan vaskulopati, yang membuat seseorang lebih rentan terhadap hipertensi dan aterosklerosis [14].

Identifikasi Variabel Riwayat Penyakit Jantung

Tabel 7. Distribusi Frekuensi Riwayat Penyakit Jantung

Riwayat Penyakit Jantung	Stroke Iskemik		Non Stroke Iskemik	
	Jumlah	%	Jumlah	%
Iya	40	21,86	1	0,55
Tidak	143	78,14	182	99,45
Total	183	100	183	100

Hasil dari tabel 3.6 menunjukkan bahwa pasien dengan riwayat penyakit jantung sebanyak 41 orang 40 diantaranya terkena penyakit stroke iskemik. Hal ini dikarenakan seseorang yang memiliki riwayat penyakit jantung memiliki risiko 5,440 kali lebih besar untuk mengalami stroke iskemik [15]

Identifikasi Variabel Riwayat Kolesterol

Tabel 8. Distribusi Frekuensi Riwayat Kolesterol

Riwayat Kolesterol	Stroke Iskemik		Non Stroke Iskemik	
	Jumlah	%	Jumlah	%
Iya	153	83,60	182	99,45
Tidak	30	16,40	1	0,55
Total	183	100	183	100

Hasil dari tabel 3.7 menunjukkan bahwa 83,60% pasien stroke iskemik memiliki riwayat kolesterol. Hal ini terjadi karena seseorang yang memiliki riwayat kolesterol beresiko terkena penyakit stroke, dikarenakan peningkatan kadar LDL mengakibatkan semakin banyaknya pembentukan plak aterosklerosis, yang lama-kelamaan akan menjadi keras dan sulit mengalami regresi. Hal ini menyebabkan disfungsi pada aliran pembuluh darah, yang

berkontribusi pada peningkatan angka kesakitan setelah serangan stroke serta memperpanjang lamanya perawatan [16].

Identifikasi Variabel Obesitas

Tabel 9. Distribusi Frekuensi Obesitas

Obesitas	Stroke Iskemik		Non Stroke Iskemik	
	Jumlah	%	Jumlah	%
Iya	16	8,74	2	1,09
Tidak	167	91,26	181	98,91
Total	183	100	183	100

Hasil dari tabel 4.10 menunjukkan bahwa dari 18 orang yang menderita obesitas 16 diantaranya terkena penyakit stroke iskemik. Hal ini dikarenakan Obesitas atau kegemukan menjadi salah satu faktor yang menyebabkan stroke. Hal ini diakibatkan oleh timbunan lemak pada tubuh yang dapat menyebabkan sumbatan pembuluh darah. Jika tidak segera ditangani, sumbatan tersebut dapat menyebabkan pecahnya pembuluh darah serebral, yang pada akhirnya mengakibatkan stroke [17].

Mengklasifikasikan stroke iskemik dan menentukan variabel yang paling berpengaruh berdasarkan klasifikasi penyakit stroke iskemik di Rumah Sakit Citra Husada Jember menggunakan algoritma C4.5

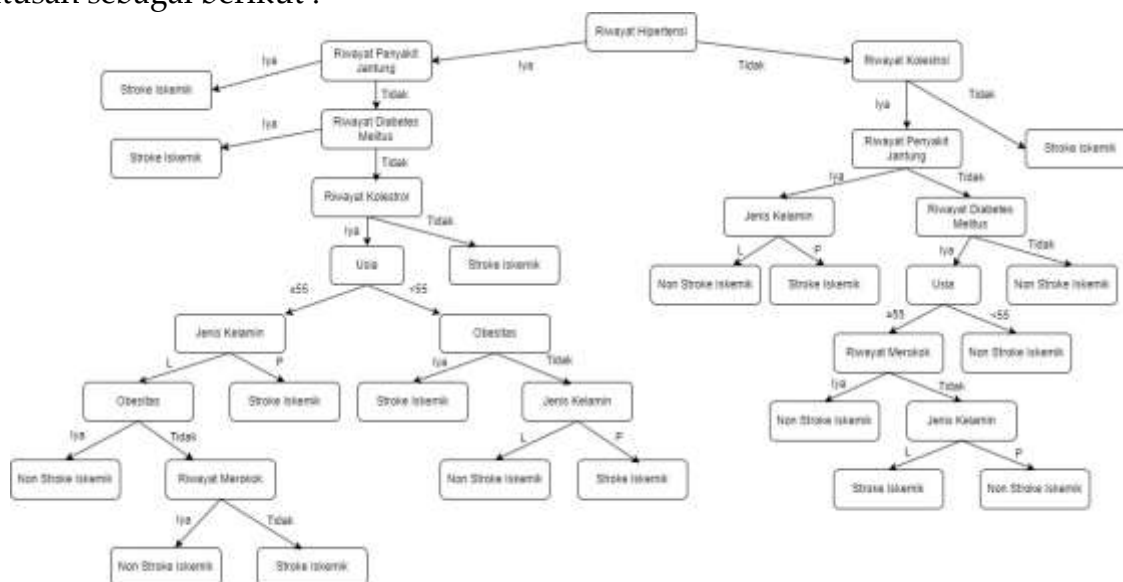
Tabel 10. Perhitungan Algoritma C4.5

	Jumlah	Stroke Iskemik	Non Stroke Iskemik	Entropy	Gain	Split Infor	Gain Ratio
Total	330	164	166	1			
Jenis Kelamin					0,003	0,997	0,003
Laki-Laki	168	81	87	0,998			
Perempuan	198	102	96	0,996			
Usia					0,054	1	0,054
≥ 55 tahun	177	117	60	0,950			
< 55 tahun	189	66	123	0,941			
Riwayat Merokok					0,000	0,935	0,000
Iya	129	67	62	1			
Tidak	237	116	121	1			
Riwayat Hipertensi					0,400	0,959	0,417
Iya	229	176	53	0,779			
Tidak	137	7	130	0,310			
Riwayat Diabetes Melitus					0,149	0,818	0,182
Iya	93	82	11	0,560			
Tidak	273	101	172	0,951			
Riwayat Penyakit Jantung					0,102	0,506	0,202
Iya	41	40	1	0,179			
Tidak	325	143	182	0,988			
Riwayat Kolesterol					0,072	0,419	0,171
Iya	335	153	182	0,994			
Tidak	31	30	1	0,222			
Riwayat Obesitas					0,023	0,280	0,084

	Jumlah	Stroke Iskemik	Non Stroke Iskemik	Entropy	Gain	Split Infor	Gain Ratio
Iya	18	16	2	0,544			
Tidak	348	167	181	0,999			

Hasil dari perhitungan pada tabel 3.9 didapatkan nilai *gain ratio* tertinggi pada variabel riwayat hipertensi yaitu sebesar 0,417 sehingga variabel tersebut dijadikan *root*. Akar pohon keputusan tersebut merupakan variabel yang memiliki pengaruh paling besar untuk menentukan bahwa seseorang terdiagnosa penyakit Stroke Iskemik. Berdasarkan data yang didapat pada Rumah Sakit Citra Husada penyakit hipertensi masuk dalam kategori 10 besar penyakit rawat jalan dan masuk pada peringkat ke-3 dengan jumlah 5.148 pasien terkena hipertensi. Hipertensi merupakan faktor risiko utama untuk stroke, karena menimbulkan ketegangan berlebihan pada pembuluh darah, membuatnya menebal dan memburuk, yang pada akhirnya bisa menyebabkan stroke. Ketika pembuluh darah menebal disertai dengan tekanan darah yang meningkat, kolesterol dan zat lemak lainnya bisa merusak dinding arteri dan menyebabkan penyumbatan di arteri otak [18].

Setelah akar pertama pohon keputusan didapatkan, maka dilakukan langkah selanjutnya untuk menentukan cabang-cabang dari pohon keputusan tersebut. Perhitungan yang dilakukan sama dengan cara sebelumnya yaitu dengan menghitung nilai *entropy*, *gain*, *split info*, dan *gain ratio* pada atribut yang tersisa yaitu jenis kelamin, usia, riwayat merokok, riwayat diabetes melitus, riwayat penyakit jantung, riwayat kolesterol dan obesitas. Dari perhitungan yang dilakukan dengan bantuan *tools* RapidMiner maka didapatkan pohon keputusan sebagai berikut :



Gambar 2. Pohon Keputusan Algoritma C4.5

Menganalisis hasil *confusion matrix* dari klasifikasi penyakit stroke iskemik menggunakan metode algoritma C4.5

Tabel 11. Perbandingan Performa Klasifikasi

Training	Rasio Testing	<i>Accuracy</i>	<i>Precision</i>	<i>Recall</i>
10	90	83,33	76,33	96,34
20	80	84,30	84,72	83,56
30	70	85,15	83,45	87,40
40	60	85,84	86,79	84,40
50	50	87,43	86,17	89,01
60	40	87,76	88,73	86,30
70	30	87,27	88,68	85,45
80	20	89,04	88,89	88,89
90	10	97,22	94,73	100

Dalam proses data *mining* menggunakan algoritma C4.5 pembagian data menggunakan operator split data dengan memilih operator *Stratified Sampling* yang dimana pengambilan sampel yang dilakukan yaitu secara acak, namun memastikan bahwa distribusi kelas dalam setiap subset sama dengan distribusi dataset asli. Oleh karena itu didapatkan hasil perhitungan dari tabel *confusion matrix* yang berupa perhitungan *accuracy* pada dataset Stroke Iskemik dengan jumlah data sebanyak 366 menggunakan algoritma C4.5 dengan split rasio data 10:90 sampai dengan 90:10 didapatkan hasil bahwa perbandingan 90:10 memiliki nilai *accuracy* yang tinggi yaitu mencapai 97,22%. Nilai akurasi menunjukkan seberapa tepat model dalam membuat klasifikasi yang benar [19]. Hal ini menunjukkan bahwa nilai *accuracy* dapat dikategorikan klasifikasi penyakit stroke iskemik menggunakan algoritma C4.5 memiliki hasil kategori yang baik (*good classification*).

Hasil perhitungan dari tabel *confusion matrix* yang berupa perhitungan *precision* pada dataset Stroke Iskemik dengan jumlah data sebanyak 366 menggunakan algoritma C4.5 dengan split rasio data 10:90 sampai dengan 90:10 didapatkan hasil bahwa perbandingan 90:10 memiliki nilai *precision* yang tinggi yaitu mencapai 94,73%. Nilai *precision* adalah ukuran ketepatan sistem dalam mengidentifikasi informasi yang benar, menunjukkan sejauh mana data yang diklasifikasikan sebagai positif atau negatif benar-benar akurat [19]. Selain itu, *precision* juga menggambarkan keakuratan data yang diinginkan dibandingkan dengan hasil prediksi yang dihasilkan oleh model klasifikasi [20].

Hasil perhitungan dari tabel *confusion matrix* yang berupa perhitungan *recall* pada dataset Stroke Iskemik dengan jumlah data sebanyak 366 menggunakan algoritma C4.5 dengan split rasio data 10:90 sampai dengan 90:10 didapatkan hasil bahwa perbandingan 90:10 memiliki nilai *recall* yang tinggi yaitu mencapai 100%. Nilai *recall* adalah ukuran yang menunjukkan tingkat keberhasilan dalam mengidentifikasi secara benar informasi tentang data yang termasuk dalam kelas positif atau negative [21]. Selain itu, *recall* juga menggambarkan bahwa keberhasilan suatu model dalam menemukan kembali informasi yang benar [19]. Meskipun hasil akurasi ini sangat tinggi algoritma C4.5 juga memiliki kekurangan yaitu pohon keputusan yang dihasilkan sering kali kompleks dan sulit diinterpretasikan, serta algoritma ini sensitif terhadap data yang tidak seimbang dan mengandung *noise*. Oleh karena itu, peneliti selanjutnya diharapkan mampu membandingkan dengan penggunaan metode lainnya.

Simpulan

- a. Identifikasi variabel Stroke Iskemik berdasarkan dokumen rekam medis pasien rawat inap tahun 2022 dan 2023 di Rumah Sakit Citra Husada Jember yang berjumlah 183 data, diketahui bahwa sebagian besar pasien Stroke Iskemik berjenis kelamin perempuan sebesar 55,74%; berusia lebih dari 55 tahun sebesar 63,93%; riwayat merokok 36,61%; riwayat hipertensi 96,17%; riwayat diabetes melitus 44,81%; riwayat penyakit jantung 21,86%; riwayat kolesterol 83,60%; obesitas 8,74%.
- b. Berdasarkan hasil klasifikasi dan perhitungan nilai *entropy*, *gain information*, *split information*, dan *gain ratio* didapatkan bahwa variabel riwayat hipertensi memiliki nilai *gain ratio* tertinggi yaitu 0,422 dan dijadikan sebagai akar pohon keputusan atau variabel yang paling berpengaruh terhadap penyakit stroke iskemik.
- c. Hasil dari perhitungan *confusion matrix* dapat diketahui bahwa nilai *accuracy*, *precision*, dan *recall* tertinggi yaitu pada split rasio 90:10. Dimana nilai *accuracy* sebesar 97,22%, nilai *precision* sebesar 94,73% dan nilai *recall* yaitu 100%.

Referensi

- A. Byna and M. Basit, "Penerapan Metode Adaboost Untuk Mengoptimasi Prediksi Penyakit Stroke Dengan Algoritma Naïve Bayes," *J. Sisfokom (Sistem Informasi dan Komputer)*, vol. 9, no. 3, pp. 407–411, 2020. doi: <https://doi.org/10.32736/sisfokom.v9i3.1023>
- A. Veronica Agustin and A. Voutama, "Implementasi Data Mining Klasifikasi Penyakit Diabetes Pada Perempuan Menggunakan Naïve Bayes," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 7, no. 2, pp. 1002–1007, 2023. doi: <https://doi.org/10.36040/jati.v7i2.6808>
- D. Hisni, M. E. Saputri, and S. Sujarni, "Faktor-Faktor Yang Berhubungan Dengan Kejadian Stroke Iskemik Di Instalasi Fisioterapi Rumah Sakit Pluit Jakarta Utara Periode Tahun 2021," *Jurnal Penelitian Keperawatan Kontemporer*, vol. 2, no. 1, pp. 140–149, 2022. doi: <https://doi.org/10.59894/jpkk.v2i1.333>
- D. W. Nugraha, A. Y. E. Dodu, and N. Chandra, "Klasifikasi Penyakit Stroke Menggunakan Metode Naive Bayes Classifier (Studi Kasus Pada Rumah Sakit Umum Daerah Undata Palu)," vol. 3, no. 2, pp. 13–22, 2017.
- F. Alghifari and D. Juardi, "Penerapan Data Mining Pada Penjualan Makanan Dan Minuman Menggunakan Metode Algoritma Naïve Bayes," *Jurnal Ilmiah Informatika*, vol. 9, no. 2, pp. 75–81, 2021. doi: <https://doi.org/10.33884/jif.v9i02.3755>
- I. Indriasari, S. Sahreni, and A. A. F. Pratama, "Hubungan Diabetes Melitus Tipe 2 Dengan Stroke Iskemik Pada Pasien Poliklinik Saraf Di Rumah Sakit Budi Kemuliaan Kota

- Batam Tahun 2021," *Zona Kedokteran: Program Studi Pendidikan Dokter Universitas Batam*, vol. 13, no. 1, pp. 342–350, 2023. **doi:** <https://doi.org/10.37776/zked.v13i1.1153>
- M. Azhari, Z. Situmorang, and R. Rosnelly, "Perbandingan Akurasi, Recall, dan Presisi Klasifikasi pada Algoritma C4.5, Random Forest, SVM dan Naive Bayes," *Jurnal Media Informatika Budidarma*, vol. 5, no. 2, p. 640, 2021. **doi:** <https://doi.org/10.30865/mib.v5i2.2937>
- M. A. Riondi, E. Kristanti, and F. The, "Hubungan Faktor Risiko dengan Luaran Pasien Stroke Iskemik di RSUD Dr. H. Chasan Boesoirie," *Jurnal Ilmiah Indonesia*, vol. 4, no. 1, pp. 30–38, 2024.
- M. Halimatunnisa' *et al.*, "Faktor Risiko Stroke di Rumah Sakit Umum Daerah Praya Tahun 2022," *Jurnal Kesehatan Qamarul Huda*, vol. 11, no. 1, pp. 371–381, 2023. **doi:** <https://doi.org/10.37824/jkqh.v11i1.2023.507>
- N. M. Daulay, N. D. Siregar, E. Sujoko, F. A. Simamora, and N. Harahap, "Faktor Faktor Yang Berhubungan Dengan Kejadian Stroke Di Rumah Sakit Umum Daerah Kota Padangsidimpuan Tahun 2022," *Jurnal Kesehatan Ilmiah Indonesia (Indonesian Health Science Journal)*, vol. 8, no. 1, pp. 106–116, 2023. **doi:** <https://doi.org/10.51933/health.v8i1.1035>
- P. Ayudia and Y. Imran, "Kadar Kolesterol LDL Sebagai Prediktor Lama Perawatan Pada Pasien Stroke Iskemik Akut," *Jurnal Penelitian dan Karya Ilmiah Lembaga Penelitian Universitas Trisakti*, vol. 8, no. 2, pp. 310–320, 2023. **doi:** <https://doi.org/10.25105/pdk.v8i2.15710>
- P. N. Puspitasari, "Hubungan Hipertensi Terhadap Kejadian Stroke," *Jurnal Ilmiah Kesehatan Sandi Husada*, vol. 12, no. 2, pp. 922–926, 2020. **doi:** <https://doi.org/10.35816/jiskh.v12i2.435>
- P. Romadloni, B. Adhi Kusuma, and W. Maulana Baihaqi, "Komparasi Metode Pembelajaran Mesin Untuk Implementasi Pengambilan Keputusan Dalam Menentukan Promosi Jabatan Karyawan," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 6, no. 2, pp. 622–628, 2022. **doi:** <https://doi.org/10.36040/jati.v6i2.5238>
- Pusdatin Kemkes, *Stroke Don't Be The One*. Pusat Data dan Informasi Kementerian Kesehatan RI, 2019.
- R. E. Pambudi, Sriyanto, and Firmansyah, "Klasifikasi Penyakit Stroke Menggunakan Algoritma Decision Tree C4.5," *Jurnal Teknik*, vol. 16, no. 2, pp. 221–226, 2022.
- R. S. Wahono, *Data Mining*, vol. 2, 2023. Available: https://www.cambridge.org/core/product/identifier/CBO9781139058452A007/type/book_part

-
- S. Clara, D. Laksmi Prianto, R. Al Habsi, E. Friscila Lumbantobing, and N. Chamidah, "Implementasi Seleksi Fitur Pada Algoritma Klasifikasi Machine Learning Untuk Prediksi Penghasilan Pada Adult Income Dataset," *Seminar Nasional Mahasiswa Ilmu Komputer dan Aplikasinya (SENAMIKA)*, vol. 2, no. 1, pp. 741–747, 2021.
- S. Saharman and E. Winarto, "Hubungan Hipertensi dengan Kejadian Stroke di RSUD Kotamobagu," vol. 14, no. 1, pp. 370–374, 2023.
- S. Ucha Putri, E. Irawan, and F. Rizky, "Implementasi Data Mining Untuk Prediksi Penyakit Diabetes Dengan Algoritma C4.5," vol. 2, no. 1, pp. 39–46, 2021.
- V. Azzahra and S. Ronoatmodjo, "Faktor-faktor yang Berhubungan dengan Kejadian Stroke pada Penduduk Usia ≥ 15 Tahun di Provinsi Daerah Istimewa Yogyakarta (Analisis Data Riskesdas 2018)," *Jurnal Epidemiologi Kesehatan Indonesia*, vol. 6, no. 2, 2023. doi: <https://doi.org/10.7454/epidkes.v6i2.6508>
- Zuriati and Qomariyah, "Klasifikasi Penyakit Stroke Menggunakan Algoritma K-Nearest Neighbor (KNN)," *Jurnal Sistem dan Teknologi Informasi*, vol. 1, no. 1, pp. 1–8, 2023.