

PENERAPAN ALGORITMA DECISION TREE DALAM KLASIFIKASI PENYAKIT STROKE OTAK

Hendriyansyah, Ade Irma Purnamasari, Tati Suprapti

Teknik Informatika, STMIK IKMI Cirebon

Jl. Perjuangan No.10B, Karyamulya, Kec. Kesambi, Kota Cirebon, Jawa Barat

hendriyansyah564@gmail.com, irma2974@yahoo.com

ABSTRAK

Stroke adalah gangguan kesehatan serius yang memiliki tingkat kematian tinggi, sehingga penanganan cepat sangat penting untuk mengurangi risiko komplikasi dan kerusakan otak. Studi ini menggunakan pendekatan Knowledge Discovery in Databases dan mengimplementasikan algoritma decision tree untuk melakukan klasifikasi penyakit stroke. Hasilnya menunjukkan tingkat akurasi sebesar 85.81%, recall sebesar 86.12%, dan presisi sebesar 85.85%. Dengan demikian, langkah-langkah pencegahan dapat diimplementasikan lebih efektif, membantu mengurangi dampak stroke pada tingkat kesehatan masyarakat.

Kata kunci : *Decision tree, Klasifikasi, Penyakit stroke otak, Knowledge Discovery in Databases (KDD).*

1. PENDAHULUAN

Stroke adalah suatu gangguan kesehatan yang dapat berdampak dalam jangka panjang dan merupakan salah satu penyakit dengan risiko kematian tertinggi. Risiko tersebut semakin meningkat, terutama jika seseorang juga mengalami gangguan pada jantung, yang dapat memperburuk kondisinya [1]. Gangguan pada aliran darah ke otak dapat menyebabkan stroke. Hal ini dapat terjadi karena stroke hemoragik, dimana pembuluh darah di otak pecah, atau stroke iskemik, dimana pembuluh darah yang mengantarkan darah ke otak tersumbat. Kondisi ini mengakibatkan terganggunya pasokan darah dan oksigen ke sel-sel otak, menyebabkan kerusakan pada jaringan otak yang dapat berakibat fatal atau menimbulkan dampak yang signifikan terhadap fungsi tubuh.

Tujuan penelitian ini adalah untuk mengembangkan dan mengimplementasikan teknik klasifikasi yang dapat memprediksi risiko seseorang terkena penyakit stroke dengan akurasi tinggi. Dengan demikian, penelitian ini bertujuan untuk memberikan informasi yang berharga bagi praktisi kesehatan dalam melakukan tindakan pencegahan atau intervensi dini. Langkah-langkah ini diharapkan dapat meningkatkan efektivitas dan tepat waktu dalam penanganan penyakit stroke, serta mengurangi dampaknya pada tingkat kesehatan masyarakat secara keseluruhan.

Dalam penelitian sebelumnya yang berkaitan dengan klasifikasi, Fida Maisa Hana telah melakukan penelitian yang membahas penggunaan metode algoritma decision tree untuk mengklasifikasikan individu yang mengalami penyakit diabetes dan individu yang sehat. Pengujian yang dilakukan menghasilkan tingkat Akurasi yang tinggi mencapai 97,12%, dengan tingkat Precision sebesar 93,02%, dan tingkat Recall sebesar 100,00%, yang menunjukkan bahwa hasil klasifikasinya sangat baik [2]. Dalam penelitian lain yang dilakukan oleh Asmaul Husnah Nasrullah, metode decision tree digunakan untuk mengklasifikasikan produk laris dengan menggunakan

data privat, yaitu data stok dan penjualan barang di PT Cipta Karya Gorontalo. Hasil penelitian menunjukkan bahwa akurasi model klasifikasi produk laris menggunakan Decision Tree mencapai 90%, dan nilai AUC sebesar 0.709, yang dapat dikategorikan sebagai klasifikasi yang baik [3]. Penelitian oleh Pambudi membahas tentang stroke sebagai gangguan fungsi otak yang serius di Indonesia. Deteksi dini gejala stroke menjadi fokus penelitian dengan menganalisis atribut seperti usia, jenis kelamin, kadar glukosa, riwayat penyakit jantung, hipertensi, tipe pekerjaan, tipe tempat tinggal, status merokok, index massa tubuh, dan status pernikahan. Penelitian menggunakan algoritma Decision Tree C4.5, mencapai akurasi sebesar 99.07%. Kesimpulannya, penelitian ini membantu memahami faktor-faktor penyebab stroke dan menyoroti keefektifan algoritma Decision Tree C4.5 dalam menganalisis data [4]. Pada penelitian yang dilakukan oleh fauziah dan zulfikar tahun 2023 membahas mengenai analisis risiko penyakit jantung dengan menggunakan metode klasifikasi, khususnya algoritma iterative dichotomiser 3 (ID3). Penelitian ini dilakukan dengan memanfaatkan dataset dari UCI Machine Learning Repository yang mencakup berbagai atribut terkait kondisi pasien. Evaluasi performa dilakukan dengan menggunakan metrik dari confusion matrix, yang menghasilkan tingkat akurasi, presisi, dan recall yang cukup tinggi. Kesimpulannya bahwa dengan penerapan metode decision tree menggunakan algoritma ID3 dapat mempercepat pengambilan keputusan dalam mengklasifikasikan risiko penyakit jantung dengan akurasi sebesar 85, 71%, presisi sebesar 84, 62% dan recal sebesar 84, 62% [5]. Penelitian oleh Wiratama & Pradnya membahas tentang diabetes mellitus (DM) sebagai penyakit kronis dengan jumlah penderita yang tinggi di seluruh dunia, termasuk Indonesia. Keterlambatan dalam diagnosis DM dapat meningkatkan angka kematian, oleh itu penting untuk deteksi dini. Penelitian ini membandingkan tiga algoritma (KNN, Naïve Bayes, dan C4.5) untuk deteksi DM dengan

optimalisasi menggunakan metode backward elimination. Hasilnya menunjukkan bahwa algoritma KNN yang dioptimasi memiliki akurasi tertinggi (97,6%), sedangkan algoritma C4.5 yang dioptimasi memiliki nilai AUC tertinggi (0,988) untuk deteksi dini DM [6].

Penelitian ini menerapkan algoritma decision tree untuk mengklasifikasikan penyakit stroke otak. Pendekatan ini telah sukses dalam penelitian sebelumnya terkait diabetes dan produk laris. Misalnya, penelitian Fida Maisa Hana mencapai akurasi 97,12% dalam mengklasifikasikan diabetes [2]. Begitu pula, Asmaul Husnah Nasrullah menggunakan decision tree untuk mengklasifikasikan produk laris dengan akurasi 90% [3]. Dengan fokus pada penyakit stroke, metode ini diharapkan memberikan prediksi yang akurat untuk mendukung penanganan dini. Hasil penelitian ini diharapkan dapat memperkuat langkah-langkah pencegahan dan intervensi, meningkatkan efektivitas penanganan penyakit stroke otak secara keseluruhan.

2. TINJAUAN PUSTAKA

2.1. Penyakit Stroke Otak

Stroke adalah gangguan fungsi sistem saraf pusat (SSP) yang disebabkan oleh terganggunya aliran darah normal ke otak. Stroke dibagi menjadi dua kategori utama: stroke iskemik dan stroke hemoragik.

Stroke iskemik meliputi stroke trombotik dan stroke emboli. Diperkirakan 85% dari seluruh stroke adalah stroke iskemik. Otak sangat bergantung pada oksigen dan glukosa untuk metabolisme. Ketika aliran darah ke area tertentu dari otak terhambat atau berkurang, itu disebut stroke iskemik. Trombosis dan emboli adalah dua penyebab menurunnya aliran darah ke otak. Berkurangnya aliran darah merusak fungsi neurologis dan lambat laun menyebabkan iskemia permanen. Pada stroke iskemik, penurunan aliran darah ke otak menyebabkan hipoksemia di area tertentu di otak, memicu reaksi berantai yang menyebabkan kematian sel otak. Hal ini menyebabkan masalah fisik, psikologis, dan sosial bagi pasien. Masalah fisik yang terjadi pada pasien stroke adalah kerusakan saraf. Di antara gejala klinis stroke adalah gangguan motorik, gangguan komunikasi verbal, gangguan sensorik, gangguan kognitif dan psikologis, dan disfungsi kandung kemih. Neuropati yang terjadi bergantung pada lokasi dan ukuran lesi [9].

2.2. Data Mining

Serangkaian prosedur yang disebut data mining, bertujuan untuk otomatis menghasilkan nilai tambahan dari database, mengungkap informasi yang tidak diketahui secara manual. Proses ini melibatkan ekstraksi data dari database dan pengenalan pola-pola yang signifikan atau menarik. Hasil informasi yang dihasilkan berasal dari analisis data yang terdapat dalam database, memungkinkan identifikasi wawasan yang tidak terlihat secara langsung. Dengan demikian, data mining memanfaatkan teknik komputasional

untuk menggali informasi berharga dan mengidentifikasi hubungan atau pola yang dapat memberikan pemahaman lebih mendalam terhadap data yang ada [8].

2.3. Klasifikasi

Metode klasifikasi adalah untuk mendefinisikan dan menunjukkan perbedaan antara satu kategori informasi dan yang lain untuk menentukan apakah suatu objek memiliki perilaku dan karakteristik kelompok tertentu. Selain itu, klasifikasi menunjukkan hubungan antar atribut [10].

2.4. Decision Tree

Pohon keputusan merupakan output dari perhitungan entropi dan perolehan informasi setiap kali atribut pada pohon mendapatkan kelas. Proses ini melibatkan iterasi berulang, di mana setiap langkah melibatkan pemilihan atribut terbaik berdasarkan entropi dan perolehan informasi. Langkah-langkah ini terus diulang hingga tidak ada lagi operasi komputasi yang dapat dilakukan, dan setiap cabang pohon memiliki kelas yang ditetapkan. Dengan demikian, pohon keputusan secara iteratif menyusun aturan keputusan berdasarkan atribut-atribut yang signifikan dalam memprediksi hasil kelas [7].

2.5. Confusion Matrix

Metode evaluasi Confusion Matrix menggunakan tabel untuk menilai kinerja model. Dataset yang terdiri dari dua kelas dianggap positif dalam matriks ini, sedangkan yang lain dianggap negatif. Menghitung nilai Akurasi, Recall, dan Presisi dibantu oleh Confusion Matrix, yang menunjukkan seberapa baik kinerja model dalam memprediksi kedua kelas tersebut.

Tabel 1 Confusion Matrix

		Kelas Sebenarnya	
		True	False
Kelas Predikdi	True	TP	FP
	False	FN	TN

Tabel 1, yang juga dikenal sebagai Confusion Matrix, merupakan alat evaluasi yang berguna dalam mengklasifikasikan model atau algoritma. True Positive (TP), True Negative (TN), False Positive (FP), dan False Negative (FN) adalah empat elemen utama Confusion Matrix, yang dapat dijelaskan sebagai berikut.

- a. True Positive (TP): Jumlah data yang telah diprediksi dan benar-benar Positif.
- b. False Positive (FP): Sebuah jumlah data yang sebenarnya Negatif, tetapi diduga Positif secara salah.
- c. True Negative (TN): jumlah data yang benar-benar Negatif dan telah diprediksi sebagai Negatif.
- d. False Negative (FN): Sebuah jumlah data yang sebenarnya Positif, tetapi diprediksi secara salah sebagai Negatif..

Penilaian menggunakan Confusion Matrix memberikan nilai Akurasi, Recall, Presisi. Perhitungan dapat dijelaskan sebagai berikut:

a. Accuracy

Merupakan rasio prediksi yang benar dengan data keseluruhan.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

b. Recall

merupakan rasio prediksi benar positif dibandingkan dengan jumlah data yang benar positif secara keseluruhan.

$$Recall = \frac{TP}{TP + FN}$$

c. Precision

Merupakan rasio prediksi benar yang positif dibandingkan dengan hasil prediksi keseluruhan yang positif.

$$Precision = \frac{TP}{TP + FP}$$

Dalam penelitian yang dilakukan oleh Ucha Putri membahas mengenai masalah global diabetes dengan sekitar 120 juta penderitanya. Dengan menggunakan Algoritma Data Mining C4.5, tujuan penelitian adalah untuk membuat model prediksi yang mengambil data seperti berat badan, usia, tekanan darah, denyut nadi, dan tingkat gula darah. Hasilnya akan digunakan untuk menilai risiko diabetes pada individu berdasarkan atribut tersebut, memungkinkan pencegahan dini [11].

Dalam penelitian yang dilakukan oleh Pambudi membahas tentang stroke sebagai gangguan fungsi otak yang serius di Indonesia. Fokus penelitian adalah untuk menemukan gejala stroke segera dengan melihat informasi seperti usia, jenis kelamin, kadar glukosa, riwayat penyakit jantung, hipertensi, jenis pekerjaan, jenis tempat tinggal, status merokok, indeks massa tubuh, dan status pernikahan. Penelitian ini menggunakan algoritma Decision Tree C4.5 dan mencapai akurasi sebesar 99.07%. Kesimpulannya, penelitian ini membantu memahami faktor-faktor yang menyebabkan stroke dan menunjukkan bahwa algoritma Decision Tree C4.5 sangat efektif dalam menganalisis data [4].

Dalam penelitian yang dilakukan oleh Supriyadi membahas tentang peningkatan tridharma dosen sebagai upaya mencapai keunggulan. Pemimpin perguruan tinggi dapat memberikan penghargaan terprogram kepada dosen yang melakukan pekerjaan yang baik. Dengan menggunakan Naive Bayes dan Decision Tree, penelitian ini mengklasifikasikan dosen sebagai berprestasi dan tidak berprestasi berdasarkan faktor-faktor seperti jenjang pendidikan, publikasi internasional, riwayat jabatan fungsional, dan lama memperoleh serdos. Dilakukan di Universitas Sebelas Maret (2018-2021), hasil pengujian menunjukkan akurasi Naive Bayes 94,80% dan Decision Tree 95,80%, memberikan kontribusi dalam identifikasi dosen berprestasi secara efektif [12].

Pada penelitian yang dilakukan oleh aziz tahun 2023 membahas mengenai pentingnya prediksi penyakit jantung sejak dini, mengingat sulitnya mendeteksi penyakit ini pada tahap awal. Angka kematian akibat serangan jantung di seluruh dunia meningkat pada tahun 2016, menurut data WHO. Penelitian ini menggunakan algoritma decision tree, tetapi algoritma ini memiliki kelemahan dalam memprediksi dengan tingkat akurasi yang tinggi. Akibatnya, penelitian ini bertujuan untuk meningkatkan keakuratan prediksi penyakit jantung dengan menggunakan metode kelompok Bootstrap Aggregating (Bagging) dan Adaptive Boosting (Adaboost). Kedua pendekatan ini dicoba untuk memprediksi kematian akibat penyakit jantung. Menurut hasil analisis perbandingan dari tiga dataset penyakit jantung, algoritma C4.5 menghasilkan nilai akurasi 84,78% untuk dataset penyakit jantung, prediksi penyakit jantung 72,96%, dan prediksi gagal jantung 79,29%. Ini menunjukkan bahwa metode optimisasi ensemble Bootstrap Agregating (Bagging) dan Adaptive Boosting (AdaBoost) mungkin lebih baik daripada performa klasifikasi algoritma base estimator Decision Tree [13].

Penelitian tahun 2023 oleh Fauziah dan Zulfikar membahas analisis risiko penyakit jantung dengan metode klasifikasi, khususnya algoritma iterative dichotomiser 3 (ID3). Penelitian ini dilakukan dengan memanfaatkan dataset dari UCI Machine Learning Repository yang mencakup berbagai atribut terkait kondisi pasien. Evaluasi performa dilakukan dengan menggunakan metrik dari confusion matrix, yang menghasilkan tingkat akurasi, presisi, dan recall yang cukup tinggi. Mereka menemukan bahwa penerapan metode decision tree dengan algoritma ID3 mempercepat pengambilan keputusan dalam mengklasifikasikan resiko penyakit jantung dengan akurasi 85, 71%, presisi 84, 62%, dan recall 84, 62% [5].

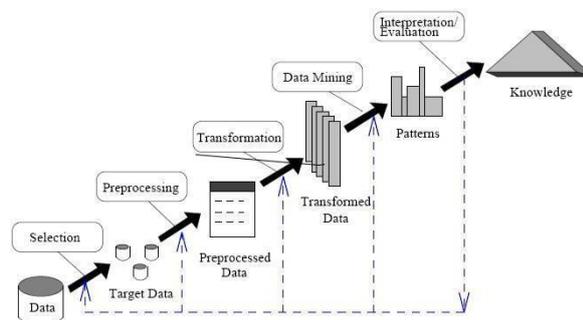
Penelitian yang dilakukan oleh Qisthiano membahas bagaimana prediksi kelulusan mahasiswa diklasifikasikan menggunakan Decision Tree dengan dataset dari beberapa perguruan tinggi di Kota Palembang. Jurusan, perguruan tinggi, jenis kelas yang dipilih siswa, dan nilai mereka dari semester pertama hingga semester ke-4 digunakan dalam penelitian ini. Selain itu, data tahun kelulusan dan tahun masuk siswa juga digunakan. Penelitian ini menggunakan alat bantu Rapidminer untuk mengelola data klasifikasi dengan model Decision Tree. Setelah menjalani 5 kali proses validasi K-Fold, penelitian ini mencapai akurasi sebesar 87.93% dalam melakukan klasifikasi terhadap prediksi kelulusan siswa [14].

Pada penelitian yang dilakukan oleh Wiratama membahas tentang diabetes mellitus (DM) sebagai penyakit kronis dengan jumlah penderita yang tinggi di seluruh dunia, termasuk Indonesia. Keterlambatan dalam diagnosis DM dapat meningkatkan angka kematian, oleh itu penting untuk deteksi dini. Penelitian ini membandingkan tiga algoritma (KNN,

Naïve Bayes, dan C4.5) untuk deteksi DM dengan optimalisasi menggunakan metode backward elimination. Hasilnya menunjukkan bahwa algoritma KNN yang dioptimasi memiliki akurasi tertinggi (97,6%), sedangkan algoritma C4.5 yang dioptimasi memiliki nilai AUC tertinggi (0,988) untuk deteksi dini DM [6].

3. METODE PENELITIAN

Penelitian ini akan menggunakan pendekatan Knowledge Discovery in Databases (KDD). Knowledge Discovery in Database Process (KDD) merupakan salah satu metode yang dapat digunakan dalam melakukan pemrosesan data atau data mining. Berikut adalah langkah-langkah yang akan diambil dalam proses penerapan metode KDD ini:



Gambar 1. Alur Knowledge Discovery in Databases (KDD)

3.1. Selection Data

Pada tahap ini, melakukan pemilihan data atribut: Karena tidak semua atribut yang ada dalam data digunakan untuk proses pengolahan data mining, proses pemilihan data ini dilakukan untuk memilih atribut atau variabel yang mempengaruhi gejala penyakit stroke otak untuk memprediksi penyakit stroke otak. Dari 11 atribut, 8 dipilih.

3.2. Preprocessing

Melakukan Penanganan Missing Value: Tahap ini bertujuan untuk meningkatkan nilai akurasi dengan menangani nilai yang hilang pada setiap atribut.

3.3. Transformation

Pada tahap ini, Melakukan transformasi data pada atribut Hypertension, Heart disease, Stroke dari bentuk angka (1 dan 0) menjadi yes dan no.

3.4. Data Mining

Pada tahap ini, Melakukan penerapan Split Data: Melakukan penerapan split data, yang berarti membagi dataset yang akan dikategorikan menjadi data training dan data testing, dimana rasio pembagiannya adalah 8:2, Melakukan penerapan algoritma Decision Tree: Melakukan penerapan algoritma Decision Tree pada dataset yang mengalami penyakit stroke otak.

3.5. Interpretation / Evaluation

Pada tahap ini, Melakukan evaluasi pada hasil penerapan algoritma Decision Tree: Menghasilkan aturan dari pemodelan pohon keputusan menggunakan algoritma Decision Tree.

4. HASIL DAN PEMBAHASAN

Data sekunder yang digunakan diambil dari Kaggle.com dalam bentuk format Comma Separated Values (CSV).

Tabel 1. Dataset Penyakit Stroke Otak

gender	age	hypertension	heart_disease	ever_married	work_type	Residence_type	avg_glucose_level	bmi	smoking_status	stroke
Male	67	0	1	Yes	Private	Urban	228.69	36.6	formerly smoked	1
Male	80	0	1	Yes	Private	Rural	105.92	32.5	never smoked	1
Female	49	0	0	Yes	Private	Urban	171.23	34.4	smokes	1
Female	79	1	0	Yes	Self-employed	Rural	174.12	24	never smoked	1
Male	81	0	0	Yes	Private	Urban	186.21	29	formerly smoked	1
Male	74	1	1	Yes	Private	Rural	70.09	27.4	never smoked	1
Female	69	0	0	No	Private	Urban	94.39	22.8	never smoked	1
Female	78	0	0	Yes	Private	Urban	58.57	24.2	Unknown	1
Female	81	1	0	Yes	Private	Urban	80.43	29.7	never smoked	1
Female	61	0	1	Yes	Govt job	Rural	120.46	36.8	smokes	1
Female	54	0	0	Yes	Private	Urban	104.51	27.3	smokes	1
Female	79	0	1	Yes	Private	Urban	214.09	28.2	never smoked	1
Female	50	1	0	Yes	Self-employed	Rural	167.41	30.9	never smoked	1
Male	64	0	1	Yes	Private	Urban	191.61	37.5	smokes	1
Male	75	1	0	Yes	Private	Urban	221.29	25.8	smokes	1
Female	60	0	0	No	Private	Urban	89.22	37.8	never smoked	1
Female	71	0	0	Yes	Govt job	Rural	193.94	22.4	smokes	1
Female	52	1	0	Yes	Self-employed	Urban	233.29	48.9	never smoked	1
Female	79	0	0	Yes	Self-employed	Urban	228.7	26.6	never smoked	1

Tabel di atas menggambarkan isi dataset penyakit stroke otak, yang terdiri dari 4982 baris dan 11 atribut diantaranya Gender, Age, Hypertension, Heart Disease, Ever Married, Work Type, Residence Type, Avg Glucose Level, BMI, Smoking Status, Stroke. Pada data tersebut, jumlah data yang terkena stroke sebanyak 248 data dan untuk yang tidak terkena stroke sebanyak 4733 data. Dari perbandingan jumlah data yang terkena stroke dan yang tidak terkena stroke, terdapat perbedaan yang sangat jauh. Selanjutnya, untuk mengimbangi data karena ada perbedaan data antara stroke dan tidak stroke, teknik SMOTE (Synthetic Minority Over-sampling Technique) digunakan. Di mana metode ini menghasilkan jumlah data yang lebih kecil menjadi jumlah data yang paling besar [15].

Karena tidak semua data yang digunakan siap digunakan begitu saja, preprocessing adalah langkah penting dalam proses analisis data. Pada tahap ini, nilai yang hilang akan ditangani, tetapi tidak ditemukan nilai yang hilang pada dataset yang digunakan, kemudian selanjutnya akan dilakukan pemilihan atribut, atribut pada dataset penyakit stroke otak sebanyak 11 atribut yaitu Gender, Age, Hypertension, Heart Disease, Ever Married, Work Type, Residence Type, Avg Glucose Level, BMI, Smoking Status, Stroke, namun atribut yang akan digunakan pada penelitian ini hanya 8 atribut yaitu Gender, Age, Hypertension, Heart Disease, Avg Glucose Level, BMI, Smoking Status, Stroke. Gejala yang paling umum dialami oleh penderita stroke otak adalah atribut yang di pilih.

Untuk mengubah data dari kumpulan ke format yang dapat diproses oleh perangkat lunak yang

digunakan, proses transformasi data terdiri dari dua tahap, yaitu mengubah atribut hypertension, heart disease, dan Stroke dari tipe data nominal menjadi numerik. Detail transformasi dari nominal ke numerik terlihat pada tabel di bawah ini.

Tabel 2. Transformasi

Atribut	Sebelum	Sesudah
	Transformasi	Transformasi
Hypertension	0	No
	1	Yes
Heart Disease	0	No
	1	Yes
Stroke	0	No

Setelah melakukan beberapa tahapan yang diantaranya data selection, preprocessing, dan transformation maka tahap selanjutnya adalah data mining. Teknik pemodelan data mining yang dipilih adalah klasifikasi menggunakan algoritma Decision Tree. Klasifikasi algoritma decision tree digunakan untuk mencapai tujuan awal yaitu meningkatkan akurasi dalam klasifikasi penyakit stroke otak. Penerapan data mining untuk klasifikasi penyakit stroke otak menggunakan algoritma decision tree ditunjukkan di sini.

Setelah melakukan proses data mining, maka selanjutnya adalah melakukan penghitungan accuracy menggunakan Confusion Matrix. Metode evaluasi Confusion Matrix menggunakan tabel untuk menilai kinerja model. Dalam matriks ini, ketika dataset memiliki dua kelas, salah satunya dianggap positif dan yang lainnya dianggap negatif. Confusion Matrix membantu menghitung nilai Accuracy, Recall, dan Presisi, yang menunjukkan seberapa baik kinerja model dalam memprediksi kedua kelas tersebut.

Tabel 3. Confusion Matrix Decision Tree

	True Yes	True No
Pred. Yes	712	165
Pred. No	82	782

Berdasarkan hasil Confusion Matrix pada tabel 1.3 dapat dilihat bahwa sebanyak 712 data diklasifikasikan “Yes” dengan valid, kemudian sebanyak 782 data diklasifikasikan “No” dengan valid.

Setelah diketahui hasil klasifikasi penyakit stroke otak menggunakan algoritma Decision Tree, maka dapat diketahui nilai Accuracy, Recall, Precision

Tabel 5. Hasil

Accuracy	85,81%
Recall	89,67%
Presisi	81,19

5. KESIMPULAN DAN SARAN

Hasil penelitian menunjukkan bahwa algoritma decision tree berhasil mengklasifikasikan penyakit stroke otak dengan baik. Dengan akurasi sebesar 85.81%, hasil tersebut menunjukkan tingkat keakuratan yang tinggi dalam memprediksi kasus-kasus penyakit stroke otak dalam dataset yang digunakan. Kesimpulannya, penerapan algoritma decision tree dalam klasifikasi penyakit stroke otak berhasil memberikan hasil yang menjanjikan dengan performa yang baik dalam memprediksi serta mengidentifikasi kasus penyakit stroke otak dalam dataset yang digunakan. Dalam penelitian selanjutnya diharapkan mencoba eksplorasi penggunaan algoritma lain seperti Random Forest, Support Vector Machines, atau Neural Networks untuk melihat apakah terdapat peningkatan performa dalam klasifikasi penyakit stroke otak.

DAFTAR PUSTAKA

- [1] M. B. S* and D. A. T, “Predictive Model for Brain Stroke in CT using Deep Neural Network,” *Int. J. Recent Technol. Eng.*, vol. 9, no. 1, pp. 2011–2017, 2020, doi: 10.35940/ijrte.f9954.059120.
- [2] F. M. Hana, “Klasifikasi Penderita Penyakit Diabetes Menggunakan Algoritma Decision Tree C4.5,” *J. SISKOM-KB (Sistem Komput. dan Kecerdasan Buatan)*, vol. 4, no. 1, pp. 32–39, 2020, doi: 10.47970/siskom-kb.v4i1.173.
- [3] A. H. Nasrullah, “Implementasi Algoritma Decision Tree Untuk Klasifikasi Data Peserta Didik,” *J. Pilar Nusa Mandiri*, vol. 7, no. 2, p. 217, 2021.
- [4] E. R. Pambudi, Sriyanto, and Firmansyah, “Teknika 16 (02): 221-226,” *Ijccs*, vol. 16, No.02, no. x, pp. 221–226, 2022.
- [5] E. Fauziah and A. F. Zulfikar, “Penerapan Metode Decision Tree Menggunakan Algoritma Iterative Dichotomiser 3 (ID3) Untuk Klasifikasi Resiko Penyakit Jantung,” *OKTAL J. Ilmu Komput. dan Sains*, vol. 2, no. 04, pp. 1207–1219, 2023, [Online]. Available: <https://journal.mediapublikasi.id/index.php/oktal/article/view/1302>
- [6] M. A. Wiratama and W. M. Pradnya, “Optimasi Algoritma Data Mining Menggunakan Backward Elimination untuk Klasifikasi Penyakit Diabetes,” *J. Nas. Pendidik. Tek. Inform.*, vol. 11, no. 1, p. 1, 2022, doi: 10.23887/janapati.v11i1.45282.
- [7] H. Sulistiani and A. A. Aldino, “Decision Tree C4.5 Algorithm for Tuition Aid Grant Program Classification (Case Study: Department of Information System, Universitas Teknokrat Indonesia),” *Edutic - Sci. J. Informatics Educ.*, vol. 7, no. 1, pp. 40–50, 2020, doi: 10.21107/edutic.v7i1.8849.
- [8] A. Fitri Boy, “Implementasi Data Mining Dalam Memprediksi Harga Crude Palm Oil (CPO) Pasar

- Domestik Menggunakan Algoritma Regresi Linier Berganda (Studi Kasus Dinas Perkebunan Provinsi Sumatera Utara)," *J. Sci. Soc. Res.*, vol. 4307, no. 2, pp. 78–85, 2020, [Online]. Available: <http://jurnal.goretanpena.com/index.php/JSSR>
- [9] D. S. Dewi and A. Asman, "Resiko Stroke Pada Usia Produktif di Ruangrawat Inap RSUD Pariaman," *J. Sci. Mandalika*, vol. 2, no. 11, pp. 576–581, 2021, [Online]. Available: <http://ojs.cahayamandalika.com/index.php/jomla/article/view/487/389>
- [10] O. N. Washilaturrizqi, "Implementasi Algoritma C4 . 5 Untuk Menentukan Penerima," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 7, no. 1, pp. 373–377, 2023.
- [11] S. Ucha Putri, E. Irawan, F. Rizky, S. Tunas Bangsa, P. A. -Indonesia Jln Sudirman Blok No, and S. Utara, "Implementasi Data Mining Untuk Prediksi Penyakit Diabetes Dengan Algoritma C4.5," *Januari*, vol. 2, no. 1, pp. 39–46, 2021.
- [12] A. Supriyadi, "Perbandingan Algoritma Naive Bayes dan Decision Tree(C4.5) dalam Klasifikasi Dosen Berprestasi," *Gener. J.*, vol. 7, no. 1, pp. 39–49, 2023, doi: 10.29407/gj.v7i1.19797.
- [13] M. I. Aziz, A. Z. Fanani, and A. Affandy, "Analisis Metode Ensemble Pada Klasifikasi Penyakit Jantung Berbasis Decision Tree," *J. Media Inform. Budidarma*, vol. 7, no. 1, pp. 1–12, 2023, doi: 10.30865/mib.v7i1.5169.
- [14] M. R. Qisthiano, P. A. Prayesy, and I. Ruswita, "Penerapan Algoritma Decision Tree dalam Klasifikasi Data Prediksi Kelulusan Mahasiswa," *G-Tech J. Teknol. Terap.*, vol. 7, no. 1, pp. 21–28, 2023, doi: 10.33379/gtech.v7i1.1850.
- [15] A. N. Kasanah, M. Muladi, and U. Pujiyanto, "Penerapan Teknik SMOTE untuk Mengatasi Imbalance Class dalam Klasifikasi Objektivitas Berita Online Menggunakan Algoritma KNN," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 3, no. 2, pp. 196–201, 2019, doi: 10.29207/resti.v3i2.945.