

IMPLEMENTASI PENERAPAN DECISION TREE DALAM KLASIFIKASI RESIKO STROKE PADA USIA MUDA

Nikodemus Christiano David¹, Muhammad Rizky Aggara², Daffa Islam Fatahillah³, Muhammad Rafi Salman⁴, Adhika Tyo Ferdiansyah⁵

Prodi Teknik Informatika, STMIK Amikom Surakarta, Sukoharjo Indonesia¹⁻⁵

Email: nikodemus.10361@mhs.amikomsolo.ac.id¹, muh.10372@mhs.amikomsolo.ac.id², daffa.10364@mhs.amikomsolo.ac.id³, muhammad.10363@mhs.amikomsolo.ac.id⁴, adhika.10366@mhs.amikomsolo.ac.id⁵

Informasi	Abstract
Volume : 2 Nomor : 10 Bulan : Oktober Tahun : 2025 E-ISSN : 3062-9624	<i>This research focuses on the application of decision tree methods for identifying the risk of stroke among young adults. Stroke is a significant health concern globally, often leading to long-term disability or death. Identifying individuals at high risk can help in early intervention and prevention strategies. We employed a decision tree algorithm to analyze various risk factors, such as hypertension, diabetes, smoking habits, and physical inactivity. The data was collected from a healthcare database, consisting of young adults aged 18 to 40 years. Our results demonstrate that the decision tree model is effective in classifying individuals with a high risk of stroke, with an accuracy rate of 67,71%. This study suggests that decision tree algorithms can be a valuable tool in clinical settings for early identification and management of stroke risk in young adults. Keywords: decision tree, stroke risk, young adults, machine learning, healthcare</i>

Keyword: Application of Decision Tree, Stroke risk, young age

Abstrak

Penelitian ini berfokus pada penerapan metode decision tree untuk mengidentifikasi risiko stroke pada usia muda. Stroke merupakan masalah kesehatan yang signifikan secara global, yang sering kali mengakibatkan kecacatan jangka panjang atau kematian. Identifikasi individu yang berisiko tinggi dapat membantu dalam intervensi dini dan strategi pencegahan. Kami menggunakan algoritma decision tree untuk menganalisis berbagai faktor risiko seperti hipertensi, diabetes, kebiasaan merokok, dan kurangnya aktivitas fisik. Data dikumpulkan dari basis data kesehatan yang terdiri dari orang dewasa muda berusia 18 hingga 40 tahun. Hasil kami menunjukkan bahwa model decision tree efektif dalam mengklasifikasikan individu dengan risiko tinggi stroke, dengan tingkat akurasi sebesar 67,71%. Studi ini menyarankan bahwa algoritma decision tree dapat menjadi alat yang berharga di lingkungan klinis untuk identifikasi dan manajemen risiko stroke pada usia muda. Kata Kunci: decision tree, risiko stroke, usia muda, pembelajaran mesin, Kesehatan

Kata Kunci: Penerapan Decision Tree, resiko Stroke, usia muda

A. PENDAHULUAN

Perkembangan teknologi dalam bidang kesehatan telah memberikan berbagai inovasi dalam metode diagnosis dan prediksi penyakit[1]. Salah satu penyakit yang menjadi perhatian serius di kalangan medis adalah stroke. Stroke merupakan penyebab utama

kecacatan dan kematian di seluruh dunia dan tidak hanya menyerang orang lanjut usia tetapi juga semakin banyak ditemukan pada usia muda. Identifikasi dini terhadap risiko stroke pada usia muda sangat penting untuk pencegahan dan penanganan yang lebih efektif[2].

Penelitian ini menggunakan metode pohon keputusan untuk mengidentifikasi risiko stroke pada usia muda dengan memanfaatkan data sekunder yang diperoleh dari jurnal dan referensi ilmiah lainnya[3]. Metode pengumpulan data sekunder ini memungkinkan peneliti untuk menganalisis dataset yang sudah ada tanpa harus melakukan pengumpulan data primer yang memerlukan waktu dan sumber daya yang lebih banyak. Dalam penelitian ini, pohon keputusan algoritma digunakan untuk menganalisis berbagai faktor risiko seperti hipertensi, diabetes, kebiasaan merokok, dan melemahkan aktivitas fisik[4].

Hasil dari penerapan algoritma pohon keputusan ini diharapkan dapat memberikan akurasi yang tinggi dalam mengklasifikasikan individu dengan risiko tinggi stroke, sebagaimana dibuktikan dalam studi-studi sebelumnya[5]. Penelitian ini juga bertujuan untuk menunjukkan bahwa algoritma pohon keputusan dapat menjadi alat yang berharga di lingkungan klinis untuk mengidentifikasi dan manajemen risiko stroke pada usia muda[6].

B. METODE PENELITIAN

Dataset

Pada penelitian ini tahap pertama yaitu mencari referensi dari studi literatur yang terkait algoritma decision tree dan penyakit stroke. Dan data yang akan digunakan pada penelitian ini adalah data sekunder yang diambil dari jurnal yang berjudul Klasifikasi Penyakit Stroke Menggunakan Algoritma Decision Tree C.4[7]. Data sekunder yakni data yang didapatkan dari penelitian sebelumnya dengan kata lain peneliti tidak harus mengambil data secara langsung ke lapangan. Dan data penelitian ini diambil dari Klasifikasi Penyakit Stroke Menggunakan Algoritma Decision Tree C.4 dengan jumlah dataset sejumlah 500 dataset[8][9].

Decision tree

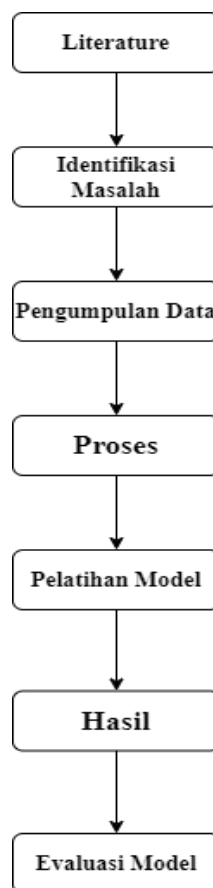
Pada penelitian ini algoritma yang digunakan yaitu algoritma decision tree. Decision tree merupakan algoritma flowchart yang berbentuk seperti struktur pohon, dimana setiap simpul internal menunjukkan tes atribut, dan setiap cabang menunjukkan hasil tes, serta pada setiap simpul daun menunjukkan kelas atau distribusi kelas[8]. Dan Algoritma decision tree merupakan metode yang paling sering digunakan dalam data mining dikarenakan cepat pada proses pengklasifikasian pohon keputusan, selain itu algoritma decision tree juga memiliki inputan samples dan training samples. Sample yaitu field-field data yang digunakan sebagai

parameter untuk klasifikasi data. Sedangkan training samples terdiri dari contoh data yang digunakan untuk membuat sebuah pohon yang telah diuji kebenarannya[10].

Rapidminer

Dalam penelitian ini untuk membantu proses mengolah data akan menggunakan software yang bernama Rapidminer. Rapidminer merupakan alat yang mendukung untuk pengembangan data dengan antarmuka visual yang mudah untuk digunakan. Rapidminer juga berfungsi sebagai software bisnis dan komersial serta dimanfaatkan juga untuk Pendidikan, penelitian, membuat prototype dengan cepat, dan untuk mengembangkan sebuah aplikasi, dan mampu mendukung semua proses pembelajaran mesin termasuk mempersiapkan data, visualisasi hasil, validasi dan pengoptimalan[11].

Flowchart Alur Penelitian



Gambar 1. Alur Penelitian

Identifikasi Masalah

Pada tahap ini peneliti akan mengidentifikasi dan merumuskan masalah penelitian. Seperti cara menggunakan algoritma decision tree dalam mengidentifikasi resiko penyakit stroke pada usia muda.

Pengumpulan Data

Data yang digunakan dalam penelitian ini merupakan data yang diambil dari jurnal Klasifikasi Penyakit Stroke Menggunakan Algoritma Decision Tree C.4. Penelitian ini mengambil dataset sejumlah 500 dataset.

Proses

Pada tahap ini data yang dikumpulkan akan diolah menggunakan aplikasi Rapidminer,serta mempersiapkan data untuk pelatihan model.

Pelatihan Model

Pada tahap ini,agloritma decision tree akan digunakan untuk melatih model dengan data yang telah diproses sebelumnya. Nantinya model ini akan digunakan untuk mengidentifikasi resiko penyakit stroke.

Hasil

Setelah melalui tahapan pelatihan model,tahap selanjutnya adalah tahap hasil,yang dimana hasil dari prediksi resiko penyakit stroke akan dihasilkan. Hasil ini nanti akan dianalisis untuk mengetahui baik tidaknya model tersebut dalam mengidentifikasi resiko penyakit stroke.

Evaluasi Model

Tahapan yang terakhir adalah tahap Evaluasi Model. Yang dimana model akan diuji serta dievaluasi untuk keakuratannya dalam mengidentifikasi resiko penyakit stroke. Hal ini penting dilakukan karena untuk mengetahui apakah model tersebut telah layak digunakan atau masih memerlukan perbaikan.

C. HASIL DAN PEMBAHASAN

Pada bagian ini, kami menyajikan hasil dari penerapan algoritma Decision Tree C4.5 dalam menganalisis dataset penyakit stroke. Hasil tersebut meliputi tingkat akurasi, presisi, sensitivitas, dan spesifisitas dari model yang dibangun. Hasil penelitian ini menunjukkan bahwa algoritma Decision Tree C4.5 memiliki performa yang baik dalam mendiagnosis penyakit stroke berdasarkan dataset yang digunakan. Berikut gambar hasil pengolahan data yang dibuat oleh peneliti.

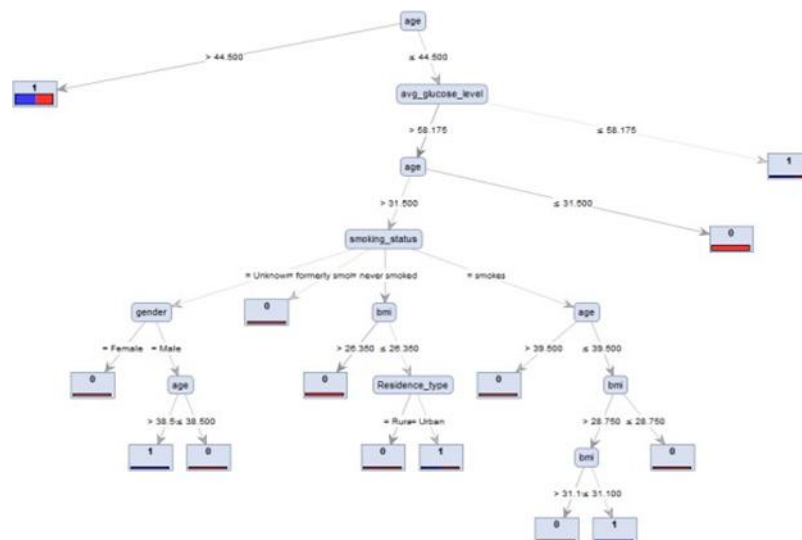
Pemodelan Algoritma Decision Tree

Pada penelitian ini, algoritma Decision Tree C4.5 digunakan untuk memodelkan data pasien dengan tujuan mengklasifikasikan kemungkinan terjadinya stroke, antara ya dan tidak.

Proses pemodelan melibatkan beberapa atribut yaitu digunakan penerapan algoritma untuk membangun pohon keputusan.

Peneliti mengambil data sejumlah 542 , yang didapat dari data asli yang berjumlah 5111. Peneliti memilih data hanya 542 dikarenakan Menggunakan seluruh data yang berjumlah 5111 bisa memerlukan sumber daya komputasi yang lebih besar, baik dari segi waktu maupun memori, dan terlebih juga terdapat nilai dari atribut yang tidak terbaca karena error.

Pada data yang diambil oleh peneliti, Pohon keputusan yang ditampilkan di atas adalah hasil dari algoritma decision tree yang digunakan untuk mengklasifikasikan data berdasarkan beberapa atribut, yaitu usia (age), tingkat glukosa rata-rata (avg_glucose_level), status merokok (smoking_status), jenis kelamin (gender), BMI (Body Mass Index), dan tipe tempat tinggal (Residence_type). Berikut Hasil penelitian menggunakan aplikasi Rapid Miner.



Gambar 2. Hasil Pohon Keputusan Klasifikasi Penyakit Stroke

Pada hasil gambar 2 diatas menyatakan bahwa terdapat jumlah 14 rule yang terolah data dengan Klasifikasi Decision Tree , yang masing-masing terdapat angka 0 berjumlah 9 yang berarti (Tidak Stroke), lalu angka 1 menandakan bahwa itu (Ya Stroke).Sedangkan Atribut yang dirancang awal yang berjumlah total 11 , setelah masuk di pohon Keputusan diatas mendapatkan hanya 6 Atribut yang terdiri dari Age, Avg_glucose_level, smoking_status, gender, bmi, Residence_type.

Hasil Akurasi Algoritma Decision Tree

Pada penelitian ini bertujuan untuk membandingkan hasil akurasi pada jurnal sebelumnya. Maka hasil akurasi yang diperoleh untuk prediksi penyakit stroke sebesar 66,71% seperti gambar 3 dibawah ini.

accuracy: 66.71% +/- 7.13% (mikro: 66.73%)

	true 1	true 0	class precision
pred. 1	185	168	52.41%
pred. 0	12	176	93.62%
class recall	93.91%	51.16%	

Gambar 3. Hasil Akurasi Prediksi Penyakit Stroke

hasil akurasi antara Jurnal sebelumnya yang menggunakan data berjumlah 4.906 dengan sampel yang dilakukan oleh peneliti berjumlah 542 memiliki perbedaan. Dataset yang lebih besar cenderung memberikan hasil yang lebih akurat karena lebih banyak informasi yang tersedia untuk pelatihan model. Dengan lebih banyak data, model dapat belajar pola yang lebih kompleks dan variatif dalam dataset.

D. KESIMPULAN

Efektivitas Model Algoritma decision tree belum terbukti efektif dalam mengklasifikasikan individu dengan risiko tinggi stroke pada usia muda, dengan hasil tingkat akurasi sebesar 66,71%. Faktor Risiko, Faktor-faktor risiko yang dianalisis meliputi hipertensi, diabetes, kebiasaan merokok, dan kurangnya aktivitas fisik. Pemilihan Data, Peneliti menggunakan dataset berjumlah 542 dari total 5111 data yang tersedia. Penggunaan dataset yang lebih besar cenderung meningkatkan akurasi model. Implementasi Klinis, Algoritma decision tree memiliki potensi besar untuk digunakan dalam setting klinis guna membantu identifikasi dini dan manajemen risiko stroke pada usia muda.

SARAN

Penggunaan Dataset yang Lebih Besar, Penelitian selanjutnya sebaiknya menggunakan dataset yang lebih besar untuk meningkatkan akurasi model. Penambahan Variabel Lain, Disarankan untuk memasukkan variabel lain yang mungkin berpengaruh terhadap risiko stroke, seperti riwayat keluarga dan pola makan. Pengembangan Model Lain, Selain decision tree, metode lain seperti random forest atau gradient boosting juga dapat dieksplorasi untuk membandingkan performanya. Implementasi di Berbagai Setting, Menguji model di berbagai setting klinis dan populasi yang berbeda untuk memastikan generalisasi hasil penelitian.

E. DAFTAR PUSTAKA

- R. J. Ridwan, A. I. Pawelloi, and A. Hastuti, "PENERAPAN ALGORITMA C4.5 DECISION TREE UNTUK MENDIAGNOSA PENYAKIT DALAM," 2024. [Online].
Available: <https://jurnal.umpar.ac.id/index.php/sylog>
- Y. Widiastiwi, I. Ernawati, J. R. Fatmawati, P. Labu, and J. Selatan, "Klasifikasi Penyakit Batu Ginjal Menggunakan Algoritma Decision Tree C4.5 Dengan Membandingkan Hasil Uji Akurasi," 2021.
- R. Ridho, "KLASIFIKASI DIAGNOSIS PENYAKIT COVID-19 MENGGUNAKAN METODE DECISION TREE," 2021. [Online]. Available:
<https://jurnal.umj.ac.id/index.php/just-it/index>
- N. N. Habibah, A. Nazir, I. Iskandar, F. Syafria, L. Oktavia, and I. Syurfi, "Pemodelan Klasifikasi Untuk Menentukan Penyakit Diabetes dengan Faktor Penyebab Menggunakan Decision Tree C4.5 Pada Wanita," *Jurnal Sistem Komputer dan Informatika (JSON)*, vol. 4, no. 4, p. 654, Jun. 2023, doi: 10.30865/json.v4i4.6202.
- Y. Azhar, A. Khoiriyah Firdausy, and P. J. Amelia, "SINTECH Journal | 191 Perbandingan Algoritma Klasifikasi Data Mining Untuk Prediksi Penyakit Stroke," 2022, [Online]. Available: <https://doi.org/10.31598>
- J. Homepage, F. Akbar, H. Wira Saputra, A. Karel Maulaya, and M. Fikri Hidayat, "MALCOM: Indonesian Journal of Machine Learning and Computer Science Implementation of Decision Tree Algorithm C4.5 and Support Vector Regression for Stroke Disease Prediction Implementasi Algoritma Decision Tree C4.5 dan Support Vector Regression untuk Prediksi Penyakit Stroke," vol. 2, pp. 61–67, 2022.
- I. M. Agus Oka Gunawan, I. D. A. Indah Saraswati, I. D. G. Riswana Agung, and I. P. Eka Putra, "Klasifikasi Penyakit Jantung Menggunakan Algoritma Decision Tree Series C4.5 Dengan Rapidminer," *Jurnal Teknologi Dan Sistem Informasi Bisnis*, vol. 5, no. 2, pp. 73–83, Apr. 2023, doi: 10.47233/jteksis.v5i2.775.
- J. Teknika and R. Estian Pambudi, "Teknika 16 (02): 221-226," *IJCCS*, vol. x, No.x, pp. 1–5.
- H. Wendy Ariono, "Sistem Pakar Pengklasifikasi Stadium Kanker Serviks Berbasis Mobile Menggunakan Metode Decision Tree," 2022. [Online]. Available:
<http://ejurnal.ubharajaya.ac.id/index.php/JKI>
- S. F. Damanik, A. Wanto, and I. Gunawan, "Penerapan Algoritma Decision Tree C4.5 untuk Klasifikasi Tingkat Kesejahteraan Keluarga pada Desa Tiga Dolok." [Online]. Available:

<https://ejournal.catuspata.com/index.php/jkdn/index>

- R. Nofitri and N. Irawati, "ANALISIS DATA HASIL KEUNTUNGAN MENGGUNAKAN SOFTWARE RAPIDMINER," JURTEKSI (Jurnal Teknologi dan Sistem Informasi), vol. 5, no. 2, pp. 199–204, Jul. 2019, doi: 10.33330/jurteksi.v5i2.365.