

PENERAPAN METODE KLASIFIKASI POHON KEPUTUSAN PADA PASIEN PENYAKIT JANTUNG MENGGUNAKAN ALGORITMA C4.5

J. Prayoga¹, Buyung Solihin Hasugian², Virdyra Tasril³, Amru Yasir⁴

1,2) Sistem Informasi, Fakultas Teknik dan Ilmu Komputer, Universitas Dharmawangsa, Medan, Indonesia

3) Teknologi Rekayasa Perangkat Lunak, Politeknik Negeri Medan, Indonesia

4) Teknologi Informasi, Fakultas Teknik dan Ilmu Komputer, Universitas Dharmawangsa, Medan, Indonesia

Article Info

Article history:

Received: 04 April 2026

Revised: 20 April 2026

Accepted: 22 April 2026

ABSTRACT

Abstrak

Data mining adalah sebuah proses analisis data dengan menggunakan perangkat lunak untuk menemukan pola dan aturan dari sejumlah data. Data mining mampu menganalisis data yang sangat besar dan kemudian mengubahnya menjadi sebuah informasi. Informasi yang memiliki arti penting untuk mendukung pengambilan keputusan. Salah satu teknik dalam data mining adalah klasifikasi, yaitu suatu proses untuk menemukan suatu model fungsi yang menjelaskan suatu konsep atau kelas data untuk memprediksi kelas dari objek yang tidak sejenis. Pada tugas akhir ini akan dibuat suatu perangkat lunak yang menerapkan metode decision tree dengan algoritma C4.5, kemudian menganalisis persentase nilai kebenaran dari pohon dan hasil klasifikasi, dan mengomputasikannya dengan RAPID MINER. Algoritma ini memiliki persentase nilai kebenaran antara 66,67% hingga 100%. Persentase nilai kebenaran pohon ini bergantung pada data training yang digunakan untuk membangun pohon.

Kata Kunci: Data Mining, Decision Tree, Algoritma C4.5, Penyakit Jantung

Abstract

Data Mining is a process of data analysis using software to find patterns and rules from a large amount of data. Data mining is able to analyze a huge amount of data and then change it into information that has an important meaning for decision support. One of the techniques in data mining is classification, which is a process to find a model of a function that explains a concept or class data in order to predict a class from an unknown object. In this research, we will make a software that applies the decision tree method with the C4.5 algorithm, then analyze the percentage of truth value of the tree and classification result, and compare it with RAPID MINER. The algorithm has a percentage of the truth value between 66.67% to 100%. This percentage of the tree's truth value depends on the training data that was used to build the tree.

Keywords: Data Mining, Decision Tree, C4.5 Algorithm, Heart Disease

Djtechno: Jurnal Teknologi Informasi oleh Universitas Dharmawangsa. Artikel ini bersifat open access yang didistribusikan di bawah syarat dan ketentuan dengan Lisensi Internasional Creative Commons Attribution NonCommercial ShareAlike 4.0 ([CC-BY-NC-SA](https://creativecommons.org/licenses/by-nc-sa/4.0/)).



Corresponding Author:

E-mail: yoga@dharmawangsa.ac.id

1. PENDAHULUAN

Rumah sakit perlu memanfaatkan sumber daya, termasuk sistem informasi, untuk bersaing secara efektif. Menurut Strahan et al. [12], sistem informasi memainkan peran penting dalam meningkatkan kinerja organisasi layanan kesehatan. Penelitian mereka menunjukkan bahwa rumah sakit dengan sistem informasi yang baik lebih mampu mengelola informasi, mendukung operasi harian, dan mendukung pengambilan keputusan strategis. Hal ini tidak hanya meningkatkan efisiensi, tetapi juga memberikan keunggulan kompetitif dalam industri layanan kesehatan.

Sistem informasi memainkan peran penting dalam meningkatkan pengalaman pasien di rumah sakit. Penggunaan teknologi informasi yang tepat dapat meningkatkan interaksi antara pasien dan tenaga medis, sehingga meningkatkan kepuasan pasien [12]. Selain itu, penggunaan natural language processing (NLP) untuk menganalisis catatan klinis, khususnya dalam mendeteksi masalah kesehatan mental dan penggunaan zat pada pasien HIV, dapat membantu rumah sakit mengidentifikasi dan mengatasi masalah kesehatan yang sebelumnya terlewatkan, sehingga memberikan perawatan yang lebih baik bagi pasien.

Penyakit jantung, termasuk penyakit jantung koroner, merupakan penyebab kematian utama di dunia, menyumbang 32% dari total kematian global [WHO, 2021]. Masalahnya adalah penyakit-penyakit ini sering tidak menunjukkan gejala pada awalnya, sehingga banyak pasien menyadarinya terlambat. Deteksi dini sangat penting agar pengobatan dapat dilakukan sebelum kondisi memburuk. Penelitian menunjukkan bahwa pencegahan adalah kunci dalam mengelola penyakit jantung dan mengurangi risiko komplikasi serius [11].

Keputusan berdasarkan analisis Data Mining cenderung lebih akurat dan efektif. Model prediktif dari Data Mining dapat meningkatkan kualitas keputusan di berbagai bidang, seperti bisnis, kesehatan, dan ilmu sosial [5]. Dengan memanfaatkan data, organisasi dapat menemukan peluang baru dan mengurangi risiko.

Data mining telah digunakan di berbagai bidang, seperti keuangan, kesehatan, dan pemasaran. Di sektor layanan kesehatan, data mining dapat membantu mendiagnosis penyakit dan mengelola perawatan pasien. Data pasien memungkinkan rumah sakit untuk menemukan pola yang dapat meningkatkan kualitas perawatan [14].

2. METODE PENELITIAN

Data Mining

Dalam konteks data mining, data mining adalah proses menemukan pola dan informasi berguna dari kumpulan data besar yang dapat digunakan untuk pengambilan keputusan yang lebih baik [8]. Selain itu, penelitian menunjukkan bahwa visualisasi dan pemetaan pengetahuan dapat meningkatkan pemahaman tentang data yang kompleks, yang sangat relevan dalam konteks data mining [16].

Data mining menggunakan berbagai teknik, seperti klasifikasi, regresi, clustering, dan asosiasi. Setiap teknik ini memiliki tujuan dan aplikasi yang berbeda, disesuaikan dengan jenis data dan informasi yang ingin diperoleh. Misalnya, klasifikasi data mining dapat digunakan untuk memprediksi kategori data baru, sedangkan clustering membantu mengelompokkan data yang serupa. Teknik-teknik data mining dalam analisis sinyal ECG dapat dimanfaatkan untuk menghasilkan urutan bit acak yang efisien [10].

Decision Tree

Decision tree adalah salah satu teknik klasifikasi yang paling populer dalam dunia analisis data. Metode ini menggambarkan struktur seperti pohon, di mana setiap node mewakili suatu atribut, cabang mewakili nilai atribut tersebut, dan daun mewakili kelas atau hasil akhir dari proses klasifikasi. Node paling atas dalam pohon keputusan disebut node akar, yang bertindak sebagai titik awal dalam proses pengambilan keputusan. Decision tree tidak hanya mudah dipahami dan diinterpretasikan, tetapi juga sangat efektif dalam mengelola data yang kompleks [6].

Metode ini bekerja dengan membagi data menjadi cabang berdasarkan atribut penting, membuat proses klasifikasi dan prediksi lebih mudah dan terstruktur [4]. Metode ini tidak hanya menyederhanakan proses pengambilan keputusan yang kompleks, tetapi juga membantu dalam mengidentifikasi faktor-faktor kunci yang memengaruhi hasil [13].

Algoritma C4.5

Algoritma C4.5 adalah salah satu algoritma populer yang dikembangkan dari ID3 untuk membangun pohon keputusan. Menurut Al-walidi [1], algoritma ini dapat diterapkan dalam sistem rekomendasi untuk menganalisis dan memprioritaskan kebutuhan pengguna. C4.5 tidak hanya meningkatkan kualitas rekomendasi, tetapi juga membantu memahami pola dan preferensi pengguna.

Dalam konteks prediksi, C4.5 sering digunakan untuk klasifikasi data. C4.5 tetap relevan karena transparansi dan kemudahan interpretasi modelnya [15]. Jika diperlukan pemahaman mendalam tentang keputusan model, C4.5 masih menjadi pilihan utama. Dalam prediksi churn, algoritma C4.5 memainkan peran penting dalam menganalisis dan mengklasifikasikan data pelanggan [7]. C4.5 mampu menghasilkan pohon keputusan yang mudah dipahami, membantu perusahaan mengidentifikasi faktor-faktor yang menyebabkan churn dan merancang strategi retensi.

Algoritma C4.5 dihitung menggunakan konsep entropi. Entropi total dihitung dengan:

$$E(S) = - \sum_{i=1}^c p_i \log_2 p_i$$

Keterangan :

S = Dataset

C= Jumlah Kelas

p_i = Proporsi Kelas ke- i

Split entropy yang membagi S dengan record menjadi himpunan S_v dan S adalah:

$$Gain(S, A) = E(S) - \sum_{v \in Values(A)} \frac{|S_v|}{|S|} \cdot E(S_v)$$

Keterangan :

A = Atribut

Values(A) = Nilai-nilai atribut A

S_v = Subset dengan Nilai v

Atribut untuk A relatif terhadap data keluaran y adalah:

$$SplitInfo(S, A) = - \sum_{i=1}^c \frac{|S_i|}{|S|} \log_2 \frac{|S_i|}{|S|}$$

Keterangan :

S = Subset ke-I dari s berdasarkan atribut A

Partisi informasi dihitung dengan:

$$GainRatio(S, A) = \frac{Gain(S, A)}{SplitInfo(S, A)}$$

Keterangan :

GainRantio digunakan C4.5 untuk menghindari bias atribut banyak nilai

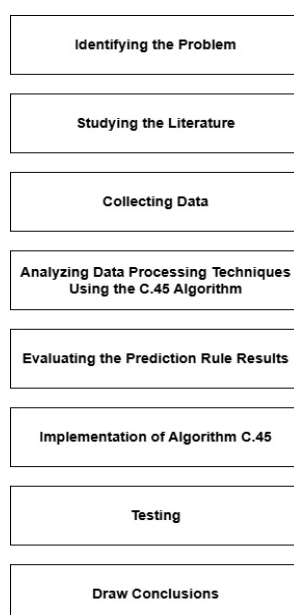
Penyakit Jantung

Belakangan ini, penelitian tentang memprediksi kelangsungan hidup pasien penyakit jantung terus berkembang, dengan berbagai metode klasifikasi yang bertujuan untuk meningkatkan akurasi diagnosis. Salah satunya adalah Naive Bayes, metode sederhana yang efektif dalam memproses data dengan berbagai atribut. Menurut Fajriati et al. [3], optimasi algoritma Naive Bayes mencapai akurasi 83% pada dataset Cleveland, menunjukkan potensi signifikan metode ini dalam dunia medis.

Pohon keputusan telah terbukti efektif untuk memprediksi penyakit jantung. Menurut Depari et al. [2], dalam penelitian menggunakan dataset 'Personal Key Indicators of Heart Disease', Decision Tree menunjukkan akurasi dan kemudahan interpretasi yang lebih baik daripada Naive Bayes.

Prediksi kelangsungan hidup pasien jantung tidak hanya bergantung pada algoritma, tetapi juga pada kualitas data. Data yang baik dan fitur yang tepat sangat memengaruhi akurasi prediksi [14].

Kerangka Penelitian



Kerangka penelitian meliputi langkah-langkah: (1) Identifikasi Masalah, yaitu menentukan dan mendefinisikan batas masalah yang akan diteliti; (2) Analisis Masalah,

yaitu memahami masalah yang telah ditentukan ruang lingkup atau batasannya; (3) Studi Literatur, yaitu mempelajari literatur yang diharapkan dapat digunakan; (4) Pengumpulan Data, yaitu melakukan observasi dan wawancara langsung; (5) Analisis Teknik Pengolahan Data menggunakan Algoritma C4.5; (6) Evaluasi Aturan Hasil Prediksi; (7) Implementasi Algoritma C4.5; dan (8) Penarikan Kesimpulan berdasarkan output hasil Data Mining.

3. HASIL DAN PEMBAHASAN

Pada penelitian ini, data pasien penyakit jantung dikumpulkan dan diproses menggunakan algoritma C4.5. Data pre-processing ditampilkan pada Tabel 1 berikut ini:

Tabel 1. Data Pre-Processing Pasien Jantung

No	Nama Pasien	Penyakit Bawaan	Usia	Berat Badan	Keputusan
1	Zubaidah	Hematematis Melana	Tua	Berat	Ya
2	Refnawati	Hypertensi	Muda	Menengah	Tidak
3	Feby Yolanda	Hematematis Melana	Muda	Menengah	Tidak
4	Hetty Manik	Hematematis Melana	Tua	Berat	Ya
5	Nanda Pratama	Hypertensi	Muda	Menengah	Tidak
6	Aisyah	Hematemesis Melana	Tua	Berat	Ya
7	Rima Natarigan	Hypertensi	Muda	Berat	Ya
8	Butet Silohoman	Sironis Hepatitis	Muda	Berat	Ya
9	Riska Ahmad	Hypertensi	Tua	Berat	Ya
10	Rahni Eka Putri	Hematematis Melana	Muda	Berat	Ya
11	Mulyadi	Sironis Hepatitis	Tua	Menengah	Tidak
12	Iis Ahmad	Sironis Hepatitis	Tua	Menengah	Tidak
13	Septri Wahyuni	Hypertensi	Muda	Ringan	Tidak
14	Widya Firman	Hematematis Melana	Muda	Menengah	Tidak
15	Deded Fery Andesta	Hypertensi	Tua	Menengah	Tidak
16	Desti Santika	Sironis Hepatitis	Muda	Berat	Ya
17	Randika Permana	Hematemesis Melana	Muda	Berat	Ya
18	Fery Muslianto	Hypertensi	Tua	Menengah	Tidak
19	Hendra Sutanto	Sironis Hepatitis	Tua	Berat	Ya
20	Ade Saputra	Hypertensi	Muda	Berat	Ya

Sumber: Data Penelitian (2024)

Perhitungan Entropi dan Gain

Untuk menghitung entropi, persamaan (1) digunakan sesuai dengan landasan teori yang telah dibahas. Menghitung total entropi dilakukan dengan menghitung jumlah keputusan Ya dan Tidak dari semua kasus yang ada.

$$\begin{aligned}
 \text{Entropy Total} &= -\left(\frac{12}{20} * \text{Log}_2\left(\frac{12}{20}\right)\right) + \left(-\left(\frac{8}{20}\right) * \text{Log}_2\left(\frac{8}{20}\right)\right) \\
 &= (0.4421 + 0.5287) = 0.9710
 \end{aligned}$$

Entropy (total) menghitung total nilai keputusan Ya sebanyak 12 dan keputusan Tidak sebanyak 8 dari 20 total kasus.

a. Atribut Usia

Entropi usia tua dihitung dari jumlah kasus dengan 6 keputusan Tidak dan 6 keputusan Ya dari total 12 kasus.

$$\begin{aligned} \text{Entropy Tua} &= \left(-\left(\frac{6}{12}\right) * \text{Log}_2\left(\frac{6}{12}\right) \right) + \left(-\left(\frac{6}{12}\right) * \text{Log}_2\left(\frac{6}{12}\right) \right) \\ &= (0.5 + 0.5) = 1 \end{aligned}$$

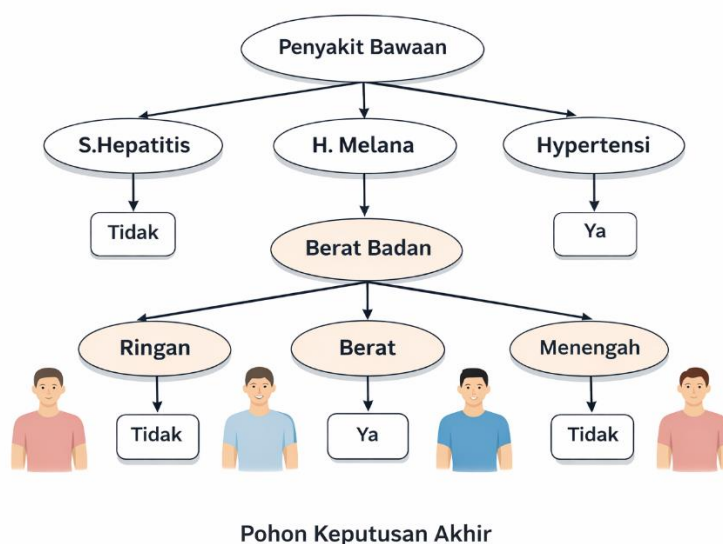
Entropi usia muda dihitung dari jumlah kasus dengan keputusan Tidak = 2 dan Ya = 6 dari 8 kasus.

$$\begin{aligned} \text{Entropy Muda} &= \left(-\left(\frac{6}{8}\right) * \text{Log}_2\left(\frac{6}{8}\right) \right) + \left(-\left(\frac{2}{8}\right) * \text{Log}_2\left(\frac{2}{8}\right) \right) \\ &= (0.3112 + 0.5) = 0.8113 \end{aligned}$$

Selanjutnya, mencari gain dengan menggunakan persamaan 2:

$$\begin{aligned} \text{Gain (Total,usia)} &= 0.9710 - \left(\left(\frac{12}{20}\right) * 1 \right) + \left(\left(\frac{8}{20}\right) * 0.8113 \right) \\ &= 0.9710 - (0.6 + 0.3244) \\ &= 0.9710 - (0.9244) = 0.4643 \end{aligned}$$

Hasil akhir dari proses kalkulasi C4.5 menghasilkan pohon keputusan akhir sebagaimana ditampilkan pada Gambar 1 berikut:



Gambar 1. Final Decision Tree

4. SIMPULAN

Penerapan metode klasifikasi Decision Tree menggunakan algoritma C4.5 untuk pasien penyakit jantung telah menunjukkan efektivitasnya dalam mendiagnosis dan memprediksi penyakit jantung. Penelitian ini berhasil memanfaatkan rekam medis pasien sebagai data input, memungkinkan pembuatan pohon keputusan yang mengidentifikasi atribut kunci yang memengaruhi diagnosis penyakit jantung. Hasil penelitian menunjukkan bahwa algoritma C4.5 memberikan model klasifikasi yang andal dengan akurasi tinggi, menjadikannya alat yang berguna untuk membantu para profesional kesehatan dalam membuat keputusan yang tepat. Kemampuan algoritma dalam menangani data kategorikal dan kontinu memastikan adaptabilitasnya dalam dataset medis. Selain itu, interpretabilitas struktur pohon keputusan memungkinkan pemahaman yang lebih baik dan transparansi dalam proses pengambilan keputusan, yang sangat penting dalam aplikasi medis. Penelitian selanjutnya dapat mengeksplorasi pendekatan hibrida atau mengintegrasikan teknik machine learning tambahan untuk meningkatkan lebih lanjut kemampuan prediktif dan ketahanan model.

REFERENCES

- [1] N. H. Al-walidi, "A Systematic Literature Review of Recommender Systems for Requirements Engineering," vol. 175, no. 14, pp. 31-41, 2020, doi: 10.5120/ijca2020920630.
- [2] D. H. Depari, Y. Widiastiwi, and M. M. Santoni, "Perbandingan Model Decision Tree, Naive Bayes dan Random Forest untuk Prediksi Klasifikasi Penyakit Jantung," vol. 4221, pp. 239-248, 2022, doi: 10.52958/iftk.v18i3.4694.
- [3] N. Fajriati and B. Prasetyo, "Optimasi Algoritma Naive Bayes dengan Diskritisasi K-Means pada Diagnosis Penyakit Jantung," vol. 10, no. 3, pp. 503-512, 2023, doi: 10.25126/jtiik.2023106510.
- [4] N. S. S. Farheen and A. Jain, "Improved routing in MANET with optimized multi-path routing fine-tuned with hybrid modeling," J. King Saud Univ. - Comput. Inf. Sci., vol. 34, no. 6, pp. 2443-2450, 2022, doi: 10.1016/j.jksuci.2020.01.001.
- [5] Z. C. Farronato, "Outsourcing Tasks Online: Matching Supply and Demand on Peer-to-Peer Internet Platforms," Management Science, vol. 67, no. 7, pp. 3985-4003, 2021, doi: 10.1287/mnsc.2020.3730.
- [6] D. Fischer, "Data Science with Julia," International Statistical Review, 2019, doi: 10.1111/insr.12345.
- [7] L. Geiler, S. Affeldt, and M. Nadif, "A survey on machine learning methods for churn prediction," 2022, doi: 10.1007/s41060-022-00312-5.
- [8] C. Hamilton, "AAAI News," vol. 24, no. 4, pp. 3-8, 2021, doi: 10.1609/aimag.v24i4.1740.

- [9] S. T. Informatika, U. Hamzanwadi, "Komparasi Metode Klasifikasi Data Mining Decision Tree dan Naive Bayes Untuk Prediksi Penyakit Diabetes," *Infotek: Jurnal Informatika dan Teknologi*, vol. 4, no. 1, pp. 63-69, 2021, doi: 10.29408/jit.v4i1.2994.
- [10] S. Islam, "Using ECG signal as an entropy source for efficient generation of long random bit sequences," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 34, no. 8, pp. 5144-5155, 2022, doi: 10.1016/j.jksuci.2022.01.001.
- [11] T. Iwano, H. Toda, and K. Nakamura, "Preventative effects of bisoprolol transdermal patches on postoperative atrial fibrillation in high-risk patients undergoing non-cardiac surgery," *J. Cardiology*, vol. 78, no. 5, pp. 349-354, 2021, doi: 10.1016/j.jjcc.2021.05.001.
- [12] J. O. Strahan and L. R., "Assessing the Performance Value of the Baldrige Journey: A Comparison of Baldrige Applicants and Nonapplicants," *J. Healthcare Management*, 2022, doi: 10.1097/JHM-D-21-00045.
- [13] A. Kocsis and G. Molnar, "Factors influencing academic performance and dropout rates in higher education," *Oxford Review of Education*, vol. 00, no. 00, pp. 1-19, 2024, doi: 10.1080/03054985.2024.2316616.
- [14] J. Santos-Pereira, L. Gruenwald, and J. Bernardino, "Top data mining tools for the healthcare industry," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 34, no. 8, pp. 4968-4982, 2022, doi: 10.1016/j.jksuci.2021.06.002.
- [15] K. Wangchuk, P. Riyamongkol, and R. Waranusast, "Next syllables prediction system in Dzongkha using long short-term memory," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 34, no. 6, pp. 3800-3806, 2022, doi: 10.1016/j.jksuci.2021.01.001.
- [16] A. H. Zyoud, "Visualization and Mapping of Knowledge and Science Landscapes in Expert Systems With Applications Journal: A 30 Years' Bibliometric Analysis," 2021, doi: 10.1177/21582440211027574.