

KLASIFIKASI PENYAKIT JANTUNG DENGAN MENGGUNAKAN ALGORITMA C5.0

Nada Aprillia¹, Jasmir², Yulia Arvita³

Program Studi Teknik Informatika Universitas Dinamika Bangsa, Jambi

Jl. Jendral Sudirman, Thehok – Jambi, Telp. 0741-35095/Fax. 0741-35093

Email : nadapriya@gmail.com, ijay_jasmir@yahoo.com, yulia_arvita@yahoo.co.id

Artikel Info :

Artikel History :

Submitted : 04-06-2025

Accepted : 27-08-2025

Published : 30-09-2025

Kata Kunci: Data Mining, Algoritma C5.0, Penyakit Jantung

Keyword: Data Mining, C5.0 Algorithm, Heart Disease

Abstrak Penyakit jantung merupakan gangguan pada sistem pembuluh darah besar yang menyebabkan organ vital makhluk hidup tidak berfungsi dengan baik. Seiring dengan perkembangan ilmu pengetahuan, prediksi penderita penyakit jantung dapat dilakukan menggunakan teknik data mining. Namun, dalam proses diagnosis, sering kali ditemukan permasalahan berupa rendahnya tingkat akurasi klasifikasi. Penelitian ini melakukan klasifikasi pada dataset penyakit jantung yang diperoleh dari situs Kaggle. Dataset tersebut terdiri dari 918 data dengan 12 atribut, antara lain: Penderita Penyakit Jantung (Heart Disease), Usia (Age), Jenis Kelamin (Sex), Jenis Sakit Dada (Chest Pain Type), Tekanan Darah Saat Istirahat (Resting Blood Pressure), Kolesterol (Cholesterol), Gula Darah (Fasting Blood Sugar), Hasil Elektrokardiografi Saat Istirahat (Resting ECG), Detak Jantung Maksimum (Max Heart Rate), Latihan Diinduksi Angina (Exercise Angina), Oldpeak, dan ST Slope. Klasifikasi dilakukan menggunakan algoritma C5.0 melalui tools RapidMiner untuk mengetahui kombinasi atribut yang optimal dan menghitung tingkat akurasi.

Kata Kunci :

Abstract Heart disease is a disorder of the large blood vessel system that causes vital organs of living things to not function properly. Along with the development of science, prediction of heart disease sufferers can be done using data mining techniques. However, in the diagnosis process, problems are often found in the form of low levels of classification accuracy. This study classifies the heart disease dataset obtained from the Kaggle site. The dataset consists of 918 data with 12 attributes, including: Heart Disease Patients, Age, Sex, Chest Pain Type, Resting Blood Pressure, Cholesterol, Fasting Blood Sugar, Resting ECG Results, Maximum Heart Rate, Exercise Induced Angina, Oldpeak, and ST Slope. Classification is carried out using the C5.0 algorithm through the RapidMiner tool to determine the optimal combination of attributes and calculate the level of accuracy.

1.PENDAHULUAN

Di era globalisasi, penyakit jantung adalah istilah yang sering digunakan untuk menggambarkan gangguan jantung. 18,6 juta orang di seluruh dunia meninggal karena penyakit jantung setiap tahun. Penyakit jantung merupakan penyakit dengan angka kematian yang tinggi. Menurut WHO (Organisasi Kesehatan Dunia) dan CDC, penyakit jantung adalah penyebab utama kematian di Inggris, Amerika Serikat, Kanada, dan Australia. Jumlah orang dewasa yang terdiagnosis penyakit jantung adalah 26,6 juta (11,3%) dari populasi orang dewasa [1], oleh karena itu, deteksi dini diperlukan terlepas dari apakah seseorang memiliki penyakit jantung atau tidak. Karena jika penanganya terlambat dapat menyebabkan kematian.

Seiring kemajuan ilmu pengetahuan, kasus-kasus yang berkaitan dengan prediksi penyakit jantung dapat diselesaikan dengan menggunakan teknik data mining. Data mining adalah ekstraksi informasi yang berguna dari kumpulan data besar [2], yang menggunakan teknik pembelajaran mesin [3] yang biasa dikenal dengan istilah database data discovery (KDD)[4]. dan Data mining memiliki beberapa teknik antara lain prediksi atau klasifikasi, clustering dan asosiasi. Masalah umum dalam proses diagnostik adalah kurangnya presisi dalam proses klasifikasi. Klasifikasi adalah rekor baru di salah satu dari beberapa kategori[5]. Teknik klasifikasi dapat

digunakan untuk mengukur tingkat akurasi suatu dataset. Klasifikasi adalah evaluasi objek data untuk menempatkannya dalam kategori tertentu dari antara kategori yang tersedia [6].

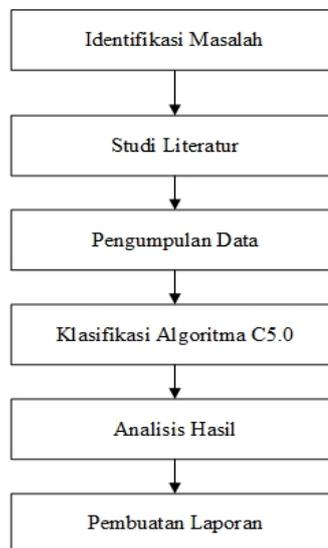
Klasifikasi adalah menemukan sekumpulan pola yang dapat menggambarkan dan membedakan antara kelas informasi, sehingga pola tersebut bisa digunakan untuk memprediksi kelas dari suatu objek yang tidak diketahui [7]. Algoritma C5.0 memiliki kemampuan untuk mengklasifikasikan ke dalam pohon keputusan [8] atau regulasi karena merupakan classifier yang mengklasifikasikan data dalam waktu yang lebih singkat dibandingkan dengan classifier lainnya [9]. Perbedaan yang terlihat secara fisik adalah pohon yang dihasilkan oleh algoritma C5.0 lebih padat dibandingkan dengan algoritma C4.5. Sementara algoritma C4.5 berhenti sampai perolehan informasi dihitung, algoritma C5.0 terus menghitung rasio perolehan menggunakan perolehan informasi dan entropi yang tersedia. Oleh karena itu, membangun pohon dengan algoritma C5.0 lebih cepat dibandingkan dengan algoritma C4.5.

Berdasarkan penjelasan di atas, maka penulis melakukan penelitian ini yaitu mengklasifikasikan data penyakit jantung yang diperoleh dari situs Kaggle, mengetahui kombinasi atribut dari setiap kumpulan data, kemudian melakukan perhitungan untuk menentukan tingkat akurasi kumpulan data menggunakan data mining. Yang berjudul “**KLASIFIKASI PENYAKIT JANTUNG DENGAN MENGGUNAKAN ALGORITMA C5.0**”

2.METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Dalam menyusun penelitian ini perlu dibuat sebuah susunan kerangka kerja penelitian dengan langkah-langkahnya yang jelas agar penelitian dapat dilakukan secara sistematis serta dapat mencapai hasil penelitian yang memuaskan.. Gambar 3.1 dibawah ini merupakan kerangka kerja yang digunakan penelitian ini.



Gambar 1. Tahapan penelitian

Berdasarkan kerangka kerja yang telah digambarkan diatas, maka dapat diuraikan pembahasan dari masing-masing tahapan penelitian yaitu:

1. Identifikasi Masalah
Pada tahap ini, peneliti melakukan peninjauan serta mengidentifikasi hal-hal yang menjadi permasalahan.
2. Studi Literatur

Pada tahap ini, peneliti melakukan kegiatan mencari referensi berbagai sumber seperti buku, artikel, ataupun internet tentang penelitian sejenis yang bertujuan untuk mendapatkan pijakan bagi penulis untuk memahami persoalan yang diteliti dengan benar dan sesuai dengan kerangka berpikir ilmiah.

3. Pengumpulan Data

Pengumpulan data dilakukan untuk memperoleh informasi yang dibutuhkan, dalam penelitian ini peneliti melakukan pengumpulan data dengan mengambil data dari kaggle.

4. Klasifikasi Algoritma C5.0

Pada tahap ini penulis melakukan perhitungan *decision tree* dengan menggunakan algoritma C5.0 dengan menghitung nilai *entropy* dan *gain* dari data-data yang telah dinormalisasikan dan digunakan untuk membuat hasil dalam bentuk *decision tree*.

5. Analisis Hasil

Pada tahap ini, penulis menganalisis hasil dari perhitungan algoritma C5.0. Hasil yang didapat guna untuk mengetahui klasifikasi penderita penyakit jantung dengan menggunakan data yang sudah dikumpulkan.

6. Pembuatan Laporan

Pada tahap ini kesimpulan diambil berdasarkan hasil penelitian yang dibuat dalam bentuk laporan penelitian.

3. HASIL DAN PEMBAHASAN

3.1 ANALISIS PERMASALAHAN

3.1.1 Gambaran Umum Penyakit Jantung

Penyakit jantung merupakan organ vital bagi makhluk hidup [10]. Penyakit jantung terjadi akibat adanya gangguan pada pembuluh darah besar yang menyebabkan fungsi jantung dan sirkulasi yang tidak normal. Penyakit yang berhubungan dengan jantung dan pembuluh darah antara lain gagal jantung, penyakit arteri koroner, dan penyakit jantung rematik [11].

3.1.2 Analisis Sistem Yang Sedang Berjalan

Gejala-gejala yang dialami penderita penyakit jantung seringkali tidak diperhatikan dan sering kali diabaikan karena gejalanya merupakan penyakit-penyakit ringan yang sering dialami. Dalam masyarakat hal ini dianggap sebagai hal yang biasa, sehingga penyakit jantung ini tidak terdeteksi dalam kehidupan sehari-hari. Oleh karena itu, deteksi dini diperlukan terlepas dari apakah seseorang memiliki penyakit jantung atau tidak. Karena jika penanganannya terlambat dapat menyebabkan kematian.

3.1.3 Solusi Pemecahan Masalah

Berdasarkan analisa sistem yang berjalan, maka penulis merancang suatu sistem yang dapat melakukan diagnosa yang mengarah pada penyakit jantung dengan mengenali gejala-gejala yang dialami penderita dengan menggunakan rapidminer. Rapidminer mengekstraksi pola dari kumpulan data besar dengan menggabungkan metode statistik, kecerdasan buatan dan basis data [12]. Dengan rapid miner memungkinkan pengguna lebih mudah melakukan perhitungan data besar [13]. Dengan adanya sistem pakar ini, akan memberikan efektifitas waktu serta efisiensi biaya bagi para pengunjungnya karena sistem ini dapat dipakai kapan saja dan dimana saja.

3.2 PERHITUNGAN METODE C 5.0 PADA KLASIFIKASI DATA PENYAKIT JANTUNG

4.2.1 Representasi Data

Data yang digunakan dalam penelitian ini menggunakan data penyakit jantung yang diambil dari situs kaggle. Dataset terdiri dari kumpulan objek dan atribut [14] dari data penderita penyakit jantung tersebut terdiri dari 12 atribut dan data tersebut dibagi ke dalam 2 kelas, yaitu "PENDERTA" dan "BUKAN PENDERITA" yang terdapat pada kolom *Heart Disease* yang diwakilkan dengan angka 1 untuk penderita dan 0 untuk yang

bukan penderita. Terdapat 508 data dalam kelas “PENDERITA” dan 410 data yang masuk ke dalam kelas “BUKAN PENDERITA”.

TABEL DATA PENYAKIT JANTUNG

No	Age	Sex	Chest Pain Type	Resting BP	Cholesterol	Fasting BS	Resting ECG	Max HR	Exercise Angina	Oldpeak	ST_Slope	Heart Disease
1	40	M	ATA	140	289	0	Normal	172	N	0	Up	0
2	49	F	NAP	160	180	0	Normal	156	N	1	Flat	1
3	37	M	ATA	130	283	0	ST	98	N	0	Up	0
4	48	F	ASY	138	214	0	Normal	108	Y	1.5	Flat	1
5	54	M	NAP	150	195	0	Normal	122	N	0	Up	0
6	39	M	NAP	120	339	0	Normal	170	N	0	Up	0
7	45	F	ATA	130	237	0	Normal	170	N	0	Up	0
8	54	M	ATA	110	208	0	Normal	142	N	0	Up	0
9	37	M	ASY	140	207	0	Normal	130	Y	1.5	Flat	1
10	48	F	ATA	120	284	0	Normal	120	N	0	Up	0

3.2.2 Perhitungan Manual C 5.0

Penulis melakukan perhitungan dengan menggunakan algoritma C5.0 dengan persamaan berikut :

$$I(S_1, S_2, \dots, S_m) = - \sum_{i=1}^m p_i * \log_2(p_i)$$

Keterangan :

S = Himpunan kasus

S₁ = jumlah sampel

Pi = proporsi kelas

Untuk mendapatkan informasi nilai subset dari atribut A tersebut maka digunakan formula dibawah ini

$$E(A) = \sum_{j=i}^y \frac{S_{1j} + \dots + S_{mj}}{S} I(S_{1j}, \dots, S_{mj})$$

Keterangan :

$\frac{S_{1j} + \dots + S_{mj}}{S}$ Jumlah subset j yang dibagi dengan jumlah sampel S

Untuk mendapatkan nilai gain selanjutnya digunakan formula

Gain A = I S₁, S₂, ..., S_m - E

Alur kerja C 5.0 adalah sebagai berikut

Dalam algoritma ini penentuan sampel awal dilakukan melalui perhitungan jumlah kasus, menghitung jumlah kasus dengan untuk keputusan “Penderita” atau di representasikan dengan angka 1 dan jumlah kasus “Bukan Penderita” yang diwakilkan dengan angka 0 pada attribute penderita penyakit jantung (*heart disease*). *Entropy* dari semua kasus dan juga gain dari setiap attribute, yaitu umur (*age*), jenis kelamin (*sex*), jenis nyeri dada (*chest pain type*), tekanan darah saat istirahat (*Resting Blood Pressure*), kolesterol (*cholesterol*), gula darah saat istirahat (*fasting blood sugar*), hasil elektrokardiografi saat istirahat (*resting ECG*), detak jantung maksimum (*max heart rate*), Latihan diinduksi (*exercise angina*), *oldpeak*, *ST Slope*.

3.5 Akurasi Perhitungan

	true N	true Y	class precision
pred. N	102	18	85.00%
pred. Y	33	122	78.71%
class recall	75.56%	87.14%	

Gambar 4. Akurasi Perhitungan

Berdasarkan langkah-langkah di atas dengan penggunaan *tools* RapidMiner, hasil perhitungan tersebut terdapat pada Gambar 4 didapatkan hasil visualisasi *Decision Tree* (pohon keputusan) Algoritma C5.0 berdasarkan Data Penyakit Jantung, yang kemudian dapat di deskripsikan menjadi berbagai Rules yang digunakan untuk memprediksi data Penyakit Jantung. Yang kemudian dilakukan validasi untuk melihat akurasi kebenaran data dengan menggunakan metode *Split Validation* yang menggunakan 70% dari total data menjadi *Data Training* dan 30% menjadi *Data Testing*. Berdasarkan hasil spit data akurasi di dapatkan 81.45. Hal ini menunjukkan bahwa algoritma ini mampu mengklasifikasi penyakit jantung dengan tepat. Sehingga Rules yang sudah didapat dapat dijadikan acuan dalam kehidupan sehari-hari untuk memprediksi apakah orang dengan ciri-ciri berdasarkan Rules merupakan penderita penyakit jantung.

4. KESIMPULAN

Berdasarkan analisis penelitian yang telah dijalankan, telah ditunjukkan bagaimana cara mengklasifikasi dan menganalisis *dataset* menggunakan algoritma C5.0 serta bagaimana cara mengevaluasi hasil tingkat akurasi prediksi algoritma C5.0 menggunakan *tools* RapidMiner pada data Penyakit Jantung dengan jumlah 918 data yang mempunyai 12 atribut, yaitu Penderita Penyakit Jantung (*Heart Disease*), Usia (*Age*), Jenis Kelamin (*Sex*), Jenis Sakit Dada (*Chest Pain Type*), Tekanan Darah Saat Istirahat (*Resting Blood Pressure*), Kolesterol (*Cholesterol*), Gula Darah (*Fasting Blood Sugar*), Hasil Elektrokardiografi Saat Istirahat (*Resting ECG*), Detak Jantung Maksimum (*Max Heart Rate*), Latihan Diinduksi Angina (*exercise angina*), *oldpeak*, *ST Slope*.

Setelah melakukan perhitungan menggunakan algoritma C5.0 dan dilakukan validasi terhadap akurasi perhitungan algoritma dengan menggunakan metode *Split Validation*, yaitu dengan membagi data menjadi 2 bagian, *Data Training* dan *Data Testing* dengan rasio persentase 70% dari total data sebagai *Data Training* dan 30% dari total data sebagai *Data Testing* didapatkan akurasi algoritma C5.0 sebesar 81.45%.

Dengan tingginya tingkat akurasi yang dihasilkan algoritma C5.0, hasil pembentukan pohon keputusan serta deskripsi atau *rules* yang dihasilkan dari pohon keputusan dapat dijadikan acuan untuk mengklasifikasikan apakah seseorang terindikasi penderita penyakit jantung atau bukan.

DAFTAR PUSTAKA

- [1] R. Annisa, "Analisis Komparasi Algoritma Klasifikasi Data Mining Untuk Prediksi Penderita Penyakit Jantung," *J. Tek. Inform. Kaputama*, vol. 3, no. 1, pp. 22–28, 2019, [Online]. Available: <https://jurnal.kaputama.ac.id/index.php/JTIK/article/view/141/156>.
- [2] D. P. Utomo, P. Sirait, and R. Yunis, "Reduksi Atribut Pada Dataset Penyakit Jantung dan Klasifikasi Menggunakan Algoritma C5.0," *J. Media Inform. Budidarma*, vol. 4, no. 4, pp. 994–1006, 2020, doi: 10.30865/mib.v4i4.2355.
- [3] "Data Penyakit Jantung" [Online]. Available: <https://www.kaggle.com/datasets/firdaus9914/penyakit-jantung>.
- [4] I. A. Nikmatun and I. Waspada, "Implementasi Data Mining untuk Klasifikasi Masa Studi Mahasiswa Menggunakan Algoritma K-Nearest Neighbor," *J. SIMETRIS*, vol. 10, no. 2, pp. 421–s432, 2019.
- [5] A. Asroni, H. Fitri, and E. Prasetyo, "Penerapan Metode Clustering dengan Algoritma K-Means pada Pengelompokan Data Calon Mahasiswa Baru di Universitas Muhammadiyah Yogyakarta (Studi Kasus: Fakultas Kedokteran dan Ilmu Kesehatan, dan Fakultas Ilmu Sosial dan Ilmu Politik)," *Semesta Tek.*, vol. 21, no. 1, pp. 60–64, 2018, doi: 10.18196/st.211211.
- [6] D. P. Utomo and M. Mesran, "Analisis Komparasi Metode Klasifikasi Data Mining dan Reduksi Atribut Pada Data Set Penyakit Jantung," *J. Media Inform. Budidarma*, vol. 4, no. 2, p. 437, 2020, doi: 10.30865/mib.v4i2.2080.
- [7] P.-A. I. M. Wahyuni Sri, Saputra Kana, "IMPLEMENTASI RAPIDMINERDALAM MENGNALISA DATA MAHASISWA DROP OUT 1Sri," vol. 10, pp. 421-437, 2017.
- [8] P. Y. Saputra D. H. Subhi, and F. Z. A Winantama, "Implementasi Sentimen Analisis Komentar Channel Video Pelayanan Pemerintah di Youtube Menggunakan Algoritma Naive Bayes," *J. Inform Polinema*, vol. 5, no. 4, pp. 209-213, 2019, doi : 10.33795/jip.v5i4.259.
- [9] D. P. Utomo, P. Sirait, and R. Yunis, "Reduksi Atribut Pada Dataset Penyakit Jantung dan Klasifikasi Menggunakan Algoritma C5.0," *J. Media Inform. Budidarma*, vol. 4, no. 4, pp. 994–1006, 2020, doi: 10.30865/mib.v4i4.2355.
- [10] M. Pardede, E. Buulolo, and E. Ndruru, "Implementasi Algoritma C5.0 Pada Kelulusan Peserta Ujian Kemahiran Berbahasa Indonesia (Ukbi) Pada Balai Bahasa Sumatera Utara," *KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer)*, vol. 3, no. 1, pp. 64–72, 2019, doi: 10.30865/komik.v3i1.1569.
- [11] A. H. Nasrullah, "Implementasi Algoritma Decision Tree Untuk Klasifikasi Data Peserta Didik," *J. Pilar Nusa Mandiri*, vol. 7, no. 2, p. 217, 2021.
- [12] M. Pardede, E. Buulolo, and E. Ndruru, "Implementasi Algoritma C5.0 Pada Kelulusan Peserta Ujian Kemahiran Berbahasa Indonesia (Ukbi) Pada Balai Bahasa Sumatera Utara," *KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer)*, vol. 3, no. 1, pp. 64–72, 2019, doi: 10.30865/komik.v3i1.1569.
- [13] B. Rahmat C.T.I. *et al.*, "Implemetasi k-means clustering pada rapidminer untuk analisis daerah rawan kecelakaan," *Semin. Nas. Ris. Kuantitatif Terap. 2017*, no. April, pp. 58–60, 2017.
- [14] N. Salim and O. Nurdiawan, "Penerapan Data Mining Pada Penjualan Barang Menggunakan Metode Metode Naive Bayes Classifier Untuk Optimasi Strategi Pemasaran," *J. Teknol. Inf. dan Komun.*, no. April, pp. 84–95, 2018.
- [15] N. Rabbi Radliya, "Data mining materi 3," no. 321, pp. 1–10, 2015.
- [19] A. Rohman, V. Suhartono, and C. Supriyanto, "Penerapan Agoritma C4.5 Berbasis Adaboost Untuk Prediksi Penyakit Jantung," *J. Teknol. Inf.*, vol. 13, pp. 13–19, 2017.

- [20] R. Firmansyah, "PENERAPAN ALGORITMA C5.0 BERBASIS BPSO UNTUK DETEKSI DINI PENYAKIT JANTUNG," *J. Ekon. Vol. 18, Nomor 1 Maret201*, vol. 2, no. 1, pp. 41–49, 2020.