

# Prediksi Penyakit Jantung Menggunakan Algoritma Decision Tree C4.5

Zaehol Fatah, M.Kom<sup>1</sup>, Muhammad Salman Al Farisy<sup>2</sup>

<sup>1,2</sup> Universitas Ibrahimy, Situbondo

Email: alfarisykhann@gmail.com

## Abstrak

Penyakit jantung masih merupakan ancaman kesehatan global yang serius, bertanggung jawab atas kematian lebih dari 17 juta orang di dunia pada tahun 2015. Diagnosis tradisional rentan terhadap variasi subjektif dan memakan waktu. Untuk meningkatkan efisiensi dan akurasi, teknik data mining dan machine learning telah muncul sebagai alat yang menjanjikan. Penelitian ini mengusulkan pengembangan model untuk memprediksi penyakit jantung menggunakan algoritma Decision Tree C4.5, yang dipilih karena kekuatannya sebagai metode klasifikasi dan kemampuannya menghasilkan model yang intuitif serta mudah diinterpretasikan. Data yang digunakan adalah 303 sampel catatan medis pasien yang diambil dari repositori publik Kaggle.com. Prosesnya meliputi pra-pemrosesan data untuk eliminasi nilai kosong dan transformasi format, diikuti dengan pembagian data menggunakan teknik Hold-Out 70%:30% (212 sampel pelatihan, 91 sampel pengujian). Implementasi model dan evaluasi kinerja dilakukan menggunakan aplikasi RapidMiner. Hasil menunjukkan bahwa model C4.5 mencapai akurasi keseluruhan sebesar 70.33% pada set pengujian. Secara spesifik, model menunjukkan kekuatan dalam mendeteksi kasus positif (Sakit) dengan nilai Recall untuk kelas 'Iya' sebesar 84.00%, yang sangat penting untuk meminimalkan False Negative dalam konteks medis. Atribut Jenis Nyeri Dada teridentifikasi sebagai faktor klinis paling dominan (node akar) dalam penentuan kemungkinan penyakit jantung. Model ini valid sebagai sistem pendukung keputusan awal, khususnya dalam mengidentifikasi pasien berisiko tinggi.

**Kata Kunci: Prediksi Penyakit Jantung , Decision Tree C4.5 , Klasifikasi , Data Mining , RapidMiner**

## Abstract

*Heart disease remains a serious global health threat, was accountable for the demise of over 17 million people worldwide in 2015. Traditional diagnosis is prone to subjective variations and is time-consuming. To improve efficiency and accuracy, data mining and machine learning techniques have emerged as promising tools. This study proposes the development of a model for predicting heart disease using the Decision Tree C4.5 algorithm, chosen for its robustness as a classification method and its ability to produce intuitive and easily interpretable models. The data used are 303 patient medical record samples taken from the public repository Kaggle.com. The process includes data preprocessing for null value elimination and format transformation, followed by data splitting using the 70%:30% Hold-Out technique (212 training samples, 91 testing samples). Model implementation and performance evaluation were performed using the RapidMiner application. The results show that the C4.5 model achieves an overall accuracy of 70.33% on the test set. Specifically, the model demonstrated robustness in detecting positive cases (Sick) with a Recall value for the 'Yes' class of 84.00%, which is crucial for minimizing False Negatives in a medical context. The Chest Pain Type attribute was identified as the most dominant clinical factor (root node) in determining the likelihood of heart disease. This model is valid as an early decision support system, particularly in identifying high-risk patients.*

**Keyword: Heart Disease Prediction, Decision Tree C4.5, Classification, Data Mining, RapidMiner**

## PENDAHULUAN

Bagian organ tubuh penting manusia yakni jantung harus berfungsi dengan baik dan teratur, jika terjadi gangguan, maka akibat fatalnya yakni serangan jantung (Muzakki et al., 2024). Penyakit jantung masih menjadi salah satu ancaman kesehatan global paling serius, Menurut statistik World Health Organization (WHO) per tahun 2015, angka mortalitas global menunjukkan bahwa kurang lebih 17 juta manusia di dunia tutup usia disebabkan oleh penyakit jantung (Riani et al., 2019). Faktor-faktor risiko seperti gaya hidup, tekanan darah, kolestrol dan denyut jantung semuanya berperan signifikan dalam peningkatan risiko penyakit jantung. Tingginya angka ini menyoroti pentingnya diagnosis dini dan intervensi preventif yang cepat dan akurat. Diagnosis penyakit jantung secara tradisional seringkali mengandalkan interpretasi dokter terhadap gejala, hasil tes laboratorium, dan rekam medis pasien, sebuah proses yang dapat memakan waktu dan rentan terhadap variasi subjektif.

Dalam upaya meningkatkan efisiensi dan akurasi proses diagnosis Metode penggalian data (data mining) dan algoritma belajar mandiri (machine learning) kini diakui sebagai teknologi sangat menjanjikan untuk berbagai aplikasi di industri kesehatan (Supriyatna & Mustika, 2018). Serangkaian cara telah dirancang untuk mengekstraksi pola dan data berguna dari berbagai himpunan data; cara ini disebut sebagai Penambangan Data (Data Mining) (Mustakim et al., 2021). Sebagai salah satu sub-bidang dalam ilmu komputer, penambangan data memanfaatkan berbagai prosedur komputasi, termasuk klasifikasi, metode statistik, dan teknik pengenalan pola untuk menganalisis kumpulan data (Haryanti et al. 2024). Proses ini mengintegrasikan berbagai teknik dari disiplin ilmu seperti statistika, matematika, pembelajaran mesin dan kecerdasan buatan (AI) (Larose, 2024). Proses ini juga bisa diartikan prosedur untuk mencari pola, hubungan, dan

kecenderungan yang bermakna dengan menganalisis kelompok data berukuran besar yang tersimpan (storage), melalui teknik identifikasi pola. Adapun beberapa metode utama yang dioperasikan dalam data mining meliputi: deskripsi (description), prediksi (prediction), asosiasi (association), klasifikasi (classification), klustering (clustering), dan estimasi (estimation) (Wahyuni, 2018).

Penelitian ini mengusulkan pengembangan model untuk memprediksi penyakit jantung menggunakan algoritma Decision Tree C.45 . Algoritma ini dipilih karena termasuk metode klasifikasi yang sangat kuat dan juga terkenal. serta kemampuannya untuk dapat menghasilkan model yang intuitif, mudah diinterpretasikan, serta representasi visual berupa pohon keputusan yang jelas, memudahkan pemahaman alur prediksi bagi praktisi medis dan peneliti (Kursorong et al., 2019). Algoritma C4.5 adalah bentuk update dari algoritma ID3 yang dilakukan Ross Quinlan (Akbar et al., 2019).

Sejumlah penelitian terdahulu telah berhasil menerapkan algoritma klasifikasi untuk prediksi penyakit jantung. Muzakki et al. (2024) juga menerapkan algoritma C4.5 dalam kerangka RapidMiner untuk prediksi penyakit jantung. Penelitian lain menggunakan metode yang berbeda, seperti Naive Bayes dan Support Vector Machine (SVM), untuk meningkatkan akurasi dalam memprediksi penyakit kutil dan stunting. Selain itu, keunggulan C4.5 dalam menghasilkan model yang mudah diinterpretasikan juga telah dikaji. Namun, sebagian besar studi seringkali hanya menekankan pada nilai akurasi keseluruhan. Dalam konteks medis, validasi model harus lebih dalam, dengan penekanan pada metrik sensitivitas (Recall) untuk memastikan kemampuan model dalam mendeteksi kasus positif secara maksimal, sebuah aspek yang krusial untuk meminimalkan risiko False Negative dalam screening pasien berisiko tinggi.

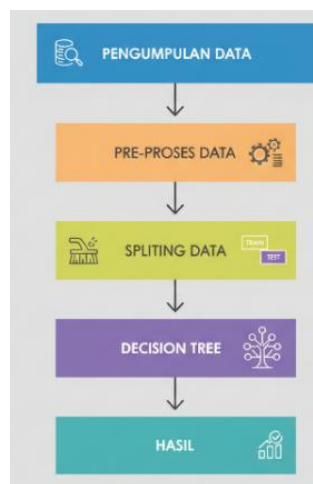
Berdasarkan research gap tersebut, penelitian ini memiliki kontribusi (novelty)

yang berfokus pada validasi klinis dan interpretasi model Decision Tree C4.5 pada dataset penyakit jantung: Validasi Kinerja Klinis (Sensitivitas Tinggi): Penelitian ini secara spesifik memvalidasi model untuk mencapai nilai Recall yang tinggi (84.00% pada kelas 'Iya' / Sakit). Tingginya nilai Recall ini merupakan pembeda utama karena menunjukkan kapabilitas model untuk meminimalkan False Negative, sehingga berfungsi sebagai alat screening awal yang sensitif dan bernilai dalam lingkungan klinis.

**Identifikasi Faktor Dominan:** Kami mengidentifikasi secara empiris bahwa Atribut Jenis Nyeri Dada teridentifikasi sebagai faktor klinis paling dominan (node akar) dalam penentuan kemungkinan penyakit jantung. Hasil interpretatif ini memberikan wawasan klinis yang jelas bagi praktisi medis di lokasi studi.

**METODE**

Prosedur dan metodologi yang diterapkan untuk memprediksi penyakit jantung, khususnya dengan metode Decision Tree C4.5, diuraikan secara sistematis dalam sub-bab ini. Rincian yang disajikan mencakup karakteristik data yang digunakan, tahapan pra-pemrosesan data, implementasi model Decision Tree C4.5, serta metrik evaluasi yang diterapkan untuk mengukur kinerja model.



**Gambar 1. Alur Metode Penelitian**

**Pengumpulan Data**

Repositori publik *Kaggle.com* menyediakan bahan yang diperlukan untuk penelitian ini, yang diambil dari data yang tersedia di platform tersebut dengan judul “Health care: Heart attack possibility”. Dataset ini berisi catatan medis pasien yang berpotensi mengidap penyakit jantung. Jumlah total sampel (record) yang digunakan dalam pengujian ini adalah 303 dataset dengan 13 atribut prediktor dan 1 label kelas target.

Atribut-atribut tersebut, yang berfungsi sebagai faktor risiko klinis, dijelaskan secara rinci dalam Tabel 1, meliputi data numerik dan kategorikal, seperti gula darah, hasil tes thallium, kolesterol serum, dan denyut jantung tertinggi. Variabel hasil (target) adalah TARGET, yang bersifat biner: Iya (Kemungkinan Lebih Besar terkena penyakit jantung) atau Tidak (Kemungkinan Lebih Kecil terkena penyakit jantung). perhatikan tabel dibawah ini.

Tabel 1. Atribut Data

<b>Atribut</b>	<b>Keterangan</b>
Jenis Nyeri Dada	Dibagi menjadi 4 kategori, dari nyeri khas (angina tipikal) hingga tidak ada gejala nyeri (asintomatik).
Jenis Kelamin	1 = Laki-laki; 0 = Perempuan.
Tekanan Darah Saat Istirahat	Tekanan darah sistolik dalam mm Hg.
Usia	Usia pasien dalam tahun.
Kolesterol Serum	Kadar kolesterol dalam darah dalam mg/dl.

Gula Darah Puasa	Indikator biner: 1 jika gula darah puasa lebih dari 120mg/dl (diabetes/berisiko); 0 jika tidak.
Hasil EKG Saat Istirahat	Hasil elektrokardiogram, diklasifikasikan menjadi 3 kategori (0, 1, atau 2) berdasarkan temuan.
Denyut Jantung	Denyut jantung tertinggi yang dicapai selama tes stres/olahraga.
Angina (Nyeri Dada) yang Dipicu Olahraga	Indikator biner: 1 jika nyeri dada terjadi setelah berolahraga; 0 jika tidak.
Oldpeak	Tingkat depresi segmen ST (perubahan gelombang EKG) yang terjadi saat berolahraga dibandingkan saat istirahat.
Kemiringan Segmen ST Puncak Olahraga	Arah kemiringan segmen ST saat puncak olahraga (misalnya, miring ke atas, datar, miring ke bawah).
Jumlah Pembuluh Darah Utama yang Terlihat dengan Fluoroskopi Thal (Hasil Tes Thallium)	Jumlah pembuluh darah utama (0 hingga 3) yang mengalami penyumbatan atau penyempitan signifikan.
TARGET (VARIABEL HASIL):	Menunjukkan hasil tes pencitraan jantung: Normal (0), Cacat Tetap (1, kerusakan permanen), atau Cacat Reversibel (2, kekurangan aliran darah hanya saat stres). Tidak = Kemungkinan Lebih Kecil terkena penyakit jantung. Iya = Kemungkinan Lebih Besar terkena penyakit jantung.

### Pre-Processing Data

Prakondisi data merupakan serangkaian proses untuk mempersiapkan data mentah sebelum dapat dioperasikan pada tahap pemrosesan berikutnya (Septhya et al., 2023). Pada fase ini, data mentah diproses melalui pembersihan data dan transformasi data. Nilai-nilai yang kosong atau null dieliminasi, dan format data diubah agar lebih terstruktur. Tujuannya adalah untuk mengoptimalkan pemodelan dengan memastikan data input siap dan berkualitas.

### Pembagian Data

Untuk mengukur kinerja generalisasi model Decision Tree C4.5, dataset total 303 sampel 8 dibagi menggunakan teknik *Hold-Out* 70%:30%, Pembagian ini menghasilkan 212 sampel untuk set pelatihan (*training\ set*) dan 91 sampel untuk set pengujian (*testing\ set*).

### Decision Tree

Algoritma ini merupakan salah satu algoritma klasifikasi yang paling diminati karena kemudahan interpretasinya (Ditendra et al., 2022). Algoritma adalah metode klasifikasi dalam disiplin ilmu data mining. Pohon keputusan didefinisikan sebagai konsep diagram alir (flowchart) yang memiliki struktur pohon (tree). Dalam struktur tersebut, setiap node (simpul internal) merepresentasikan suatu atribut,

Setiap cabang pada struktur bertindak sebagai titik uji yang memilah data berdasarkan nilai atribut yang diberikan. Setelah semua pengujian dilalui, daun menjadi label keluaran yang menunjukkan kelas yang menjadi hasil prediksi atau sebaran kelas yang terkait. (Rohman & Rufiyanto, 2019).

Algoritma ini juga merupakan metode yang kuat dalam dunia pembelajaran mesin yang mampu mengubah dataset berlabel menjadi pohon keputusan. Pohon ini berfungsi sebagai model prediktif, yang kemudian dapat diuji keakuratannya menggunakan data uji tak terlihat untuk mengukur kemampuan generalisasinya. Algoritma ini memanfaatkan konsep entropi informasi untuk membangun pohon keputusan. Proses ini dimulai dengan menganalisis data pelatihan, secara cerdas menentukan atribut yang memiliki tingkat informasi tertinggi sebagai dasar pembagian data dan membentuk struktur pohon yang optimal (Ditendra et al., 2022).

Berikut adalah alur dalam penerapan algoritma Decision Tree (Putri et al., 2023).

1. Persiapkan Dataset: Sediakan kumpulan data pelatihan (*training dataset*) yang lengkap. Data ini akan menjadi dasar untuk melatih dan

membangun struktur Pohon Keputusan.

2. Hitung Entropi Dataset: Tentukan tingkat ketidakmurnian atau keacakan pada himpunan data saat ini ( $S$ ). Gunakan rumus Entropi standar untuk mengukur seberapa heterogen kelas target dalam set data.

$$Entropy(S) = \sum_{i=1}^N -\pi \times \log 2\pi$$

Keterangan :

Himpunan  $S$  adalah total kasus yang sedang dianalisis,  $N$  menunjukkan banyaknya partisi yang mungkin dalam himpunan  $S$ ,  $\pi$  didefinisikan sebagai rasio (proporsi) antara sub-himpunan data ( $S_i$ ) dengan total himpunan  $S$

3. Tentukan Node Akar dengan Nilai Gain: Hitung nilai Information Gain (Gain) untuk setiap atribut yang tersedia. Gain menunjukkan seberapa besar pengurangan Entropi (peningkatan homogenitas) yang dicapai jika atribut tersebut digunakan untuk membagi data.

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^N \frac{|S_i|}{|S|} \times Entropy(S_i)$$

Keterangan :

Dalam proses konstruksi pohon, Node Akar (dan node pemisah berikutnya) dipilih berdasarkan atribut yang menghasilkan nilai Gain tertinggi, karena ia menawarkan pemisahan paling informatif.

4. Hitung Split Information (SI): Sebelum menghitung Gain Ratio, tentukan nilai Split Information untuk menormalkan Gain dengan mempertimbangkan sebaran data yang dihasilkan dari pembagian.

$$Split\ Information(S, A) = - \sum_{i=1}^n \frac{|S_i|}{|S|} \log_2 \frac{|S_i|}{s}$$

Keterangan:

$|S_i|$  adalah jumlah kasus di partisi  $i$

5. Hitung Gain Ratio: Jika menggunakan C4.5, hitung Gain Ratio untuk memilih atribut pemisah terbaik yang telah dinormalisasi.
6. Ulangi Proses Rekursif: Terapkan

$$Gain\ Ratio(a) = \frac{Gain(a)}{Entropy(a)} \dots \dots \dots (2.4)$$

langkah 2 hingga 5 secara berulang (rekursif) pada setiap subset data yang dihasilkan dari pemisahan sebelumnya. Proses ini terus berlanjut untuk membangun sub-pohon dan cabangnya.

7. Hentikan Pembentukan Pohon: Proses konstruksi dihentikan ketika semua node pada suatu cabang menjadi homogen, artinya semua record dalam node tersebut termasuk dalam kelas target yang sama. Titik ini akan menjadi Daun (Leaf). Kriteria penghentian lain mungkin termasuk kehabisan atribut atau mencapai batas kedalaman pohon yang telah ditentukan.

### RapidMiner

Perangkat lunak ilmu data bernama RapidMiner dibuat oleh perusahaan yang memiliki nama identik (Nofitri & Irawati 2019). Dibangun dengan arsitektur inti terbuka, perangkat lunak ini mencakup seluruh siklus machine learning mulai dari pra-pemrosesan data, tugas data mining (seperti predictive dan descriptive mining), hingga validasi, optimasi, dan visualisasi hasil. Fitur-fiturnya menjadikannya ideal untuk menggambarkan data secara visual (A. R., A., & O. H. I. 2018).

### HASIL DAN PEMBAHASAN Dataset

Himpunan data yang digunakan untuk studi ini diambil dari *Kaggle.com* dan disajikan dalam format file “.csv”. Pengujian ini menggunakan 303 dataset dengan 13 atribut dan 1 label. Adapun

sebagian dataset yang digunakan untuk menyusun laporan ini dapat dilihat pada tabel 2.

Tabel 2. Dataset

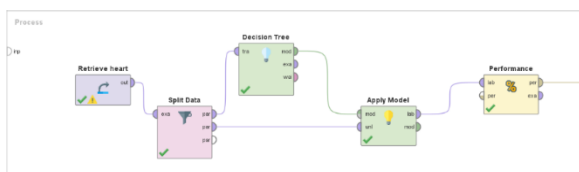
Ro...	TAR...	Usia	Jenis Kelamin	Jenis ...	Tekanan D...	Kolester...	Gula Dara...	Hasil EKG S...	Denyut Jant...	Angina (Nye...	oldpe...	Kemiring...	Jumla...	Thal (Hasil
1	lya	63	1	3	145	233	1	0	150	0	2.300	0	0	1
2	lya	37	1	2	130	250	0	1	187	0	3.500	0	0	2
3	lya	41	0	1	130	204	0	0	172	0	1.400	2	0	2
4	lya	56	1	1	120	236	0	1	178	0	0.800	2	0	2
5	lya	57	0	0	120	354	0	1	163	1	0.600	2	0	2
6	lya	57	1	0	140	192	0	1	148	0	0.400	1	0	1
7	lya	56	0	1	140	294	0	0	153	0	1.300	1	0	2
8	lya	44	1	1	120	263	0	1	173	0	0	2	0	3
9	lya	52	1	2	172	199	1	1	162	0	0.500	2	0	3
10	lya	57	1	2	150	168	0	1	174	0	1.600	2	0	2
11	lya	54	1	0	140	239	0	1	160	0	1.200	2	0	2
12	lya	48	0	2	130	275	0	1	139	0	0.200	2	0	2
13	lya	49	1	1	130	266	0	1	171	0	0.600	2	0	2
14	lya	64	1	3	110	211	0	0	144	1	1.800	1	0	2
15	lya	58	0	3	150	283	1	0	162	0	1	2	0	2
16	lya	50	0	2	120	219	0	1	158	0	1.600	1	0	2
17	lya	58	0	2	120	340	0	1	172	0	0	2	0	2
18	lya	66	0	3	150	226	0	1	114	0	2.600	0	0	2
19	lya	43	1	0	150	247	0	1	171	0	1.500	2	0	2

**Implementasi**

Dalam penelitian ini, implementasi model Decision Tree C4.5 dilakukan menggunakan aplikasi *RapidMiner*. Perangkat lunak ini digunakan untuk mencari struktur pohon keputusan, melatih model, dan menghasilkan metrik evaluasi kinerja

**Proses**

Proses penelitian ini dirancang menggunakan perangkat lunak *RapidMiner*, dan digambarkan secara sistematis melalui Gambar dibawah.



**Gambar 2. Alur Proses Klasifikasi Data Penyakit Jantung.**

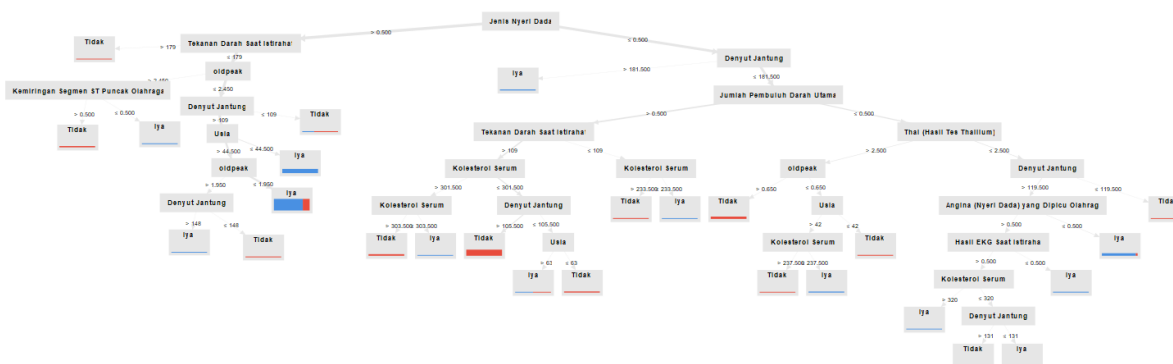
Alur proses dimulai dengan tahap Retrieve data set, yaitu data penyakit jantung

(Retrieve heart), yang berfungsi sebagai sumber masukan data (input) awal. Selanjutnya, data tersebut dialirkan ke operator Split Data untuk untuk dipecah menjadi set pelatihan (training) dan set pengujian (testing). Model klasifikasi dibuat oleh algoritma Decision Tree menggunakan data pelatihan.

Model yang telah dihasilkan oleh Decision Tree kemudian dialirkan ke operator Apply Model bersama dengan data pengujian. Operator Apply Model berfungsi untuk mengaplikasikan model klasifikasi tersebut pada data pengujian.

Tahap terakhir adalah evaluasi kinerja model menggunakan operator Performance. Indikator performa seperti akurasi, presisi, dan recall, serta metrik kinerja lainnya, dihitung oleh operator ini. Dan dari hasil prediksi yang didapat pada tahap Apply Model untuk mendapatkan hasil akhir penelitian (res)."

## Pohon Keputusan



Gambar 3. Pohon Keputusan

Gambar diatas memperlihatkan proses dihasilkan oleh Algoritma C4.5 memperlihatkan proses klasifikasi data uji secara visual. Dari Gambar 1, dapat diidentifikasi bahwa atribut Jenis Nyeri Dada menjadi node akar (pembelah) utama. Hal ini menunjukkan bahwa jenis nyeri dada adalah faktor klinis yang paling informatif (dengan Gain Ratio tertinggi) dalam menentukan kemungkinan penyakit jantung pada dataset ini.

Percabangan utama pada pohon menunjukkan bahwa:

- Cabang Kanan (Jenis Nyeri Dada > 0.500): Mengarah ke node Denyut Jantung dan kemudian Jumlah Pembuluh Darah Utama, menunjukkan bahwa pada pasien dengan jenis nyeri dada atipikal atau non-angina, tingkat denyut jantung dan kondisi pembuluh darah menjadi prediktor kuat.
- Cabang Kiri (Jenis Nyeri Dada < 0.500): Mengarah ke node Tekanan Darah Saat Istirahat, menunjukkan bahwa pada pasien dengan nyeri dada khas (angina tipikal), faktor tekanan darah menjadi penentu penting.

Pohon ini membentuk serangkaian aturan yang dapat diinterpretasikan. Sebagai

contoh, salah satu jalur keputusan yang mengarah ke prediksi Iya (Sakit) adalah: Jika Jenis Nyeri Dada < 0.500 DAN Tekanan Darah Saat Istirahat < 175 dan Oldpeak < 2.400, maka pasien diprediksi memiliki kemungkinan lebih besar terkena penyakit jantung.

### Kinerja Model

Kinerja model Decision Tree C4.5 diukur menggunakan testing set (91 sampel) dari pembagian 70:30. Hasil evaluasi disajikan dalam Matriks Konfusi dan metrik kinerja (Gambar 2 ):

Gambar 4. Akurasi

accuracy: 70.33%			
	true Iya	true Tidak	class precision
pred Iya	42	19	68.85%
pred Tidak	8	22	73.33%
class recall	84.00%	53.66%	

- Akurasi (*Accuracy*): Nilai akurasi keseluruhan model adalah 70.33%. Hal ini menunjukkan bahwa 70.33% dari total sampel data uji berhasil diklasifikasikan dengan benar (42 True Iya + 22 True Tidak).
- Presisi (*Precision*):
  - Presisi untuk kelas Iya adalah 68.85%. Artinya, dari semua pasien yang diprediksi sakit oleh model, 68.85% di antaranya benar-benar sakit.

- Presisi untuk kelas Tidak adalah 73.33%. Artinya, dari semua pasien yang diprediksi sehat oleh model, 73.33% di antaranya benar-benar sehat.
- *Recall* (Sensitivitas):
  - *Recall* untuk kelas Iya (Sakit) adalah 84.00%. Ini menunjukkan kemampuan model yang kuat dalam mengidentifikasi pasien yang benar-benar sakit (Sensitivitas).
  - *Recall* untuk kelas Tidak (Sehat) adalah 53.66%. Ini berarti model kurang efektif dalam mengenali pasien yang benar-benar sehat.

Hasil *Recall* yang tinggi (84.00%) untuk kelas 'Iya' (Sakit) sangat penting dalam konteks medis, karena meminimalkan *False Negative* (pasien sakit diklasifikasikan sebagai sehat). Meskipun akurasi keseluruhan (70.33%) moderat, *recall* yang tinggi menunjukkan bahwa model ini berharga sebagai alat *screening* awal.

### SIMPULAN (PENUTUP)

Penelitian ini berhasil mengusulkan dan mengimplementasikan model prediksi penyakit jantung menggunakan Algoritma Decision Tree C4.5. Pembangunan (atau Penyusunan) model didasarkan pada 303 dataset yang diperoleh dari Kaggle, dan informasi yang tercakup di dalamnya adalah 13 atribut prediktor klinis dan 1 variabel target. Model C4.5 menunjukkan kinerja klasifikasi dengan akurasi keseluruhan sebesar 70.33% pada set pengujian 70%:30. Evaluasi metrik menunjukkan kekuatan model dalam mendeteksi kasus positif (Sakit), dibuktikan dengan nilai *Recall* untuk kelas 'Iya' sebesar 84.00%. Selain itu, atribut Jenis Nyeri Dada teridentifikasi sebagai faktor klinis paling dominan dalam penentuan kemungkinan penyakit jantung, karena dipilih sebagai node akar pohon keputusan. Hasil ini menunjukkan bahwa C4.5 adalah alat yang valid untuk dijadikan sistem pendukung keputusan awal, khususnya dalam mengidentifikasi

sebagian besar pasien yang berisiko tinggi.

### SARAN

Peningkatan Akurasi: Penelitian selanjutnya disarankan untuk menerapkan teknik *Feature Selection* yang lebih canggih atau metode *Ensemble Learning* (seperti *Random Forest*) untuk meningkatkan akurasi keseluruhan model, mengingat akurasi 70.33% masih moderat

Validasi Data: Melakukan pengujian terhadap *dataset* penyakit jantung yang berbeda atau lebih besar, termasuk *dataset* yang lebih seimbang, untuk memverifikasi kemampuan generalisasi model C4.5 dan meningkatkan nilai *Recall* untuk kelas 'Tidak' (Sehat) yang saat ini masih 53.66%.

Implementasi Klinis: Mengembangkan model Decision Tree yang didapatkan menjadi aplikasi berbasis *web* yang dapat diakses oleh praktisi medis, sehingga dapat dimanfaatkan sebagai alat bantu visual dan diagnostik yang cepat.

### DAFTAR PUSTAKA

- Muzakki, F., Ubaydillah, I., Assyiami, N. R., & Soleha, S. (2024). Penerapan Algoritma C4.5 Untuk Prediksi Penyakit Jantung Menggunakan Rapidminer. *Jurnal Komputer Antartika*, 2(2), 71–79.
- Riani, A., Susianto, Y., & Rahman, N. (2019). Implementasi Data Mining untuk Memprediksi Penyakit Jantung Menggunakan Metode Naive Bayes. *JINITA J. Innov. Inf. Technol. Appl.*, 1(1), 25–34.
- Supriyatna, A., Mustika, W. P., & J. (2018). Komparasi Algoritma Naive bayes dan SVM Untuk Memprediksi Keberhasilan Imunoterapi Pada Penyakit Kutil. [Tersedia online]. Tersedia: <http://tunasbangsa.ac.id/ejurnal/index.php/jsakti>
- Mustakim, Rosdina, Ramadhani, D., Afdal, M., & Rahmawita, M. (2021). The Classification Status of River Water Quality in Riau Province Using Modified K-Nearest Neighbor

- Algorithm with STORET Modeling and Water Pollution Index. *Journal of Physics: Conference Series*, 1783(1), 012020.
- Haryanti, M. F., Fauzi, A., Jelita, A. A., Setiyowati, A., Octarina, A., Edina, E. P., Aulia, R. Z., & Fitriana, S. (2024). Pengaruh Data Mining, Strategi Perusahaan Terhadap Laporan Kinerja Perusahaan. *Jurnal Manajemen dan Bisnis (PORTOFOLIO)*, 3(1), 71.
- Larose, D. T., & Larose, C. D. (2014). *Discovering knowledge in data: An introduction to data mining* (2nd ed.). Wiley.
- Wahyuni, S. (2018). Implementation of Data Mining to Analyze Drug Cases Using C4. 5 Decision Tree. In *Journal of Physics: Conference Series*, 970(1), 012030.
- Kusrorong, N. S. B., Sina, D. R., & Rumlaklak, N. D. (2019). Kajian machine learning dengan komparasi klasifikasi prediksi dataset tenaga kerja non-aktif. *Jurnal Ilmu Komputer*, 7(1), 37–49.
- Akbar, F., Saputra, H. W., Maulaya, A. K., Hidayat, M. F., & Rahmaddeni. (n.d.). Implementasi algoritma Decision Tree C4.5 dan Support Vector Regression untuk prediksi penyakit stroke. Program Studi Teknik Informatika, STMIK AMIK Riau, Indonesia.
- Septhya, D., Rahayu, K., Rabbani, S., Fitria, V., Rahmaddeni, I., Irawan, Y., & Hayami, R. (2023). Implementation of Decision Tree Algorithm and Support Vector Machine for Lung Cancer Classification. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 3(1), 15–19.
- Ditendra, E., Suryani, S., Romelah, S., Tanjung, M. H. A., & Sarah, M. (2022). Perbandingan Algoritma Klasifikasi untuk Analisis Sentimen Islam Nusantara di Indonesia: Comparison of Classification Algorithms for Sentiment Analysis of Islam Nusantara in Indonesia. *Malcom: Indonesian Journal of Machine Learning and Computer Science*, 2(1), 71–77.
- Rohman, A., & Rufiyanto, A. (2019). Implementasi Data Mining Dengan Algoritma Decision Tree C4.5 Untuk Prediksi Kelulusan Mahasiswa Di Universitas Pandaran. *Proceeding SINTAK 2019*, 134–139.
- Putri, A. I., Syarif, Y., Jayadi, P., Arrazak, F., & Salisah, F. N. (2023). Implementation of Decision Tree and Support Vector Machine (SVM) Algorithm for Stunting Risk Prediction. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 3(2), 349–357.
- Nofitri, R., & Irawati, N. (2019). ANALISIS DATA HASIL KEUNTUNGAN MENGGUNAKAN SOFTWARE RAPIDMINER. *JURTEKSI (Jurnal Teknologi dan Sistem Informasi)*, 5(2), 199–204. <https://doi.org/10.33330/jurteksi.v5i2.365>
- A. R., A., & O. H. I. (2018). PERFORMANCE EVALUATION OF SELECTED DISTANCE-BASED AND DISTRIBUTION-BASED CLUSTERING ALGORITHMS. *International Journal of Software Engineering and Computer Systems*, 4(2), 38–48. <https://doi.org/10.15282/ijsecs.4.2.2018.3.0047>