Seminar Nasional Informatika (SENATIKA)
Prosiding Senatika 2025 **QRSBN:62-0151-02942-8**

Penerapan Algoritma C4.5 Dalam Mengklasifikasi Penyakit Jantung

Andre Firmansyah[1], Reny Wahyuning Astuti[2], Novhirtamely Kahar[3]
[1,2,3]Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Nurdin Hamzah, Jambi, Indonesia
e-mail: [1]andrebby0612@gmail.com, [2]r3ny4stuti@gmail.com, [3]novmely@gmail.com

Abstrak

Penyakit jantung merupakan penyebab utama kematian di dunia sehingga diperlukan metode deteksi dini yang cepat dan akurat. Penelitian ini bertujuan mengembangkan model klasifikasi penyakit jantung menggunakan algoritma C4.5 sebagai alat bantu diagnosis berbasis data klinis pasien. Data diperoleh dari rekam medis Rumah Sakit Umum Kambang Jambi tahun 2024 serta dataset pendukung dari Kaggle. Atribut yang digunakan meliputi usia, jenis kelamin, tekanan darah, kolesterol, indeks massa tubuh, detak jantung, elektrokardiogram, nyeri dada, kebiasaan merokok, dan riwayat penyakit. Proses penelitian mencakup pra-pemrosesan data, pembagian data menjadi training dan testing, pembangunan model melalui perhitungan entropy dan gain, serta evaluasi menggunakan confusion matrix, akurasi, precision, dan k-fold cross validation. Hasil menunjukkan bahwa algoritma C4.5 mampu menghasilkan pohon keputusan yang mudah dipahami dan efektif dalam memprediksi penyakit jantung. Faktor tekanan darah, kolesterol, dan nyeri dada terbukti dominan dalam klasifikasi. Model ini berpotensi mendukung diagnosis dini penyakit jantung dan dapat dikembangkan lebih lanjut melalui metode ensemble atau optimasi parameter untuk meningkatkan akurasi.

Kata kunci: *: Algoritma C4.5, Data Mining, Penyakit Jantung, Pohon Keputusan, Prediksi*

Abstract

Heart disease remains the leading cause of death worldwide, making it essential to develop early detection methods that are fast and accurate. This study aims to develop a heart disease classification model using the C4.5 algorithm as a decision-support tool based on patients' clinical data. The data were obtained from medical records at Kambang General Hospital Jambi in 2024 and additional datasets from Kaggle. The attributes used include age, gender, blood pressure, cholesterol, body mass index, heart rate, electrocardiogram, chest pain, smoking habits, and medical history. The research process involved data preprocessing, splitting the data into training and testing sets, building the model through entropy and gain calculations, and evaluating it using a confusion matrix, accuracy, precision, and k-fold cross-validation. The results indicate that the C4.5 algorithm can generate an interpretable decision tree that effectively predicts heart disease. Blood pressure, cholesterol, and chest pain were found to be the most dominant factors in the classification. This model shows strong potential as a diagnostic support system for early heart disease detection and can be further improved through ensemble methods or parameter optimization to enhance accuracy.

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

---

1. Pendahuluan

Kemajuan teknologi informasi telah membawa dampak signifikan dalam dunia kesehatan, terutama dalam upaya deteksi dini dan penanganan penyakit. Pemanfaatan teknologi memungkinkan pengolahan data medis secara cepat dan akurat, sehingga dapat membantu

menyelamatkan nyawa manusia melalui peningkatan kualitas diagnosis [1]. Salah satu penyakit yang menjadi perhatian global adalah penyakit jantung. Penyakit ini ditandai oleh gangguan fungsi jantung, termasuk penyakit kardiovaskular, jantung koroner, hingga serangan jantung, yang menurut laporan WHO (*World Health Organization*) masih menjadi penyebab kematian nomor satu di dunia dengan angka sekitar 17,9 juta kematian setiap tahunnya [2].

Faktor risiko penyakit jantung meliputi pola hidup tidak sehat, kurangnya aktivitas fisik, diet tinggi lemak dan garam, kebiasaan merokok, stres, hingga faktor genetik. Selain itu, tekanan darah tinggi dan kadar kolesterol yang tidak terkontrol juga menjadi pemicu utama terjadinya penyakit jantung [3]. Oleh karena itu, diperlukan metode prediksi yang mampu membantu tenaga medis dalam mendeteksi risiko penyakit jantung secara dini dan memberikan hasil analisis yang dapat diinterpretasikan dengan mudah.

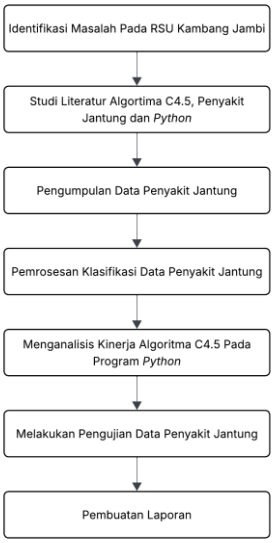
Salah satu pendekatan yang dapat digunakan adalah teknik data mining, yaitu proses penggalian pola dan pengetahuan dari data dalam jumlah besar untuk menghasilkan informasi yang bermanfaat[4]. Dalam konteks penelitian ini, algoritma C4.5 digunakan sebagai salah satu metode utama dalam data mining karena kemampuannya membangun pohon keputusan (*decision tree*) yang mampu mengklasifikasikan data pasien berdasarkan atribut tertentu. Hubungan antara data mining dan algoritma C4.5 sangat erat, di mana algoritma C4.5 berperan sebagai alat analisis dalam tahap klasifikasi data mining untuk menemukan pola hubungan antara faktor risiko dan kemungkinan penyakit jantung.

Beberapa penelitian terdahulu menunjukkan efektivitas algoritma *decision tree* dalam menganalisis data kesehatan. Misalnya, algoritma C4.5 mampu memprediksi penyakit jantung dengan akurasi 65,25% pada skenario split data 90:10 [5]. Penelitian lain membandingkan metode *K-Nearest Neighbor* dan *Logistic Regression*, di *mana Logistic Regression* memperoleh akurasi lebih tinggi sebesar 88%, sementara KNN mencapai 64,03% [6][7]. Selain itu, penelitian lain juga menunjukkan bahwa penggabungan beberapa algoritma melalui metode *ensemble* dapat meningkatkan akurasi prediksi penyakit jantung dibandingkan model tunggal seperti C4.5 [8][9].Berdasarkan hasil-hasil tersebut, masih terdapat peluang untuk meningkatkan performa algoritma C4.5 melalui pemilihan atribut yang lebih relevan, penggunaan data aktual dari instansi medis, serta penerapan teknik evaluasi model yang lebih komprehensif seperti *cross-validation*[10].

Oleh karena itu, penelitian ini bertujuan untuk mengembangkan model klasifikasi penyakit jantung menggunakan algoritma C4.5 dengan memanfaatkan data rekam medis RSU Kambang Jambi tahun 2024 dan data pendukung dari Kaggle. Pembeda utama penelitian ini terletak pada penggunaan data klinis lokal yang lebih representatif terhadap kondisi pasien di Indonesia serta penerapan evaluasi model yang lebih menyeluruh untuk meningkatkan akurasi dan interpretabilitas hasil prediksi. Penelitian ini diharapkan dapat menghasilkan model prediksi yang efisien, mudah dipahami, dan bermanfaat sebagai alat bantu keputusan bagi tenaga medis dalam deteksi dini penyakit jantung.

2. Metode Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan metode studi kasus. Studi dilakukan pada RSU Kambang Jambi dengan memanfaatkan data rekam medis pasien tahun 2024 serta data pendukung dari Kaggle. Atribut yang digunakan dalam penelitian meliputi usia, jenis kelamin, tekanan darah, kolesterol, indeks massa tubuh (imt), nyeri dada, detak jantung, hasil elektrokardiogram (ekg), kebiasaan merokok, riwayat penyakit, dan status diagnosis (target). Pada tahapan penelitian meliputi beberapa langkah utama yaitu :



Gambar 1. Kerangka kerja penelitian

Berdasarkan kerangka kerja penelitian yang telah digambar di atas, maka dapat diuraikan pembahasan masing-masing tahapan dalam penelitian adalah sebagai berikut :

1. Identifikasi Masalah Pada RSU Kambang Jambi
   Mencari tahu kesulitan RSU Kambang Jambi dalam mendiagnosis penyakit jantung secara cepat dan akurat.
2. Studi Literatur Algoritma C4.5, Penyakit Jantung dan Python
   Mempelajari teori tentang algoritma C4.5, faktor risiko penyakit jantung, dan cara pemrograman dengan Python.
3. Pengumpulan Data Penyakit Jantung
   Mengambil data pasien dari rekam medis RSU Kambang dan dataset online (Kaggle).
4. Pemrosesan Klasifikasi Data Penyakit Jantung
   Membersihkan data, menggabungkan data, transformasi data, reduksi data dan melakukan pembagian data menjadi data latih dan data uji untuk melatih dan menguji model.
5. Menganalisis Kinerja Algoritma C4.5 Pada Program Python
   Membangun model prediksi pohon keputusan C4.5 menggunakan Python dan mengukur seberapa akurat model tersebut.
6. Melakukan Pengujian Data Penyakit Jantung
   Menguji model dengan data baru (data uji) untuk memastikan keakuratannya dalam memprediksi penyakit jantung.
7. Pembuatan Laporan
   Tahap ini Menyusun hasil penelitian dalam bentuk laporan dan visualisasi grafik dan memberikan interpretasi tentang hasil klasifikasi penyakit jantung yang ditemukan.

Implementasi penelitian dilakukan dengan bahasa pemrograman Python melalui pustaka *Scikit-learn*, sedangkan perhitungan manual *entropy* dan *gain* dilakukan menggunakan Microsoft Excel sebagai pembanding. Output yang dihasilkan berupa model pohon keputusan, visualisasi struktur pohon, serta hasil prediksi status pasien yang menunjukkan apakah terindikasi penyakit jantung atau tidak.

2.1. Lokasi Penelitian

Penelitian ini dilakukan di RSU Kambang Jambi, dengan data penelitian berupa rekam medis pasien penyakit jantung pada tahun 2024. Selain itu, digunakan juga dataset pendukung dari Kaggle untuk melengkapi atribut dan memperkuat model prediksi.

2.2. Dataset

Dataset yang digunakan mencakup beberapa atribut seperti tabel 1 dibawah ini :

Tabel 1. Dataset

No	Variabel	Sub Kriteria
1	Usia	Dewasa / Lansia
2	Tekanan Darah	Normal / Pre Hipertensi / Hipertensi Stadium 1 / Hipertensi Stadium 2
3	Kolestrol	Normal / Sedang / Tinggi
4	Indeks Massa Tubuh	Kurang / Normal / Kelebihan / Obesitas Kelas 1 / Obesitas Kelas 2 / Obesitas Kelas 3
5	Merokok	Ya / Tidak
6	Riwayat Penyakit	Ada / Tidak
7	Detak Jantung	Bradikardia / Normal / Takikardia
8	Jenis Kelamin	Pria / Wanita
9	Nyeri Dada	Angina Tipikal / Angina Atipikal / Nyeri Non-Angina / Asimptomatik
10	Elektrodiagram	Normal / Abnormalitas / Hipertrofi Ventikel Kiri
11	Target	Ya / Tidak (terindikasi penyakit jantung / tidak terindikasi penyakit jantung)

2.3. Algoritma C4.5

Algoritma C4.5 adalah suatu pengembangan induksi dari ID3 (*Iterative Dichotomiser* 3) karena C4.5 merupakan sebuah algoritma yang dapat digunakan untuk membuat sebuah pohon keputusan (*decision tree*). Pohon-pohon keputusan dibangun berdasarkan kriteria pembentuk suatu keputusan. Pengembangan yang terdapat dalam C4.5 adalah dapat mengatasi nilai suatu atribut yang hilang (*missing value*), dapat mengolah data berjenis diskret maupun *numerik*, menghasilkan aturan yang mudah dalam pengintepretasian serta *prunning* atau pemangkasan [1].

Keunggulan algoritma C4.5 dibandingkan metode lain adalah kemampuannya menangani data dengan atribut numerik maupun kategorikal, serta dapat mengatasi nilai atribut yang hilang (*missing values*). Selain itu, algoritma ini memiliki mekanisme pruning untuk memangkas cabang pohon yang tidak relevan sehingga dapat mengurangi kompleksitas model dan mencegah terjadinya *overfitting* [2].

Beberapa penelitian terdahulu menunjukkan bahwa C4.5 memiliki tingkat akurasi antara 65,25% hingga 94,4%, tergantung pada atribut yang digunakan serta metode validasi yang diterapkan [5][6][7][10]. Meskipun terdapat algoritma lain seperti *K-Nearest Neighbor*, *Random Forest*, atau *Gradient Boosting* yang kadang menghasilkan akurasi lebih tinggi, algoritma C4.5 tetap banyak digunakan karena struktur pohonnya mudah dipahami dan diinterpretasikan, terutama pada konteks medis yang membutuhkan transparansi hasil klasifikasi. Untuk menghitung nilai *entropy* dan *gain* dapat digunakan menggunakan rumus seperti berikut :

1. Menghitung nilai *entropy* dengan menggunakan rumus yang ditunjukkan oleh rumus (i) berikut ini :

$$Entropy\ (S) = \sum_{i=1}^{n} - pi * log2pi \qquad (i)$$

   Keterangan :
   S : himpunan kasus
   n : jumlah partisi himpunan S
   Pi : Kasus Partisi ke - i
2. Mencari nilai *gain* dengan menggunakan rumus yang telah ditunjukkan pada persamaan (ii) berikut ini :

$$Gain(S,A) = -\sum_{i=1}^{n} \frac{|Si|}{|S|} * Entropy\ (Si) \quad (ii)$$

Keterangan :
S : kasus
A : atribut
n : jumlah partisi atribut ke-A
|Si| : jumlah kasus partisi

Setelah selesai menghitung nilai *entropy* dan *gain* maka akan ada atribut yang memiliki nilai gain tertinggi, kemudian nilai *gain* tertinggi akan dipillih menjadi node atau akar untuk membagun pohon keputusan.

3. Hasil dan Pembahasan

Ada beberapa tahapan yang harus dilakukan sebelum melakukan klasifikasi menggunakan algortima c4.5, tahap awal yang dilakukan adalah *preprocessing data* untuk memastikan data yang digunakan bersih, lengkap dan siap diolah. Proses ini terdiri dari beberapa langkah utama, yaitu:

1. *Data Cleaning* (Pembersihan Data)
   Menghapus data duplikat, memperbaiki nilai yang hilang (*missing value*), serta menyesuaikan data yang tidak konsisten agar seluruh atribut memiliki format seragam.
2. *Data Integration* (Penggabungan Data)
   Menggabungkan dua sumber data, yaitu rekam medis pasien RSU Kambang Jambi tahun 2024 dan dataset pendukung dari Kaggle, agar data yang digunakan lebih bervariasi dan representatif.
3. *Data Transformation* (Transformasi Data)
   Mengubah data numerik seperti tekanan darah, kolesterol, dan indeks massa tubuh (IMT) ke dalam bentuk kategorikal (misalnya *normal*, *tinggi*, *rendah*) agar sesuai dengan karakteristik algoritma C4.5.
4. *Data Splitting* (Pembagian Data)
   Dataset dibagi menjadi dua bagian, yaitu *data training* dan *data testing* dengan beberapa rasio perbandingan (50:50 hingga 90:10) untuk mengevaluasi stabilitas model.

Tahapan *preprocessing* ini penting agar model yang dihasilkan tidak bias dan memiliki performa klasifikasi yang optimal.

Setelah data siap, model dibangun menggunakan algoritma c4.5 yang bekerja dengan menghitung nilai *entropy* dan *gain* untuk menentukan atribut terbaik sebagai akar pohon keputusan. Atribut dengan nilai *gain* tertinggi menjadi simpul utama, kemudian proses berlanjut hingga seluruh data terklasifikasi. Model diimplementasikan menggunakan bahasa pemrograman Python dan pustaka *Scikit-learn*. Proses pelatihan menghasilkan pohon keputusan (*decision tree*) yang merepresentasikan hubungan antar atribut seperti tekanan darah, kolesterol, dan nyeri dada terhadap risiko penyakit jantung.

Pengujian dilakukan menggunakan data testing untuk menilai performa model. Hasil pengujian menunjukkan bahwa tidak seluruh data testing diklasifikasikan dengan benar, yang berarti masih terdapat tingkat kesalahan prediksi (*misclassification*). Evaluasi model dilakukan menggunakan *confusion matrix*, akurasi. Pada tabel 2 menunjukkan hasil klasifikasi pada data testing.

Tabel 2. Hasil klasifikasi testing

No	Usia	Jk	Nyeri dada	Tekanan darah	Kolestrol	Ekg	Denyut jantung	Imt	Merokok	Riwayat	target
1	dewasa	pria	nyeri non-angina	hipertensi stadium 2	tinggi	abnormalitas gelombang st-t	takikardia	normal	tidak	ada	tidak
2	dewasa	pria	nyeri non-angina	normal	normal	normal	takikardia	obesitas kelas 3	ya	tidak	tidak
3	dewasa	pria	nyeri non-angina	hipertensi stadium 2	tinggi	abnormalitas gelombang st-t	takikardia	kelebihan	tidak	ada	tidak
4	dewasa	pria	nyeri non-angina	hipertensi stadium 1	normal	abnormalitas gelombang st-t	takikardia	kelebihan	ya	tidak	ya
5	dewasa	wanita	nyeri non-angina	pre hipertensi	tinggi	abnormalitas gelombang st-t	takikardia	normal	tidak	ada	ya
…	…	…	…	…	…	…	…	…	…	…	…
…	…	…	…	…	…	…	…	…	…	…	…
20	20	lansia	wanita	nyeri non-angina	hipertensi stadium 2	normal	Hipertrofi ventrikel kiri	takikardia	normal	ya	tidak

Berdasarkan hasil evaluasi, algoritma C4.5 memperoleh tingkat akurasi terbaik sebesar 83,33% pada perbandingan data training dan testing 70:30. Pada Hasil menunjukkan bahwa model memiliki kemampuan klasifikasi yang baik dan stabil. Pada Tabel 3 berikut menampilkan hasil perbandingan akurasi pada berbagai rasio pembagian data.

Tabel 3. Akurasi beberapa rasio data training – testing

Rasio Training	Rasio Testing	Akurasi Training	Akurasi Testing
50	50	100 %	70 %
60	40	98.33 %	72.50 %
70	30	98.57 %	83.33 %
80	20	98.75 %	75 %
90	10	98.89 %	80 %

Dari hasil diatas terlihat bahwa model memiliki kemampuan belajar yang sangat baik pada data training (akurasi di atas 98%), namun dapat terjadi *overfitting* jika porsi data testing terlalu kecil. Oleh karena itu, rasio 70:30 dianggap paling ideal karena menghasilkan keseimbangan antara kemampuan model memahami pola dan kemampuan generalisasi terhadap data baru. Selanjutnya dilakukan pengujian terhadap data baru (prediksi) untuk menguji kemampuan model dalam mengklasifikasikan data tanpa label target. Data prediksi ditunjukkan pada Tabel 4 berikut.

Tabel 4. Data Prediksi

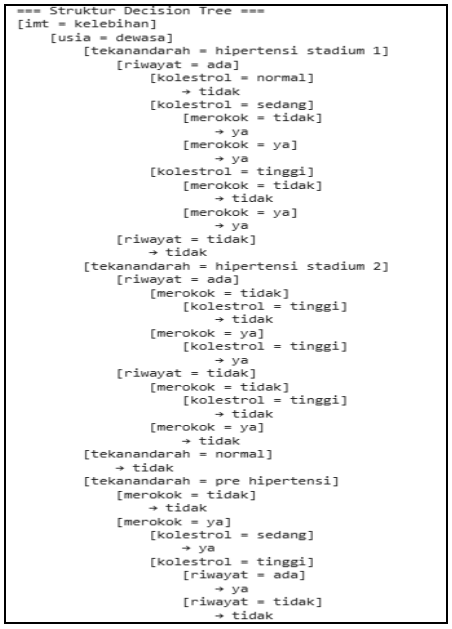
No	Usia	Jk	Nyeri Dada	Tekanan Darah	Kolestrol	Ekg	Denyut Jantung	Imt	Merokok	Riwayat	Target
1	lansia	pria	angina atipikal	pre hipertensi	tinggi	normal	takikardia	obesitas kelas 1	tidak	tidak	?
2	dewasa	wanita	nyeri non-angina	normal	sangat tinggi	normal	takikardia	normal	tidak	tidak	?

Hasil prediksi dari data ditas ditunjukkan pada Tabel 5 berikut.

Tabel 5. Hasil Prediksi

No	Usia	Jk	Nyeri Dada	Tekanan Darah	Kolestrol	Ekg	Denyut Jantung	Imt	Merokok	Riwayat	Target
1	lansia	pria	angina atipikal	pre hipertensi	tinggi	normal	takikardia	obesitas kelas 1	tidak	tidak	ya
2	dewasa	wanita	nyeri non-angina	normal	sangat tinggi	normal	takikardia	normal	tidak	tidak	tidak

Berdasarkan hasil prediksi, pasien pertama (lansia, pria, tekanan darah *pre-hipertensi*, kolesterol tinggi) diprediksi berisiko tinggi terkena penyakit jantung (Ya), sedangkan pasien kedua (dewasa, wanita, tekanan darah normal, kolesterol sangat tinggi) diprediksi tidak berisiko signifikan (Tidak). Hasil analisis menunjukkan bahwa tekanan darah dan kadar kolesterol merupakan faktor paling dominan dalam menentukan hasil klasifikasi, diikuti oleh jenis nyeri dada. Atribut-atribut tersebut menghasilkan pola keputusan yang kemudian divisualisasikan dalam bentuk pohon keputusan seperti pada Gambar 2.



Gambar 2. Pohon Keputusan

Berdasarkan hasil pembentukan model algoritma c4.5, diperoleh struktur pohon keputusan seperti terlihat pada gambar 2. Pohon ini menunjukkan bahwa atribut Indeks Massa Tubuh (IMT) menjadi akar utama (root) dari pohon, dengan cabang pertama menuju kategori kelebihan berat badan. Hal ini menunjukkan bahwa kondisi berat badan merupakan faktor awal yang signifikan dalam menentukan risiko penyakit jantung.

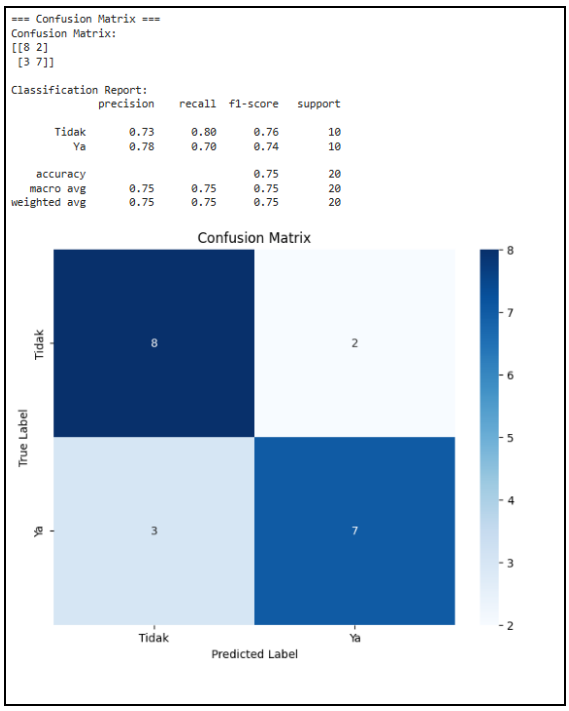
Dari atribut *IMT*, cabang berikutnya menunjukkan bahwa faktor usia juga berperan penting, terutama pada kelompok usia dewasa. Pada cabang usia dewasa, atribut tekanan darah (tekanandarah) menjadi faktor pembeda berikutnya dengan beberapa kategori, antara lain *hipertensi stadium 1*, *hipertensi stadium 2*, *pre-hipertensi*, dan *normal*. Setiap kategori tekanan darah kemudian bercabang ke atribut riwayat penyakit, kolesterol, serta kebiasaan merokok (merokok). Struktur keputusan tersebut menunjukkan bahwa:

1. Jika IMT = *kelebihan*, Usia = *dewasa*, Tekanan Darah = *hipertensi stadium 1*, Riwayat = *ada*, Kolesterol = *tinggi*, dan Merokok = *ya*, maka hasil klasifikasi adalah “Ya” (terindikasi penyakit jantung).
2. Jika IMT = *kelebihan*, Usia = *dewasa*, Tekanan Darah = *normal*, dan Kolesterol = *sedang*, maka hasil klasifikasi adalah “Tidak” (tidak terindikasi penyakit jantung).

3. Jika Tekanan Darah = *hipertensi stadium 2* dan Kolesterol = *tinggi*, baik dengan atau tanpa riwayat penyakit, maka model cenderung memberikan hasil "Ya", menandakan bahwa kolesterol tinggi dan tekanan darah tinggi memiliki kontribusi besar terhadap risiko penyakit jantung.
4. Jika Tekanan Darah = *pre-hipertensi* dan Merokok = *tidak*, maka hasil klasifikasi cenderung "Tidak", namun jika kolesterol tinggi dan ada riwayat penyakit, maka hasilnya "Ya".

Dari pola tersebut, dapat disimpulkan bahwa atribut tekanan darah, kolesterol, riwayat penyakit, dan kebiasaan merokok merupakan atribut dengan *information gain* tertinggi setelah IMT dan usia. Pohon keputusan ini menunjukkan interaksi antar faktor risiko yang realistis dan mudah diinterpretasikan oleh tenaga medis. Pola cabang yang berulang pada atribut tekanan darah dan kolesterol menunjukkan bahwa kedua variabel tersebut merupakan determinan utama dalam klasifikasi. Selain itu, kondisi merokok juga sering muncul sebagai faktor pembeda akhir, yang memperkuat temuan bahwa kebiasaan merokok memperbesar peluang seseorang memiliki penyakit jantung. Dengan demikian, model yang dihasilkan tidak hanya memberikan hasil klasifikasi, tetapi juga dapat dijadikan dasar pengambilan keputusan medis. Pohon keputusan ini mampu menjelaskan secara transparan bagaimana setiap atribut berkontribusi terhadap hasil akhir, menjadikannya alat bantu diagnosis yang interpretatif dan berbasis data.

Berikut tampilan hasil dari *confusion matrix* yang dihasilkan seperti gambar 3 dibawah ini:



Gambar 3. *Confusion matrix*

4. Kesimpulan

Penelitian ini berhasil mencapai tujuannya dengan menghasilkan model klasifikasi penyakit jantung menggunakan algoritma C4.5 yang akurat dan mudah diinterpretasikan. Model memberikan hasil terbaik dengan akurasi 83,33%, menunjukkan bahwa algoritma C4.5 efektif digunakan sebagai alat bantu diagnosis dini penyakit jantung. Namun, penelitian ini masih memiliki keterbatasan yang perlu diperbaiki. Disarankan untuk meningkatkan kualitas dan jumlah data, menambah variasi atribut seperti aktivitas fisik dan pola makan, serta menerapkan metode *ensemble* atau algoritma lain seperti *Naïve Bayes*, *SVM*, dan *Jaringan Saraf Tiruan* untuk meningkatkan akurasi dan mengurangi *overfitting*. Selain itu, penggunaan *cross-*

validation dan perbandingan dengan perangkat lunak *data mining* lain seperti Weka, Orange, atau Matlab perlu dilakukan guna memastikan konsistensi dan keandalan model.

Daftar Pustaka

[1] S. R. Cholil, A. F. Dwijayanto, and T. Ardianita, "PREDIKSI PENYAKIT DEMAM BERDARAH DI PUSKESMAS NGEMPLAK SIMONGAN MENGGUNAKAN ALGORITMA C4.5," Jurnal Sistem Informasi, vol. 9, p. 529, Sep. 2020.

[2] A. Sepharni, I. E. Hendrawan, and C. Rozikin, "KLASIFIKASI PENYAKIT JANTUNG DENGAN MENGGUNAKAN ALGORITMA C4.5," STRING (Satuan Tulisan Riset dan Inovasi Teknologi), vol. 7, Dec. 2022, Accessed: Mar. 11, 2025. [Online]. Available: https://www.kaggle.com/fedesoriano/heart-failure-prediction

[3] W. S. Naomi, I. Picauly, and S. M. Toy, "FAKTOR RISIKO KEJADIAN PENYAKIT JANTUNG KORONER (Studi Kasus di RSUD Prof. Dr. W. Z. Johannes Kupang)," Media Kesehatan Masyarakat, vol. 3, 2021.

[4] S. Setyaningtyas, B. I. Nugroho, and Z. Arif, "TINJAUAN PUSTAKA SISTEMATIS: PENERAPAN DATA MINING TEKNIK CLUSTERING ALGORITMA K-MEANS," Jurnal Teknoif Teknik Informatika Institut Teknologi Padang, vol. 10, no. 2, pp. 52–61, Oct. 2022, doi: 10.21063/jtif.2022.v10.2.52-61.

[5] F. Muzakki, I. Ubaydillah, N. R. Assyiami, and S. Soleha, "Penerapan Algoritma C4.5 Untuk Prediksi Penyakit Jantung Menggunakan Rapidminer," Jurnal Komputer Antartika, vol. 2, pp. 71–79, 2024, [Online]. Available: https://ejournal.mediaantartika.id/index.php/jka

[6] D. Sitanggang, N. Nicholas, V. Wilson, A. R. A. Sinaga, and A. D. Simanjuntak, "IMPLEMENTASI DATA MINING UNTUK MEMPREDIKSI PENYAKIT JANTUNG MENGGUNAKAN METODE K-NEAREST NEIGHBOR DAN LOGISTIC REGRESSION," Jurnal Teknik Informasi dan Komputer (Tekinkom), vol. 5, no. 2, p. 493, Dec. 2022, doi: 10.37600/tekinkom.v5i2.698.

[7] A. Yogianto, A. Homaidi, and Z. Fatah, "Implementasi Metode K-Nearest Neighbors (KNN) untuk Klasifikasi Penyakit Jantung," G-Tech: Jurnal Teknologi Terapan, vol. 8, no. 3, pp. 1720–1728, Jul. 2024, doi: 10.33379/gtech.v8i3.4495.

[8] S. R. Patil and A. K. Birajdar, "Heart Disease Prediction using Machine Learning and Data Mining Techniques: A Review," *International Journal of Computer Applications*, vol. 183, no. 32, pp. 1–6, 2022, doi: 10.5120/ijca2022912365.

[9] R. A. Nugraha, F. R. Saputra, and M. Kurniawan, "Penerapan Metode Ensemble Random Forest untuk Prediksi Penyakit Jantung," *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK)*, vol. 11, no. 2, pp. 150–157, 2024, doi: 10.25126/jtiik.112.150157.

[10] C. N. Syahputri and M. S. Hasibuan, "OPTIMASI KLASIFIKASI DECISION TREE DENGAN TEKNIK PRUNING UNTUK MENGURANGI OVERFITTING," JSiI (Jurnal Sistem Informasi), vol. 11, no. 2, pp. 87–96, Sep. 2024, doi: 10.30656/jsii.v11i2.9161.