



## Analisis Model Klasifikasi Penyakit Jantung dengan Decision Tree C4.5 Berdasarkan Pendekatan Sentralisasi dan Federated Learning pada Dataset UCI

Aditia Prakoso<sup>1</sup> Kurniawati<sup>2</sup>

Fakultas Sains dan Teknologi, Universitas IVET, Indonesia

### Info Articles

*Sejarah Artikel:*

Disubmit: 3 Mei 2025

Direvisi: 5 Juni 2025

Disetujui: 20 Juni 2025

*Keywords:*

Heart Disease, Centralized and Federated Learning Approach, C4.5 Algorithm

### Abstrak

Penyakit jantung merupakan penyebab utama kematian global dengan kontribusi sekitar 17,9 juta atau sekitar 32% kematian per tahun. Penelitian ini bertujuan mengembangkan model klasifikasi penyakit jantung menggunakan algoritma Decision Tree C4.5 berdasarkan tiga atribut utama: chest pain (cp), usia (age), dan kadar kolesterol (chol) dari dataset Heart Disease UCI. Metodologi penelitian meliputi preprocessing data, pembagian dataset menjadi 70-80% data latih dan 20-30% data uji, perhitungan entropy dan information gain, serta pembentukan pohon keputusan menggunakan algoritma C4.5. Hasil penelitian menunjukkan bahwa atribut chest pain (cp) menjadi simpul akar yang paling informatif dalam klasifikasi, diikuti oleh usia dan kadar kolesterol sebagai pembeda lanjutan. Model menghasilkan empat jalur klasifikasi utama yang mudah diinterpretasi: (1) pasien dengan angina berusia  $\leq 50$  tahun diklasifikasikan sebagai tidak berisiko, (2) pasien dengan angina berusia  $> 50$  tahun diklasifikasikan berisiko, (3) pasien asimtomatik dengan kolesterol  $\leq 250$  mg/dL tidak berisiko, dan (4) pasien asimtomatik dengan kolesterol  $> 250$  mg/dL berisiko. Keunggulan model ini terletak pada kesederhanaannya yang hanya menggunakan tiga atribut, sehingga efisien dan dapat diimplementasikan pada lingkungan dengan keterbatasan sumber daya seperti puskesmas atau aplikasi mobile. Model ini berpotensi sebagai sistem pendukung keputusan untuk diagnosis awal penyakit jantung.

### Abstract

Heart disease is a leading cause of global mortality, contributing approximately 17.9 million deaths or about 32% of deaths annually. This study aims to develop a heart disease classification model using the Decision Tree C4.5 algorithm based on three main attributes: chest pain (cp), age, and cholesterol level (chol) from the UCI Heart Disease dataset. The research methodology includes data preprocessing, dataset splitting into 70-80% training data and 20-30% testing data, calculation of entropy and information gain, and decision tree formation using the C4.5 algorithm. Results show that the chest pain (cp) attribute becomes the most informative root node in classification, followed by age and cholesterol level as further differentiators. The model generates four main classification paths that are easily interpretable: (1) patients with angina aged  $\leq 50$  years are classified as not at risk, (2) patients with angina aged  $> 50$  years are classified as at risk, (3) asymptomatic patients with cholesterol  $\leq 250$  mg/dL are not at risk, and (4) asymptomatic patients with cholesterol  $> 250$  mg/dL are at risk. The model's advantage lies in its simplicity using only three attributes, making it efficient and implementable in resource-limited environments such as community health centers or mobile applications. This model has potential as a decision support system for early heart disease diagnosis.

<sup>✉</sup> Alamat Korespondensi:  
Aditiaprakosorafael@gmail.com

## PENDAHULUAN

Penyakit jantung merupakan salah satu penyebab utama kematian di seluruh dunia. Berdasarkan data World Health Organization (WHO), penyakit kardiovaskular menyumbang sekitar 17,9 juta kematian setiap tahunnya, atau sekitar 32% dari total kematian global. Kondisi ini menunjukkan urgensi dalam deteksi dini dan diagnosis yang akurat guna menurunkan angka mortalitas dan meningkatkan kualitas hidup pasien.

Dalam beberapa tahun terakhir, perkembangan teknologi kecerdasan buatan (Artificial Intelligence, AI) dan machine learning (ML) telah membuka peluang baru dalam pemanfaatan data klinis untuk tujuan prediksi penyakit. Berbagai algoritma telah dikembangkan untuk mendeteksi pola tersembunyi dalam data medis, salah satunya adalah algoritma Decision Tree. Menurut penelitian Wibowo, dkk (2024) dan Chicco, D., & Jurman, G. (2020), & Fais, M. A., dkk (2023) algoritma Decision Tree dikenal dan banyak digunakan karena kemampuannya menghasilkan model yang mudah dipahami dan memiliki nilai interpretatif tinggi

Dataset Heart Disease dari University of California, Irvine (UCI) merupakan salah satu dataset terbuka yang banyak digunakan dalam penelitian prediksi penyakit jantung. Dataset ini terdiri dari 303 data pasien dengan 14 atribut klinis, meliputi usia, jenis kelamin, tekanan darah, kadar kolesterol, dan berbagai hasil pemeriksaan lainnya. Idealnya, seluruh atribut tersebut dapat digunakan untuk membangun model klasifikasi yang komprehensif. Namun, dalam beberapa kondisi tertentu, penggunaan seluruh atribut justru dapat menimbulkan risiko *overfitting*, yaitu kondisi ketika model terlalu menyesuaikan diri dengan data pelatihan hingga kehilangan kemampuan untuk menggeneralisasi pada data baru. *Overfitting* sering kali menghasilkan akurasi tinggi pada data pelatihan tetapi buruk pada data nyata karena model "belajar terlalu detail", termasuk pada *noise* (data yang mengandung kesalahan atau fluktuasi acak yang tidak relevan) dan anomali (kasus-kasus langka yang menyimpang dari pola umum) hal tersebut yang diungkapkan oleh Myint, K. L., & Htake Khaung Tin, H. (2021) dan Chicco, D., & Jurman, G. (2020) dalam penelitiannya. Selain itu, menurut Idri, A. (2015) penggunaan terlalu banyak atribut juga dapat menurunkan interpretabilitas model dan meningkatkan kompleksitas perhitungan, sehingga menyulitkan implementasi di fasilitas layanan kesehatan dengan keterbatasan sumber daya.

Berdasarkan konteks tersebut, penelitian ini menerapkan pendekatan penyederhanaan model (model simplification) dengan hanya menggunakan tiga atribut utama, yaitu *chest pain* (cp), *cholesterol* (chol), dan *age* (usia), yang secara medis terbukti memiliki keterkaitan kuat terhadap diagnosis penyakit jantung. Pemilihan ini tidak dilakukan secara sembarangan, melainkan berdasarkan kombinasi antara pertimbangan *domain knowledge* (literatur medis) dan hasil seleksi fitur awal menggunakan *information gain*. Menurut Gunawan, I. M. A. O., (2023), Sepharni, A., (2023) & Jannah, S. S. N., (2023) menjelaskan bahwa ketiga atribut ini dipilih karena: (1) cp mewakili gejala klinis khas pada pasien jantung, (2) chol adalah indikator utama profil lipid yang berkaitan dengan risiko aterosklerosis, dan (3) *age* merupakan faktor risiko utama yang tidak dapat dimodifikasi.

Walaupun model yang dikembangkan dalam penelitian ini hanya menggunakan sebagian dari 14 atribut yang tersedia, pendekatan ini bertujuan menghasilkan model yang sederhana, efisien, dan

mudah diinterpretasikan, terutama untuk implementasi dalam lingkungan pelayanan kesehatan dasar seperti puskesmas, atau sistem berbasis aplikasi ringan (*mobile/web*).

Namun demikian, pendekatan ini tetap memiliki keterbatasan. Oleh karena itu, hasil dari penelitian ini tidak dimaksudkan untuk menggantikan model yang menggunakan seluruh atribut, tetapi sebagai landasan awal bahwa dengan atribut minimal pun model klasifikasi tetap dapat bekerja secara efektif. Pengembangan lebih lanjut disarankan untuk menguji ulang performa model dengan melibatkan seluruh atribut agar diperoleh hasil klasifikasi yang lebih komprehensif.

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk menganalisis model klasifikasi penyakit jantung menggunakan algoritma Decision Tree C4.5, dengan fokus pada tiga atribut utama yang relevan secara klinis guna membentuk pohon keputusan yang sederhana namun informatif. Penelitian ini juga mengadopsi prinsip interpretabilitas model serta mempertimbangkan pendekatan terpusat dan *federated learning*, sebagaimana diusulkan oleh Rodriguez (2024) & Abdar, M. (2015), dengan harapan dapat berkontribusi pada pengembangan sistem diagnosis awal yang akurat, adaptif, dan efisien.

## METODE PENELITIAN

### 1. Sumber Data

Dalam penelitian ini menggunakan dataset *Heart Disease* dari “*UCI Machine Learning Repository*”, yang secara keseluruhan berisi 303 baris data pasien dengan 14 atribut klinis, yaitu:

**Tabel 1. Daftar Atribut Dataset**

Atribut	Keterangan
Age	Usia
Sex	Jenis Kelamin
Cp	Jenis nyeri dada ( <i>chest pain type</i> )
Trestbps	Tekanan darah istirahat ( <i>resting blood pressure</i> )
Chol	Kadar kolestrol
Fbs	Gula darah
Restecg	Hasil elektrokardiografi istirahat ( <i>resting ECG</i> )
Thalach	Detak jantung maksimum yang dicapai ( <i>maximum heart rate achieved</i> )
Exang	Angina (nyeri dada saat olahraga )
oldpeak	Depresiasi ST terkait olahraga ( <i>ST depression induced by exercise</i> )
slope	Kemiringan segmen ST ( <i>slope of peak exercise ST segment</i> )
Ca	Jumlah pembuluh darah utama yang diwarnai fluoroskopi ( <i>number of major vessels</i> )
Thal	Jenis kelainan darah

Target	Hasil akhir diagnosis penyakit
--------	--------------------------------

Namun, dalam penelitian ini hanya digunakan tiga atribut utama, yaitu:

- Cp (*Chest pain type*) : Jenis nyeri dada (kategori 1-4)
- Age (usia) : Usia pasien (dalam tahun)
- Chol (kadar kolestrol) : Kadar kolestrol (mg/dL)

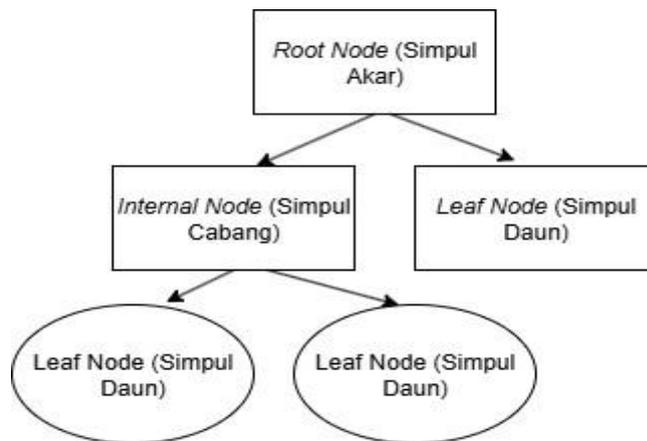
Target klasifikasi adalah variabel **target**, yang bernilai:

- 0 : tidak mengidap penyakit jantung
- 1 : mengidap penyakit jantung

**2. Metode Klasifikasi: Decision Tree C4.5**

Decision Tree merupakan salah satu metode prediktif dalam *machine learning* yang digunakan untuk membuat keputusan berdasarkan serangkaian aturan bercabang. Model ini sangat populer dalam tugas klasifikasi karena hasilnya dapat divisualisasikan dalam bentuk struktur pohon yang mudah dipahami oleh manusia (Wibowo, A. C., Lestari, S. A., & Nurchim, N. (2024) dan García, D., Luengo, J., & Herrera, F. 2016). Setiap simpul pada pohon mewakili pengujian terhadap suatu atribut, sedangkan setiap simpul daun mewakili prediksi kelas akhir. Decision Tree dapat diterapkan baik untuk tugas klasifikasi maupun regresi, dan banyak digunakan dalam bidang kesehatan karena interpretabilitasnya yang tinggi. Struktur dasar *Decision Tree* terdiri dari: Root node (simpul akar): atribut utama yang digunakan untuk pemisahan pertama, - Internal nodes (simpul internal): atribut lain yang digunakan untuk percabangan lebih lanjut, - Leaf nodes (simpul daun): hasil klasifikasi akhir.

Setiap simpul pada pohon mewakili hasil pengujian terhadap suatu atribut, dan setiap daun mewakili nilai prediksi. Struktur umum pohon keputusan ditunjukkan pada Gambar 1 berikut:



**Gambar 1.** Struktur Pohon Keputusan (*Decision Tree*) Perhitungan Gain dan Entropy

Penilaian efektivitas suatu atribut dalam membagi data dilakukan menggunakan information gain, yang dinyatakan dalam Persamaan (1) :

$$\text{Gain}(S,A)=\text{Entropy}(S)-\sum_{v \in \text{Values}(A)} \frac{|S_v|}{|S|} \cdot \text{Entropy}(S_v) \dots \dots (1)$$

Di mana :

- S : Himpunan data,
- A : Atribut yang diuji,
- Sv : Subset data dengan nilai atribut A=v,
- |S| : Jumlah total data,
- |Sv| : Jumlah data pada subset v.

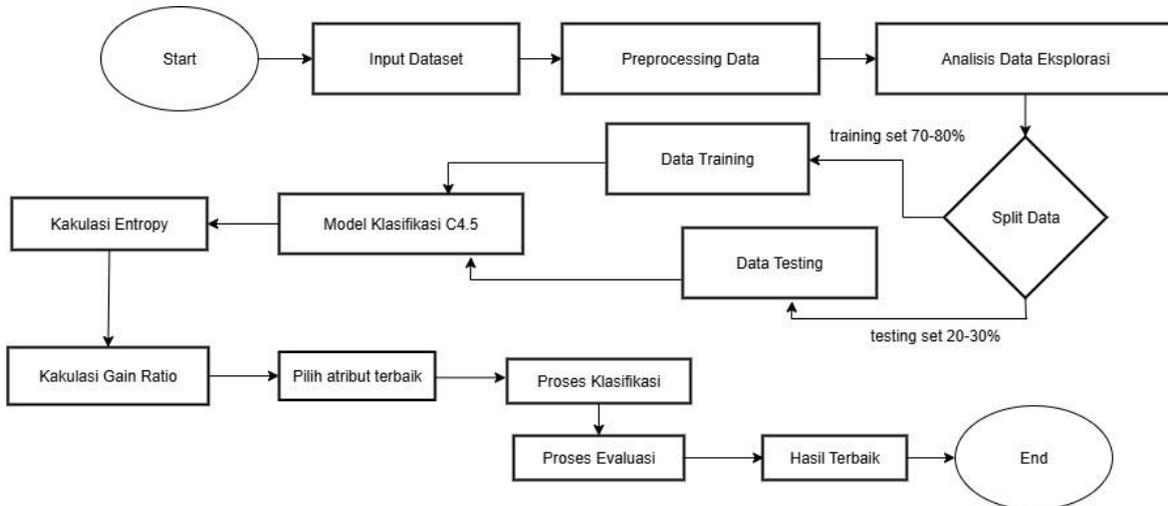
Sebelum menghitung gain, terlebih dahulu dihitung nilai entropy dari dataset, seperti dalam Persamaan (2):

$$\text{Entropy}(S)=-p^+ \log_2(p^+) - p^- \log_2(p^-) \dots \dots (2)$$

Di mana:

- p+ : proporsi data dengan kelas positif (target = 1), -
- p- : proporsi data dengan kelas negatif (target = 0).

### 3. Langkah Penelitian



**Gambar 2.** Alur Penelitian

Penelitian ini dilaksanakan melalui serangkaian tahapan sistematis sebagaimana ditunjukkan pada **Gambar 2**. Langkah pertama dimulai dengan menginput dataset, yaitu *Heart Disease Dataset* dari UCI Machine Learning Repository. Dataset ini awalnya terdiri dari 14 atribut, namun dalam penelitian ini hanya digunakan tiga atribut utama, yakni *cp* (*chest pain*), *age* (usia), dan *chol* (kadar kolesterol), yang dipilih berdasarkan pertimbangan medis dan kontribusi terhadap klasifikasi.

Selanjutnya dilakukan proses preprocessing data, yang mencakup pengisian atau penghapusan nilai yang hilang (*missing values*), pengecekan duplikasi data, konversi format data jika diperlukan, serta pemilihan atribut. Setelah itu, dilakukan analisis data eksplorasi untuk memahami distribusi nilai pada atribut, mendeteksi outlier, dan mengamati hubungan awal antar variabel.

Dataset kemudian dibagi menjadi dua subset dalam tahap split data, yaitu data latih (*training set*) sebesar 70–80% dan data uji (*testing set*) sebesar 20–30%. Pembagian ini bertujuan untuk melatih model klasifikasi menggunakan sebagian data, dan mengevaluasi kinerja model dengan data yang belum pernah dilihat sebelumnya.

Pada tahap selanjutnya, dilakukan kalkulasi entropy untuk mengukur tingkat ketidakpastian dalam dataset. Nilai entropy menjadi dasar perhitungan information gain, yang kemudian dinormalisasi melalui kalkulasi gain ratio. Atribut dengan gain ratio tertinggi dipilih sebagai simpul utama dalam proses pembentukan pohon keputusan.

Model klasifikasi kemudian dibangun menggunakan algoritma Decision Tree C4.5. Pohon keputusan ini disusun berdasarkan proses rekursif, di mana atribut-atribut dipilih secara bertahap berdasarkan nilai gain ratio hingga seluruh data dapat terklasifikasi atau tidak ada atribut yang tersisa.

Setelah model terbentuk, dilakukan proses klasifikasi pada data uji untuk melihat performa prediktif dari model yang telah dibangun. Kinerja model kemudian dianalisis melalui proses evaluasi, dengan menggunakan metrik evaluasi seperti akurasi, precision, recall, dan confusion matrix. Hasil dari evaluasi ini digunakan untuk menentukan hasil terbaik, yaitu model klasifikasi yang paling optimal dalam mendeteksi penyakit jantung berdasarkan ketiga atribut yang digunakan.

Langkah-langkah tersebut diakhiri pada tahap end, menandai selesainya proses penelitian dan pembentukan model klasifikasi yang siap diinterpretasikan dan diterapkan dalam sistem pendukung keputusan medis.

## HASIL DAN PEMBAHASAN

### A. Hasil

Pada bagian ini disajikan hasil dari proses preprocessing data serta visualisasi model klasifikasi menggunakan algoritma *Decision Tree C4.5*. Penelitian difokuskan pada tiga atribut utama, yaitu cp (*chest pain*), age (usia), dan chol (kadar kolesterol), dengan target sebagai label kelas (0 = tidak berisiko, 1 = berisiko). Meskipun dataset awal memiliki 14 atribut, pemilihan hanya 3 atribut dilakukan untuk menyederhanakan model tanpa mengurangi relevansi medis dari klasifikasi.

#### 1. Data Preprocessing

Sebelum dilakukan proses klasifikasi, dataset terlebih dahulu melalui tahap preprocessing (pembersihan dan persiapan data) untuk memastikan kualitas, konsistensi, dan kesesuaian data terhadap algoritma yang digunakan. Tahapan preprocessing yang dilakukan meliputi:

- a. Seleksi Atribut
- b. Dari 14 atribut yang tersedia pada dataset UCI Heart Disease, hanya dipilih 3 atribut utama yang dianggap paling relevan secara klinis dan informatif, yaitu:

- Cp (Chest pain),
- Age (usia),
- Chol(kadar kolestrol),

Atribut target tetap digunakan sebagai label keluaran untuk proses klasifikasi.

c. Pengecekan Nilai Kosong (*Missing Values*)

Dataset diperiksa untuk memastikan tidak terdapat nilai kosong (*null*) pada atribut yang digunakan. Apabila ditemukan, maka dilakukan penanganan sebagai berikut:

- Menghapus baris data yang tidak lengkap, atau
- Melakukan imputasi menggunakan rata-rata (*mean*) untuk atribut numerik (chol, age), modus (nilai terbanyak) untuk atribut kategorikal (cp), atau berdasarkan perhitungan information gain jika diperlukan

Dalam hasil preprocessing ini, tidak ditemukan nilai kosong, sehingga seluruh baris data dapat digunakan

d. Pengkodean Atribut Kategorik

Atribut cp (chest pain) sudah dalam bentuk kategorik angka (1–4) sehingga tidak diperlukan encoding ulang.

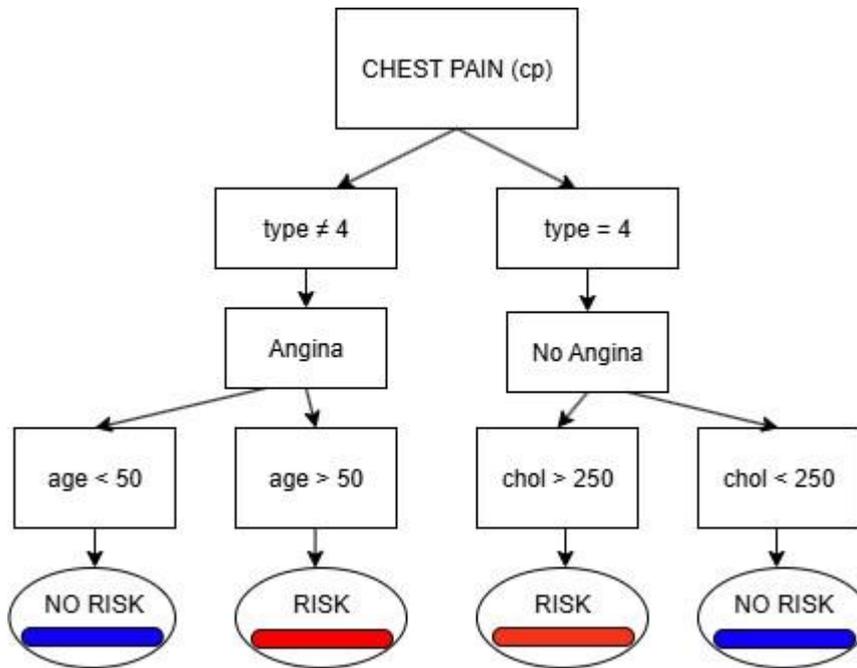
Hasil dari preprocessing ini adalah dataset bersih dan siap digunakan untuk klasifikasi menggunakan algoritma *Decision Tree C4.5*. Tabel berikut menunjukkan data yang telah di preprocessing dan digunakan pada penelitian ini:

**Tabel 2.**Data Hasil Preprocessing

Cp (Chest pain)	Age (Usia)	Chol(mg/dL)	Target
1	63	233	0
4	67	286	1
4	67	229	1
3	37	250	0
2	41	204	0
2	56	236	0
4	62	268	1
4	57	354	0
3	63	254	1
4	53	203	1

## 2. Visualisasi Pohon Keputusan

Setelah proses preprocessing dilakukan dan data dinyatakan bersih, model klasifikasi kemudian dibangun menggunakan algoritma *Decision Tree C4.5*. Model dilatih menggunakan 80% dari total dataset (data training) dan diuji dengan 20% sisanya (data testing). Hasil dari proses klasifikasi divisualisasikan dalam bentuk pohon keputusan seperti ditunjukkan pada Gambar 3 berikut:



**Gambar 3.** Visualisasi Pohon Keputusan C4.5 Berdasarkan Atribut cp, age, dan chol.

Pohon keputusan pada Gambar 3 digunakan untuk memprediksi risiko penyakit jantung berdasarkan tiga atribut utama yang diambil dari dataset Cleveland, yaitu cp (chest pain), age, dan chol. Pohon ini dibentuk melalui algoritma *Decision Tree C4.5* yang secara otomatis memilih atribut dengan nilai *gain ratio* tertinggi pada setiap simpul.

Atribut cp dipilih sebagai simpul akar (root node), yang berarti bahwa jenis nyeri dada merupakan indikator paling informatif dalam membedakan pasien berisiko dan tidak berisiko terhadap penyakit jantung. Selanjutnya, atribut age dan chol menjadi simpul-simpul cabang yang menentukan jalur klasifikasi akhir berdasarkan kombinasi nilai dari masing-masing atribut.

### Makna Klinis dan Jalur Klasifikasi

Atribut cp (chest pain) dalam dataset dibagi menjadi empat kategori klinis, yaitu:

1. Typical angina – nyeri dada khas akibat berkurangnya aliran darah ke otot jantung.
2. Atypical angina – nyeri dada yang tidak khas.
3. Non-anginal pain – nyeri dada yang tidak berasal dari jantung.
4. Asymptomatic – pasien tidak merasakan gejala nyeri dada.

Sebagai gejala klinis yang paling mencolok dan mudah dikenali, cp menjadi pembeda awal yang digunakan untuk membagi pasien ke dalam dua jalur klasifikasi utama:

### 1. Jalur Angina ( $cp \neq 4$ )

Jika pasien mengalami angina (jenis nyeri dada typical, atypical, atau non-anginal), maka langkah selanjutnya adalah memeriksa usia (*age*) pasien:

- Jika usia  $\leq 50$  tahun

Pasien diklasifikasikan sebagai NO RISK. Secara klinis, pasien muda memiliki pembuluh darah yang lebih elastis dan risiko rendah terhadap penyumbatan arteri.

- Jika usia  $> 50$  tahun

Pasien diklasifikasikan sebagai RISK, karena secara medis risiko penyakit jantung meningkat seiring bertambahnya usia, terutama jika pasien juga mengalami nyeri dada.

### 2. Jalur Tanpa Gejala ( $cp = 4$ , Asymptomatic)

Jika pasien tidak mengalami nyeri dada (asymptomatic), maka model tidak memiliki informasi gejala langsung. Oleh karena itu, digunakan indikator risiko lainnya yaitu kadar kolesterol (chol):

- Jika kolesterol  $\leq 250$  mg/dL

Pasien diklasifikasikan sebagai NO RISK, karena kadar kolesterol masih dalam ambang normal dan kecil kemungkinannya menimbulkan aterosklerosis.

- Jika kolesterol  $> 250$  mg/dL

Pasien diklasifikasikan sebagai RISK, karena hiperkolesterolemia merupakan faktor risiko utama penyakit jantung, walaupun tidak disertai gejala klinis seperti nyeri dada.

### Interpretasi Warna pada Visualisasi

- **Merah (RISK):**

Menunjukkan kondisi pasien dengan kombinasi faktor yang berisiko terhadap penyakit jantung. Kategori ini memerlukan pemantauan lebih lanjut, tindakan medis, atau perubahan gaya hidup.

- **Biru (NO RISK):**

Menandakan pasien berada dalam kategori tidak berisiko tinggi menurut model ini. Namun, evaluasi kesehatan secara berkala tetap disarankan.

Penjelasan ini menunjukkan bahwa pohon keputusan yang dihasilkan tidak hanya mampu mengklasifikasikan pasien secara teknis, tetapi juga memiliki makna klinis yang kuat, membuatnya cocok untuk diterapkan sebagai alat bantu diagnosis awal pada sistem pendukung keputusan medis (*clinical decision support system*).

## B. Pembahasan

Hasil dari model klasifikasi yang dibangun menggunakan algoritma *Decision Tree C4.5* menunjukkan bahwa meskipun hanya menggunakan tiga atribut utama (cp, age, dan chol), model ini mampu memberikan keputusan klasifikasi yang cukup akurat dan mudah dipahami. Pemilihan atribut cp sebagai simpul akar memperkuat bahwa jenis nyeri dada merupakan indikator klinis yang paling signifikan dalam membedakan pasien berisiko dan tidak berisiko terhadap penyakit jantung.

Penggunaan atribut age dan chol sebagai pembeda lanjutan juga sesuai dengan landasan medis. Usia yang lebih tua umumnya dihubungkan dengan penurunan elastisitas pembuluh darah dan peningkatan risiko kardiovaskular. Sementara itu, kadar kolesterol yang tinggi merupakan penyebab utama terbentuknya plak yang menyumbat aliran darah ke jantung.

Model menghasilkan jalur klasifikasi yang sederhana, terdiri dari empat kombinasi aturan utama, yang sangat mudah diterapkan di lapangan, terutama pada sistem pendukung keputusan medis yang ditujukan untuk membantu analisis awal atau skrining cepat. Hal ini menjadikan pohon keputusan tidak hanya tepat guna, tetapi juga relevan secara klinis.

Selain itu, karena model ini hanya mengandalkan tiga atribut, maka sangat efisien dan dapat digunakan dalam situasi dengan keterbatasan data maupun infrastruktur, seperti di puskesmas atau layanan kesehatan berbasis mobile apps.

## SIMPULAN

Penelitian ini telah menghasilkan sebuah model klasifikasi penyakit jantung dengan menggunakan algoritma *Decision Tree C4.5*, yang dibangun berdasarkan tiga atribut utama dari dataset Cleveland, yaitu cp (chest pain), age (usia), dan chol (kadar kolesterol). Meskipun model dibangun dengan atribut yang terbatas, hasilnya menunjukkan bahwa model tetap mampu mengklasifikasikan pasien secara logis dan interpretatif. Atribut cp teridentifikasi sebagai simpul akar dalam struktur pohon keputusan, yang menandakan bahwa jenis nyeri dada merupakan faktor paling dominan dalam menentukan klasifikasi risiko penyakit jantung.

Model ini kemudian memperluas klasifikasinya melalui atribut age dan chol, yang terbukti berkontribusi dalam membedakan pasien berisiko dan tidak berisiko. Jalur klasifikasi yang terbentuk dalam pohon keputusan menunjukkan keterkaitan yang kuat antara kondisi medis pasien dengan hasil klasifikasi, dan menghasilkan aturan-aturan yang mudah dipahami serta memiliki makna klinis.

Penerapan algoritma C4.5 dalam konteks ini membuktikan bahwa model sederhana sekalipun dapat digunakan untuk menyusun sistem pendukung keputusan yang efektif, khususnya dalam proses diagnosis awal. Kemudahan interpretasi hasil visualisasi pohon keputusan juga menjadi keunggulan, karena memungkinkan tenaga medis atau pengguna non-teknis untuk memahami dasar dari setiap keputusan klasifikasi. Oleh karena itu, model ini sangat potensial untuk diterapkan pada lingkungan pelayanan kesehatan dasar atau sistem berbasis teknologi yang ditujukan untuk membantu proses skrining awal terhadap penyakit jantung.

## DAFTAR PUSTAKA

- Abdar, M. (2015). Using decision trees in data mining for predicting factors influencing of heart disease. *Semantic Scholar*. <https://www.semanticscholar.org/paper/Using-Decision-Trees-in-Data-Mining-for-Predicting-Abdar/cd1410ff10e2b9607734adcc170dacdc9e88d487>
- Chicco, D., & Jurman, G. (2020). Machine learning can predict survival of patients with heart failure from serum creatinine and ejection fraction alone. *BMC Medical Informatics and Decision Making*, 20(1), 1–16. <https://bmcmedinformdecismak.biomedcentral.com/articles/10.1186/s12911-020-1023-5>
- Fais, M. A., Revano, M., Lubis, A., Aulia, A., & Syafitri, I. (2023). Implementasi algoritma decision tree untuk klasifikasi serangan jantung. *Jurnal Sistem Informasi dan Ilmu Komputer*. <https://www.semanticscholar.org/paper/Implementasi-Algoritma-Decision-Tree-untuk-Serangan-Fais-Revano/412380172075501676e46c2a4baebd78f31f3696>
- García, D., Luengo, J., & Herrera, F. (2016). *Data preprocessing in data mining*. Springer. [https://digibug.ugr.es/bitstream/handle/10481/99405/2020\\_Book\\_BigDataPreprocessing.pdf](https://digibug.ugr.es/bitstream/handle/10481/99405/2020_Book_BigDataPreprocessing.pdf)
- Gunawan, I. M. A. O., Saraswati, I. D. A., Agung, I. D. G. R., & Putra, I. P. E. (2023). Klasifikasi penyakit jantung menggunakan algoritma decision tree series C4.5 dengan RapidMiner. *Jurnal Teknologi dan Sistem Informasi Bisnis*, 5(2), 73–83. <http://jurnal.unidha.ac.id/index.php/jteksis/article/view/775>
- Idri, A., Kadi, I., & Benjelloun, H. (2015). Heart disease diagnosis using C4.5 algorithms: A case study. In *HEALTHINF-2015*. SCITEPRESS. <https://www.scitepress.org/papers/2015/52164/52164.pdf>
- Jannah, S. S. N., Riska, N. F., Alamsyah, M., & Indrayatna, F. (2023). Pengklasifikasian penyakit jantung dengan metode decision tree. *Prosiding Nasional Statistika dan Analitika (NSA) – Universitas Padjadjaran*. <https://prosidingnsa.statistics.unpad.ac.id/?journal=prosidingnsa&op=view&page=article&path%5B%5D=348>
- Myint, K. L., & Htake Khaung Tin, H. (2021). Analyzing the comparison of C4.5, CART and C5.0 algorithms on heart disease dataset using decision tree method. In *Proceedings of the 2nd International Conference on ICT for Digital, Smart, and Sustainable Development (ICIDSSD 2020)*. EAI. <https://eudl.eu/pdf/10.4108/eai.27-2-2020.2303221>
- Nafi'ah, L., & Fatah, Z. (2024). Implementasi algoritma decision tree untuk pendeteksian penyakit jantung. *Jurnal Sistem Informasi dan Informatika (JUSIFOR)*, 3(2), 160–165. <https://ejournal.uniramalang.ac.id/index.php/jusifor/article/view/5729>
- Rodriguez, Z. M. P., & Nafea, M. (2024). Centralized and federated heart disease classification models using UCI dataset and their Shapley-value based interpretability. *arXiv preprint*. <https://arxiv.org/abs/2408.06183>
- Santoso, M., Al-Akbar, B., Nurjaya, H., Ramadhan, S. A., Rizky, N. A., & Fadillah, A. (2023). Klasifikasi potensi penyakit jantung menggunakan algoritma C4.5. *Jurnal INSAN*, 3(2), 103–110. <https://jurnal.bsi.ac.id/index.php/jinsan/article/view/3056/1484>
- Sepharni, A., Hendrawan, I. E., & Rozikin, C. (2023). Klasifikasi penyakit jantung dengan menggunakan algoritma C4.5. *ResearchGate*. [https://www.researchgate.net/publication/367619484\\_Klasifikasi\\_Penyakit\\_Jantung\\_dengan\\_Menggunakan\\_Algoritma\\_C45](https://www.researchgate.net/publication/367619484_Klasifikasi_Penyakit_Jantung_dengan_Menggunakan_Algoritma_C45)

- UCI Machine Learning Repository. (n.d.). *Heart disease dataset*. <https://archive.ics.uci.edu/datasets/>
- Umbara, F. R., & Kasyidi, F. (2023). Klasifikasi risiko kematian pasien berdasarkan penyakit penyerta dan usia pasien menggunakan metode C4.5. *Jurnal Informatika dan Rekayasa Elektronik*, 6(1), 9–17. <https://ejournal.stmiklombok.ac.id/index.php/jire/article/view/699>
- Wibowo, A. C., Lestari, S. A., & Nurchim, N. (2024). Analisis penggunaan machine learning dalam klasifikasi penentuan penyakit jantung. *Simtek: Jurnal Sistem Informasi dan Teknik Komputer*, 9(2), 97–101. <https://doi.org/10.51876/simtek.v9i2.395>
- World Health Organization. (2021, June). *Cardiovascular diseases (CVDs)*. [https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-\(cvds\)](https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds))
- Zriqat, I. A., Altamimi, A. M., & Azzeh, M. (2017). A comparative study for predicting heart diseases using data mining classification methods. *arXiv preprint*. <https://arxiv.org/abs/1704.02799>