

# Implementasi Algoritma C4.5 untuk Mendiagnosis Penyakit Jantung Koroner pada RapidMiner

Zaehol Fatah<sup>1</sup>, Mayla Dini Zakkia<sup>2</sup>  
<sup>1,2</sup> Universitas Ibrahimy, Situbondo  
Email : [mayladinizakia@gmail.com](mailto:mayladinizakia@gmail.com)

## Abstrak

Penyakit kardiovaskular, khususnya penyakit jantung koroner, menjadi alasan utama terjadinya kematian di seluruh dunia dan menduduki posisi kedua di Indonesia setelah stroke. Tingginya angka kematian akibat penyakit ini menuntut inovasi dalam proses diagnosis yang lebih cepat dan akurat. Studi ini mengaplikasikan metode penggalian data untuk menemukan faktor risiko utama yang terkait dengan penyakit jantung koroner, dengan menerapkan metode klasifikasi Decision Tree menggunakan algoritma C4.5. Algoritma C4.5 dipilih karena kemampuannya untuk memproses data berkelanjutan dengan efisiensi tinggi dan dapat menghasilkan model prediksi yang mudah dipahami. Proses analisis dilakukan dengan memanfaatkan perangkat lunak RapidMiner, yang menawarkan visualisasi terstruktur untuk membantu memahami hasil. Temuan dari penelitian ini diharapkan bisa menjadi pedoman dalam pengembangan sistem yang mendukung keputusan untuk tenaga medis, baik dalam meningkatkan mutu diagnosis maupun penanganan penyakit jantung.

**Kata Kunci:** Algoritma C4.5, Data Mining, Penyakit Jantung, Klasifikasi

## Abstract

*Cardiovascular disease, particularly coronary heart disease, is the leading cause of death worldwide and ranks second in Indonesia after stroke. The high mortality rate from this disease demands innovation in faster and more accurate diagnostic processes. This study applies data mining methods to identify key risk factors associated with coronary heart disease, using the Decision Tree classification method using the C4.5 algorithm. The C4.5 algorithm was chosen for its ability to process continuous data with high efficiency and can produce easy-to-understand predictive models. The analysis process was carried out using RapidMiner software, which offers structured visualizations to aid in understanding the results. The findings of this study are expected to guide the development of decision-support systems for medical personnel, both in improving the quality of diagnosis and treatment of heart disease.*

*Keywords: C4.5 Algorithm, Data Mining, Heart Disease, Classification*

## PENDAHULUAN

Penyakit kardiovaskular, yang umum disebut penyakit jantung, adalah penyebab teratas kematian di seluruh dunia. Di Indonesia, penyakit jantung menduduki posisi kedua sebagai penyebab kematian terbanyak setelah serangan stroke. Menurut laporan Organisasi Kesehatan Dunia (WHO), pada tahun 2019 sekitar 17,9 juta orang meninggal karena penyakit jantung. Serangan jantung tergolong sebagai salah satu dari 10 penyakit dengan risiko tinggi menyebabkan kematian. Penyakit jantung

koroner adalah penyakit jantung yang terjadi disebabkan oleh penyempitan pembuluh darah koroner. Terjadi gangguan pada arteri yang mengalirkan darah dari otot jantung sehingga jantung tidak dapat memompa darah untuk mencapai organ penting (Jantung et al., 2021). Penyakit ini termasuk dalam penyakit kardiovaskular. Penyakit kardiovaskular (CVDs) mencakup berbagai macam masalah yang berkaitan dengan jantung dan pembuluh darah, seperti penyakit jantung koroner, penyakit otak yang berhubungan dengan pembuluh

darah, penyakit jantung akibat rematik, serta berbagai gangguan lainnya. Penyakit jantung memiliki hubungan yang kuat dengan sejumlah faktor risiko, termasuk hipertensi, kadar kolesterol, glukosa dalam darah, asam urat, usia, pola hidup, faktor umur, jenis kelamin, serta tekanan darah dan kolesterol. yang merupakan penyebab utama kematian di seluruh dunia (Wijayadhi et al., 2023).

Proses untuk mengidentifikasi penyakit jantung koroner biasanya masih dilakukan dengan cara tradisional, yaitu dengan berkonsultasi secara langsung kepada dokter spesialis. Oleh karena itu, kami menganalisis data ini untuk mengidentifikasi faktor-faktor apa yang memiliki kemungkinan terbesar terkait dengan penyakit jantung (Maspiyanti et al., 2015). Metode yang diterapkan dalam penelitian ini untuk pengolahan data mining adalah Decision Tree dengan algoritma C4. 5. yaitu Salah satu algoritma populer dalam klasifikasi. Algoritma C4. 5 adalah metode yang dapat mengelola data berkelanjutan yang memiliki berbagai keunggulan(Muzakki et al., 2024), termasuk kecepatan proses yang lebih tinggi, tingkat akurasi yang tinggi, serta hasil yang lebih mudah dipahami dan diterapkan. Dengan pemanfaatan perangkat lunak RapidMiner, proses klasifikasi menggunakan algoritma C4.5 dapat dilakukan secara sistematis dan visual, sehingga mempermudah tenaga medis maupun peneliti dalam mengolah dan memahami hasil analisis prediksi penyakit jantung.

Dengan perkembangan teknologi, data mining menjadi solusi potensial untuk membantu proses diagnosis berdasarkan data medis historis. Penelitian ini menggunakan algoritma C4.5 karena memiliki akurasi tinggi, cepat dalam pemrosesan, serta menghasilkan model yang mudah ditafsirkan. Penelitian sebelumnya banyak menggunakan dataset standar seperti Cleveland Heart Disease Dataset. Penelitian ini menggunakan dataset Kaggle dengan 300 data dan 12

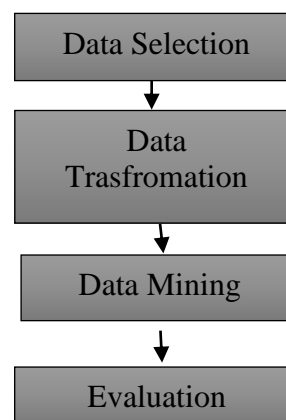
variabel, sehingga pola korelasi antaratribut dan hasil model berbeda dari studi-studi terdahulu. Penelitian terdahulu banyak yang hanya menampilkan akurasi. Penelitian ini mengukur akurasi, presisi, dan recall, sehingga evaluasi performa model lebih komprehensif.

## METODE

### Proses Pengumpulan Data dan Atribut

Informasi yang terdapat dalam studi ini diambil dengan cara mengumpulkan data sekunder. Data tersebut berasal dari situs web publik Kaggle, yang menawarkan dataset terbuka yang berhubungan dengan catatan medis pasien yang mengalami penyakit jantung. Penggunaan data sekunder dilakukan karena informasi tersebut sudah tersedia sehingga tidak perlu lagi melakukan observasi, wawancara, atau survey(accessed Apr. 30, 2023). Dataset yang digunakan dapat ditemukan melalui tautan: <https://www.kaggle.com/code/alkidiarete/heart-disease-roc-0-98>.

penjelasan Dataset tersebut berisi gejala-gejala klinis yang berkaitan dengan penyakit jantung koroner. Informasi yang digunakan meliputi lebih dari 300 catatan pasien, di mana setiap catatan memiliki 12 variabel, yang terdiri dari 11 variabel bebas dan 1 variabel terikat sebagai label untuk klasifikasi. Tabel dibawah ini merupakan atribut keseluruhan dari databes yang diperoleh dari *Kaggle*.



Gambar 1. Tahapan KDD

**Data Selection**

Dalam naskah, Tahap awal dalam proses *Knowledge Discovery in Database (KDD)* adalah pemilihan data (data selection)(D. L. Rianti, Y. Umaidah, 2021). Di fase ini, peneliti melakukan pemilihan variabel-variabel yang dianggap relevan dan signifikan untuk dianalisis lebih lanjut. Pemilihan dilakukan berdasarkan pertimbangan medis serta relevansi atribut terhadap tujuan penelitian, yaitu memprediksi kemungkinan penyakit jantung berdasarkan gejala yang ada.

**Tabel 1.** Tabel yang Dipilih

N0	Nama	Keterangan
1	Jenis Kelamin	Atribut Fitur
2	Usia	Atribut Fitur
3	Tekanan Darah	Atribut Fitur
4	Kolesterol	Atribut Fitur
5	Nyeri Dada	Atribut Fitur

6	Hasil EKG	Atribut Fitur
7	Deta Jantung Maksimal	Atribut Fitur
8	ST Depresson	Atribut Fitur
9	Kemiringan ST	Atribut Fitur
10	Penyakit Jantung	Atribut Fitur

Ateribut yang terpilih adalah Jenis kelamin, Usaia, Tekanan Darah, Usia, Tekanan Darah, Kolesterol, Nyeri Dada, Hasil EKG, Deta Jantung Maksimal, ST Depresson, Kemiringan ST, Penyakit Jantung. Dikarnakan atribut tersebut berhubungan dengan gejala penyakit jantung. Atribut tersebut diolah menggunakan algoritma C4. 5 dan perangkat lunak RapidMiner yang menghasilkan pohon keputusan serta tingkat akurasi dalam memprediksi penyakit jantung.

Tabel 2. Databases pasien

Jenis Kelamin	Usia	Tekanan Darah	Kolesterol	Nyeri Dada	Hasil EKG	Deta Jantung Maksimal	ST Depresson	Kemiringan ST	Penyakit Jantung
M	40	140	289	ATA	Normal	172	0	Up	0
F	49	160	180	NAP	Normal	156	0	Flat	1
M	37	130	283	ATA	ST	98	1	UP	0
F	48	138	214	ASY	Normal	108	1,5	Flat	1
M	54	150	195	NAP	Normal	122	0	Up	0
M	39	120	339	ATA	Normal	170	0	Up	0
F	45	130	327	ATA	Normal	170	0	Flat	0
M	54	110	208	ASY	ST	142	2	Up	0
M	37	140	207	ATA	Normal	130	0	Up	1
F	48	115	284	NAP	ST	120	1	Up	0

Dari table 2 dapat dideskripsikan sebagai berikut:

a. Jenis Kelamin

Jenis Kelamin adalah atribut yang mencangkup gender pasien dengan kategori Pria dan wanita.

b. Usia

Usia merupakan atribut yang menunjukkan usia pasien yang dibagi dalam berbagai kategori seperti anak-anak, remaja, orang dewasa, pra-lanjut usia, dan lanjut usia.

- c. Tekanan Darah
- d. Kolesterol
- e. Kolesterol adalah atribut yang memberikan informasi mengenai kadar kolesterol total pasien dalam mg/dL.
- f. Nyeri Dada  
"Nyeri Dada" merupakan atribut menggambarkan tingka keparahan nyeri dada yang dialami pasien.
- g. Hasil EKG  
"Hasil EKG" merupakan atribut hasil pemeriksaan aktivitas listrik jantung pasien.
- h. Deta Jantung Maksimal  
"Deta Jantung Maksimal" merupakan atribut yang menunjukkan detak jantung maksimum pasien saat uji fisik.
- i. ST Depresson
- j. "ST Depresson" merupakan atribut penurunan segemen ST pada EKG saat olahraga.
- k. Kemiringan ST
- l. "Kemiringan ST" merupakan atribut yang menggambarkan arah perubahan segmen ST saat uji fisik.
- m. Penyakit Jantung
- n. "Penyakit Jantung" merupakan atribut lebel atau target dari data yang menunjukkan apakah pasien memiliki penyakit jantung.

Dalam penentuan klasifikasi potensi penyakit jantung berdasarkan atribut Jenis kelamin, Usia, Tekanan Darah, Usia, Tekanan Darah, Kolesterol, Nyeri Dada, Hasil EKG, Deta Jantung Maksimal, ST Depresson, Kemiringan ST, Penyakit Jantung.

3.

Tabel 3. Atribut Dataset Penyakit Jantung

N0	Kriteria	Keterangan
1.	Jenis Kelamin	Peria (m), Wanita (f)
2.	Usia	Anak(0-9), remaja(10-18), dewasa(19-44), pra lanjut usia(45-59), lanjut usia (60-185)
3.	Tekanan Darah	Normal(19-120), prehipertensi (121-139), hipertensi stadium 1 (140-159),

		hipertensi stadium 2 (160)
4.	Kolesterol	Normal (0-199), ambang batas (200-239), tinggi(240-500)
5.	Nyeri Dada	ATA:ringan, Nap:sedang, ASY:berat
6.	Hasil EKG	Noraml
7.	Deta Jantung Maksimal	Nilai numrik
8.	ST Depresson	Nilai numerik
9.	Kemiringan ST	Up(menanjak), flat(datar), down(menurun)
10.	Penyakit Jantung	0 = tidak ada penyakit, dan 1=ada penyakit

## HASIL DAN PEMBAHASAN

Penelitian ini menerapkan metode algoritma C4.5 yang mencakup beberapa tahap, yaitu pemilihan data, pemrosesan data, transformasi data, dan evaluasi. Tahap-tahap ini akan dijadikan panduan dalam memprediksi gejala terkena penyakit jantung dengan hasil yang diperoleh, seperti nilai akurasi, presisi, dan recall. Tabel 2 di bawah ini menunjukkan seluruh atribut dari dataset yang diambil dari Kaggle.

### Data Transformation

Tahap ini memperbaiki data dengan maksud agar data menjadi lebih baik dan siap untuk proses pemodelan data mining(Jantung et al., 2021). Proses pada tahap transformasi ini adalah mengkonversi data ke dalam format yang tepat agar bisa diproses lebih lanjut dalam data mining. Perubahan pada dataset ini dilakukan menggunakan RapidMiner yang mengubah tipe data dari polinomial menjadi binomial pada atribut yang digunakan, seperti yang terlihat pada Gambar 2.

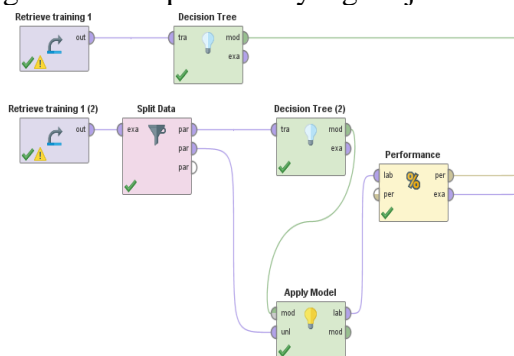
lar kole... per	Gula saat p... integer	Hasil EKG ... polynomial	Detak Jant... integer	nyeri dada... polynomial	perbedaan ... real	Kemiringa... polynomial	Penyakit ... polynomial
1	0	Normal	172	N	0.000	Up	0
2	0	Normal	156	N	1.000	Flat	1
3	0	ST	98	N	0.000	Up	0
4	0	Normal	138	Y	1.500	Flat	1
5	0	Normal	122	N	0.000	Up	0
6	0	Normal	170	N	0.000	Up	0
7	0	Normal	170	N	0.000	Up	0
8	0	Normal	142	N	0.000	Up	0
9	0	Normal	130	Y	1.500	Flat	1
10	0	Normal	120	N	0.000	Up	0
11	0	Normal	142	N	0.000	Up	0
12	0	ST	99	Y	2.000	Flat	1
13	0	Normal	145	N	0.000	Up	0
14	0	Normal	140	Y	1.000	Flat	1
15	0	ST	137	N	0.000	Up	0
16	0	Normal	150	N	1.500	Flat	0
17	0	Normal	166	N	0.000	Flat	1

Gambar 2. Transformation pada dataset Data Mining

Data mining merupakan metode pengolahan informasi yang berasal dari big data, yang sering kali diperoleh dari internet atau database, kemudian diolah menjadi informasi yang bermanfaat. Teknik ini banyak dimanfaatkan di berbagai sektor untuk meningkatkan pemahaman, dan juga dapat berfungsi untuk memperbesar angka penjualan(Santoso et al., 2023). Langkah ini menerapkan cara pengolahan informasi dengan menggunakan metode atau algoritma tertentu. pada penelitian ini memanfaatkan algoritma C4. 5. Algoritma C4. 5 diterapkan dalam proses data mining ini untuk meramalkan pasien yang menderita penyakit paru-paru berdasarkan kumpulan data yang telah melalui langkah pemilihan dan transformasi(Ilmu & Ruru, 2025).

**Evaluation**

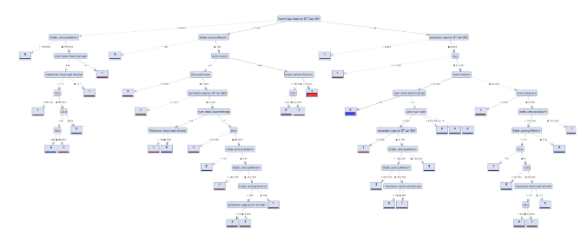
Tahap terakhir adalah evaluasi, di mana pada fase ini keberhasilan model yang diterapkan dapat diukur dengan tingkat akurasi pada data yang diuji.



Gambar 3. Pengujian menggunakan Rapidminer

Gambar 3 menggambarkan langkah-langkah dalam pengujian yang dilakukan dengan RapidMiner, diawali dengan

mengimpor dataset dan kemudian membagi data untuk melatih model. Selanjutnya, ada proses pengujian untuk mengevaluasi kinerja model, diikuti dengan pemilihan Apply Model, di mana model Decision Tree telah dipilih dalam pengujian ini. Kinerja digunakan untuk menghitung tingkat akurasi yang diperoleh dari hasil prediksi. Semua komponen tersebut terhubung, sehingga menghasilkan pohon keputusan (Decision Tree) untuk prediksi pasien dengan penyakit jantung, yang terlihat pada Gambar 4.



Gambar 4. Tampilan pohon Keputusan Nilai Akurasi

Nilai akurasi yang diperoleh berasal dari pemisahan data menjadi data pelatihan dan data pengujian. Gambar 5 di bawah ini menunjukkan perbandingan beberapa pemisahan data yang diuji berdasarkan nilai akurasi dari algoritma C4.5.

	trial 0	trial 1	class prediction
pred 0	100	87	89.00%
pred 1	48	143	74.80%
class total	70.00%	73.75%	

Gambar 5. Nilai Akurasi Nilai Recall

Beris Nilai akurasi didapatkan melalui pemisahan data menjadi data pelatihan dan pengujian. Gambar 5 di bawah ini menunjukkan perbandingan antara beberapa pemisahan data yang diuji berdasarkan nilai akurasi dari algoritma C4.5.

	trial 0	trial 1	class prediction
pred 0	100	87	89.00%
pred 1	48	143	74.80%
class total	70.00%	73.75%	

Gambar 6. Nilai Recall Nilai Precision

BerisNilai *accuracy* diperoleh dari hasil pembagian data (*split data*) menjadi data *training* datang *testing*, Berikut

gambar 5 merupakan perbandingan dari beberapa split data yang diuji dari hasil nilai akurasi algoritma C4.5.

split	akurasi	split	akurasi
split 1	0.85	split 2	0.85
split 3	0.85	split 4	0.85

Gambar 7. Nilai Precision

**Pengolahan Data Menggunakan Algoritma C4.5**

Tahap ini sangat penting untuk menghitung akar pohon dengan menggunakan rumus entropi berdasarkan data dari Tabel. 1, yang menghasilkan nilai entropi untuk setiap atribut yang ada di Tabel. 4. Data Terening sudah disiapkan terlebih dahulu sebelum diproses, kemudian atribut akan dievaluasi dengan menghitung nilai entropi(No et al., 2023). Rumus 1 adalah rumus yang digunakan untuk menghitung nilai Entropy.

$$Entropy = \sum n - p \log_2(P_i)$$

Keterangan:

- S = sekumpulan data atau himpunan kasus
- n = umlah kelas atau partisi dalam S
- pi = proporsi jumlah data pada kelas ke-i terhadap total data dalam S

Dapat menggunakan Rumus 2 untuk menghitung nilai Information Gain sebagai berikut:

$$Gain(S, A) = entropy(S) - \sum_{i=1}^n |S_i| / |S| Entropy(S_i)$$

Keterangan:

- S = himpunan keseluruhan data
- A = atribut atau fitur yang dievaluasi
- n = banyaknya partisi berdasarkan atribut A
- |S<sub>i</sub>| = jumlah data pada partisi ke-i
- |S| = jumlah total data pada himpunan S

Tabel 4. Entropy Potensi Penyakit Jantung

Atribut	Keterangan	Jumlah	Iy	Tid	Entro
	an	ah	a	ak	py
Jenis	Pria(M)	6	1	5	0.650
Kelami	Wanita	4	2	2	0.22
n	(F)				1.000
					0.00

Katago	Dewasa	3	0	3	0.000
ri	PraLanju	6	3	3	0.00
t	Usia	1	0	1	0.000
	Lanjut				0.00
	Usia				1.000
					0.00
					0.000
					0.00
Tekana	Normal	3	0	3	0.000
n	Prehipert	2	0	2	0.00
Darah	ensi	4	0	2	0.000
	Prehipert				0.00
	ensi St.1	1	1	0	1.000
	Hiperten				0.00
	si St.2				0.000
					0.00
Kolest	Normal	6	0	0	0.000
erol	Ambang	0	0	0	0.00
	Tinggi	9	3	6	-
					0.918
					296
Nyeri	Ringan	5	1	4	0.721
Dada	Sedang	3	1	2	0.928
	Berat	2	1	1	0.918
					296
					1.000
					0.00

Perhitungan entropi untuk setiap node dalam Tabel 4 digunakan untuk menghitung nilai information gain, yang nantinya akan diterapkan untuk menentukan akar pohon tertinggi. Hasil dari perhitungan nilai information gain dapat ditemukan di Tabel 5.

Tabel 5. Nilai Information gain

Atribut	Information Gain	Gain Ration
Jenis Kelamin	0.020978	0.021553
Katagori Usia	0.059588	0.027577
Kolestrol	0.124511	0.042755
Nyeri Dada	0.252202	0.112839

Pada Tabel 5, terlihat bahwa atribut dengan nilai information gain yang paling tinggi adalah Nyeri Dada, yang akan

dijadikan sebagai cabang utama pohon. Pada beberapa penelitian sebelumnya, atribut dengan gain tertinggi biasanya chol (kolestrol), age, atau cp (chest pain) dengan nilai kecil. Kemudian proses ini diulang pada langkah kedua hingga semua data tercakup. Pada tahap ini, langkah kedua diulang kembali. Setelah nilai entropi untuk setiap atribut diperoleh, selanjutnya dapat dihitung nilai information gain untuk masing-masing atribut.

### SIMPULAN (PENUTUP)

Jantung adalah organ yang sangat krusial untuk dirawat karena peranannya yang esensial bagi kesehatan manusia. Pengenalan penyakit jantung sering kali sulit dilakukan, yang dapat mengakibatkan banyak orang meninggal secara mendadak tanpa mereka ketahui, baik saat menderita penyakit jantung ataupun ketika mengalami serangan jantung. Penyakit jantung merupakan faktor utama yang menyebabkan kematian di seluruh dunia. Menurut World Health Organization pada tahun 2019 jumlah kematian diperkirakan sekitar 17,9 juta orang. Penelitian ini mengambil data melalui situs penyedia dataset yaitu Kaggle sebanyak 300 data yang memiliki 12 variabel, tetapi hanya 10 variabel saja yang terpilih karena dinilai sesuai dengan prediksi penyakit Jantung. Metode pohon keputusan menggunakan algoritma C4.5 dipilih sebagai langkah uji dalam penelitian ini, di mana hasilnya mencakup nilai akurasi, presisi, dan recall dengan cara membandingkan pembagian data pada data pelatihan dan data pengujian. Penerapan algoritma C4.5 dalam penelitian ini menunjukkan akurasi yang memuaskan untuk memprediksi penyakit jantung. Analisis ini juga menghasilkan model pohon keputusan yang menggambarkan hubungan antara atribut dengan kemungkinan terjadinya penyakit jantung. Model yang dibuat mampu memberikan nilai akurasi, presisi, dan recall yang cukup baik. RapidMiner memfasilitasi visualisasi proses serta hasil klasifikasi. Penelitian ini

memiliki perbedaan dengan penelitian terdahulu, terutama dalam penggunaan dataset Kaggle dengan struktur atribut yang berbeda, seleksi variabel yang lebih terfokus, serta verifikasi manual terhadap perhitungan entropi dan information gain sebelum divalidasi menggunakan RapidMiner. Perbedaan tersebut menghasilkan model yang unik dengan atribut dominan yang tidak selalu sama dengan penelitian sebelumnya. Untuk studi selanjutnya, bisa dicoba menggunakan algoritma lain dari teknik pengolahan data untuk mengetahui hasil akurasi yang diperoleh dengan algoritma lainnya sebagai perbandingan terhadap penelitian ini.

### UCAPAN TERIMA KASIH

Ucapan terima kasih ditujukan kepada semua yang telah memberikan dukungan dalam pelaksanaan penelitian ini, sehingga proses penelitian dapat berjalan dengan baik tanpa kendala.

### DAFTAR PUSTAKA

- d. l. rianti, y. umaidah, & a. v. (2021). *Marketplace trends based on customer review classification using kernel support vector machine comparison*. STRING (Satuan Tulisan Riset dan Inovasi Teknologi).
- ilmu, j., & ruru, k. (2025). *Peningkatan kinerja decision tree C4.5 melalui pendekatan mean*. 2(1), 24–32.
- jantung, p., haryadi, d., marini, d., & atmaja, u. (2021). *Penerapan algoritma k-means clustering untuk pengelompokan tingkat risiko*. 1089, 1–16.
- maspiyanti, f., informatika, t., pancasila, u., & gatc, j. (2015). *Diagnosa penyakit jantung pada ponsel menggunakan*. 1(1), 13–20.
- muzakki, f., ubaydillah, i., assyiami, n. r., & soleha, s. (2024). *Penerapan algoritma C4.5 untuk prediksi penyakit jantung menggunakan RapidMiner*. 2, 71–79.
- no, v., hal, a., agus, i. m., gunawan, o., ayu, i. d., saraswati, i., gede, i. d., agung, r., & eka, i. p. (2023). *Klasifikasi penyakit*

- jantung menggunakan algoritma decision tree series C4.5 dengan RapidMiner.* 5(2), 73–83.
- santoso, m., al-akbar, b., nurjaya, h., ramadhan, s. a., rizky, n. a., & fadillah, a. (2023). *Klasifikasi potensi penyakit jantung menggunakan algoritma C4.5.* 3(2), 96–103.
- wijayadhi, a., effendi, m. m., & rahardjo, s. b. (2023). *Prediksi penyakit jantung dengan algoritma regresi linier.* 4(1), 15–28.
- dqlab. (2022). *Pentingnya implementasi data sekunder dalam sebuah penelitian.* <https://dqlab.id/pentingnya-implementasi-data-sekunder-dalam-sebuah-penelitian>
- safuan, s., & firmansyah, r. (2020). *Penerapan algoritma C4.5 untuk klasifikasi penyakit jantung berbasis data mining.* Jurnal Teknologi Informasi, 7(2), 55–63.
- patel, j., & tejal, m. (2019). *Heart disease prediction using decision tree (C4.5) algorithm.* International Journal of Computer Applications, 182(25), 10–15.
- kumar, s., & goyal, p. (2020). *A study on heart disease prediction using data mining techniques.* International Journal of Engineering Research, 9(3), 40–45.