Jurnal Informatika dan Sistem Informasi (JIFoSI)
Vol. 5, No. 3. Desember 2024
e-ISSN : 2722-130
DOI: https://doi.org/10.33005/jifosi.v5i3.471

PENERAPAN ALGORITMA C4.5 UNTUK KLASIFIKASI PENYAKIT JANTUNG

Marchel Adias Pradana[1)], Riza Satria Putra[2)], Muhammad Misbachuddin[3)], Eva Yulia Puspaningrum[4)]
E-mail : [1)]21081010084@student.upnjatim.ac.id , [2)]21081010010@student.upnjatim.ac.id , [3)] 21081010024@student.upnjatim.ac.id , [4)]evapuspaningrum.if@upnjatim.ac.id
[1]departemen Informatika, Fakultas Ilmu Komputer, Institusi Universitas Pembangunan Nasional "Veteran" Jawa Timur

(Naskah masuk: 23 Desember 2024, diterima untuk diterbitkan: 31 Desember 2024)

Abstrak

Penyakit Jantung Koroner (PJK) menjadi salah satu penyebab utama kematian di Indonesia setiap tahunnya. Kurangnya akses informasi kesehatan yang memadai menjadi satu diantara banyaknya faktor penyebabnya, yang berujung pada keterlambatan diagnosis dini. Keterlambatan ini meningkatkan risiko komplikasi serius dan angka kematian akibat PJK. Penelitian ini bertujuan untuk mengembangkan model prediksi berbasis algoritma C4.5, yang membangun pohon keputusan berdasarkan data historis pasien. Algoritma C4.5 dipilih karena kemampuannya dalam mengolah data untuk menghasilkan klasifikasi yang akurat. Evaluasi model dilakukan menggunakan teknik validasi silang (k-fold cross validation) untuk memastikan keandalan hasil. Dalam penelitian ini, 10 skenario pengujian dilakukan, dengan performa terbaik diperoleh pada skenario K=3. Hasil evaluasi menunjukkan akurasi sebesar 0.8941, presisi 0.9000, dan recall 0.9184. Temuan ini menunjukkan bahwa algoritma C4.5 efektif dalam mendukung prediksi awal PJK dengan tingkat akurasi yang tinggi. Model prediksi ini diharapkan dapat membantu dalam mengurangi angka kematian akibat PJK melalui deteksi dini yang lebih baik dan berbasis data.

Kata kunci: *penyakit jantung, klasifikasi, algoritma C4.5*

1. PENDAHULUAN

Penyakit Jantung Koroner (PJK) merupakan salah satu penyebab kematian tertinggi di Indonesia yang terjadi setiap tahun. Penyakit kardiovaskuler, seperti kanker, stroke, gagal ginjal, serta gangguan jantung dan pembuluh darah, semakin meningkat jumlahnya. Penyempitan dan penyumbatan pada dinding pembuluh darah koroner menghambat aliran darah ke otot jantung, yang pada akhirnya mengganggu fungsi jantung[1].
Data Riskesdas 2018 [2] menunjukkan bahwa prevalensi penyakit jantung berdasarkan diagnosis dokter di Indonesia mencapai 1,5%, dengan prevalensi penyakit jantung koroner tetap 1,5% dari 2013 hingga 2018, hipertensi meningkat dari 25,8% (2013) menjadi 34,1% (2018), stroke meningkat dari 12,1 per mil (2013) menjadi 10,9 per mil (2018), dan penyakit gagal ginjal kronis meningkat dari 0,2% (2013) menjadi 0,38% (2018). Di antara delapan provinsi tersebut, Jawa Tengah 16%, Jawa Barat 16%, DKI Jakarta 19%, Sulawesi Utara (1,8 persen), Sulawesi Tengah (1,9 persen), Kalimantan Barat 16%, Aceh 16%, dan Kalimantan Timur (1,9 persen),
Berdasarkan data yang tersedia, Banyak orang masih belum mengetahui penyebab penyakit ini. Dokter sering menemukan penyakit dalam stadium lanjut setelah melakukan pemeriksaan medis. Ada berbagai pilihan pengobatan untuk mencegah dan mengobati

penyakit ini, seperti operasi, terapi radiasi, dan kemoterapi. Namun, faktor utama yang menyebabkan penderita menunggu untuk berkonsultasi dengan dokter adalah kekurangan informasi dan media. Hubungan antara kurangnya akses informasi dengan keterlambatan diagnosis awal penyakit jantung berdampak pada meningkatnya angka kematian setiap tahunnya.[3].

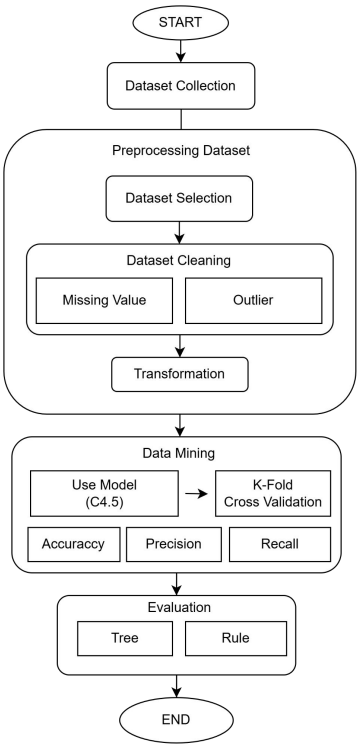
Oleh karena itu, diperlukan sistem klasifikasi yang mampu memberikan informasi tentang penyakit jantung dan memungkinkan deteksi dini. Sistem ini membutuhkan metode yang tepat untuk mengelola pengetahuan dari pakar, guna memberikan hasil yang akurat.[3]. Salah satu metode yang dapat digunakan adalah Algoritma C4.5, yang termasuk dalam supervised learning.[4].

Algoritma C4.5 sangat efektif dalam klasifikasi, menggunakan pendekatan pohon keputusan. Bentuk pohon keputusan menyerupai flowchart, dengan setiap simpul internal mengevaluasi atribut, cabang-cabangnya mewakili hasil, dan simpul daun menandakan kelas. Algoritma ini menggunakan konsep gain atau pengurangan entropi untuk membentuk pembagian yang optimal, serta mampu menghasilkan aturan yang mudah diinterpretasikan dan bekerja lebih cepat dibandingkan algoritma lainnya.[4].

Penelitian ini bertujuan untuk mengevaluasi efektivitas algoritma C4.5 dalam memprediksi penyakit jantung, dengan akurasi sebagai parameter utama untuk menilai kinerja algoritma tersebut. Selain itu, penelitian ini juga akan mempertimbangkan kecepatan dan kemampuannya dalam menghasilkan model klasifikasi yang akurat dan mudah diinterpretasikan. Hasil dari penelitian diharapkan dapat berkontribusi pada pengembangan sistem deteksi dini penyakit jantung yang lebih efektif.

2. METODOLOGI

Metodologi yang digunakan pada penelitian ini dilakukan dengan beberapa tahapan mulai dari teknik pengumpulan data hingga teknik pemrosesan data untuk memperoleh hasil akhir sesuai dengan keinginan.



Gambar 1. Flowchart Pengerjaan

2.1 Teknik Pengumpulan Data

Pada penelitian ini menggunakan teknik pengambilan data sekunder yang diambil dari website Kaggle.com dengan kontributor username FEDESORIANO dengan nama

dataset Heart Failure Disease Dataset yang dapat dilihat pada https://www.kaggle.com/datasets/fedesoriano/heart-failure-prediction. Dataset berformat csv yang berisi 918 data meliputi 11 atribut dan 1 output.

2.2 Teknik Pemrosesan Data

Selanjutnya data yang telah diperoleh akan dilakukan pemrosesan dengan bahasa pemrograman Python pada platform Google Collaboratory. Proses ini melibatkan beberapa tahapan penting yang saling terkait, mulai dari seleksi data hingga penyajian data secara bersih yang dapat digunakan untuk proses selanjutnya[5]. Berikut adalah tahapan pemrosesan data:

a. Selection Data

Seleksi data adalah tahap pengumpulan data yang diperlukan untuk mengidentifikasi informasi dari basis data operasional untuk analisis lebih lanjut[6]. Tahap pertama dalam proses data mining adalah menentukan basis data yang akan digunakan untuk prediksi. saat tahap ini dilakukan, data dari dataset Kaggle akan dipilih. Setelah dipilih, data ini akan dianalisis dan dilanjutkan ke tahapan berikutnya.

b. Cleaning Data

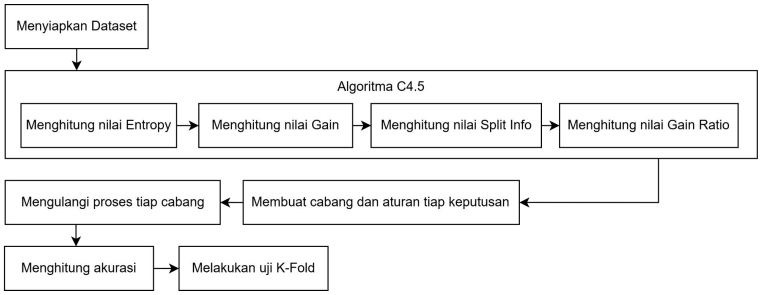
Salah satu tahapan dalam proses pengolahan data adalah data cleaning, yang bertujuan untuk membersihkan dataset yang memiliki noise. Pemrosesan awal data terdiri dari transformasi nilai data dari kumpulan data tertentu, bertujuan untuk mengoptimalkan perolehan dan proses informasi serta menangani data yang tidak diperlukan, seperti data yang tidak memiliki nilai (missing value), outlier, atau ketidakseimbangan (imbalance).[7] Untuk menangani data yang tidak seimbang, metode resampling akan digunakan. Sebaliknya, sejumlah record akan dihapus untuk menangani data yang memiliki nilai yang hilang atau outlier ini.

c. Transformation

Pada tahap transformasi data ini, diubah ke dalam bentuk yang sesuai dengan proses pemrosesan data supaya mempengaruhi hasil ujicoba[8]. Mengubah tipe data kontinu menjadi tipe data diskrit adalah salah satu teknik yang dapat digunakan dalam proses transformasi data. Dengan membuat label interval dan membagi range atribut, discretize juga dapat digunakan untuk mengurangi jumlah atribut numerik.

2.2 Data Mining

Saat data telah melewati tahap preprocessing dan transformasi, langkah selanjutnya adalah pengolahan data menggunakan metode atau algoritma data mining. Data mining adalah cara untuk menemukan pola dan tren tersembunyi dalam banyak data, Ini dimulai dengan memilih data yang relevan melakukan pemodelan data untuk memprediksi[9]. Dalam penelitian ini, algoritma C4.5 digunakan untuk menganalisis gejala penyakit jantung. Berikut algoritma dari Decision Tree C4.5 pada gambar 2.



Gambar 2. Proses data mining

1. Dataset harus disiapkan dengan variabel fitur (X) dan target (Y) untuk membedakan kelas atau kategori yang ingin diprediksi.
2. Entropy mengukur ketidakpastian atau kekacauan dalam data. Entropy (S) untuk himpunan data *S* dihitung dengan rumus:

$$\text{Entropy(s)} = \sum_{i=1}^{n} \quad - p_i * log_2\, p_i$$

Dimana:
S = Dataset kasus
n = banyaknya label
p_i= probabilitas dari kelas i.

3. Gain mengukur seberapa besar informasi yang diperoleh dari pemilihan atribut tertentu. Gain adalah selisih antara entropy total dan entropy setelah pemisahan oleh atribut tersebut:

$$\text{Gain}(A) = \text{entropy(s)} - \sum_{i=1}^{n} \frac{|S_i|}{|S|} * entropy(S_i)$$

Dimana:
S = Dataset kasus
A = atribut
$|S_i|$= jumlah kasus pada partisi ke-i
$|S|$ = jumlah kasus pada S

4. Split Info menghitung jumlah informasi yang digunakan untuk melakukan pembagian berdasarkan atribut tertentu, sehingga dapat memperhitungkan pembagian yang kurang berguna. Rumusnya:

$$\text{SplitInfo}(S, A) = -\sum_{i=1}^{n} \frac{S_i}{S} log_2 \frac{S_i}{S}$$

dimana S_i adalah subset yang dihasilkan oleh nilai *i* pada atribut A.

5. Gain Ratio adalah rasio antara Gain dan Split Info. Digunakan untuk menormalkan Gain agar tidak terlalu berpihak pada atribut dengan banyak nilai unik:

$$\text{GainRation} = \frac{Gain(A)}{SplitInfo(S, A)}$$

6. Setelah atribut terbaik dipilih berdasarkan Gain Ratio tertinggi, algoritma membagi data menjadi cabang-cabang berdasarkan nilai-nilai atribut tersebut. Aturan terbentuk dari setiap cabang menuju ke daun pohon.
7. Langkah-langkah ini berulang untuk setiap subset hingga semua data pada cabang memiliki kelas yang sama atau tidak ada atribut lagi yang dapat digunakan.
8. Langkah selanjutnya melakukan pembagian menjadi data uji dan data latih dengan menggunakan *k-fold cross validation* dengan k=10, yang nantinya melibatkan 10 skenario pengujian dengan data uji dan data latih yang berbeda.
9. Langkah terakhir adalah menghasilkan nilai akurasi, ketepatan (precision), dan recall dari pengujian:

Nilai akurasi diperoleh dengan rumus:

$$\text{Akurasi} = \frac{TP + TN}{TP+TN+FP+FN}$$

Sedangkan nilai precision dihitung dengan:

$$\text{Precision} = \frac{TP}{TP+FP}$$

Dan untuk recall dihitung dengan:

$$\text{Recall} = \frac{TP}{TP+FN}$$

Dalam tahap interpretasi ini, atribut menampilkan pola informasi yang dapat dipahami oleh orang yang memakainya.

2.4 Evaluasi

Dengan menggunakan algoritma C4.5 berupa model pohon keputusan, evaluasi yang dihasilkan dari model tersebut dapat menghasilkan *decision tree classifier* dan pembagian pembagian keputusan berdasarkan *decision rule* yang tercipta. Dengan menggunakan satu aturan (rule) keputusan, pohon keputusan dapat membagi kumpulan data yang sangat besar menjadi kumpulan data yang lebih kecil[10].*Decision rule* merupakan dasar dari keputusan node pohon yang dapat mewakilkan atribut yang telah diuji dari tiap cabang yang menjadikan suatu pembagian hasil uji serta node daun (leaf) dari kelompok kelas tertentu[11].

3. HASIL DAN PEMBAHASAN

Pada proses sebelumnya telah dijelaskan beberapa tahapan sebelum melakukan proses data mining, termasuk pemilihan data, preprocessing, transformasi, dan penilaian. Penelitian ini menggunakan algoritma C4.5 sebagai penerapan data mining untuk memprediksi gejala penyakit jantung. Algoritma ini menghasilkan rules dari pohon keputusan seperti nilai akurasi, recall , dan precision.

3.1 Pengumpulan Data

Sex, Age, RestingBP, Chestpaintype, FastingBS, Cholesterol, MaxHR, RestingECG, Oldpeak, ExerciseAngina, Heart Disease, dan STSlope adalah beberapa atribut dan label yang telah dikumpulkan dari situs Kaggle dengan total data 918 yang dilabeli dengan Heart Disease terbagi dengan 1 untuk True berjumlah 508 data dan 0 untuk False berjumlah 410 data. Data yang didapat dapat dilihat pada Tabel 1.

Tabel 1. Deskripsi Data

No	Atribut	Tipe Data	Keterangan
1	*Age*	Numerik	Usia pasien
2	*Sex*	Kategorik	Jenis kelamin pasien
3	*ChestPainType*	Kategorik	Tipe sakit dada pasien
4	*RestingBP*	Numerik	Tekanan darah pasien
5	*Cholesterol*	Numerik	Kolesterol pada pasien
6	*FastingBS*	Numerik	Gula darah puasa pada pasien
7	*RestingECG*	Kategorik	Elektrokardiogram saat istirahat pada pasien
8	*MaxHR*	Numerik	Detak jantung maksimal pasien
9	*ExerciseAngina*	Kategorik	Rasa nyeri dada pada pasien
10	*Oldpeak*	Numerik	Tipe depresi pada pasien
11	*ST_Slope*	Kategorik	Kemiringan detak jantung pada pasien dengan EKG
12	*HeartDisease*	Kategorik	Label pasien menderita jantung atau tidak

3.2 Pemrosesan Data

Pemrosesan yang dilakukan pada data melalui beberapa tahapan dengan metode dimulai dari awal hingga akhir antara lain:

a. Data Selection

Data selection atau seleksi data dilakukan untuk menyaring data yang akan digunakan pada proses data mining yang nantinya akan mempengaruhi hasil dengan signifikan. Proses seleksi dengan memilih atribut atau variabel yang dinilai berpengaruh dengan penyakit jantung dilakukan dengan menghitung nilai korelasi dari tiap fitur, kemudian mengambil 5 fitur teratas dengan nilai korelasi tertinggi. Beberapa atribut beserta label yang dipilih nantinya akan digunakan antara lain pada Tabel 2.

Tabel 2. Atribut dan Label Terpilih

Atribut Terpilih	Label Terpilih
RestingBP	*HeartDisease*
Cholesterol	
Max HR	
Exercise Angina	
ST Slope	

Atribut yang terpilih yaitu RestingBP, Cholesterol, Max HR, Exercise Angine, dan ST Slope dianggap berpengaruh terhadap gejala yang sering dialami oleh para pasien penderita penyakit jantung. HeartDisease merupakan label yang menentukan pasien menderita penyakit jantung atau tidak. Data yang tidak terpakai akan dihapus, sedangkan data yang digunakan akan diolah menggunakan algoritma C4.5 yang nantinya menghasilkan sebuah rules yang berasal dari decision tree untuk digunakan sebagai prediksi penyakit jantung.

b. Data Cleaning

Data cleaning merupakan tahap pembersihan data yang sangat penting untuk digunakan pada proses data mining, sebab data yang tidak dalam kondisi bersih nantinya akan sangat mempengaruhi proses pelatihan dan akan menyebabkan permasahan nantinya. Beberapa cara yang dilakukan pada tahapan preprocessing atau pembersihan data antara lain adalah penanganan khusus untuk outlier dan missing value. Untuk itu perlu dilakukan kedua penanganan tersebut sebelum dilakukan proses data mining.

1. Penanganan Outliers

Tabel 3. Outlier Data

Nama Atribut	Min	Max	Average
RestingBP	0	200	132,4
Cholesterol	0	600	198,4

Tabel 3 menunjukkan outliers dalam dataset yang sudah diseleksi. Untuk menangani outliers data, range yang terlalu jauh dihapus. Misalnya, nilai tertinggi untuk atribut resting BP dan cholesterol adalah 200–600, kemudian, atribut yang melebihi range tersebut dihapus.

2. Penanganan Missing Value

```
# Penanganan Missing Value
dataset.isnull().any()
```

	0
RestingBP	False
Cholesterol	False
MaxHR	False
ExerciseAngina	False
ST_Slope	False
HeartDisease	False

dtype: bool

Gambar 3. Missing Value

Berdasarkan Gambar diatas dapat disimpulkan bahwa Dataset yang telah melewati proses seleksi data, Pada atribut RestingBP, Cholesterol, MaxHR, ExerciseAngina, STSlope, dan Heart Disease ternyata tidak memiliki missing value

c. Transformation

Tahapan yang dilakukan untuk mengubah data ke dalam bentuk yang sesuai dengan kebutuhan disebut transformasi data. Atribut *ExerciseAngina* dan *STSlope* dilakukan transformasi data karena pada python pelaksanaan proses evaluasi, atribut harus berisikan tipe data numerik.

Tabel 4. Transformasi ExerciseAngine

Atribut *ExerciseAngina*	Detail Penanganan
Tidak	0
Ya	1

Transformasi dilakukan pada class "Ya" diganti dengan nilai 1 dan class "Tidak" diganti dengan nilai 0 pada atribut *ExerciseAngina.*

Tabel 5. Transformasi *STSlope*

Atribut *STSlope*	Detail Penanganan
Up	0
Flat	1
Down	2

Transformasi dilakukan pada class "Up" diganti dengan nilai 0 dan class "Flat" diganti dengan nilai 1 serta class “Down”diganti dengan Nilai 2 pada atribut *STSlope.*

3.3 Data Mining

Setelah melalui tahapan data selection, preprocessing, dan transformation data lanjut diolah seperti tujuan awal, yaitu klasifikasi prediksi menggunakan algoritma C 4.5. tahap pegerjaan awal yaitu menentukan nilai entropy, Gain, SplitInfo , dan GainRatio. berikut hasil perhitungan decision tree pada percobaan pertama:

Tabel 6. Hasil Entropy

Hasil Entropy	
Entropy	0.9919

setelah dilakukan perhitungan ketidakpastian atau kekacauan dalam data, hasil entropy yaitu 0.9919.

Hasil perhitungan Gain pada atribut-atribut dapat dilihat pada tabel 7 dibawah ini:

Tabel 7. Hasil Gain

Information Gain	
RestingBP	0.0906
Cholesterol	0.3212
Max HR	0.2327
Exercise Angina	0.1895
ST Slope	0.2990

Hasil perhitungan Split Info dapat dilihat pada tabel 8 dibawah ini:

Tabel 8. Split Info

Split Info	
RestingBP	4.7000
Cholesterol	6.7458
Max HR	6.3459
Exercise Angina	0.9729
ST Slope	1.2886

Tabel 9 dibawah menampilkan hasil perhitungan Gain Ratio.

Tabel 9. Gain Ratio

Gain Ratio	
RestingBP	0.0192
Cholesterol	0.0476
Max HR	0.0367
Exercise Angina	0.1948
ST Slope	0.2320

Dari tabel 9 di atas, terlihat bahwa atribut ST Slope memiliki nilai Gain Ratio tertinggi, yaitu 0.2320, diikuti oleh Exercise Angina dengan nilai Gain Ratio sebesar 0.1948.

Berdasarkan algoritma C4.5, atribut ST Slope dipilih sebagai atribut terbaik untuk menjadi node pertama (atau root) dalam pohon keputusan karena nilai Gain Ratio-nya paling tinggi dibandingkan atribut lainnya. langkah selanjutnya yaitu membuat model algoritma decission tree, kemudian data dibagi menjadi data uji dan data latih dimana menggunakan k-folds cross validation dengan nilai k=10. berikut hasil perhitungan dari skrenario pertama.

Akurasi = $\frac{TP+TN}{TP+TN+FP+FN} = \frac{51+23}{51+23+8+7} = \frac{74}{89} = 0.8315$

Precision = $\frac{TP}{TP+FP} = \frac{51}{51+8} = 0.8644$

Recall = $\frac{TP}{TP+FN} = \frac{51}{51+7} = 0.8793$

pada hasil pengujian skenario pertama didapatkan nilai akurasi sebesar 0.8315, presisi sebesar 0.8644, dan nilai recall sebesar 0.8793. pengujian kali ini dilakukan sebanyak 10 skenario, berikut hasil dari 10 pengujian yang telah dilakukan.

Tabel 10. Hasil Pengujian

No	Akurasi	Presisi	Recall
1	0.8315	0.8644	0.8793
2	0.8276	0.8627	0.8462
3	0.8941	0.9000	0.9184
4	0.8941	0.8837	0.9048
5	0.8795	0.8431	0.9556
6	0.8000	0.8519	0.8364
7	0.7889	0.7544	0.8958
8	0.8391	0.7778	0.9545
9	0.8554	0.8444	0.8837
10	0.8182	0.7755	0.8837

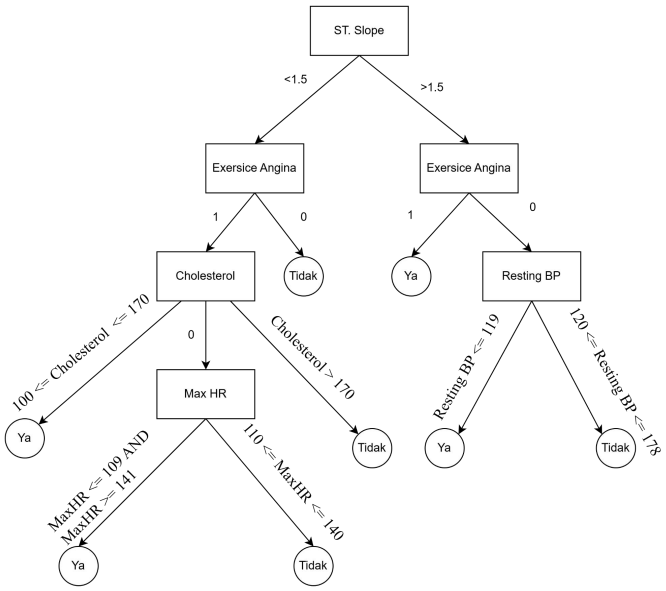
Dilihat dari tabel 10 diatas, penulis telah melakukan sebanyak 10 skenario pengujian yang dimana mendapatkan hasil akurasi tertinggi pada K=3 dengan nilai Akurasi : 0.8941, Presisi: 0.9000 dan, Recall: 0.9184. Sedangkan, mendapat rata - rata nilai Akurasi : 0.8428, Presisi : 0.8358 dan, Recall : 0.8958.

3.4 Evaluasi

Dari pengujian yang telah dilakukan dengan menggunakan algoritma C 4.5, penulis mendapatkan aturan yang didapatkan dari pemodelan sebuah pohon keputusan (*rules*). Pohon keputusan dapat dilihat pada Gambar 4.Berdasarkan pohon keputusan diatas menghasilkan beberapa aturan atau rules sebagai berikut:

a. IF ST_Slope < 1.5 AND ExerciseAngina = 0 THEN HeartDisease = Tidak
b. IF ST_Slope < 1.5 AND ExerciseAngina = 1 AND Cholesterol >= 100 AND Cholesterol <= 325 THEN HeartDisease = Tidak
c. IF ST_Slope < 1.5 AND ExerciseAngina = 1 AND Cholesterol = 0 AND MaxHR <= 110 AND MaxHR >= 141 THEN HeartDisease = Ya
d. IF ST_Slope < 1.5 AND ExerciseAngina = 1 AND Cholesterol = 0 AND MaxHR > 110 AND MaxHR < 141 THEN HeartDisease = Tidak
e. IF ST_Slope < 1.5 AND ExerciseAngina = 1 AND Cholesterol > 170 THEN HeartDisease = Ya
f. IF ST_Slope > 1.5 AND ExerciseAngina = 1 THEN HeartDisease = Ya
g. IF ST_Slope > 1.5 AND ExerciseAngina = 0 AND RestingBP <= 119 THEN HeartDisease = Ya

h. IF ST_Slope > 1.5 AND ExerciseAngina = 0 AND RestingBP >= 120 AND RestingBP <= 178 THEN HeartDisease = Tidak



Gambar 4. Pohon Keputusan

4. KESIMPULAN DAN SARAN

Setelah melakukan penelitian bisa disimpulkan bahwa algoritma C4.5 dapat digunakan untuk mengklasifikasikan penderita penyakit jantung. penelitian menggunakan *k-fold cross validation* dengan nila k =10, dimana setelah melakukan 10 kali pengujian didadapatkan hasil akurasi terbaik pada K=3 dengan nilai Akurasi 0.8941, Presisi 0.9000 dan, Recall 0.9184. Sementara itu, rata-rata hasil keseluruhan pengujian menunjukkan akurasi sebesar 0.8428, presisi 0.8358, dan recall 0.8958. Hasil ini menunjukkan potensi algoritma C4.5. Untuk penelitian selanjutnya sebaiknya melakukan penentuan bobot untuk penanganan pemilihan variabel, serta membandingkan data yang memiliki perbandingan di label yang sama.

5. DAFTAR RUJUKAN

[1] Tampubolon, L. F., Ginting, A., & Saragi Turnip, F. E. 2023. Gambaran Faktor Yang Mempengaruhi kejadian Penyakit Jantung Koroner (PJK) di Pusat Jantung Terpadu (PJT). Jurnal Ilmiah Permas: Jurnal Ilmiah STIKES Kendal, 13(3), 1043–1052.

[2] Rokom. (2021, September 28). Penyakit Jantung Koroner Didominasi masyarakat kota. Sehat Negeriku. [Online] (Updated 28 September 2021) Available at: https://sehatnegeriku.kemkes.go.id/baca/umum/20210927/5638626/penyakit-jantung-koroner-didominasi-masyarakat-kota/ [Accessed 4 November 2024]

[3] Bianto, M. A., Kusrini, K., & Sudarmawan, S. 2020. Perancangan Sistem Klasifikasi Penyakit Jantung Menggunakan Naïve Bayes. Creative Information Technology Journal, 6(1), 75-83.

[4] Sepharni, A., Hendrawan, I. E., & Rozikin, C. 2022. Klasifikasi Penyakit Jantung dengan Menggunakan Algoritma C4. 5. STRING (Satuan Tulisan Riset dan Inovasi Teknologi), 7(2), 117-126.

[5] Prasetio, R. T., & Ripandi, E. 2019. Optimasi Klasifikasi jenis hutan menggunakan deep learning berbasis optimize selection. Jurnal Informatika, 6(1), 100-106.

[6] Suci, W., R, N., & M. Basysyar, F. 2022. Klasifikasi Data Bantuan Sosial Pada Desa Sindangpano Dengan menggunakan algoritma K-nearest neighbor. Jurnal Accounting Information System (AIMS), 5(2), 167–174.
[7] Larriva-Novo, X., Villagrá, V. A., Vega-Barbas, M., Rivera, D., & Sanz Rodrigo, M. 2021. An IOT-focused intrusion detection system approach based on preprocessing characterization for cybersecurity datasets. Sensors, 21(2), 656.
[8] Widaningsih, S. 2022. Penerapan data mining untuk memprediksi Siswa Berprestasi dengan menggunakan algoritma k nearest neighbor. JATISI (Jurnal Teknik Informatika Dan Sistem Informasi), 9(3), 2598–2611.
[9] Ameliana, N., Suarna, N., & Prihartono, W. 2024. Analisis Data mining Pengelompokkan UMKM Menggunakan algoritma k-means clustering di provinsi Jawa Barat. JATI (Jurnal Mahasiswa Teknik Informatika), 8(3), 3261–3268.
[10] Setio, P. B. N., Saputro, D. R. S., & Winarno, B. (2020, February). Klasifikasi Dengan Pohon Keputusan Berbasis Algoritme C4. 5. In *PRISMA, Prosiding Seminar Nasional Matematika* (Vol. 3, pp. 64-71).
[11] Zega, S. A. (2014, June). Penggunaan pohon keputusan untuk klasifikasi tingkat kualitas mahasiswa berdasarkan jalur masuk kuliah. In *Seminar Nasional Aplikasi Teknologi Informasi (SNATI)*.