

Penerapan Algoritma C4.5 Untuk Prediksi Penyakit Jantung Menggunakan Rapidminer

Farhan Muzakki^{1*}, Iqbal Ubaydillah², Nur Rahma Assyiami³, Sugiyarti Soleha⁴

^{1,2,3,4}Program Studi Sistem Informasi, Universitas Bina Sarana Informatika
Jl. Kramat Raya No 98, Jakarta Pusat, Indonesia

e-mail korespondensi: 19215033@bsi.ac.id

Informasi Artikel: Submit: 14-06-2024 | Revisi : 20-06-2024 | Terima : 24-06-2024

Abstrak - Jantung merupakan organ yang sangat penting untuk dijaga kesehatannya, karena begitu pentingnya fungsi jantung bagi tubuh manusia. Penyakit jantung sangat sulit untuk dideteksi, sehingga menyebabkan banyak orang yang meninggal secara tiba-tiba tanpa menyadari jika seseorang terkena penyakit jantung atau serangan jantung. Penyakit jantung merupakan penyebab kematian nomor satu di dunia. Data ini kami olah untuk mengetahui faktor apa saja yang paling berpotensi pada penyakit jantung, dengan hasil pengolahan data ini kami harapkan dapat menjadi pengingat bagi siapa saja untuk menjaga pola hidup sehat dan mengantisipasi potensi penyakit jantung sejak dini. Teknik data mining yang digunakan dalam penelitian ini menggunakan metode Decision Tree melalui algoritma C4.5 hasil yang diperoleh mulai dari nilai *accuracy*, *precision*, dan *recall* dengan membandingkan *split data* pada *data training* dan *data testing* yang terbesar adalah *split data* 90 – 10 yaitu *accuracy* = 65.25%, *recall* = 70.87% , dan *precision* = 62.46%. Dengan menggunakan algoritma C4.5 ini diharapkan dapat menghasilkan akurasi yang tepat untuk memprediksi penyakit jantung.

Kata kunci: Penyakit jantung, Klasifikasi, Data mining, Algoritma C4.5

Abstract - The heart is an organ that is very important to maintain its health, because it is so important the function of the heart for the human body. Heart disease is very difficult to detect, causing many people to die suddenly without realizing if someone has heart disease or a heart attack. Heart disease is the number one cause of death in the world. We process this data to find out what factors have the most potential in heart disease, with the results of this data processing we hope to be a reminder for anyone to maintain a healthy lifestyle and anticipate the potential for heart disease early on. The data mining technique used in this study uses the Decision Tree method through the C4.5 algorithm. The results obtained start from the accuracy, precision, and recall values by comparing the split data on training data and testing data. The largest is a data split of 90 – 10, namely accuracy = 65.25%, recall = 70.87%, and precision = 62.46%. By using this C4.5 algorithm, it is expected to produce the right accuracy to predict heart disease.

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

1. Pendahuluan

Organ utama atau jantung harus berfungsi dengan benar dan teratur jika gangguan tersebut terjadi, maka akibat yang fatal, yaitu timbulnya serangan jantung akan terjadi. Serangan jantung termasuk 10 penyakit dengan risiko tinggi infeksi kematian. Karena kurangnya kesadaran atau pemahaman tentang fungsi jantung, gejala penyakit dan gangguan sering tidak dilaporkan oleh masyarakat umum. Ini juga akibat biaya pemeriksaan Kesehatan yang mahal dan kurangnya waktu luang karena komitmen kerja. Ada empat gejala umum penyakit jantung atau kardiovaskulitis, termasuk nyeri dada (terutama dada sebelah kiri), ritme atau detak jantung yang tidak teratur, detak jantung yang sangat cepat dan kadang-kadang menyebabkan hasil positif palsu untuk denyut nadi sama sekali, dan sesak napas. Cara mencegah penyakit jantung maka keempat gejala tersebut tidak boleh diabaikan dan pasien harus segera melakukan pemeriksaan ke dokter atau ahli jantung terdekat [1].

Jantung ialah organ tubuh yang sangat penting untuk dijaga kesehatannya, karena begitu penting fungsi jantung untuk tubuh manusia. Penyakit jantung sangat sulit untuk dideteksi, sehingga menyebabkan banyak orang yang meninggal dunia secara mendadak tanpa disadari jika seseorang mengidap penyakit jantung atau terkena serangan jantung. Penyakit jantung adalah penyebab kematian nomor satu di dunia. Menurut World Health Organization (2017) jumlah kematian diperkirakan sekitar 17, 9 juta orang [2].

PJK merupakan jenis penyakit yang multifaktorial. Salah satu faktor yang sering dikaitkan terhadap PJK ialah kebiasaan merokok. WHO menyebutkan bahwa rokok menyebabkan sekitar 7 juta kematian setiap tahunnya. Hal ini diprediksi akan mengalami peningkatan sampai 8 juta kematian setiap tahunnya pada tahun 2030 (CDC, 2018). Lebih dari 6 juta orang meninggal sebagai perokok aktif dan sekitar 890.000 lainnya meninggal akibat



terkena paparan asap rokok atau yang dikenal sebagai perokok pasif Sebanyak 80% dari 1,1 milyar perokok di seluruh dunia berasal dari negara berpenghasilan rendah dan menengah (WHO, 2018) [3].

Penyakit jantung koroner merupakan penyakit degeneratif. Penyebab penyakit degeneratif diduga akibat perubahan gaya hidup. Selain hiperkolesterolemia, kasus yang banyak di temukan di masyarakat yaitu kelainan pada tekanan darah. Tekanan darah adalah gaya (dorongan) darah ke arteri saat darah dipompa keluar dari jantung ke seluruh tubuh [4].

Klasifikasi penyakit jantung sangat penting dalam penelitian medis, pengembangan obat, dan perancangan intervensi kesehatan masyarakat. Dengan memahami perbedaan antar jenis penyakit jantung, para ahli kesehatan dapat merancang strategi pengobatan yang lebih efektif dan memberikan perawatan yang lebih spesifik sesuai dengan kebutuhan masing-masing pasien (Amalia Hilda & Evicienna, n.d.) . Dengan demikian, upaya klasifikasi ini merupakan langkah penting dalam meningkatkan deteksi dini, pengobatan, dan pencegahan penyakit jantung untuk meningkatkan kesehatan jantung secara keseluruhan, penelitian ini dibuat agar dapat memecahkan masalah dalam menentukan penyebab utama penyakit jantung [5].

Untuk itu, kami mengolah data ini untuk mengetahui faktor apa yang paling berpotensi dalam penyakit jantung , dengan adanya hasil dari olah data ini kami berharap dapat menjadi pengingat bagi siapapun agar tetap menjaga pola hidup yang sehat dan mengantisipasi potensi penyakit jantung sejak dini. Berdasarkan penjelasan yang sudah diuraikan di atas, teknik data mining yang digunakan pada penelitian ini menggunakan metode *Decision Tree* melalui algoritma C4.5. Dengan menggunakan algoritma C4.5 ini, diharapkan akan menghasilkan akurasi yang tepat untuk memprediksi penyakit jantung.

Tabel 1. Literatur Penelitian

<i>Research Problem (RP)</i>	<i>Literatur Support</i>
RP1 Mengklasifikasi penyakit jantung	<p>Klasifikasi adalah proses mengajar dan mendidik fungsi target dengan memetakan setiap set karakteristik (atribut) ke satu jumlah label kelas yang tersedia. Dengan menggunakan 909 kasus dan 6 karakteristik, memprediksi kelangsungan hidup penyakit jantung.</p> <p>Hasil penelitian menunjukkan bahwa metode pohon keputusan memiliki nilai terbaik. Selain itu, S.B. Kotsiantis menjelaskan dalam paper rivew bahwa metode decision tree memiliki kelebihan dalam pengolahan kumpulan data penyakit jantung, seperti transparansi pengetahuan atau klasifikasi, kecepatan klasifikasi, dan sifat diskrit, binari, atau kontinu [6],[7]..</p>
RP2 Menentukan Atribut atau dataset yang menjadi penyebab penyakit jantung	<p>Pada tahap ini, tipe data akan diubah untuk menyesuaikan dengan set data jantung yang akan diproses menjadi informasi atau lebih sederhana dilakukan encoding atau konversi data untuk membuatnya dapat digunakan dalam format tertentu, seperti yang ditunjukkan di bawah ini: 1. Usia (Fitur) = Integer 2. Jenis Kelamin (Fitur) = Binominal 3. Tipe Nyeri Dada (Fitur) = Polynominal 4. RestingBP (Fitur) = Integer 5. Kolesterol (Fitur) = Integer 6. EKG istirahat (Fitur) = Polynominal 7. MaxHR (Fitur) = Integer 8. Latihan Angina (Fitur) = Binominal 9. Puncak Tua (Fitur) = Real 10. ST_Lereng (Fitur) = Poly Penyakit Jantung (Label) adalah biner [8].</p>

2. Metode Penelitian

Penelitian ini menggunakan tahapan atau metode KDD (*Knowledge Discovery in Database*). KDD adalah proses untuk mencari informasi yang memiliki nilai, mudah dipahami dan baru dari penyimpanan data yang besar dan kompleks [9]. KDD sendiri merupakan metode yang digunakan untuk menghasilkan sebuah informasi melalui data yang sudah ada [10].

2.1 Proses Pengumpulan Data dan Atribut

Dalam penelitian ini, proses pengambilan data menggunakan teknik pengambilan data sekunder yang sudah tersedia pada salah satu situs penyedia dataset yaitu *Kaggle*. Data sekunder digunakan karena sumber data sudah ada sehingga tidak perlu melakukan observasi, wawancara maupun *survey* untuk mendapatkan datanya [15]. Dataset yang digunakan dapat diakses melalui link <https://www.kaggle.com/code/alkidiarete/heart-disease-roc-0->

98. Tabel 2 di bawah merupakan atribut keseluruhan dari dataset yang diperoleh dari *Kaggle*.



Gambar 1. Tahapan KDD

2.2 Data Selection

Tahapan pertama pada metode KDD ini adalah *data selection*, tahapan ini melakukan pengambilan dan pemilihan data serta melakukan pelabelan data [11]. untuk diolah menjadi informasi yang berguna. Tahap ini dilakukan karena ada beberapa atribut yang tidak terlalu penting untuk digunakan, proses ini dilakukan untuk memilih atribut yang sesuai pada prediksi gejala penyakit Jantung. Tabel 1 di bawah merupakan atribut yang akan digunakan dan diolah pada software RapidMiner.

Tabel 1. Atribut Terpilih

No	Atribut Terpilih	Keterangan
1.	Jenis Kelamin	Atribut Fitur
2.	Merokok	Atribut Fitur
3.	Kategori Usia	Atribut Fitur
4.	Kolesterol	Atribut Fitur
5.	Tekanan Darah	Atribut Fitur
6.	Nyeri Dada	Atribut Fitur
7.	Potensi	Atribut Label

Atribut yang terpilih adalah Jenis Kelamin, Merokok, Kategori Usia, Kolesterol, Tekanan Darah, Nyeri Dada, dan Potensi. Hal itu dikarenakan atribut tersebut sesuai dengan gejala terserang penyakit Jantung. Atribut tersebut akan diolah menggunakan algoritma C4.5 dan software RapidMiner yang akan menghasilkan pohon keputusan serta akurasi untuk memprediksi penyakit Jantung.

Tabel 2. Dataset Pasien

Jenis Kelamin	Merokok	Kategori Usia	Kolesterol	Tekanan Darah	Nyeri Dada	Potensi
Wanita	Ya	Pra Lansia	Diatas Normal	Tinggi	Sangat Nyeri	Ya
Wanita	Tidak	Lansia	Diatas Normal	Normal	Berat	Tidak
Pria	Ya	Lansia	Diatas Normal	Normal	Ringan	Tidak

Jenis Kelamin	Merokok	Kategori Usia	Kolesterol	Tekanan Darah	Nyeri Dada	Potensi
Wanita	Tidak	Lansia	Diatas Normal	Normal	Ringan	Tidak
Wanita	Tidak	Dewasa	Tinggi	Normal	Tidak	Tidak
Wanita	Ya	Lansia	Normal	Tinggi	Tidak	Tidak
Pria	Ya	Lansia	Tinggi	Tinggi	Ringan	Ya
Pria	Ya	Lansia	Tinggi	Normal	Ringan	Tidak
Wanita	Tidak	Lansia	Normal	Tinggi	Berat	Tidak
Pria	Ya	Lansia	Normal	Tinggi	Berat	Ya

Dari tabel 2 dapat dideskripsikan sebagai berikut:

- Jenis_Kelamin
"Jenis_Kelamin" adalah atribut yang mencakup jenis kelamin pasien dengan nilai Pria dan Wanita.
- Merokok
"Merokok" merupakan atribut yang diberikan kepada pasien sebagai status apakah merupakan perokok aktif atau pasif, nilai yang diberikan adalah Pasif dan Aktif.
- Kategori Usia
Atribut "Kategori Usia" adalah data usia pasien yang ada pada dataset ini yang diberi nilai Tua dan Muda.
- Kolesterol
"Kolesterol" merupakan atribut yang diberikan kepada pasien sebagai status apakah pasien mempunyai status Diatas Normal/Tinggi/Normal, , nilai yang diberikan adalah Ya dan Tidak.
- Tekanan Darah
"Tekanan Darah" merupakan atribut yang diberikan kepada pasien apakah pasien merokok, kelebihan berat badan, kurangnya olahraga, stres, dan konsumsi alkohol, nilai yang diberikan adalah Ya dan Tidak.
- Nyeri Dada
"Nyeri Dada" merupakan atribut yang diberikan kepada pasien apakah pasien sering mengalami aliran darah yang tersumbat seringkali dari bekuan darah ke otot jantung, nilai yang diberikan adalah Sangat Nyeri/Nyeri Berat/Sedang/ Ringan
- Potensi
Atribut "Potensi" merupakan label yang mencakup pasien yang terkena penyakit Jantung dengan nilai Ya (terkena penyakit Jantung) dan Tidak (Tidak terkena penyakit Jantung).

Dalam penentuan klasifikasi potensi penyakit jantung dengan atribut umur, jenis kelamin, merokok, tekanan darah, kadar kolesterol, dan nyeri dada memiliki beberapa kriteria seperti yang terlihat didalam Tabel 3.

Tabel 3. Atribut Dataset Penyakit Jantung

No	Kriteria	Keterangan
1.	Jenis Kelamin	Pria, Wanita
2.	Merokok	Ya(Aktif), Tidak(Pasif)
3.	Usia	Anak (0-9), Remaja (10-18), Dewasa (19-44), Pra Lanjut Usia (45- 59), Lanjut Usia (60-185)
4.	Kolesterol	Normal (0-199), Ambang Batas (200-239), Tinggi (240-500)
5.	Tekanan Darah	Normal(90-120), Tinggi(121-139) PreHipertensi, (140-159)Hipertensi stadium 1, (>160) Hipertensi Stadium 1
6.	Nyeri Dada	Sangat Nyeri(10),Nyeri Berat(7-0), Nyeri Sedang(4-6), Nyeri Ringan(1-3)

3. Hasil dan Pembahasan

Penelitian ini menggunakan metode algoritma C4.5 dengan alur proses seperti *data selection*, *data preprocessing*, *data transformation*, dan *evaluation*. Alur tersebut akan digunakan sebagai acuan dalam memprediksi gejala terserang penyakit jantung dengan hasil akhir yang didapat seperti nilai *accuracy*, *precision*

dan recall. Tabel 2 di bawah merupakan atribut keseluruhan dari dataset yang diperoleh dari *Kaggle*.

3.1 Data Transformation

Tahap ini mengembangkan data dengan tujuan *supaya* data lebih baik lagi dan siap pada tahap pemodelan data mining [12]. Tahap pada *transformation* ini adalah mengubah data ke dalam format yang sesuai untuk selanjutnya dapat diolah pada *data mining*, proses pengubahan pada dataset ini menggunakan RapidMiner dengan mengubah tipe data dari *polynomial* menjadi *binomial* pada atribut yang digunakan seperti yang ditunjukkan pada Gambar 2.

Jenis Kelamin binomial	Merokok binomial	Kategori Usia polynomial	Kolesterol polynomial	Tekanan Darah binomial	Nyeri Dada polynomial	POTENSI binomial label
Wanita	Ya	Pra Lansia	Diatas Normal	Tinggi	Heavy	YA
Wanita	Tidak	Lansia	Diatas Normal	Normal	Tinggi	TIDAK
Pria	Ya	Lansia	Diatas Normal	Normal	Ringan	TIDAK
Wanita	Tidak	Lansia	Diatas Normal	Normal	Ringan	TIDAK
Wanita	Tidak	Dewasa	Tinggi	Normal	Tidak	TIDAK
Wanita	Ya	Lansia	Normal	Tinggi	Tidak	TIDAK
Pria	Ya	Lansia	Tinggi	Tinggi	Ringan	YA
Pria	Ya	Lansia	Tinggi	Normal	Ringan	TIDAK
Wanita	Tidak	Lansia	Normal	Tinggi	Signifikan	TIDAK
Pria	Ya	Lansia	Normal	Tinggi	Signifikan	YA
Wanita	Ya	Pra Lansia	Diatas Normal	Tinggi	Heavy	YA
Wanita	Ya	Pra Lansia	Diatas Normal	Normal	Signifikan	TIDAK
Wanita	Tidak	Dewasa	Diatas Normal	Normal	Ringan	TIDAK
Pria	Tidak	Dewasa	Diatas Normal	Normal	Ringan	TIDAK
Pria	Tidak	Lansia	Tinggi	Normal	Tidak	TIDAK
Wanita	Ya	Dewasa	Normal	Tinggi	Tidak	YA
Pria	Tidak	Pra Lansia	Tinggi	Tinggi	Ringan	YA
Wanita	Tidak	Dewasa	Tinggi	Normal	Ringan	YA
Wanita	Ya	Dewasa	Normal	Normal	Signifikan	YA

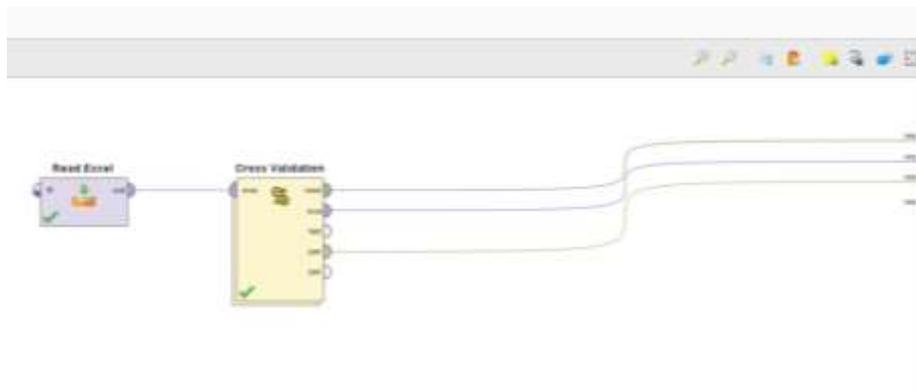
Gambar.2 Transformation pada dataset

3.2 Data Mining

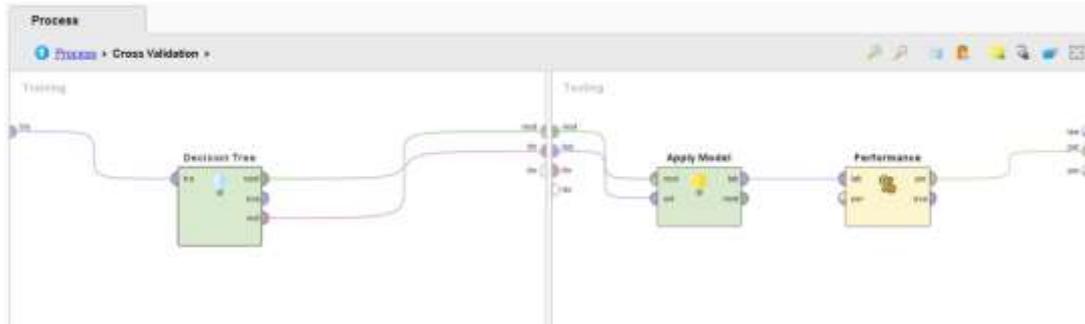
Data mining adalah teknik pengolahan suatu pengetahuan yang didasarkan pada sebuah *big data*, data yang biasa digunakan diambil dari *database*, *warehouse data*, *web* dan lain-lain untuk diproses menjadi informasi yang menarik. *Data mining* sendiri banyak digunakan di beberapa bidang untuk membantu peningkatan dalam hal pengetahuan, selain itu juga data mining digunakan untuk meningkatkan penjualan [13]. Tahap ini menerapkan proses pengolahan data menggunakan metode atau algoritma tertentu [14], pada penelitian ini menggunakan algoritma C4.5. Algoritma C4.5 digunakan dalam penerapan data mining ini dengan tujuan untuk memprediksi pasien penyakit paru-paru berdasarkan dataset yang sudah melewati tahap *selection* dan *transformation*.

3.3 Evaluation

Tahap yang terakhir yaitu *evaluation*, dimana tahap ini adalah keberhasilan dari model yang digunakan seperti menghasilkan nilai akurasi pada data diuji.

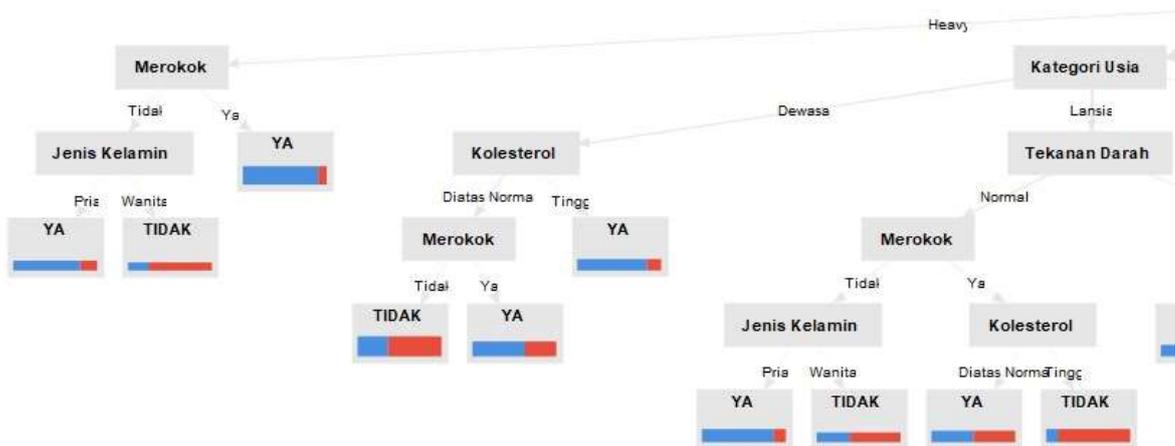


Gambar 3. Pengujian menggunakan Rapidminer



Gambar 4. Pengujian menggunakan Rapidminer

Gambar 3 dan 4 menjelaskan tentang proses pengujian menggunakan RapidMiner dimulai dengan melakukan import dataset kemudian *Cross Validation* untuk melatih model dan subprocess testing untuk pengujian sekaligus mengukur kinerja model, kemudian memilih *Apply Model* yang dimana model *Decision Tree* telah dipilih sebagai pengujian ini. *Performance* digunakan untuk menghitung akurasi yang diperoleh dari hasil prediksi, semua *socket* tersebut dihubungkan sehingga menghasilkan pohon keputusan (*Decision Tree*) pada prediksi pasien penyakit Jantung yang ditunjukkan pada Gambar 5.



Gambar 5. Tampilan pohon Keputusan

3.4 Nilai Accuracy

Nilai *accuracy* diperoleh dari hasil membagi data (*split data*) menjadi *data training* dan *data testing*, berikut Gambar 6 merupakan perbandingan dari beberapa *split data* yang diuji dari hasil nilai akurasi algoritma C4.5.

accuracy: 65.26% +/- 8.67% (micro average: 65.26%)

	true YA	true TIDAK	class precision
pred. YA	73	36	66.97%
pred. TIDAK	30	51	62.96%
class recall	70.87%	58.62%	

Gambar. 6 Nilai Accuracy

3.5. Nilai Recall

Recall adalah jumlah data yang true positive dibagi dengan jumlah data yang sebenarnya positif (true positive + true negative). Untuk nilai recall ditampilkan pada gambar 7 yaitu 70.87% pada potensi YA dan potensi TIDAK yaitu 58.62%.

recall: 58.19% +/- 17.20% (micro average: 58.62%) (positive class: TIDAK)

	true YA	true TIDAK	class precision
pred. YA	73	36	66.97%
pred. TIDAK	30	51	62.96%
class recall	70.87%	58.62%	

Gambar 7. Nilai Recall

3.6 Nilai Precision

Precision adalah jumlah data yang true positive (jumlah data positif yang dikenali secara benar sebagai positif) dibagi dengan jumlah data yang dikenali sebagai positif. Dari hasil pengujian yang ditampilkan pada gambar 8, nilai precision yaitu 70.87% untuk class YA dan 62.96% untuk class TIDAK.

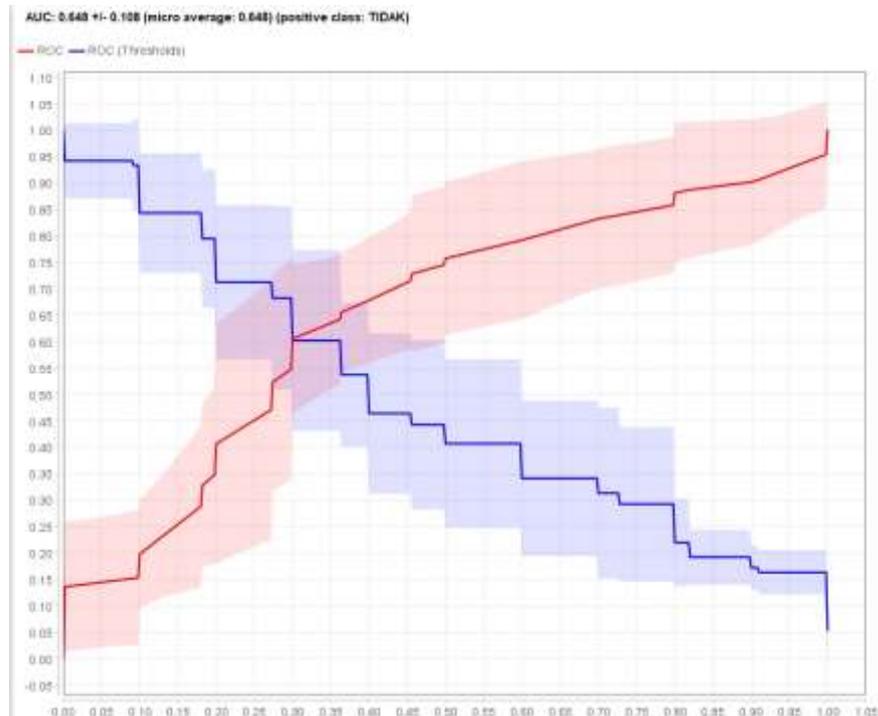
precision: 62.46% +/- 10.73% (micro average: 62.96%) (positive class: TIDAK)

	true YA	true TIDAK	class precision
pred. YA	73	36	66.97%
pred. TIDAK	30	51	62.96%
class recall	70.87%	58.62%	

Gambar 8. Nilai Precision

3.7. Nilai AUC

Selain confusion matrix, kurva Receiver Operating Characteristic (ROC) dihasilkan oleh rapid miner 10.3 kurva tersebut dapat dilihat pada gambar 9.



Gambar 9. Nilai AUC(Area Under Curva)

3.8 Pengolahan Data Menggunakan Algoritma C4.5

Dalam tahap ini dilakukan perhitungan untuk menentukan akar pohon berdasarkan rumus entropi dengan menggunakan dataset pada Tabel. 1 sehingga dihasilkan nilai entropi dari masing- masing atribut seperti yang terlihat dalam Tabel 4. Data training harus dipersiapkan terlebih dahulu sebelum di proses lalu atribut akan

dievaluasi dengan menghitung nilai Gain dan Entropy [9]. Rumus 1 adalah rumus untuk menghitung nilai Entropy.

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

Keterangan:

S = Himpunan kasus

n = Jumlah partisi S

p_i = Proporsi S_i terhadap S

Untuk menghitung nilai Gain dapat menggunakan Rumus 2

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} Entropy(S_i) \quad (2)$$

Tabel 4. Entropy Potensi Penyakit Jantung

Atribut	Keterangan	Jumlah	Iya	Tidak	Entropy
Jenis Kelamin	Pria	77	41	36	0,996956
	Wanita	113	62	51	0,993154
Merokok	Ya	99	60	39	0,967295
	Tidak	91	43	48	0,997821
Kategori Usia	Dewasa	71	40	31	0,988278
	Pra Lansia	45	25	20	0,991076
	Lansia	74	38	36	0,999473
Kolesterol	Normal	57	33	24	0,981941
	Diatas Normal	76	43	33	0,987475
Tekanan Darah	Tinggi	57	27	30	0,998001
	Normal	95	43	52	0,993516
	Tinggi	95	60	35	0,949452
Nyeri Dada	Sangat Nyeri	19	14	5	0,831474
	Nyeri Berat	57	32	25	0,989093
	Nyeri Sedang	76	45	31	0,975382
	Nyeri Ringan	38	12	26	0,899744

Sumber : Hasil Penelitian (2024)

Dari hasil perhitungan entropi dari setiap node pada Tabel 4 akan digunakan untuk menentukan nilai information gain. Dengan menggunakan rumus information gain pada Tabel 4 didapatkan nilai information gain yang nantinya akan digunakan untuk menentukan akar pohon yang paling atas. Hasil dari perhitungan nilai information gain dapat dilihat pada Tabel 5.

Tabel 5. Nilai *information gain*

Atribut	Information Gain	Gain Ratio
Jenis Kelamin	0,000184	0,000189
Merokok	0,012963	0,01298
Kategori Usia	0,00154	0,000992
Kolesterol	0,005906	0,00376
Tekanan Darah	0,023394	0,023394
Nyeri Dada	0,404799	0,219232

Sumber : Hasil Penelitian (2024)

Berdasarkan Tabel 5 diketahui atribut yang memiliki nilai information gain tertinggi adalah atribut Tekanan Darah Sistolik yang akan dijadikan sebagai akar pohon. Mengulangi langkah ke 2 sampai record dapat terpartisi Pada tahap ini diulang kembali langkah ke dua, namun perhitungan entropi hanya berdasarkan dari atribut jenis buku dikarenakan sebagai akar pohon. Setelah didapat nilai masing-masing entropi kemudian ditentukan nilai information gain atributnya.

4. Kesimpulan

Jantung ialah organ tubuh yang sangat penting untuk dijaga kesehatannya, karena begitu penting fungsi jantung untuk tubuh manusia. Penyakit jantung sangat sulit untuk dideteksi, sehingga menyebabkan banyak orang yang meninggal dunia secara mendadak tanpa disadari jika seseorang mengidap penyakit jantung atau terkena

serangan jantung. Penyakit jantung adalah penyebab kematian nomor satu di dunia. Menurut World Health Organization (2017) jumlah kematian diperkirakan sekitar 17, 9 juta orang. Penelitian ini mengambil data melalui situs penyedia dataset yaitu Kaggle sebanyak 30.000 data yang memiliki 190 atribut, tetapi hanya 10 atribut saja yang terpilih karena dinilai sesuai dengan prediksi penyakit Jantung. Metode Decision Tree dengan algoritma C4.5 dipilih sebagai proses pengujian pada penelitian ini, hasil yang diperoleh mulai dari nilai accuracy, precision, dan recall dengan membandingkan split data pada data training dan data testing yang terbesar adalah split data 90 – 10 yaitu $accuracy = 65.25\%$, $recall = 70.87\%$, dan $precision = 62.46\%$. Algoritma C4.5 yang diterapkan pada penelitian menghasilkan akurasi yang baik untuk memprediksi penyakit Jantung, untuk penelitian selanjutnya dapat dilakukan dengan mencoba algoritma lain dari teknik data mining agar mengetahui hasil akurasi yang di peroleh melalui algoritma lain sebagai perbandingan dengan penelitian ini.

Referensi

- [1] Pangaribuan, J. J., Tanjaya, H., & Kenichi, K. (2021). Mendeteksi penyakit jantung menggunakan machine learning dengan algoritma logistic regression. *Journal Information System Development (ISD)*, 6(2), 1-10.
- [2] Rokhayati, A., & Rumahorbo, H. (2020). Gambaran efikasi diri dalam pengelolaan faktor risiko dan pemeliharaan fungsi kesehatan pasien penyakit jantung koroner. *Jurnal riset kesehatan poltekkes depkes bandung*, 12(2), 285-296.
- [3] Malaeny, C. S., Katuuk, M., & Onibala, F. (2017). Hubungan Riwayat Lama Merokok Dan Kadar Kolesterol Total Dengan Kejadian Penyakit Jantung Koroner Di Poliklinik Jantung Rsu Pancaran Kasih Gmim Manado. *Jurnal Keperawatan*, 5(1).
- [4] Erawati, E. (2018). Hubungan tekanan darah dengan kadar kolesterol ldl (low density lipoprotein) pada penderita penyakit jantung koroner di rsup. dr. m. djamil padang. *Jurnal Kesehatan Perintis*, 5(2), 129-132.
- [5] Novitasari, F. (2023). Sistem Klasifikasi Penyakit Jantung Menggunakan Teknik Pendekatan SMOTE Pada Algoritma Modified K-Nearest Neighbor. *Jurnal Building of Informatics, Technology and Science (BITS)*, 5(1), 274-284.
- [6] Anggraini, R., Haerani, E., Jasril, J., & Afrianty, I. (2022). Pengelompokan Penyakit Pasien Menggunakan Algoritma K-Means. *JURIKOM(Jurnal Riset Komputer)*, 9(6), 1840. <https://doi.org/10.30865/jurikom.v9i6.5145>
- [7] Komarudin, R., Vicenna, P., Yudha, B., Maulana, Y. I., Afni, N., Salim, A., & Carolina, I. (2021). Bianglala Informatika Penerapan Metode Algoritma C4.5 Dalam Klasifikasi Diagnosa Penyakit Umum Menggunakan WEKA. *Bianglala Informatika*, 9(2).
- [8] 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.
- [9] Widaningsih, S. (2019). Perbandingan Metode Data Mining Untuk Prediksi Nilai Dan Waktu Kelulusan Mahasiswa Prodi Teknik Informatika Dengan Algoritma C4, 5, Naïve Bayes, Knn Dan Svm. *Jurnal Tekno Insentif*, 13(1), 16-25.
- [10] A. Yoga Pratama et al., “Analisis Sentimen Media Sosial Twitter Dengan Algoritma K- Nearest Neighbor Dan Seleksi Fitur Chi-Square (Kasus Omnibus Law Cipta Kerja),” *J. Sains Komput. Inform. (J-SAKTI)*, vol. 5, no. 2, pp. 897–910, 2021.
- [11] D. L. Rianti, Y. Umaidah, and A. Voutama, “Tren Marketplace Berdasarkan Klasifikasi Ulasan Pelanggan Menggunakan Perbandingan Kernel Support Vector Machine,” *STRING (Satuan Tulisan Ris. dan Inov. Teknol.*, vol. 6, no. 1, p. 98, 2021, doi: 10.30998/string.v6i1.9993.
- [12] Sofyan, F. M. A., Voutama, A., & Umaidah, Y. (2023). PENERAPAN ALGORITMA C4. 5 UNTUK PREDIKSI PENYAKIT PARU-PARU MENGGUNAKAN RAPIDMINER. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 7(2), 1409-1415.
- [13] S. Febriani and H. Sulistiani, “Analisis Data Hasil Diagnosa Untuk Klasifikasi Gangguan Kepribadian Menggunakan Algoritma C4.5,” *89Jurnal Teknol. dan Sist. Inf.*, vol. 2, no. 4, pp. 89–95, 2021.
- [14] Abdussalam Amrullah, Intam Purnamasari, Betha Nurina Sari, Garno, and Apriade Voutama, “Analisis Cluster Faktor Penunjang Pendidikan Menggunakan Algoritma K-Means (Studi Kasus: Kabupaten Karawang),” *J. Inform. dan Rekayasa Elektron.*, vol. 5, no. 2, pp. 244–252, 2022, doi: 10.36595/jire.v5i2.701
- [15] DQLab, “Pentingnya Implementasi Data Sekunder dalam Sebuah Penelitian,” 2022. <https://dqlab.id/pentingnya-implementasi-data- sekunder-dalam-sebuah-penelitian#:~:text=pembahasannya dibawah ini!-,1.,teknik pengumpulan data tertentu lainnya.> (accessed Apr. 30, 2023).