

Klasifikasi Penyakit Jantung Menggunakan Algoritma Decision Tree Series C4.5 Dengan Rapidminer

I Made Agus Oka Gunawan^{1*}, I Dewa Ayu Indah Saraswati², I Dewa Gede Riswana Agung³, I Putu Eka Putra⁴

^{1,2,3,4}Program Studi Sistem Informasi, Fakultas Sains dan Teknologi, Universitas Tabanan
Email : agusokagunawan@gmail.com¹, dewayuindahsaraswati@gmail.com², riswanaagung1002@gmail.com³, ekaputra1997@gmail.com⁴

Submitted: 27-02-2023, Reviewed: 08-03-2023, Accepted 26-03-2023
<https://doi.org/10.47233/jteksis.v5i2.775>

Abstract

Heart disease is still one of the leading causes of death in Indonesia and the world for both men and women of all ages. To reduce the number of deaths from heart disease, it is, therefore, necessary to conduct research to analyze data related to the causes of heart disease. In this study, the decision tree series C4.5 algorithm was used to classify heart disease data. The decision tree series C4.5 algorithm is processed in rapidminer version 9.10 tools. through the stages of Pre-processing, Set roles, modeling the decision tree series C4.5 algorithm on training data, applying the model to data testing, and testing to calculate the accuracy of the model on data testing. Testing using the confusion matrix resulted in an accuracy rate of 80.43% and a classification error of 19.57% was obtained. As well as the results of the Visualization of AUC (Area Under Curve) from the ROC curve, the value of AUC: 0.798 (Positive class): Heart Disease.

Keywords: classification, algorithm decision tree series C4.5, rapidminer, heart disease.

Abstrak

Penyakit jantung masih menjadi salah satu penyebab utama kematian tertinggi di Indonesia dan dunia baik pria maupun wanita pada segala usia. Untuk menekan jumlah kematian akibat penyakit jantung, maka dari itu perlunya dilakukan penelitian untuk menganalisa data terkait dengan penyebab penyakit jantung terjadi. Pada penelitian ini digunakan algoritma decision tree series C4.5 dalam melakukan proses klasifikasi data penyakit jantung. Algoritma decision tree series C4.5 ini diproses pada tools rapidminer version 9.10. melalui tahapan Pre-processing, Set role, pemodelan algoritma decision tree series C4.5 terhadap data training, apply model terhadap data testing, dan pengujian untuk menghitung keakuratan model terhadap data testing. Pengujian menggunakan confusion matrix menghasilkan tingkat Accuracy sebesar 80,43% dan didapatkan error classification sebesar 19,57%. Serta diperoleh hasil Visualisasi AUC (Area Under Curve) dari kurva ROC diperoleh nilai AUC:0,798 (Positive class): Penyakit Jantung.

Keywords: klasifikasi, algoritma decision tree series C4.5, rapidminer, penyakit jantung.

This work is licensed under Creative Commons Attribution License 4.0 CC-BY International license



PENDAHULUAN

Pemanfaatan teknologi hampir menjangkau segala jenis bidang yang ada [1], [2]. Salah satu bidang yang memanfaatkan teknologi adalah pengolahan data bidang kesehatan. Namun sebagian besar data tersebut tidak diolah dengan baik sehingga informasi yang dihasilkan kurang efektif untuk digunakan, seperti dalam mengambil keputusan kesehatan pasien [3]. Pengambilan keputusan atas dasar data dan informasi yang akurat akan menghasilkan keputusan ataupun prediksi terkait kesehatan atau penyakit seorang pasien menjadi lebih tepat [4]. Pemanfaatan teknologi yang dapat dimanfaatkan dalam permasalahan ini adalah sebuah sistem pendukung keputusan yang baik. Sistem pendukung keputusan merupakan sistem informasi interaktif yang menyediakan informasi, pemodelan, dan manipulasi data yang mendorong keputusan yang lebih cepat dan lebih cerdas berdasarkan data objektif, bukan berdasarkan kriteria subjektif atau naluri pribadi [5].

Kesehatan merupakan hal yang paling berharga bagi manusia, karena siapa saja dapat mengalami gangguan kesehatan [6]. Dewasa ini, banyak penyakit yang menyebabkan angka kematian meningkat, salah satunya adalah penyakit *Cardiovascular Diseases* (CVD). Berdasarkan laporan *World Health Organization* (WHO) angka kematian yang disebabkan oleh penyakit CVD mencapai 31% (17,7 juta orang setiap tahunnya) dari data seluruh kematian di dunia. Penyakit jantung merupakan salah satu penyakit *Cardiovascular Diseases*. Angka kejadian penyakit jantung dan pembuluh darah semakin meningkat dari tahun ke tahun. Berdasarkan data riset tahun 2018, menunjukkan bahwa 15 dari 1000 orang (sekitar 2.784.064 individu) di Indonesia menderita penyakit jantung [7]. Pada umumnya Infeksi jantung (*endokarditis*), disebabkan oleh virus atau bakteri. Bakteri yang paling umum ditemui sebagai penyebabnya adalah infeksi yaitu bakteri *Streptococcus beta hemolyticus* [8], [9].

Jantung merupakan hal yang penting untuk dijaga kesehatannya karena berfungsi memompa darah ke seluruh tubuh lewat pembuluh darah. Penyakit Jantung merupakan sebutan umum yang digunakan untuk menggambarkan gangguan terhadap fungsi kerja jantung [10]. Saat ini terdapat serangkaian gangguan yang mempengaruhi fungsi jantung, seperti: penyakit jantung koroner, kelainan irama jantung, penyakit jantung bawaan, kelainan katup atau klep jantung, dan gagal jantung [11]. Beberapa faktor gejala yang terdiagnosa dan dapat dijadikan dasar keputusan penyakit jantung diantaranya seperti jenis sakit dada (*chest Pain*), tekanan darah tinggi (*trebbs*), kolesterol (*chol*), nilai tes EKG (*resting electrodiagraphic(restecg)*), denyut jantung (*thalach*), dan kadar gula (*fastingBS*) [12]. Prediksi dari penyakit jantung perlu dilakukan karena keputusan klinis sering kali dibuat oleh dokter berdasarkan intuisi dan bukan pengalaman pengetahuan yang didapat atas data yang tersembunyi dalam *database*. Dengan keputusan klinis yang buruk akan menyebabkan ketidaktepatan diagnose, biaya medis yang berlebihan sehingga mempengaruhi kualitas pelayanan dan perawatan pada pasien. Akan Tetapi proses diagnosa merupakan hal yang cukup menantang [10] dikarenakan keterkaitan yang cukup kompleks diantara atribut-atribut (variabel bebas) yang mempengaruhi (variabel terikat (variabel yang diukur atau diuji dalam percobaan)) yakni penyakit jantung itu sendiri. Sehingga perlu diketahui sebuah *root node* (titik awal) yang dipakai oleh *user* untuk mengambil tindakan. Pada hasil akhirnya adalah sebuah *decision tree* dengan setiap cabangnya menunjukkan kemungkinan sekenario dari keputusan yang diambil serta hasilnya. Dan permasalahan yang sering kali dihadapi adalah kurangnya akurasi pada proses klasifikasi [13].

Klasifikasi dapat didefinisikan sebagai pekerjaan yang melakukan pelatihan/pembelajaran terhadap fungsi target yang memetakan setiap set atribut (fitur) ke satu jumlah label kelas yang tersedia [10]. Penelitian yang dilakukan oleh Anbarasi dkk dalam memprediksi kelangsungan hidup penyakit jantung dengan berdasarkan 909 kasus dan 6 atribut dengan menggunakan metode *naïve bayes*, *decision tree*, dan *clasifcation via clustering*. Hasil penelitian tersebut metode *decision tree* menghasilkan nilai terbaik [4]. Selain itu juga S.B. Kotsiantis dalam *rivew* papernya menjelaskan, bahwa metode *decision tree* mempunyai kelebihan-kelebihan dalam mengolah *dataset* penyakit jantung, baik dari segi kecepatan dalam klasifikasi, tiap atribut bersifat diskrit, binari, dan kontinue, serta transparansi pengetahuan atau klasifikasi [14].

Pohon keputusan (*Decision tree*) adalah salah satu bagian algoritma metode klasifikasi dari data mining dengan model klasifikasi menggunakan

struktur pohon, yang mana metode ini sangat populer, karena mudah untuk diinterpretasikan oleh manusia. Dan data mining itu sendiri merupakan proses untuk menggali (*mining*) pengetahuan dan informasi baru dari data yang berjumlah banyak pada data *warehouse*, dengan menggunakan kecerdasan buatan (*artificial intelligence*), statistik, dan matematika [15]. Untuk meringkas data dengan cara baru yang baik dimengerti dan bermanfaat bagi pemilik data [4].

Berdasarkan latar belakang di atas, maka pada penelitian ini akan dilakukan klasifikasi penyakit jantung menggunakan algoritma *decision tree*. Hasil pengujian dan evaluasi dalam mengukur *performance* atau hasil pada data testing berupa *Accuracy* yang didapat dari *confusion matrix* dan berupa AUC (*Area Under Curve*) yang di dapat dari kurva ROC.

METODE PENELITIAN

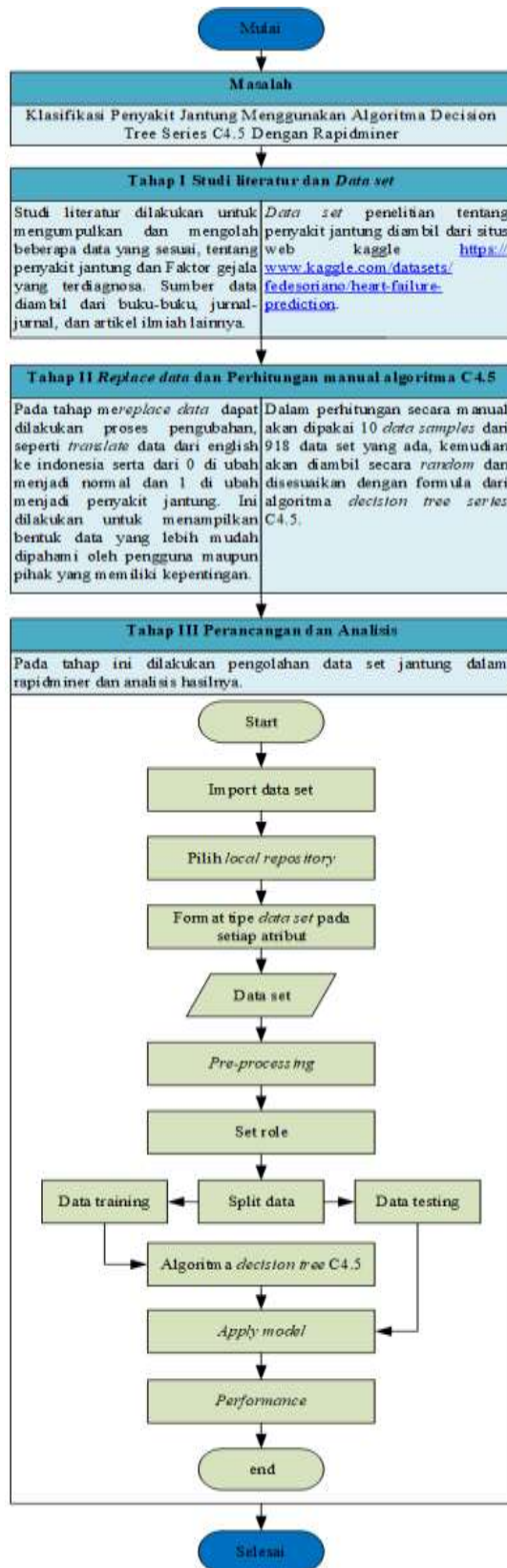
Pada tahap pertama akan dilakukan pencarian referensi dari studi literatur terkait algoritma *decision tree*, *rapidminer* dan penyakit jantung. Pada penelitian ini sumber data yang akan digunakan merupakan data sekunder dari yang diambil dari situs web *kaggle*. Situs ini menyediakan data-data *open source* terkait bidang *data science* [16], [17]



Gambar 1. Heart Disease Data Collection On Kaggle Web
(Sumber :

<https://www.kaggle.com/datasets/fedesoriano/heart-failure-prediction.>)

Algoritma *decision tree series* C4.5 digunakan pada penelitian ini untuk melakukan klasifikasi penyakit jantung. Algoritma C4.5 merupakan pengembangan dari algoritma ID3 [18]. Algoritma C4.5 merupakan salah satu *series* algoritma yang dipergunakan untuk membangun sebuah *decision tree* atau pohon keputusan dalam metode klasifikasi data mining yang mana bersifat prediktif [19], [20]. Elemen-elemen yang terdapat di algoritma C4.5 antara lain: nilai dari *Entropy*, *Gain*, *Split info*, dan *Gain ratio*. Dimana bagian yang menentukan dari simpul akar (*root*) sebuah *model tree* yang dihasilkan ini didasarkan pada nilai dari *Gain ratio* tertinggi [21].



Gambar 2. Kerangka Berpikir

Dalam melakukan proses pengolahan data, disini penulis menggunakan sebuah *tools* yaitu *rapidminer version 9.10*. *Rapidminer* adalah alat

penambangan data yang kuat untuk membuat model prediktif dengan cepat [22]. Alat *all-in-one* menampilkan ratusan persiapan data dan algoritma pembelajaran mesin untuk mendukung semua proyek *data mining*. Program ini memungkinkan pengguna untuk memasukkan data mentah, termasuk *database*, dan teks, yang kemudian dianalisis secara otomatis dan cerdas dalam skala besar. *Rapidminer* digunakan untuk penelitian, pendidikan, pelatihan, pembuatan prototipe cepat, pengembangan aplikasi, dan aplikasi industri [23].

Beberapa tahapan dilakukan pada penelitian ini dalam mengolah data set jantung. Data penelitian ini diambil dari Kaggle Dataset berjumlah 918 data dengan tipe data numerik dan nominal, dimana terdapat 12 atribut yang terdiri dari 11 atribut fitur dan 1 atribut label. Dari 918 data yang ada akan diambil 10 samples data set untuk perhitungan secara manual sesuai formula dari algoritma *decision tree series C4.5* untuk memberikan pemahaman bagaimana proses perhitungan yang terjadi dalam melakukan metode klasifikasi data menggunakan algoritma *decision tree series C4.5* sampai pada tahap menentukan dari simpul akar (*root*) sebuah model *tree* yang dihasilkan ini didasarkan pada nilai dari *Gain ratio* tertinggi, kemudian pada tahap sebenarnya dari 918 data yang ada akan diproses intern pada *tools rapidminer*. Melalui kerangka pemikiran yang penulis buat untuk penelitian ini, maka tahapan dan alur kerja sistem dapat dilihat pada Gambar 2.

HASIL DAN PEMBAHASAN

Hasil dari penelitian ini berupa pengolahan data kualitatif dan data kuantitatif dengan perhitungan yang dilakukan pada sebuah *public heart data set*, baik dilakukan perhitungan secara manual sesuai formula algoritma *decision tree series C4.5* maupun pada nantinya diproses intern pada *tools rapidminer*, yang mana nantinya akan membantu dalam menyelesaikan laporan penelitian ini.

1. Perhitungan manual algoritma *decision tree series C4.5*

ID	Age	Sex	cp	trestrest	trestrest	trestrest	chol	fastfiv	maxvd	exersa1	exersa2	exersa3	angina	restng	sttobp
1	69	M	1	150	160	Normal	234	0	163	0	0	0	0	0	136
2	35	F	1	150	160	Normal	234	0	163	0	0	0	0	0	136
3	76	F	1	150	160	Normal	234	0	163	0	0	0	0	0	136
4	63	F	1	150	160	Normal	234	0	163	0	0	0	0	0	136
5	56	F	1	150	160	Normal	234	0	163	0	0	0	0	0	136
6	52	F	1	150	160	Normal	234	0	163	0	0	0	0	0	136
7	55	F	1	150	160	Normal	234	0	163	0	0	0	0	0	136
8	67	F	1	150	160	Normal	234	0	163	0	0	0	0	0	136
9	66	F	1	150	160	Normal	234	0	163	0	0	0	0	0	136
10	56	F	1	150	160	Normal	234	0	163	0	0	0	0	0	136

Gambar 3. Contoh Data Sampel

Dari 10 *samples data set* yang diambil secara *random* pada Gambar 3, akan diperoleh perhitungan sebagai berikut.

- a. Langkah I, yaitu memisahkan data antara atribut fitur nominal dan atribut fitur numerik.

		Jumlah	Normal	Penyakit Jantung	ENTROPY
Total		10	4	6	0.9710
Jenis Kelamin	Laki-laki	7	1	6	0.9917
	Wanita	3	3	0	0.0000
Tipe Nyeri Dada	Tanpa Gejala	4	0	4	0.0000
	Angina Atipikal	4	3	1	0.8113
	Nyeri Non-Anginal	1	0	1	0.0000
Kadar Gula	Angina Khas	1	1	0	0.0000
	Normal	7	4	3	0.9852
EKG Istirahat	Pre-diabetes	3	0	3	0.0000
	ST Gelombang gelombang ST-T	3	0	3	0.0000
Normal	Normal	7	4	3	0.9852
	LVH (hipertrofi ventrikel kiri)	1	0	1	0.0000
Latihan Angina	Ya	4	0	4	0.0000
	Tidak	6	4	2	0.9183
	Menanjak	6	4	2	0.9183
ST_Lereng	Menanjak	6	4	2	0.9183
	Datar	4	0	4	0.0000

Gambar 4. Menghitung Jumlah Data Dan Nilai Entropy Data Nominal

- b. Langkah II, yaitu menghitung jumlah seluruh data, baik dari atribut label maupun masing-masing data berdasarkan atribut fitur, kemudian dari jumlah yang diperoleh, hitung kembali berapa data yang termasuk label “Normal” dan “Penyakit Jantung”. Kemudian bagian ini, dilanjutkan dengan menghitung nilai *entropy* total seluruh data berdasarkan atribut label dan menghitung nilai *entropy* untuk masing-masing data berdasarkan atribut fitur nominal.

Cara menghitung nilai *entropy*, sebagai berikut.

- *Entropy label*: Penyakit Jantung.

$$= (-\frac{4}{10} \log_2(\frac{4}{10})) + (-\frac{6}{10} \log_2(\frac{6}{10}))$$

$$= (-0,4 * -1,321) + (-0,6 * -0,736)$$

$$= (0,5284) + (0,4416)$$

$$= 0,97$$
- *Entropy* masing-masing data atribut fitur: Jenis Kelamin (Laki-laki dan Wanita), Tipe Nyeri Dada (Tanpa gejala, Angina atipikal, Nyeri non-anginal, dan Angina khas), Kadar Gula (Normal dan Pre-diabetes), EKG Istirahat (ST (kelainan gelombang ST-T), Normal, dan LVH (hipertrofi ventrikel kiri), Latihan Angina (Ya dan Tidak), dan ST_Lereng (Menanjak, Menurun, dan Datar).

Untuk perhitungan disini, akan diambil salah satu contoh pada bagian atribut fitur, yaitu Jenis Kelamin (Laki-laki dan Wanita).

$$= (-\frac{1}{7} \log_2(\frac{1}{7})) + (-\frac{6}{7} \log_2(\frac{6}{7}))$$

$$= (-0,142 * -2,807) + (-0,857 * -0,222)$$

$$= (0,398594) + (0,190254)$$

$$= 0,59 \text{ (Laki-laki)}$$

$$= (-\frac{3}{3} \log_2(\frac{3}{3})) + (-\frac{0}{3} \log_2(\frac{0}{3}))$$

$$= (-1 * 0) + (0 * 0)$$

$$= (0) + (0)$$

$$= 0 \text{ (Wanita)}$$

Pada perhitungan nilai *entropy* berikutnya untuk masing-masing data atribut fitur akan diselesaikan sama halnya dengan atribut fitur Jenis Kelamin (Laki-laki dan Wanita).

- c. Langkah III, yaitu menghitung dengan cara yang sama dengan langkah ke-II, akan tetapi bedanya pada bagian ini terdapat data atribut fitur numerik yang mana nantinya sebelum melakukan perhitungan nilai *entropy* terlebih dahulu dicari rata-rata dari masing-masing atribut, kemudian dari jumlah yang diperoleh, hitung kembali berapa data yang ada pada masing-masing atribut dan berapa yang termasuk label “Normal” dan “Penyakit Jantung”. Kemudian bagian ini, dilanjutkan dengan menghitung nilai *entropy* untuk masing-masing data berdasarkan atribut fitur numerik.

		Jumlah	Normal	Penyakit Jantung	ENTROPY
Total		10	4	6	0.9710
Umur	>48.2	4	1	3	0.8113
	<=48.2	6	3	3	1.0000
Resting BP	>130	4	2	2	1.0000
	<=130	6	2	4	0.9183
Kolesterol	>150.1	7	4	3	0.9852
	<=150.1	3	0	3	0.0000
MaxHR	>131.1	5	4	1	0.7219
	<=131.1	5	0	5	0.0000
Puncak Tua	>0.59	4	0	4	0.0000
	<=0.59	6	4	2	0.9183

Gambar 5. Menghitung Jumlah Data Dan Nilai Entropy Data Numerik

Cara menghitung nilai *entropy*, sebagai berikut.

- *Entropy* masing-masing data atribut fitur:
Umur(>48,2 dan <=48,2),
RestingBP(>130 dan <=130),
Kolesterol(>150,1 dan <=150,1),
MaxHR(>131,1 dan <=131,1), dan
Puncak Tua(>0,59 dan <=0,59).

Untuk perhitungan disini, akan diambil salah satu contoh pada bagian atribut fitur, yaitu Umur(>48,2 dan <=48,2).

$$\begin{aligned} &= -(1/4)*\log_2(1/4)+-(3/4)*\log_2(3/4) \\ &= (-0,25*-2)+(-0,75*-0,41) \\ &= (0,5)+(0,3075) \\ &= 0,81 (>48,2) \end{aligned}$$

$$\begin{aligned} &= -(3/6)*\log_2(3/6)+-(3/6)*\log_2(3/6) \\ &= (-0,5*-1)+(-0,5*-1) \\ &= (0,5)+(0,5) \\ &= 1 (<=48,2) \end{aligned}$$

Pada perhitungan nilai *entropy* berikutnya untuk masing-masing data atribut fitur akan diselesaikan sama halnya dengan atribut fitur Umur (>48,2 dan <=48,2).

- d. Langkah IV, yaitu menghitung nilai *gain* pada seluruh atribut fitur baik nominal maupun numerik.

		Jumlah	Normal	Prevalensi	ENTROPY	GAIN
Total		10	4	6	0.9710	
Jenis Kelamin						0.5688
	Laki-laki	7	1	6	0.5917	
	Wanita	3	3	0	0.0000	
Tipe Merokok						0.6464
	Tanpa Gejala	4	0	4	0.0000	
	Angina Atipikal	4	3	1	0.8113	
	Myo Non-Anginal	1	0	1	0.0000	
	Angina Khas	1	1	0	0.0000	
Prevalensi						0.2613
	Normal	7	4	3	0.9852	
	Pre-diabetes	3	0	3	0.0000	
RISIKO serangan						0.261
	ST (Serangan gelombang ST-T)	2	0	2	0.0000	
	Normal	7	4	3	0.9852	
	LVH (Papeteris ventrikel kiri)	1	0	1	0.0000	
Angina Angina						0.42
	Ya	4	0	4	0.0000	
	Tidak	6	4	2	0.9183	
Prevalensi						0.61
	Meningkat	5	4	1	0.7219	
	Merkuri	2	0	2	0.0000	
	Costar	3	0	3	0.0000	

Gambar 6. Menghitung Nilai *Gain* Data Nominal

		Jumlah	Normal	Prevalensi	ENTROPY	GAIN
Total		10	4	6	0.9710	
Umur						0.6464
	>48.2	4	1	3	0.8113	
	<=48.2	6	3	3	1.0000	
Resting BP						0.0200
	>130	4	2	2	1.0000	
	<=130	6	2	4	0.9183	
Kolesterol						0.2813
	>150.1	7	4	3	0.9852	
	<=150.1	3	0	3	0.0000	
MaxHR						0.6100
	>131.1	5	4	1	0.7219	
	<=131.1	5	0	5	0.0000	
Puncak Tua						0.41997
	>0.59	4	0	4	0.0000	
	<=0.59	6	4	2	0.9183	

Gambar 7. Menghitung Nilai *Gain* Data Numerik

Untuk perhitungan disini, akan diambil salah satu contoh pada bagian atribut fitur nominal, yaitu pada Jenis Kelamin dan atribut fitur numerik, yaitu Umur.

$$\begin{aligned} &= 0,97-((7/10*0,59)+(3/10*0)) \\ &= 0,97-((0,413)+(0)) \\ &= 0,97-((0,413) \\ &= 0,55 (\text{Jenis kelamin}) \end{aligned}$$

$$\begin{aligned} &= 0,97-((4/10*0,81)+(6/10*1)) \\ &= 0,97-((0,324)+(0,6)) \\ &= 0,97-((0,924) \\ &= 0,046 (\text{Umur}) \end{aligned}$$

Pada perhitungan nilai *gain* berikutnya untuk masing-masing data atribut fitur nominal maupun data atribut fitur numerik, akan diselesaikan sama halnya dengan atribut fitur Jenis kelamin dan Umur.

		Jumlah	Normal	Prevalensi	ENTROPY	GAIN	SPLIT INFO
Total		10	4	6	0.9710		
Jenis Kelamin						0.5688	0.88126
	Laki-laki	7	1	6	0.5917		
	Wanita	3	3	0	0.0000		
Tipe Merokok						0.6464	1.27183
	Tanpa Gejala	4	0	4	0.0000		
	Angina Atipikal	4	3	1	0.8113		
	Myo Non-Anginal	1	0	1	0.0000		
	Angina Khas	1	1	0	0.0000		
Prevalensi						0.2613	0.88126
	Normal	7	4	3	0.9852		
	Pre-diabetes	3	0	3	0.0000		
RISIKO serangan						0.261	1.18878
	ST (Serangan gelombang ST-T)	2	0	2	0.0000		
	Normal	7	4	3	0.9852		
	LVH (Papeteris ventrikel kiri)	1	0	1	0.0000		
Angina Angina						0.42	0.91080
	Ya	4	0	4	0.0000		
	Tidak	6	4	2	0.9183		
Prevalensi						0.61	1.38888
	Meningkat	5	4	1	0.7219		
	Merkuri	2	0	2	0.0000		
	Costar	3	0	3	0.0000		
Umur						0.646	0.91080
	>48.2	4	0	4	0.0000		
	<=48.2	6	4	2	0.9183		
Resting BP						0.020	1.38888
	>130	4	0	4	0.0000		
	<=130	6	0	6	0.0000		
Kolesterol						0.281	1.38888
	>150.1	7	0	7	0.0000		
	<=150.1	3	0	3	0.0000		
MaxHR						0.610	1.38888
	>131.1	5	0	5	0.0000		
	<=131.1	5	0	5	0.0000		
Puncak Tua						0.420	1.38888
	>0.59	4	0	4	0.0000		
	<=0.59	6	4	2	0.9183		

Gambar 8. Menghitung nilai *split info* data nominal

- e. Langkah V, yaitu menghitung nilai *split info* pada seluruh atribut fitur baik nominal maupun numerik.

	Jumlah	Minim	Maksimal	ENTROPI	GAIN	SPLIT INFO
Total	10	4	6	0.9710		
Umur					0.0468	0.97095
>=48.2	4	1	3	0.8113		
<48.2	6	4	4	1.0000		
Jenis Kelamin					0.0200	0.97095
>=130	4	2	2	1.0000		
<130	6	2	4	0.9183		
Marital Status					0.2818	0.88129
>=150.1	7	4	3	0.9852		
<150.1	3	0	3	0.0000		
Education					0.0100	1
>=131.1	3	1	1	0.7219		
<131.1	7	0	5	0.0000		
Income Tax					0.41097	0.97095
>=0.59	4	0	4	0.0000		
<0.59	6	4	2	0.9183		

Gambar 9. Menghitung Nilai *Split Info* Data Numerik

Untuk perhitungan disini, akan diambil salah satu contoh pada bagian atribut fitur nominal, yaitu pada Jenis Kelamin dan atribut fitur numerik, yaitu Umur.

$$= ((7/10 * (\log_2(7/10))) + (3/10 * (\log_2(3/10))))$$

$$= (-0,7 * -0,514) + (-0,3 * -1,736)$$

$$= (0,3598) + (0,5208)$$

$$= 0,88 \text{ (Jenis kelamin)}$$

$$= ((4/10 * (\log_2(4/10))) + (6/10 * (\log_2(6/10))))$$

$$= (-0,4 * -1,321) + (-0,6 * -0,736)$$

$$= (0,5284) + (0,4416)$$

$$= 0,97 \text{ (Umur)}$$

Pada perhitungan nilai *split info* berikutnya untuk masing-masing data atribut fitur nominal maupun data atribut fitur numerik, akan diselesaikan sama halnya dengan atribut fitur Jenis kelamin dan Umur.

- f. Langkah VI, yaitu menghitung nilai *gain ratio* pada seluruh atribut fitur baik nominal maupun numerik.

	Jumlah	Minim	Maksimal	ENTROPI	GAIN	SPLIT INFO	GAIN RATIO
Total	10	4	6	0.9710			
Umur					0.0468	0.97095	0.05
>=48.2	4	1	3	0.8113			
<48.2	6	3	3	1.0000			
Jenis Kelamin					0.0200	0.97095	0.02
>=130	4	2	2	1.0000			
<130	6	2	4	0.9183			
Marital Status					0.2818	0.88129	0.33
>=150.1	7	4	3	0.9852			
<150.1	3	0	3	0.0000			
Education					0.0100	1	0.01
>=131.1	3	1	1	0.7219			
<131.1	7	0	5	0.0000			
Income Tax					0.41097	0.97095	0.43
>=0.59	4	0	4	0.0000			
<0.59	6	4	2	0.9183			

Gambar 10. Menghitung Nilai *Gain Ratio* Data Nominal

Untuk perhitungan disini, akan diambil salah satu contoh pada bagian atribut fitur nominal, yaitu pada Jenis Kelamin dan atribut fitur numerik, yaitu Umur.

$$= (0,55/0,88)$$

$$= 0,63 \text{ (Jenis kelamin)}$$

$$= (0,046/0,97)$$

$$= 0,05 \text{ (Umur)}$$

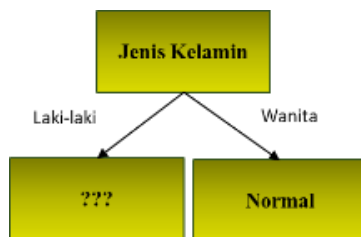
Pada perhitungan nilai *gain ratio* berikutnya untuk masing-masing data atribut fitur nominal maupun data atribut fitur numerik, akan diselesaikan sama halnya dengan atribut fitur Jenis kelamin dan Umur.

	Jumlah	Minim	Maksimal	ENTROPI	GAIN	SPLIT INFO	GAIN RATIO
Total	10	4	6	0.9710			
Umur					0.0468	0.97095	0.05
>=48.2	4	1	3	0.8113			
<48.2	6	3	3	1.0000			
Jenis Kelamin					0.0200	0.97095	0.02
>=130	4	2	2	1.0000			
<130	6	2	4	0.9183			
Marital Status					0.2818	0.88129	0.33
>=150.1	7	4	3	0.9852			
<150.1	3	0	3	0.0000			
Education					0.0100	1	0.01
>=131.1	3	1	1	0.7219			
<131.1	7	0	5	0.0000			
Income Tax					0.41097	0.97095	0.43
>=0.59	4	0	4	0.0000			
<0.59	6	4	2	0.9183			

Gambar 11. Menghitung Nilai *Gain Ratio* Data Numerik

Sampai pada tahap ini, dapat ditentukan simpul akar (*root*) sebuah model *tree* yang dihasilkan berdasarkan pada nilai *Gain ratio* tertinggi yang dilihat dari

seluruh data, baik bagian atribut fitur nominal maupun numerik. Dan nilai *gain ratio* tertinggi terdapat di bagian atribut fitur nominal, yaitu Jenis kelamin. Sebagai berikut.



Gambar 12. Model Pohon Keputusan

Pada model tree yang dihasilkan, dimana adanya penjelasan bahwa “Apakah orang tersebut terdapat penyakit jantung atau normal?”, maka dapat ditentukan dari Jenis kelamin orang tersebut, jika dia seorang wanita tentu hasilnya “Normal” ini bisa dilihat dari jumlah data wanita, dimana terdapat 3 (Normal) dan 0 (Penyakit jantung). Akan tetapi, jika dia seorang laki-laki, maka akan dicari lagi perhitungannya. Ini bisa dilihat dari jumlah data laki-laki, dimana terdapat 1 (Normal) dan 6 (Penyakit jantung) tentu dapat dikatakan setelah dipengaruhi oleh fitur “Jenis kelamin” akan ada lagi fitur yang mempengaruhinya. Untuk itu dilakukan kembali perhitungan dari langkah ke-II dengan menghilangkan hanya pada data “Jenis kelamin”, karena sudah terpakai untuk menentukan perhitungan di awal. Dan proses ini dilakukan sampai menemukan kembali nilai *gain ratio* tertinggi.

2. Rancangan data mining pada *tools rapidminer version 9.10*

a. Format tipe *data set* jantung

Pada tahap ini akan dilakukan perubahan tipe data untuk menyesuaikan pada data set jantung yang akan diolah menjadi suatu informasi atau lebih sederhana dilakukan konversi atau encoding untuk mengubah data menjadi format tertentu agar dapat digunakan, seperti berikut:

1. Umur (Fitur) = *Integer*
2. Jenis Kelamin (Fitur) = *Binominal*
3. Tipe Nyeri Dada (Fitur) = *Polynomial*

4. RestingBP (Fitur) = *Integer*
5. Kolesterol (Fitur) = *Integer*
6. Kadar gula (Fitur) = *Binominal*
7. EKG istirahat (Fitur) = *Polynomial*
8. MaxHR (Fitur) = *Integer*
9. Latihan Angina (Fitur) = *Binominal*
10. Puncak Tua (Fitur) = *Real*
11. ST_Lereng (Fitur) = *Polynomial*
12. Penyakit Jantung (Label) = *Binominal*



Gambar 13. Format Tipe *Data Set* Pada Setiap Atribut

b. *Pre-Processing*

Pembersihan data seperti adanya data *missing*, *noise*, dan *duplicate* maupun *select attributes*.



Gambar 14. *Pre-Processing*

c. *Set role*

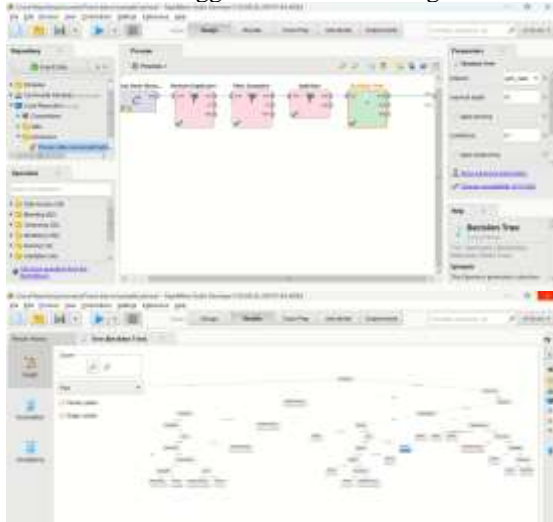
Pada tahap ini dilakukan pengaturan peran pada *data set* jantung, dimana data dipisahkan otomatis dengan operator *split data* dengan perbandingan 90:10 untuk data *training* dan data *testing*.



Gambar 15. *Set role*

d. Pemodelan algoritma *decision tree series* C4.5

Pada tahap ini dilakukan pemodelan algoritma *decision tree series* C4.5 dengan *criterion gain_ratio* pada *data set* jantung yang terdiri dari 11 atribut fitur yang terdiagnosa penyakit jantung dan 1 label yang merupakan hasil akhir prediksi. Pemodelan berupa pohon keputusan ini ditentukan menggunakan data training.



Gambar 16. Model Pohon Keputusan Klasifikasi Penyakit Jantung Menggunakan Algoritma *Decision Tree Series* C4.5

Rule yang sudah di hitung di atas pada Gambar 7 akan terlihat, seperti berikut.

ST_Lereng = Datar
 | MaxHR > 183: **Normal** {Normal=2, Penyakit Jantung=0}***Rule 1**
 | MaxHR ≤ 183
 | | Jenis Kelamin = Laki-laki
 | | | MaxHR > 176.500
 | | | | Umur > 40.500: **Normal** {Normal=3, Penyakit Jantung=0}***Rule 2**
 | | | | Umur ≤ 40.500: **Penyakit Jantung** {Normal=0, Penyakit Jantung=2}***Rule 3**
 | | | | MaxHR ≤ 176.500
 | | | | MaxHR > 159.500
 | | | | | Puncak Tua > 0.800
 | | | | | | Kolesterol > 256: **Penyakit Jantung** {Normal=0, Penyakit Jantung=4}***Rule 4**
 | | | | | | Kolesterol ≤ 256
 | | | | | | | Umur > 57: **Penyakit Jantung** {Normal=1, Penyakit Jantung=2}***Rule 5**
 | | | | | | | Umur ≤ 57: **Normal** {Normal=6, Penyakit Jantung=0}***Rule 6**

| | | | | Puncak Tua ≤ 0.800: **Penyakit Jantung** {Normal=0, Penyakit Jantung=10}***Rule 7**
 | | | | | MaxHR ≤ 159.500
 | | | | | | RestingBP > 105.500
 | | | | | | | MaxHR > 95.500
 | | | | | | | MaxHR > 97.500: **Penyakit Jantung** {Normal=21, Penyakit Jantung=245}***Rule 8**
 | | | | | | | MaxHR ≤ 97.500
 | | | | | | | | Umur > 55.500: **Normal** {Normal=2, Penyakit Jantung=0}***Rule 9**
 | | | | | | | | Umur ≤ 55.500: **Penyakit Jantung** {Normal=0, Penyakit Jantung=3}***Rule 10**
 | | | | | | | | MaxHR ≤ 95.500: **Penyakit Jantung** {Normal=0, Penyakit Jantung=31}***Rule 11**
 | | | | | | | | RestingBP ≤ 105.500
 | | | | | | | | | Kolesterol > 162.500: **Normal** {Normal=3, Penyakit Jantung=1}***Rule 12**
 | | | | | | | | | Kolesterol ≤ 162.500: **Penyakit Jantung** {Normal=0, Penyakit Jantung=8}***Rule 13**
 | | | | | | | | | Jenis Kelamin = Wanita
 | | | | | | | | | | Kadar gula = Normal
 | | | | | | | | | | RestingBP > 155: **Penyakit Jantung** {Normal=0, Penyakit Jantung=5}***Rule 14**
 | | | | | | | | | | RestingBP ≤ 155
 | | | | | | | | | | | Umur > 38.500
 | | | | | | | | | | | Puncak Tua > 2.300: **Penyakit Jantung** {Normal=0, Penyakit Jantung=2}***Rule 15**
 | | | | | | | | | | | Puncak Tua ≤ 2.300
 | | | | | | | | | | | | Latihan Angina = Tidak
 | | | | | | | | | | | | RestingBP > 148: **Penyakit Jantung** {Normal=0, Penyakit Jantung=2}***Rule 16**
 | | | | | | | | | | | | RestingBP ≤ 148: **Normal** {Normal=28, Penyakit Jantung=3}***Rule 17**
 | | | | | | | | | | | | | Latihan Angina = Ya
 | | | | | | | | | | | | | | Umur > 46.500: **Penyakit Jantung** {Normal=5, Penyakit Jantung=14}***Rule 18**
 | | | | | | | | | | | | | | Umur ≤ 46.500: **Normal** {Normal=2, Penyakit Jantung=0}***Rule 19**
 | | | | | | | | | | | | | | | Umur ≤ 38.500: **Penyakit Jantung** {Normal=0, Penyakit Jantung=2}***Rule 20**
 | | | | | | | | | | | | | | | Kadar gula = Pre-diabetes: **Penyakit Jantung** {Normal=0, Penyakit Jantung=10}***Rule 21**
 ST_Lereng = menanjak
 | | | | | | | | | | | Puncak Tua > 2.350: **Penyakit Jantung** {Normal=0, Penyakit Jantung=7}***Rule 22**
 | | | | | | | | | | | Puncak Tua ≤ 2.350
 | | | | | | | | | | | | Kolesterol > 42.500
 | | | | | | | | | | | | | Latihan Angina = Tidak

| | | | MaxHR > 193.500: **Normal**
 {Normal=1, Penyakit Jantung=1}***Rule 23**
 | | | | MaxHR ≤ 193.500
 | | | | Umur > 56.500
 | | | | | MaxHR > 151.500
 | | | | | MaxHR > 171.500: **Normal**
 {Normal=7, Penyakit Jantung=0}***Rule 24**
 | | | | | MaxHR ≤ 171.500
 | | | | | Puncak Tua > 0.850:
Normal {Normal=4, Penyakit
 Jantung=0}***Rule 25**
 | | | | | Puncak Tua ≤ 0.850:
Penyakit Jantung {Normal=6, Penyakit
 Jantung=9}***Rule 26**
 | | | | | MaxHR ≤ 151.500: **Normal**
 {Normal=30, Penyakit Jantung=2}***Rule**
27
 | | | | | Umur ≤ 56.500: **Normal**
 {Normal=201, Penyakit Jantung=6}***Rule**
28
 | | | Latihan Angina = Ya
 | | | | Tipe Nyeri Dada = Angina
 Atipikal: **Normal** {Normal=6, Penyakit
 Jantung=0}***Rule 29**
 | | | | Tipe Nyeri Dada = Angina Khas:
Normal {Normal=2, Penyakit
 Jantung=0}***Rule 30**
 | | | | Tipe Nyeri Dada = Nyeri Non-
 Anginal: **Normal** {Normal=7, Penyakit
 Jantung=0}***Rule 31**
 | | | | Tipe Nyeri Dada = Tanpa Gejala
 | | | | | Puncak Tua > 0.700: **Penyakit**
Jantung {Normal=0, Penyakit
 Jantung=12}***Rule 32**
 | | | | | Puncak Tua ≤ 0.700
 | | | | | RestingBP > 131: **Normal**
 {Normal=6, Penyakit Jantung=0}***Rule 33**
 | | | | | RestingBP ≤ 131
 | | | | | Kolesterol > 287: **Normal**
 {Normal=2, Penyakit Jantung=0}***Rule 34**
 | | | | | Kolesterol ≤ 287: **Penyakit**
Jantung {Normal=0, Penyakit
 Jantung=4}***Rule 35**
 | | | Kolesterol ≤ 42.500
 | | | | Kadar gula = Normal
 | | | | | MaxHR > 102.500
 | | | | | MaxHR > 119
 | | | | | Umur > 67: **Penyakit Jantung**
 {Normal=0, Penyakit Jantung=2}***Rule 36**
 | | | | | Umur ≤ 67
 | | | | | | Umur > 52.500: **Normal**
 {Normal=7, Penyakit Jantung=0}***Rule 37**
 | | | | | | Umur ≤ 52.500
 | | | | | | Tipe Nyeri Dada = Nyeri
 Non-Anginal: **Normal** {Normal=2,
 Penyakit Jantung=0}***Rule 38**

| | | | | | Tipe Nyeri Dada = Tanpa
 Gejala: **Penyakit Jantung** {Normal=0,
 Penyakit Jantung=3}***Rule 39**
 | | | | | MaxHR ≤ 119: **Normal**
 {Normal=4, Penyakit Jantung=0}***Rule 40**
 | | | | | MaxHR ≤ 102.500: **Penyakit**
Jantung {Normal=0, Penyakit
 Jantung=3}***Rule 41**
 | | | | Kadar gula = Pre-diabetes: **Penyakit**
Jantung {Normal=0, Penyakit
 Jantung=18}***Rule 42**
 ST_Lereng = Menurun
 | | | | MaxHR > 184.500: **Normal** {Normal=2,
 Penyakit Jantung=0}***Rule 43**
 | | | | MaxHR ≤ 184.500
 | | | | Latihan Angina = Tidak
 | | | | | RestingBP > 122.500
 | | | | | Umur > 60
 | | | | | Umur > 62.500: **Normal**
 {Normal=3, Penyakit Jantung=1}***Rule 44**
 | | | | | Umur ≤ 62.500: **Penyakit Jantung**
 {Normal=0, Penyakit Jantung=3}***Rule 45**
 | | | | | Umur ≤ 60: **Normal** {Normal=4,
 Penyakit Jantung=0}***Rule 46**
 | | | | | RestingBP ≤ 122.500: **Penyakit**
Jantung {Normal=0, Penyakit
 Jantung=7}***Rule 47**
 | | | | Latihan Angina = Ya
 | | | | | Kolesterol > 299.500: **Penyakit**
Jantung {Normal=1, Penyakit
 Jantung=2}***Rule 48**
 | | | | | Kolesterol ≤ 299.500
 | | | | | MaxHR > 140.500
 | | | | | Umur > 57: **Penyakit Jantung**
 {Normal=0, Penyakit Jantung=2}***Rule 49**
 | | | | | Umur ≤ 57: **Normal** {Normal=1,
 Penyakit Jantung=1}***Rule 50**
 | | | | | MaxHR ≤ 140.500: **Penyakit**
Jantung {Normal=0, Penyakit
 Jantung=30}***Rule 51**

Jumlah rule akhir adalah 51 yang dihasilkan dari metode klasifikasi algoritma *decision tree series* C4.5, dengan jumlah label sebanyak 24 rule (Normal) dan 27 rule (Penyakit jantung).

e. Apply model

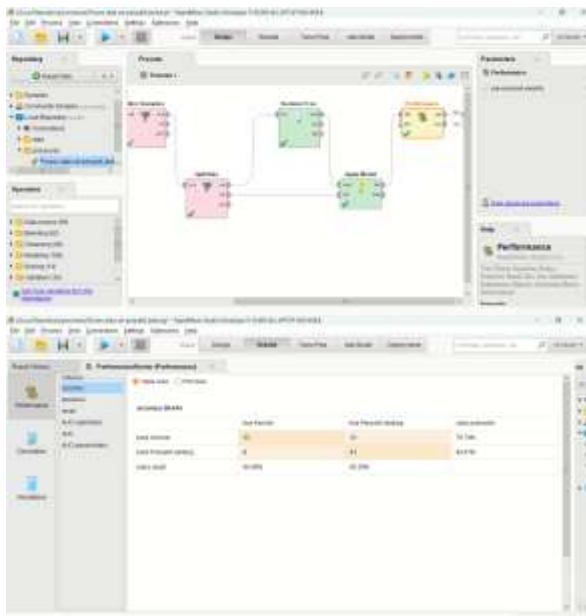
Pada tahap ini dilakukan penerapan model yang telah dilatih sebelumnya menggunakan data *training* pada *unlabeled data* (Data *testing*) dengan tujuan untuk mendapatkan prediksi pada *unlabeled data* (Data *testing*) yang belum memiliki label atau keterangan apakah orang tersebut memiliki peluang mengalami penyakit jantung atau normal.



Gambar 17. Apply Model

f. Performance

Pada tahap ini dilakukan pengujian untuk menghitung keakuratan model untuk data *testing* yang mana akan menampilkan *Accuracy* yang didapat dari *confusion matrix* dan berupa *AUC (Area Under Curve)* yang di dapat dari kurva *ROC* pada proses pengolahan *data set* jantung menggunakan *algoritma decision tree series C.45*.



Gambar 18. Hasil Pengujian Tingkat Accuracy Dari Confusion Matrix

Dari hasil perhitungan *confusion matrix* yang sudah dilakukan mendapatkan hasil, seperti berikut.

Pred. Normal-True Normal: Jumlah data yang diprediksi normal dan kenyataan normal (TP)

Pred. Penyakit Jantung-True Penyakit Jantung: Jumlah data yang diprediksi penyakit jantung dan kenyataan penyakit jantung (TN)

Pred. Normal-True Penyakit Jantung: Jumlah data yang diprediksi normal tapi kenyataan penyakit jantung (FP)

Pred. Penyakit Jantung-True Normal: Jumlah data yang diprediksi penyakit jantung tapi kenyataan normal (FN), Kemudian hasil prediksi *class recall true* normal sebesar 80,49% dan *true* penyakit jantung sebesar 80,39%. Hasil dari *class precision* sebesar *pred* normal sebesar 76,74% dan *pred* penyakit jantung sebesar 83,67% dan *accuracy* prediksi, dimana ini diperoleh berdasarkan perhitungan $accuracy = \frac{TP+TN}{TP+TN+FP+FN} = \frac{33+41}{33+41+10+8} = \frac{74}{92} = 80,43\%$ dan didapatkan *error classification*, dimana ini diperoleh dari $100\% - 80,43\% = 19,57\%$.



Gambar 19. Hasil Visualisasi AUC (Area Under Curve) dari Kurva ROC

Hasil dari prediksi penyakit jantung juga dapat dilihat dalam kurva *ROC (Receiver Operating Characteristic)* pada Gambar 19, dimana dijelaskan bahwa nilai *AUC:0,798 (Positive class)*: Penyakit Jantung, garis merah merupakan *ROC* dan garis biru merupakan *ROC (Thresholds)*.

SIMPULAN

Dari penjelasan di atas dapat diambil kesimpulan penelitian ini, yaitu dari dataset yang tersedia berjumlah 918 data dengan tipe data numerik dan nominal, dimana terdapat 12 atribut yang terdiri dari 11 atribut fitur dan 1 atribut label ini di lakukan proses pengolahan data dengan metode klasifikasi menggunakan algoritma *decision tree series C4.5* sebagai versi perbaikan dari ID3 memperoleh perhitungan yang bagus dalam tahap sebenarnya yang di proses pada tools Rapidminer version 9.10. melalui *Pre-processing*, *Set role (Split data = data training : data testing)*, pemodelan algoritma *decision tree series C4.5* terhadap data training, *apply model* terhadap data testing, dan adanya *performance* dilakukan sebagai pengujian untuk menghitung keakuratan model terhadap data *testing*. Kemudian dari proses ini dapat diperoleh apa yang menyebabkan penyakit jantung tersebut terjadi dengan mengetahui atribut mana yang dijadikan atribut utama (*root node*) yang digunakan sebagai proses pengambilan keputusan atau proses klasifikasi

pada penyakit jantung yang memiliki nilai *gain ratio* tertinggi, yakni atribut nominal “ST_Lereng”, dimana diperoleh rule akhir adalah 51 dengan jumlah label sebanyak 24 rule (Normal) dan 27 rule (Penyakit jantung). Kemudian setelah menemukan *model tree*, maka diperoleh hasil pengujian tingkat *Accuracy* dari *confusion matrix* sebesar sebesar 80,43% dan didapatkan *error classification*, sebesar 19,57%. Serta diperoleh hasil *Visualisasi AUC (Area Under Curve)* dari *kurva ROC* diperoleh nilai *AUC:0,798 (Positive class)*: Penyakit Jantung. Adapun tahapan perhitungan secara manual yang dilalui dari algoritma *decision tree series C4.5*, seperti mencari nilai *entropy*, nilai *gain*, nilai *split info*, dan terakhir mencari nilai tertinggi dari nilai *gain ratio*. Disamping kelebihan yang telah dibahas sebelumnya, adapun kekurangan algoritma *decision tree*, yaitu bersifat tidak stabil, ini menjadi salah satu keterbatasan dari algoritma *decision tree* ketika terdapat perubahan kecil pada data dapat menghasilkan perubahan besar dalam struktur pohon keputusan dan kurang efektif dalam memprediksi hasil dari variabel kontinu [24]. Maka untuk itu, penelitian selanjutnya dapat dilakukan pengembangan terhadap metode klasifikasi dengan algoritma yang lain sebagaimana diharapkan nantinya akan memberikan akurasi yang lebih bagus.

UCAPAN TERIMAKASIH

Penulis mengucapkan terima kasih kepada semua mitra yang membantu atas kelancaran pelaksanaan penelitian dan penulisan artikel “Klasifikasi Penyakit Jantung Menggunakan Algoritma Decision Tree Series C4.5 Dengan Rapidminer”. Semoga artikel ini dapat menambah wawasan bagi setiap pembacanya.

DAFTAR PUSTAKA

- [1] A. Vadreas, “Sistem Pendukung Keputusan Pemberian Reward Penyiar Radio Dengan Metode Multifactor Evaluation Process,” *J. Teknol. Dan Sist. Inf. Bisnis*, vol. 3, no. 2 SE-Articles, Jul. 2021.
- [2] H. Kurnia, V. Ariandi, and A. I. Jamhur, “Decicion Suport System Pemilihan Pestisida Terbaik Pada UD. Pupuk Jasa Tani Menggunakan Metode Weighted Product,” *J. Teknol. Dan Sist. Inf. Bisnis*, vol. 5, no. 1 SE-Articles, Jan. 2023.
- [3] S. Haryati, A. Sudarsono, and E. Suryana, “Implementasi Data Mining Untuk Memprediksi Masa Studi Mahasiswa Menggunakan Algoritma C4.5 (Studi Kasus: Universitas Dehasen Bengkulu),” *J. Media Infotama*, vol. 11, no. 2, pp. 130–138, 2015.
- [4] A. Rohman, “Prediksi Penyakit Jantung Menggunakan Algoritma C4.5 Berbasis Adaboost,” vol. Cetak 1, p. 92, 2021.
- [5] A. Andoyo, E. Y. Angraeni, and A. Khumaidi, *Sistem Pendukung Keputusan Konsep, Implementasi & Pengembangan - Google Books*. Penerbit Adab, 2021.
- [6] D. Dona, H. Maradona, and M. Masdewi, “Sistem Pakar Diagnosa Penyakit Jantung Dengan Metode Case Based Reasoning (Cbr),” *Zo. J. Sist. Inf.*, vol. 3, no. 1, pp. 1–12,

- 2021.
- [7] A. Nurmasani and Y. Pristyanto, “Algoritme Stacking Untuk Klasifikasi Penyakit Jantung Pada Dataset Imbalanced Class,” *Pseudocode*, vol. 8, no. 1, pp. 21–26, 2021.
- [8] F. R. Makarim, “7 Penyakit yang Dapat Ditimbulkan karena Infeksi Streptococcus,” 2022. [Online]. Available: <https://www.halodoc.com/kesehatan/infeksi-streptococcus>.
- [9] P. Kabo, *Mengungkap Pengob. Peny. Jantung Kmr*. Jakarta: Gramedia Pustaka Utama, 2008.
- [10] D. P. Utomo and M. Mesran, “Analisis Komparasi Metode Klasifikasi Data Mining dan Reduksi Atribut Pada Data Set Penyakit Jantung,” *J. Media Inform. Budidarma*, vol. 4, no. 2, p. 437, 2020.
- [11] L. P. C. Dewi, “Jenis, Gejala, dan Penyebab Penyakit Jantung,” 2021. [Online]. Available: <https://rs-soewandhi.surabaya.go.id/jenis-gejala-dan-penyebab-penyakit-jantung/>.
- [12] W. Aulia, “Sistem Pakar Diagnosa Penyakit Jantung Koroner Dengan Metode Probabilistic Fuzzy Decision Tree,” *J. Sains dan Inform. Res. Sci. Inform.*, vol. 4, no. 2, pp. 106–117, 2018.
- [13] D. Saripurna, A. Calam, Y. Yusnidah, and Z. Lubis, “Sistem Cerdas Pemanggang Jagung Semi Otomatis Berbasis Mikrokontroler Menggunakan Metode PWM (Pulse Width Modulation),” *J. SAINTIKOM (Jurnal Sains Manaj. Inform. dan Komputer)*, vol. 18, no. 1, p. 82, 2019.
- [14] M. Soysal and E. G. Schmidt, “Machine learning algorithms for accurate flow-based network traffic classification: Evaluation and comparison,” *Perform. Eval.*, vol. 67, no. 6, pp. 451–467, 2010.
- [15] D. Jollyta, W. Ramdhan, and M. Zarlis, “Konsep Data Mining Dan Penerapan Jollyta, D, W Ramdhan, and M Zarlis. Konsep Data Mining Dan Penerapan. Deepublish, 2020.
- [16] N. Rahmalia, “Kaggle, Tempat Belajar dan Berkompetisi di Bidang Data Science,” 2021. [Online]. Available: https://glints.com/id/lowongan/kaggle-adalah/#.Y_yEFnZBzIU.
- [17] A. Fahrezi, “Oprec Ristek 2018 - Data Science | Kaggle,” 2018. [Online]. Available: <https://www.kaggle.com/competitions/oprecristekds>.
- [18] J. Windy, Y. Rika, and K. S. Mochammad, “Algoritma C4. 5 Untuk Penilaian Kinerja Karyawan,” *Scan J. Teknol. Inf. dan Komun.*, vol. 9, no. 2, pp. 33–39, 2014.
- [19] S. Febriani and H. Sulistiani, “Analisis Data Hasil Diagnoga Untuk Klasifikasi Gangguan Kepribadian Menggunakan Algoritma C4. 5,” *J. Teknol. Dan Sist. Inf.*, vol. 2, no. 4, pp. 89–95, 2021.
- [20] I. Junaedi, N. Nuswantari, and V. Yasin, “Perancangan Dan Implementasi Algoritma C4. 5 Untuk Data Mining Analisis Tingkat Risiko Kematian Neonatum Pada Bayi,” *JISICOM (Journal Inf. Syst. Informatics Comput.*, vol. 3, no. 1, pp. 29–44, 2019.
- [21] B. Novianti, T. Rismawan, and S. Bahri, “Implementasi Data Mining Dengan Algoritma C4. 5 Untuk Penjurusan Siswa (Studi Kasus: Sma Negeri 1 Pontianak),” *Coding J. Komput. dan Apl.*, vol. 4, no. 3, 2016.
- [22] R. Nofitri and N. Irawati, “Analisis Data Hasil Keuntungan Menggunakan Software Rapidminer,” *JURTEKSI*, vol. 5, no. 2, pp. 199–204, 2019.
- [23] D. Jollyta, M. Siddik, H. Mawengkeng, and S. Efendi, *Teknik Evaluasi Cluster Solusi Menggunakan Python Dan Rapidminer - Google Books*. Deepublish, 2021.
- [24] R. Tineges, “Pahami Metode Decision Tree Sebagai Algoritma Data Science,” 2022. [Online]. Available: <https://dqlab.id/pahami-metode-decision-tree-sebagai-algoritma-data-science>.