

Komparasi Dalam Prediksi Gagal Jantung Dengan Menggunakan Metode C4.5 dan Naïve Bayes

Julia Triani¹, Yovi Pratama², Elvi Yanti³

¹Fakultas Ilmu Komputer, Teknik Informatika, Universitas Dinamika Bangsa, Kota Jambi, Indonesia

Email : juliatriani2707@gmail.com, yovi.pratama@gmail.com, elvovote92@gmail.com

Email Penulis Korespondensi : elvovote92@gmail.com

Abstrak– Industri kesehatan memiliki data kesehatan yang cukup besar, seperti dataset penyakit gagal jantung. Pada penelitian ini penulis memutuskan untuk memprediksi penyakit gagal jantung menggunakan metode C4.5 dan Naïve Bayes. Data yang dipakai diambil dari website kaggle.com yang berjumlah 918 data dengan 12 atribut. Adapun tujuan dari penelitian ini adalah untuk mengetahui perbandingan antara metode C4.5 dan Naïve Bayes dalam mengukur tingkat akurasi. Manfaat dari penelitian ini adalah dapat membantu pihak tenaga kesehatan dalam memprediksi pasien yang berkemungkinan terkena penyakit gagal jantung sehingga dapat menjadi informasi bagi pembaca untuk mengetahui resiko pasien terkena penyakit gagal jantung dengan tingkat akurasi yang tinggi. Proses implementasi metode C4.5 dan yang digunakan meraih tingkat akurasi sebesar 83.67%, presisi sebesar 85.01%, recall sebesar 86.04% dan f1-score sebesar 85.02%. untuk metode C4.5 dengan outlier. Kemudian metode C4.5 tanpa outlier meraih tingkat akurasi sebesar 85.02%, presisi sebesar 86.02%, recall sebesar 87.02% dan f1-score sebesar 86.52%. Dan metode Naïve Bayes yang digunakan meraih tingkat akurasi sebesar 85.30%, presisi sebesar 86.31%, recall sebesar 87.40% dan f1-score sebesar 86.85% untuk metode Naïve Bayes dengan outlier. Kemudian metode Naïve Bayes tanpa outlier meraih tingkat akurasi tertinggi yaitu sebesar 85.57%, presisi sebesar 86.54%, recall sebesar 87.57% dan f1-score sebesar 87.05%. Sehingga dapat disimpulkan bahwa nilai perhitungan akurasi, presisi, recall dan f1-score tertinggi yaitu metode Naïve Bayes tanpa outlier.

Kata Kunci : Prediksi, Gagal Jantung, Metode, C4.5, Naïve Bayes

Abstract– The health industry has quite large health data, such as the heart failure dataset. In this study the authors decided to predict heart failure using the C4.5 and Naïve Bayes methods. The data used was taken from the kaggle.com website which totaled 918 data with 12 attributes. The purpose of this study is to find out the comparison between the C4.5 and Naïve Bayes methods in measuring the level of accuracy. The benefit of this research is that it can help health workers in predicting patients who are likely to get heart failure so that it can be information for readers to know the risk of patients getting heart failure with a high degree of accuracy. The implementation process of the C4.5 method and those used achieved an accuracy rate of 83.67%, a precision of 85.01%, a recall of 86.04% and an f1-score of 85.02% for the C4.5 method with outliers. Then the C4.5 method without outliers achieves an accuracy rate of 85.02%, a precision of 86.02%, a recall of 87.02% and an f1-score of 86.52%. And the Naïve Bayes method used achieves an accuracy rate of 85.30%, a precision of 86.31%, a recall of 87.40% and an f1-score of 86.85% for the Naïve Bayes method with outliers. Then the Naïve Bayes method without outliers achieves the highest accuracy rate of 85.57%, precision of 86.54%, recall of 87.57% and f1-score of 87.05%. So it can be concluded that the highest calculation value for accuracy, precision, recall and f1-score is the Naïve Bayes method without outliers.

Keywords : Prediction, Heart Failure, Method, C4.5, Naïve Bayes

1. PENDAHULUAN

1.1 Latar Belakang Penelitian

Salah satu penyakit yang serius dan tidak menular adalah penyakit jantung, penyakit jantung merupakan penyakit dimana kondisi pembuluh darah utama yang mentransfer darah ke jantung mengalami kerusakan dan tidak dapat bekerja dengan baik [1]. Federasi Jantung Dunia memperkirakan angka kematian akibat penyakit gagal jantung di Asia Tenggara mencapai 1,8 juta kasus pada tahun 2014. Di Indonesia sendiri pada tahun 2013 tercatat ada setidaknya 883.447 orang yang terdiagnosis penyakit gagal jantung dengan mayoritas penderita berusia 55-64 tahun. Angka kematian akibat penyakit jantung pun menjadi cukup tinggi, yakni sekitar 45% dari seluruh angka kematian di Indonesia [2].

Salah satu bagian penting dari pengobatan atau tindakan medis adalah pengambilan keputusan dan proses klasifikasi atau prediksi pada suatu hal yang menjadi fokus seperti pendeteksi penyakit, namun klasifikasi medis biasanya merupakan proses yang sangat kompleks dan sulit dilakukan jika tidak dapat mengetahui metode yang tepat dan terbaik dalam memberikan solusinya [3]. Masalah yang dihadapi oleh pihak tenaga kesehatan adalah mendiagnosa pasien dengan benar, prediksi yang buruk dapat menyebabkan konsekuensi yang mendatangkan permasalahan yang kemudian tidak dapat diterima. Beberapa penelitian telah dilakukan dalam bidang kesehatan untuk mendapatkan prediksi penyakit dengan lebih akurat, namun belum diketahui metode apa yang paling akurat dalam memprediksi penyakit pasien. Oleh karena itu, saat ini merupakan waktu yang tepat untuk

mengembangkan sistem pengetahuan berbasis komputer yang modern, efektif dan efisien dalam mendiagnosa masalah penyakit gagal jantung [4].

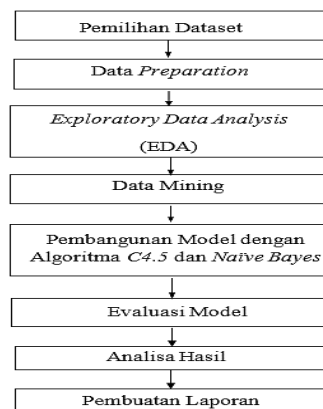
Industri kesehatan memiliki data kesehatan yang cukup besar, namun sebagian besar data tersebut tidak diolah agar mengetahui informasi tersembunyi untuk dijadikan pengambilan keputusan yang efektif oleh para praktisi kesehatan. Pengambilan keputusan data dan informasi yang akurat akan menghasilkan prediksi penyakit menjadi tepat sasaran. Penyakit jantung di Indonesia merupakan penyakit nomor satu dengan angka kematian yang cukup tinggi [5]. Pada tahun 2016 data dari *World Health Organization* (WHO) tercatat sebanyak 17,5 juta orang di dunia meninggal akibat penyakit jantung. Berdasarkan data Riset Kesehatan Dasar (Riskesdas) di Indonesia pada tahun 2018 jumlah keseluruhan penyakit jantung sebesar 1,5% atau sekitar 1.017.290 penduduk. Penyebab utama gagal jantung adalah penyakit kardiovaskuler seperti hipertensi, kardiomiopati dan penyakit jantung koroner, yang dapat menyebabkan disfungsi ventrikel kiri berupa penurunan kemampuan kontraksi, relaksasi atau keduanya yang berdampak pada penurunan curah jantung [6].

Gagal jantung merupakan masalah kesehatan yang progresif dengan angka kematian dan penyakit dalam suatu populasi yang tinggi di negara maju maupun negara berkembang termasuk Indonesia. Di Indonesia, usia pasien gagal jantung relatif lebih muda dibanding Eropa dan Amerika disertai dengan tampilan klinis yang lebih berat. Prevalensi dari gagal jantung sendiri semakin meningkat karena pasien yang mengalami kerusakan jantung yang bersifat akut dapat berlanjut menjadi gagal jantung kronik. *World Health Organization* (WHO) menggambarkan bahwa meningkatnya jumlah penyakit gagal jantung di dunia, termasuk Asia diakibatkan oleh meningkatnya angka perokok, tingkat obesitas, *dyslipidemia*, dan diabetes [7]. Oleh karena itu, penelitian ini dilakukan untuk membantu menyelesaikan permasalahan tersebut dengan data mining klasifikasi untuk prediksi penyakit jantung, diperlukan suatu metode atau teknik yang dapat mengolah data yang sudah ada. Penggunaan data mining Algoritma C4.5 dan Naive Bayes dalam Klasifikasi Penderita Gagal Jantung sebagai pilihan untuk prediksi penyakit gagal jantung dapat menjadi alternatif pilihan yang tepat, tetapi sampai saat ini belum diketahui algoritma klasifikasi yang paling akurat. Dalam melakukan penelitian ini penulis memilih algoritma C4.5 dan *Naive Bayes* sebagai perbandingan tingkat akurasi, Algoritma C4.5 membentuk pohon keputusan yang merupakan metode klasifikasi dan prediksi yang terkenal. Algoritma *Naive Bayes* yang merupakan prediksi berbasis probabilistik yg dikemukakan oleh ilmuwan Inggris Thomas Bayes. Maka penulis tertarik untuk mengangkat penelitian dengan judul “KOMPARASI DALAM PREDIKSI GAGAL JANTUNG DENGAN MENGGUNAKAN METODE C4.5 DAN NAÏVE BAYES” Untuk mengetahui tingkat akurasi dari masing-masing metode.

2. METODOLOGI PENELITIAN

2.1 Kerangka Kerja Penelitian

Kerangka kerja penelitian merupakan proses atau tahapan yang akan dilakukan selama mengerjakan penelitian dengan tujuan agar dapat menyelesaikan penelitian tepat waktu sesuai dengan yang diharapkan. Adapun kerangka kerja penelitian yang digunakan dapat dilihat pada gambar 2.1 berikut:



Gambar 2.1 Kerangka Kerja Penelitian

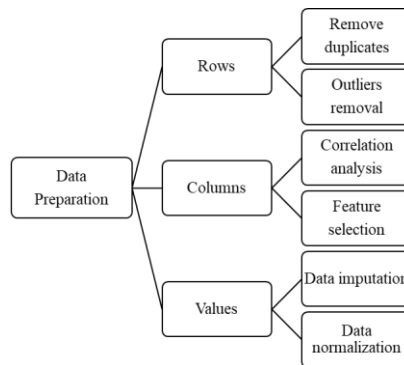
Berikut uraian kerangka kerja penelitian:

1. Pemilihan Dataset

Dataset yang digunakan yaitu *Heart Failure Prediction Dataset* yang berasal dari *website* Kaggle.com (<https://www.kaggle.com/datasets/fedoriano/heart-failure-prediction>) yang mempunyai 918 data dan 11 atribut. Atribut yang digunakan yaitu : *Age, Gender, Chest Pain Type, Resting BP, Cholesterol, Fasting BS,, Resting ECG, Max HR, Exercise Angina, Old peak, ST_Slope* dan 1 kelas keluaran yaitu *Heart Disease*.

2. Data Preparation

Pada tahap ini data *preparation* perlu dibuat kerangka kerja dengan tujuan agar proses persiapan data dapat dilakukan secara sistematis. Beberapa kelompok metode persiapan data yang akan dilakukan dapat dilihat pada gambar 2.2 berikut:



Gambar 2.2 Data Preparation

3. Exploratory Data Analysis (EDA)

Pada tahap ini penulis melakukan *Exploratory Data Analysis* menggunakan *tools Rapid Miner*. *Exploratory Data Analysis* dapat diterapkan untuk menganalisis data di segala bidang. Penulis menggunakan EDA untuk mengidentifikasi *Heart Disease* (penyakit jantung) pada dataset.

4. Data Mining

Pada tahap ini data akan diseleksi untuk menentukan variabel apa saja yang akan diambil agar tidak terjadi kesamaan dan perulangan yang tidak diperlukan dalam pengolahan teknik data mining. Ada beberapa pengelompokan dalam memprediksi yaitu *accuracy, precision, recall, dan F1-score* yang akan dilakukan perhitungan menggunakan *tools Rapid Miner*.

5. Evaluasi Model

Pada tahap ini penulis telah menggunakan Algoritma *C4.5* dan *Naïve Bayes* dengan tujuan untuk memprediksi tingkat akurasi penyakit gagal jantung pada salah satu dataset yang ada di kaggle.com dan penulis menggunakan metrik dari salah satu model evaluasi yaitu *Confusion matrix* tabel yang sering digunakan untuk menggambarkan kinerja model klasifikasi pada sekumpulan data uji yang nilai sebenarnya diketahui.

6. Analisis Hasil

Pada tahap ini akan dijelaskan hasil analisis terhadap dataset yang telah dipilih dari kaggle.com. Analisis akan dilakukan dengan menggunakan alat bantu atau *tools Rapid Miner* versi 9.10. Sesuai dengan kerangka kerja penelitian bahwa ada beberapa kelompok analisis yang akan dilakukan, yaitu *data preparation, Exploratory Data Analysis (EDA), Pembangunan model algoritma C4.5 dan Naïve Bayes* dan yang terakhir evaluasi model.

2.2 Data Mining

Data mining merupakan bidang ilmu yang digunakan untuk menangani masalah pengembalian informasi dari *database* yang besar dengan menggabungkan teknik dari statistik, pembelajaran mesin, visualisasi data, pengenalan pola, dan *database*. Tujuan dari data mining adalah untuk mengekstrak informasi dengan metode cerdas dari kumpulan data kemudian mengubah informasi menjadi struktur yang dapat dipahami untuk penggunaan lebih lanjut [8]. Didalam data mining Terdapat metode Klasifikasi yang fungsinya merupakan teknik pengolahan data yang membagi objek menjadi beberapa kelas sesuai dengan jumlah kelas yang diinginkan. Klasifikasi merupakan suatu teknik menemukan suatu pola yang mampu memisahkan kelas data yang satu dengan yang lainnya untuk menentukan objek yang masuk dengan kategori tertentu dengan melihat kelakuan dan atribut dari kelompok yang telah didefinisikan. Teknik ini mampu mengklarifikasi data baru dengan menggunakan hasilnya untuk memberikan sejumlah aturan [9].

2.3 Algoritma C4.5

Algoritma *C4.5* merupakan algoritma klasifikasi pohon keputusan yang digunakan karena memiliki keunggulan utama. Kelebihan dari algoritma *C4.5* dapat menghasilkan pohon keputusan yang mudah diinterpretasikan, memiliki tingkat akurasi yang dapat diterima dan efisien dalam menangani dua atribut yang bertipe diskrit dan numerik. Dalam membangun pohon, algoritma *C4.5* membaca semua sampel data pelatihan dari memori dan memuat nya ke dalam memori secara keseluruhan pada waktu yang bersamaan [10].

2.4 Naïve Bayes

Naïve Bayes adalah suatu metode yang digunakan memperkirakan atau memprediksi suatu *class* dari suatu objek yang kelas nya tidak diketahui dari masing-masing kelompok atribut yang ada, dan menentukan *class* mana yang paling optimal berdasarkan pengaruh yang didapat dari hasil pengamatan. Klasifikasi *Naive Bayes* adalah klasifikasi statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu *class* [11].

3. HASIL DAN PEMBAHASAN

1. Correlation Analysis

Digunakan untuk menganalisis hubungan antara dua variabel atau lebih, penulis melakukan teknik ini menggunakan *tools rapid miner*, penulis juga melakukan dengan *outliers* dan tidak dengan *outliers* yang dapat dilihat pada gambar 3.1 dengan *outliers* dan 3.2 tanpa *outliers*.

Attribut...	Age	Gender	Cholesterol	RestingBP	Cholesterol	FastingBS	MaxHR	Oldpeak	ST_Slope	HeartDisease
Age	1	-0.056	?	0.234	-0.095	0.198	?	-0.382	0.216	?
Gender	-0.056	1	?	-0.005	0.200	-0.120	?	0.189	-0.191	-0.106
Cholesterol	?	?	1	?	?	?	?	?	?	?
RestingBP	0.234	-0.005	?	1	0.101	0.070	?	-0.112	0.155	0.165
Cholesterol	-0.095	0.200	?	0.101	1	-0.251	?	0.236	-0.034	0.050
FastingBS	0.198	-0.120	?	0.070	-0.251	1	?	-0.131	0.060	0.053
MaxHR	?	?	?	?	?	?	1	?	?	?
Oldpeak	-0.382	0.189	?	-0.112	0.236	-0.131	?	1	-0.370	-0.161
ST_Slope	0.216	-0.191	?	0.155	-0.034	0.050	?	-0.370	1	0.409
HeartDisease	0.259	-0.106	?	0.165	0.050	0.053	?	-0.161	0.409	1
ST_Slope	?	?	?	?	?	?	?	?	?	?
HeartDisease	0.262	-0.305	?	0.168	-0.233	0.267	?	-0.403	0.404	0.404

Gambar 3.1 Dengan Outliers

Attribut...	Age	RestingBP	Cholesterol	FastingBS	MaxHR	Oldpeak	Gender	Cholesterol	RestingBP	Exerc...	ST_Slope	HeartD...
Age	1	0.208	-0.099	0.201	-0.389	0.203	-0.962	?	?	0.222	?	0.262
RestingBP	0.208	1	-0.187	0.053	-0.113	0.174	-0.905	?	?	0.154	?	0.107
Cholesterol	-0.099	0.187	1	-0.280	0.244	0.061	0.234	?	?	-0.837	?	-0.231
FastingBS	0.201	0.053	-0.280	1	-0.120	0.095	-0.120	?	?	0.098	?	0.270
MaxHR	-0.389	-0.113	0.244	-0.120	1	-0.167	0.189	?	?	-0.378	?	-0.409
Oldpeak	0.203	0.174	0.061	0.095	-0.167	1	-0.127	?	?	0.423	?	0.407
Gender	-0.962	-0.905	0.234	-0.120	0.189	-0.127	1	?	?	-0.187	?	-0.211
Cholesterol	?	?	?	?	?	?	?	1	?	?	?	?
RestingBP	?	?	?	?	?	?	?	?	1	?	?	?
Exerc...	0.222	0.154	-0.837	0.098	-0.378	0.423	-0.187	?	?	1	?	0.499
ST_Slope	?	?	?	?	?	?	?	?	?	?	1	?
HeartD...	0.262	0.107	-0.231	0.270	-0.409	0.407	-0.211	?	?	0.499	?	1

Gambar 3.2 Tanpa Outliers

2. Data Normalization

Merupakan penskalaan nilai agar rentang datanya sama, penulis melakukan teknik ini menggunakan *tools rapid miner*. Hanya ada beberapa atribut yang dilakukan *Normalization* diantaranya atribut *Age*, *Resting BP*, *Cholesterol*, *Max HR* dan *Old Peak*, karena untuk mentransformasi sebuah atribut numerik diskalakan dalam range yang lebih kecil. Penulis melakukan data *normalization* dengan nilai -1 sampai 1 karena pada tabel dataset perbedaan rentang nilai dari variabel *Age*, *Resting BP*, *Cholesterol*, *Max HR* dan *Old Peak* terlihat sangat jauh dibandingkan dengan nilai variabel *Fasting BS* yaitu 0 dan 1.

Row No.	Age	RestingBP	Cholesterol	MaxHR	Oldpeak	Gender	Cholesterol	FastingBS	RestingBP	Exerc...	Oldpeak	ST_Slope	HeartDisease
1	-0.510	0.400	-0.513	0.577	-0.400	0	ATX	0	Normalized	N	0	0	N
2	-0.141	0.600	-0.461	0.760	-0.300	1	UAP	0	Normalized	N	0	0	N
3	-0.510	0.300	-0.584	-0.464	-0.400	0	ATX	0	ST	N	0	0	N
4	-0.184	0.300	-0.580	-0.504	-0.500	1	ATV	0	Normalized	Y	0	0	N
5	0.664	0.600	-0.583	-0.497	-0.400	0	UAP	0	Normalized	N	0	0	N
6	-0.691	0.300	0.071	0.760	-0.400	0	UAP	0	Normalized	N	0	0	N
7	0.300	0.300	-0.514	0.560	-0.400	1	ATX	0	Normalized	N	0	0	N
8	0.664	0.300	-0.510	0.955	-0.400	0	ATX	0	Normalized	N	0	0	N
9	-0.633	0.400	-0.513	-0.014	-0.500	0	ATV	0	Normalized	Y	0	0	N
10	-0.184	0.300	-0.583	-0.595	-0.400	1	ATX	0	Normalized	N	0	0	N
11	-0.633	0.300	-0.583	0.955	-0.400	1	UAP	0	Normalized	N	0	0	N
12	0.234	0.300	-0.460	-0.461	0.547	0	ATX	0	ST	Y	0	0	N
13	-0.651	0.300	-0.523	0.947	-0.400	0	ATX	0	Normalized	N	0	0	N
14	-0.141	0.400	-0.524	0.527	-0.400	0	ATV	0	Normalized	Y	0	0	N

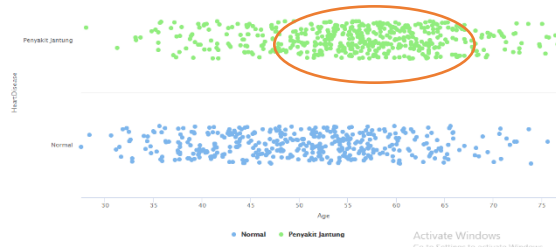
Gambar 3.3 Data Normalization

Row No.	Age	Gender	Cholesterol	RestingBP	Cholesterol	FastingBS	RestingBP	MaxHR	Oldpeak	ST_Slope	HeartDisease	
1	40	0	ATX	140	100	0	Normalized	172	N	0	0	N
2	40	1	UAP	140	100	0	Normalized	168	N	0	0	N
3	37	0	ATX	130	100	0	ST	98	N	0	0	N
4	48	1	ATV	130	214	0	Normalized	160	Y	1.000	0	N
5	54	0	UAP	150	190	0	Normalized	132	N	0	0	N
6	39	0	UAP	120	130	0	Normalized	170	N	0	0	N
7	40	1	ATX	130	237	0	Normalized	170	N	0	0	N
8	54	0	ATX	110	200	0	Normalized	140	N	0	0	N
9	37	0	ATV	140	207	0	Normalized	130	Y	1.000	0	N
10	40	1	ATX	130	201	0	Normalized	170	N	0	0	N
11	37	1	UAP	130	211	0	Normalized	140	N	0	0	N
12	38	0	ATX	130	164	0	ST	98	Y	0	0	N
13	39	0	ATX	120	204	0	Normalized	140	N	0	0	N
14	40	0	ATV	140	224	0	Normalized	140	Y	0	0	N

Gambar 3.4 Data Tanpa Normalization

3. Exploratory Data Analysis (EDA)

Penulis menggunakan teknik *Multivariate Graphical* karena *Exploratory Data Analysis* ini menggunakan grafik untuk menunjukkan hubungan antara dua dataset atau lebih. Penulis melakukan pengujian menggunakan *tools rapid miner*. Untuk lebih jelasnya dapat dilihat pada gambar dibawah ini:



Gambar 3.5 Atribut Age

Kesimpulan yang penulis dapatkan dari gambar 3.5 maka usia dibawah 30 tahun tergolong normal, akan tetapi rata-rata di semua usai bisa tergolong normal dan tergolong penyakit jantung, seperti pasien yang berusia 76 tahun ada yang tergolong normal dan ada juga yang tergolong penyakit gagal jantung. Dari hasil pengamatan usia yang rentan terkena penyakit gagal jantung mulai dari usai 50 tahun sampai dengan 65 tahun.



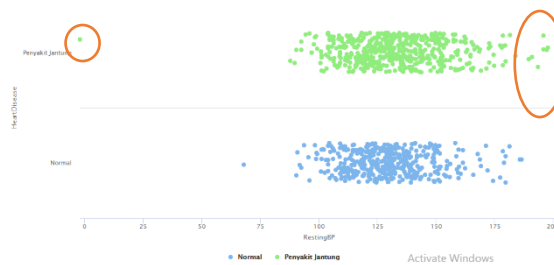
Gambar 3.6 Atribut Gender

Kesimpulan yang penulis dapatkan dari gambar 3.6 maka kedua *gender* ada yang tergolong normal dan ada juga yang tergolong penyakit gagal jantung. Dari hasil pengamatan *gender male* lebih rentan terkena penyakit gagal jantung dan *gender female* lebih tergolong normal.



Gambar 3.7 Atribut Chest Pain Type

Kesimpulan yang penulis dapatkan dari gambar 3.7 maka keempat *type* ini ada yang tergolong normal dan ada juga yang tergolong penyakit gagal jantung. Dari hasil pengamatan *type ASY* (tipe nyeri dada yang sudah positif namun tanpa gejala) lebih rentan terkena penyakit gagal jantung dibanding dengan *type* lainnya seperti *type ATA* (tipe nyeri dada akibat gangguan panik) tidak berhubungan dengan jantung sudah pasti lebih tergolong normal, namun ada juga yang tergolong penyakit gagal jantung.



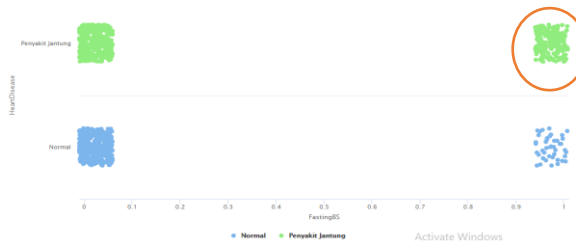
Gambar 3.8 Atribut Resting BP

Kesimpulan yang penulis dapatkan dari gambar 3.8 maka terdapat satu data tergolong penyakit gagal jantung jika *Resting BP* nya 0 mm Hg (merupakan data pencilan). Dari hasil pengamatan jika *Resting BP* nya 100 mm Hg sampai dengan 170 mm Hg maka hasil nya kemungkinan sama (ada yang tergolong normal dan ada yang tergolong penyakit gagal jantung).



Gambar 3.9 Atribut Cholesterol

Kesimpulan yang penulis dapatkan dari gambar 3.9 maka akan rentan terkena penyakit gagal jantung jika *cholesterol* nya 0 mm/dl sampai dengan 50 mm/dl. Dari hasil pengamatan jika *cholesterol* nya 150 mm/dl sampai dengan 300 mm/dl maka hasilnya kemungkinan sama (ada yang tergolong normal dan ada yang tergolong penyakit gagal jantung).



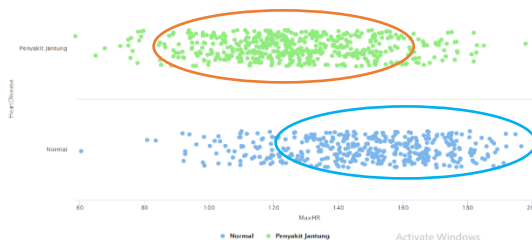
Gambar 3.10 Atribut Fasting BS

Kesimpulan yang penulis dapatkan dari gambar 3.10 dari hasil pengamatan jika nilai mendekati 0 atau puasa BS < 120 mg/dl maka hasilnya akan sama (ada yang tergolong normal dan ada yang tergolong penyakit gagal jantung). Namun jika nilai mendekati 1 atau puasa BS > 120 mg/dl maka akan lebih tergolong penyakit gagal jantung.



Gambar 3.11 Atribut Resting ECG

Kesimpulan yang penulis dapatkan dari gambar 3.11 maka ketiga tipe ini ada yang tergolong normal dan ada juga yang tergolong penyakit gagal jantung. Dari hasil pengamatan tipe normal hasil nya sama dalam artian ada yang tergolong normal dan ada yang tergolong penyakit gagal jantung. Sedangkan tipe ST (Gelombang depresi) dan LVH (pembesaran bilik kiri jantung) lebih rentan terkena penyakit gagal jantung namun tidak sedikit yang tergolong normal.



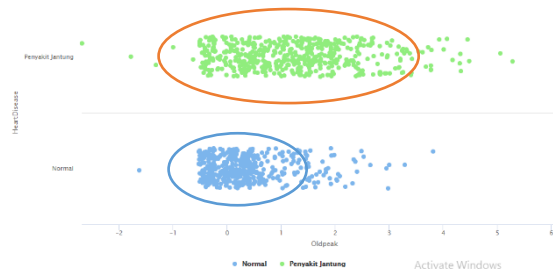
Gambar 3.12 Atribut Max HR

Kesimpulan yang penulis dapatkan dari gambar 3.12 maka akan rentan terkena penyakit gagal jantung jika $max HR$ nya 100 bpm sampai 140 bpm dan jika semakin tinggi $max HR$ nya kemungkinan akan tergolong normal dapat dilihat pada lingkaran berwarna biru $max HR$ 140 bpm sampai 190 bpm tergolong normal. Dapat disimpulkan juga bahwa ada yang tergolong normal dan ada juga yang tergolong penyakit gagal jantung.



Gambar 3.13 Atribut Exercise Angina

Kesimpulan yang penulis dapatkan dari gambar 3.13 Dari hasil pengamatan jika pasien mengalami nyeri dada akibat olahraga (Y) maka akan tergolong penyakit gagal jantung, namun tidak sedikit yang tergolong normal. Dan jika pasien tidak mengalami nyeri dada akibat olahraga (N) maka akan tergolong normal, namun tidak banyak yang tergolong penyakit gagal jantung.



Gambar 3.14 Atribut Old Peak

Kesimpulan yang penulis dapatkan dari gambar 3.14 dari hasil pengamatan atribut *old peak* lebih banyak pasien tergolong penyakit gagal jantung dan bayak pasien tergolong normal jika *old peak* nya mendekati 0 – 1 seperti gambar pada lingkaran biru. Pada lingkaran merah pasien banyak tergolong penyakit gagal jantung jika *old peak* nya mendekati 0 – 4.



Gambar 3.15 Atribut ST Slope

Kesimpulan yang penulis dapatkan dari gambar 3.15 maka ketiga tipe ini ada yang tergolong normal dan ada juga yang tergolong penyakit gagal jantung. Dari hasil pengamatan tipe *flat* (kemiringan detak jantung datar) dan *down* (kemiringan detak jantung miring ke bawah) lebih rentan terkena penyakit gagal jantung namun tidak sedikit yang tergolong normal.

4. Evaluasi Model

Pada tahap ini penulis melakukan perbandingan antara:

- a. Algoritma C4.5 dan Algoritma Naïve Bayes dengan data outliers.
- b. Algoritma C4.5 dan Algoritma Naïve Bayes tanpa outliers.

Accuracy didefinisikan sebagai tingkat kedekatan antara nilai prediksi dengan nilai aktual, precision adalah tingkat ketepatan antara label oleh pengguna dengan hasil yang didapatkan oleh sistem, recall adalah tingkat keberhasilan sistem dalam menemukan kembali sebuah informasi dan f1-score merupakan perbandingan rata-rata precision dan recall yang dibobotkan.

- 1. Algoritma C4.5 dengan Outliers

Tabel 3.1 Algoritma C4.5 dengan Outliers

	True Normal	True Penyakit Jantung	Class Precision
Prediksi Normal	331	71	82.34%
Prediksi Penyakit Jantung	79	437	84.69%
Class Recall	80.73%	86.02%	

Pada tabel 3.1 *true positive* (diagnosa penyakit jantung, kenyataan penyakit jantung) berjumlah 437 data, *true negatif* (diagnosa normal, kenyataan normal) berjumlah 331 data, *false positive* (diagnosa penyakit jantung, kenyataan normal) berjumlah 79 data, dan *false negatif* (diagnosa normal, kenyataan penyakit jantung) berjumlah 71 data. Sehingga didapat nilai *accuracy* 83.67%, nilai *precision* 85.01%, nilai *recall* 86.04% dan nilai *f1-score* 85.02%.

2. Algoritma Naïve Bayes dengan Outliers

Tabel 3.2 Algoritma Naïve Bayes dengan Outliers

	True Normal	True Penyakit Jantung	Class Precision
Prediksi Normal	339	64	84.12%
Prediksi Penyakit Jantung	71	444	86.21%
Class Recall	82.68%	87.40%	

Pada tabel 3.2 *true positive* (diagnosa penyakit jantung, kenyataan penyakit jantung) berjumlah 444 data, *true negatif* (diagnosa normal, kenyataan normal) berjumlah 339 data, *false positive* (diagnosa penyakit jantung, kenyataan normal) berjumlah 71 data, dan *false negatif* (diagnosa normal, kenyataan penyakit jantung) berjumlah 64 data. Sehingga didapat nilai *accuracy* 85.30%, nilai *precision* 86.31%, nilai *recall* 87.40% dan nilai *f1-score* 86.85%.

3. Algoritma C4.5 tanpa Outliers

Tabel 3.3 Algoritma C4.5 Tanpa Outliers

	True Normal	True Penyakit Jantung	Class Precision
Prediksi Normal	337	65	83.83%
Prediksi Penyakit Jantung	71	435	85.97%
Class Recall	82.60%	87.00%	

Pada tabel 3.3 *true positive* (diagnosa penyakit jantung, kenyataan penyakit jantung) berjumlah 435 data, *true negatif* (diagnosa normal, kenyataan normal) berjumlah 337 data, *false positive* (diagnosa penyakit jantung, kenyataan normal) berjumlah 71 data, dan *false negatif* (diagnosa normal, kenyataan penyakit jantung) berjumlah 65 data. Sehingga didapat nilai *accuracy* 85.02%, nilai *precision* 86.02%, nilai *recall* 87.02% dan nilai *f1-score* 86.52%.

4. Algoritma Naïve Bayes Tanpa Outliers

Tabel 3.4 Algoritma Naïve Bayes Tanpa Outliers

	True Normal	True Penyakit Jantung	Class Precision
--	-------------	-----------------------	-----------------

Prediksi Normal	339	62	84.54%
Prediksi Penyakit Jantung	69	438	86.39%
Class Recall	83.09%	87.60%	

Pada tabel 3.4 *true positive* (diagnosa penyakit jantung, kenyataan penyakit jantung) berjumlah 438 data, *true negatif* (diagnosa normal, kenyataan normal) berjumlah 339 data, *false positive* (diagnosa penyakit jantung, kenyataan normal) berjumlah 69 data, dan *false negatif* (diagnosa normal, kenyataan penyakit jantung) berjumlah 62 data. Sehingga didapat nilai *accuracy* 85.57%, nilai *precision* 86.54%, nilai *recall* 87.57% dan nilai *f1-score* 87.05%.

4. KESIMPULAN

Kesimpulan yang didapat dari hasil analisis evaluasi model berdasarkan *cross validation* bahwa Algoritma *C4.5* dengan hasil prediksi tertinggi adalah data tanpa *outliers* dengan nilai *accuracy* sebesar 85.02%, nilai *precision* sebesar 86.02%, nilai *recall* sebesar 87.02%. dan nilai *f1-score* sebesar 86.52%. Begitupun dengan Algoritma *Naïve Bayes* dengan hasil prediksi tertinggi adalah data tanpa *outliers* dengan nilai *accuracy* sebesar 85.57%, nilai *precision* sebesar 86.54%, nilai *recall* sebesar 87.57% dan nilai *f1-score* sebesar 87.05%.

Perbandingan antara hasil prediksi penyakit gagal jantung dengan menggunakan Algoritma *C4.5* dan Algoritma *Naïve Bayes* terdapat dua bagian data yaitu data dengan *outliers* dan data tanpa *outliers*. Jika adanya data *outliers* pada sebuah dataset maka akan membuat analisis terhadap serangkaian data menjadi bias. Dari hasil yang telah dijelaskan diatas maka prediksi menggunakan Algoritma *Naïve Bayes* data tanpa *outliers* merupakan prediksi hasil *accuracy*, *presisi*, *recall* dan *f1-score* yang tertinggi di bandingkan dengan Algoritma *C4.5* data tanpa *outliers*.

REFERENCES

- [1] K. Algoritma, C. Dan, and N. Bayes, "Komparasi algoritma c4.5 dan naive bayes dalam prediksi penderita penyakit gagal jantung," vol. 5, no. 2, pp. 60–68, 2022.
- [2] P. Subarkah, W. R. Damayanti, and R. A. Permana, "Comparison of Correlated Algorithm Accuracy Naive Bayes Classifier and Naive Bayes Classifier for Classification of heart failure," *Ilk. J. Ilm.*, vol. 14, no. 2, pp. 120–125, 2022, doi: 10.33096/ilkom.v14i2.1148.120-125.
- [3] Q. Hasanah, H. Oktavianto, and Y. D. Rahayu, "Analisis Algoritma Gaussian Naive Bayes Terhadap Klasifikasi Data Pasien Penderita Gagal Jantung," *J. Smart Teknol.*, vol. 3, no. 4, pp. 382–389, 2022.
- [4] D. Derisma, "Perbandingan Kinerja Algoritma untuk Prediksi Penyakit Jantung dengan Teknik Data Mining," *J. Appl. Informatics Comput.*, vol. 4, no. 1, pp. 84–88, 2020, doi: 10.30871/jaic.v4i1.2152.
- [5] A. Rohman and M. Rochcham, "Model Algoritma C4.5 Untuk Prediksi Penyakit Jantung," *Neo Tek.*, vol. 4, no. 2, pp. 52–55, 2018, doi: 10.37760/neoteknika.v4i2.1228.
- [6] B. L. Nopitasari, B. Nurbaety, and H. Zuhroh, "Evaluasi Penggunaan Obat Antihipertensi pada Pasien Gagal Jantung Rawat Jalan di Rumah Sakit Umum Daerah Provinsi Nusa Tenggara Barat," *Lambung Farm. J. Ilmu Kefarmasian*, vol. 1, no. 2, p. 66, 2020, doi: 10.31764/lf.v1i2.2542.
- [7] *PEDOMAN TATA LAKSANA GAGAL JANTUNG*, Ke-2. Jakarta Barat: Siswanto BB, Hersunarti N, Erwinanto, Barack R, Pratikto RS, Nauli SE, Lubis AC, 2020.
- [8] I. Werdiningsih, Barriy Nuqoba, and Muhammadun, "DATA MINING MENGGUNAKAN ANDROID, WEKA DAN SPSS," Surabaya: Airlangga University Press, 2020, p. 209.
- [9] I. Romli and A. T. Zy, "Penentuan Jadwal Overtime Dengan Klasifikasi Data Karyawan Menggunakan Algoritma C4.5," *J. Sains Komput. Inform. (J-SAKTI)*, vol. 4, no. 2, pp. 694–702, 2020.
- [10] D. Darmansyah, S.Kom M.Kom, "Data Mining Menggunakan Aplikasi Rapid Miner," in *Data Mining Menggunakan Aplikasi Rapid Miner*, M. S. Yani, Ed. Sumatera Barat: PT INSAN CENDEKIA MANDIRI, 2020, p. 86.
- [11] M. R. Fanani, "Algoritma Naïve Bayes Berbasis Forward Selection Untuk Prediksi Bimbingan Konseling Siswa," *J. DISPROTEK*, vol. 11, no. 1, pp. 13–22, 2020, doi: 10.34001/jdpt.v11i1.952.