PENERAPAN SELEKSI FITUR GAIN RATIO PADA PREDIKSI PENYAKIT JANTUNG BERBASIS NAÏVE BAYES

Alifa Lutfia[1], Gunawan[2], Ramdhan Saepul Rohman[3], A Gunawan[4]

[1]Universitas Bina Sarana Informatika
e-mail: alifalutfia4@gmail.com

[2]Universitas Bina Sarana Informatika
*e-mail korespondensi: gunawan.gnz@bsi.ac.id

[3]Universitas Bina Sarana Informatika
e-mail: ramdhan.rpe@bsi.ac.id

[4]Universitas Bina Sarana Informatika
e-mail: a.gunawan.agn@bsi.ac.id

Abstrak

Penyakit jantung, yang menjadi ancaman serius bagi berbagai kelompok usia dan jenis kelamin, menuntut upaya pencegahan dan deteksi dini. Penelitian ini fokus pada prediksi penyakit jantung menggunakan algoritma Naïve Bayes. Meski efektif, Naïve Bayes memiliki keterbatasan performa, terutama dalam seleksi atribut. Oleh karena itu, penelitian ini mengintegrasikan metode seleksi fitur Gain Ratio untuk mengatasi kelemahan tersebut. Tujuannya adalah meningkatkan akurasi prediksi dengan memilih atribut yang relevan dan mengurangi bias pada data besar. Hasil penelitian menunjukkan bahwa kombinasi Naïve Bayes dan seleksi fitur Gain Ratio memberikan akurasi sebesar 91.2%, melampaui kinerja Naïve Bayes tanpa seleksi fitur. Temuan ini menegaskan pentingnya seleksi fitur dalam konteks prediksi penyakit jantung, memberikan kontribusi pada pemahaman lebih baik tentang faktor-faktor yang signifikan. Implikasinya dapat digunakan dalam pengembangan strategi pencegahan dan manajemen penyakit jantung, membuka peluang penggunaan teknologi untuk analisis dataset guna mendukung deteksi dini dan perawatan yang lebih efektif.

Kata Kunci: Jantung, Seleksi Fitur, Naïve Bayes, Fitur.

Abstract

Heart disease, a serious threat across various age groups and genders, demands preventive efforts and early detection. This study focuses on predicting heart disease using the Naïve Bayes algorithm. Despite its effectiveness, Naïve Bayes exhibits performance limitations, particularly in attribute selection. Therefore, this research integrates the Gain Ratio feature selection method to address these weaknesses, aiming to enhance prediction accuracy by selecting relevant attributes and reducing bias in large datasets. The research findings reveal that the combination of Naïve Bayes and Gain Ratio feature selection yields an accuracy of 91.2%, surpassing Naïve Bayes performance without feature selection. These results underscore the significance of feature selection in the context of heart disease prediction, contributing to a better understanding of significant factors. The implications extend to the development of prevention and management strategies for heart disease, presenting opportunities for utilizing technology in dataset analysis to support early detection and more effective treatment.

Keywords: *Heart, Feature Selection, Naïve Bayes, Features.*

1. Pendahuluan

Jantung merupakan organ vital yang berfungsi sebagai pemompa darah, untuk memenuhi suplai kebutuhan oksigen dan nutrisi ke seluruh tubuh. Pada saat jantung terganggu, peredaran darah dalam tubuh bisa terganggu. Oleh karena itu, menjaga kesehatan jantung sangat penting untuk menghindari berbagai penyakit jantung (Pradana et al., 2022).

Naskah dikirim 10 November 2023; direvisi 26 Januari 2024; diterbitkan 21 Februari 2024

Penyakit jantung merupakan penyakit yang merenggut banyak nyawa. Penyakit ini mengancam orang-orang dari segala penyakit ini merupakan penyakit dengan angka kematian yang tinggi dalam dunia kesehatan. Serta Penyebab penyakit ini umumnya kurang dipahami (Riani et al., 2019)

Menurut WHO (World Health Organization) dan CDC (Centers for Disease Control Frevention), penyakit jantung merupakan factor kematian terbesar di Inggris, Amerika Serikat, Kanada, dan Australia. Di Amerika Serikat, 26,6 Juta (11,3% dari populasi orang dewasa) di diagnosis penyakit jantung Indonesia memiliki jumlah penderita penyakit jantung tertinggi yang dibiayai oleh BPJS Kesehatan yaitu 11.592.990 kasus dan menjadi penyebab kematian nomor dua di Indonesia, sehingga penyakit ini ditakuti masyarakat hingga saat ini (Gligorijevic et al., 2019).

Beberapa faktor risiko penyakit jantung antara lain faktor keturunan, usia, jenis kelamin, stres, kurang olahraga, merokok, kolesterol tinggi, tekanan darah tinggi, diabetes, dan obesitas (Haryadi & Atmaja, 2021) Memprediksi penyakit jantung dapat berkontribusi pada pencegahan dan pengobatan yang lebih efektif (Wijayadhi et al., 2023).

Pada penelitian ini bertujuan untuk memprediksi penyakit jantung pada Manusia berdasarkan pada data sekunder yang diperoleh dari UCI Repository pada link https://archive.ics.uci.edu/dataset/45/heart+ disease dengan mengguanakan dataset klasifikasi penyakit jantung, pada penelitian ini diusulkan penggunakan metode machine learning dengan algoritma Naïve bayes.

Metode data mining digunakan untuk menganalisis sejumlah besar data dalam basis data. Di sektor kesehatan, potensi data mining diakui secara luas (Larassati et al., 2022) sedangkan untuk algotitma yang digunakan pada penelitian ini menggunakan Algoritma Naïve Bayes. Algoritma ini diimplementasikan dengan tujuan menghitung probabilitas pasien berdasarkan informasi dari rekam medis pasien. (Alhabib, 2022) Naive Bayes bekerja dengan baik pada dataset besar dan berdimensi tinggi. Cocok untuk aplikasi real-time dan kebal terhadap Noise.

Algoritma Naive Bayes dipilih karena lebih cocok untuk menerapkan probabilitas dari berbagai ukuran data untuk memprediksi distribusi penyakit jantung di masa depan.(Anggara & Halim, 2020) Naive Bayes didasarkan pada teorema Bayes, yang memiliki sifat klasifikasi mirip dengan pohon keputusan dan jaringan saraf. (Pradana et al., 2022) akan tetapi masalah pada penyakit jantung adalah ketidakseimbangan kelas. Suatu kelas dikatakan tidak seimbang ketika kelas positif hanya diwakili oleh beberapa tupel, sedangkan sebagian besar tupel mewakili kelas negatif (Prasetyo, 2020).

Naive bayes memiliki kelemahan-kelemahan yang menghasilkan kinerja yang kurang maksimal, sehingga diperlukan metode seleksi fitur untuk meningkatkan kinerja dari Naïve Bayes dalam menanggulangi ketidak seimbangan kelas (Buani, 2021) untuk metode seleksi fitur yang yang diusulkan pada penelitai ini guna meningkatkan akurasi pada prediksi penyakit jantung yakni dengan menggunakan seleksi fitur gain ratio, Gain Ratio dapat meningkatkan kinerja algoritma naive bayes dalam klasifikasi data (Septianingrum & Irawan, 2021).

Penyakit Jantung

Penyakit jantung adalah suatu kondisi di mana jantung tidak berfungsi dan dapat menyerang secara tiba-tiba dan menyebabkan kematian, penyakit ini tanpa memandang usia atau jenis kelamin. Tingginya angka kematian akibat penyakit jantung disebabkan ketidaktahuan masyarakat akan pemeriksaan jantung secara rutin dan gaya hidup masyarakat miskin. Penyakit jantung penyebab utama kematian di negara berpenghasilan tinggi dan rendah, mempengaruhi pria dan wanita hampir sama, penyakit ini tidak mudah diprediksi dan mungkin sebagian orang tidak sadar bahwa dirinya mengidap penyakit ini (Aziz et al., 2023).

Pada prinsipnya penyakit jantung bisa dicegah dengan berbagai cara salah satunya yakni gaya hidup sehat. Selain itu, serta deteksi dini pada penyakit jantung , salah satu cara yang sangat mungkin untuk deteksi dini adalah penerapan teknologi Informasi (Nurmasani & Pristyanto, 2021).

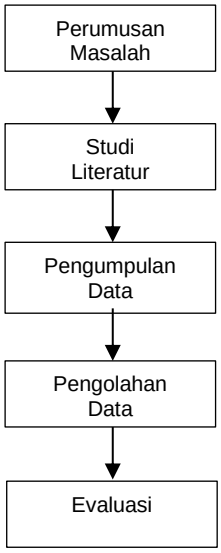
Data mining

Data Mining, yaitu analisis basis data untuk menemukan pengetahuan dalam bentuk model atau hubungan data yang valid, yang disebut *Knowledge Discovery in Databases (KDD)*. *Data Mining* adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan pembelajaran mesin untuk mengekstraksi dan mengidentifikasi informasi yang berguna

dan relevan dari basis data besar (Majid & Raharja, 2020).

2. Metode Penelitian

Agar penelitian ini dapat dikerjakan secara sistematis maka harus dibuatkan tahapan alur penelitian. Tahapan penelitian meliputi perumusan masalah, studi literatur, pengumpulan data, pengolahan data dan evaluasi. Berikut alur penelitian yang dilakukan dalam penelitian ini:



Gambar 1. Tahapan Penelitian

Perumusan Masalah

Tahap awal adalah menentukan permasalahan penelitian mengenai penyakit jantung dan datasetnya yang sangat besar serta rentan terhadap imbalance data.

Studi Literatur

Pada tahap ini dilakukan pengkajian mengenai penelitian yang dilakukan berdasarkan referensi dari jurnal dan buku.

Pengumpulan Data

Data yang digunakan dalam penelitian ialah data publik yang didapatkan dari website UCI Repository https://archive.ics.uci.edu/dataset/45/heart+disease, dataset dapat dilihat pada tabel

Pengolahan Data

Pada tahapan ini dataset kemudian diolah dengan aplikasi weka, fitur data dataset akan dilakukan seleksi menggunakan gain ratio dan dataset hasil selleksi fitur diolah menggunakan algoritma naive bayes.

Seleksi Fitur

Pemilihan fitur adalah proses pemilihan atribut yang sesuai dari kumpulan atribut yang besar. Algoritme memproses data lebih cepat dengan mengurangi atribut atau fitur pucat. Oleh karena itu, pemilihan fitur yang tepat akan menghasilkan hasil klasifikasi yang lebih cepat (Rukmana et al., 2022). Pemilihan fitur digunakan untuk mengurangi atribut data yang asing atau tidak relevan yang dihasilkan dari ekstraksi yang dapat memengaruhi hasil kinerja pembelajaran mesin (Fitriani et al., 2020).

Gain Ratio

Gain Ratio merupakan variasi dari dari Information gain yang bias. Penguatan memperhitungkan jumlah dan ukuran cabang pada waktu pemilihan parameter. Peningkatan akses ke informasi dengan memasukkan informasi internal (distribusi entropi) dalam perhitungan (Wintana, 2020). merupakan seleksi fitur yang dipakai dalam memilih fitur terbaik diantara fitur metode prediksi. Penggunaan Gain Ratio diharapkan menjadi meningkatkan akurasi perkiraan. Atribut dari relasi pengenalan itu sendiri adalah hasil bagi dari informasi timbal balik dan entrophy (Akbar & Rochimah, 2018) serta pemilihan selesi fitur gain ratio mampu meningkatkan akurasi klasifikasi penyakit (Indrayanti et al., 2017).

Rumus Gain rasio, sebagai alat pembobotan yang diterapkan dalam penelitian ini, memperlihatkan aplikabilitasnya dengan mengalokasikan bobot kepada setiap atribut. Pendekatan ini melibatkan perhitungan Gain rasio pada tiap atribut, yang kemudian diperkuat dengan perkalian jumlah data (n) dan rata-rata Gain rasio dari semua atribut. Proses selanjutnya melibatkan perhitungan bobot Naive Bayes, yang dilakukan dengan menambahkan bobot (wi) ke masing-masing atribut. Keseluruhan formula yang digunakan dapat diungkapkan dalam persamaan matematis yang terstruktur, menambah dimensi keakuratan dan kecanggihan analisis dalam kerangka penelitian ini. Berikut persamaannya:

$$\mathrm{w}_i = \frac{Gain\ ratio\ (i)}{\frac{1}{a}\sum_{i=1}^{a} Gain\ ratio(i)} \quad (1)$$

Untuk menentukan atribut terbaik dilakukan perhitungan dengan menggunakan rumus persamaan

$$\mathrm{MI}(x_i, y) = \sum_y \sum_{x1} P(x_1, y) log \frac{P(x_1 y)}{P(x_1)P(y)} \quad (2)$$

Setelah berhasil mengidentifikasi atribut terbaik, langkah berikutnya melibatkan penentuan entropy untuk mengukur kekuatan

dari input atribut. Proses ini merinci kemampuan atribut dalam menghasilkan output fitur, yang dicapai melalui perhitungan probabilitas menggunakan suatu rumus matematis yang terstruktur. Rumus tersebut menyajikan landasan matematis yang komprehensif, memungkinkan penghitungan probabilitas yang terperinci dan akurat, sehingga meningkatkan tingkat ketelitian dan keandalan dalam penentuan output fitur dari input atribut yang telah diseleksi. Berikut rumus persamaannya:

$$E(x_i) = \sum_{x1} P(x_1) log \frac{1}{P(x_1)} \quad (3)$$

Proses perhitungan guna menentukan nilai gain ratio melibatkan serangkaian langkah yang sistematis. Pertama-tama, dilakukan perhitungan nilai entropy sebagai langkah awal yang menyajikan ukuran ketidakpastian dari suatu atribut. Setelahnya, dilakukan perhitungan nilai information gain, yang menggambarkan sejauh mana atribut tersebut memberikan informasi baru yang signifikan. Selanjutnya, melibatkan perhitungan gain ratio, sebuah parameter yang mempertimbangkan ukuran information gain relatif terhadap nilai split information. Keseluruhan proses ini dirancang dengan cermat untuk memberikan landasan yang solid dalam menentukan pemilihan atribut yang dianggap memiliki pengaruh yang sangat signifikan, terutama dalam konteks klasifikasi penyakit jantung.

Naïve Bayes

Naive Bayes adalah metode klasifikasi yang menggunakan metode probabilitas yang dikembangkan oleh ilmuwan Inggris Thomas Bayes yang memprediksi kemungkinan masa depan berdasarkan pengalaman masa lalu, yang dikenal sebagai teorema Bayes (Anggara & Halim, 2020). Kelebihan metode Naive Bayes hanya membutuhkan sedikit data pelatihan untuk menentukan estimasi parameter yang dibutuhkan dalam proses klasifikasi. Dengan kinerja yang lebih baik (Gligorijevic et al., 2019).

Algoritma Naïve Bayes memiliki kemampuan dalam memprediksi didasarkan pada probabilitas P atribut x dari setiap kelas y data. Berikut merupakan model probabilitas setiap kelas dan jumlah atribut pada kelas k dan a

$$P(y_k \mid X_1, X_2, X_3 \ldots x_a) \quad (4)$$

Perhitungan probabilitas pada algoritma Naïve Bayes didasarkan pada kemunculan dokumen XA pada setiap kelas dari kelas yk (xa|yk), dikalikan dengan probabilitas dari kelas kelas P(yk), dari hasil perkalian, kemudian membagi probabilitas dengan kemunculan dokumen (xa) sehingga nantinya dalam perhitungan untuk naive bayes dapat diformulasikan menurut persamaan berikut.

$$P(y_k \mid x_a \frac{P(y_k)\, P(x_a \mid y_k)}{p(x_a)} \quad (5)$$

Setelah probabilitas dihitung, dilakukan proses pemilihan kelas yang optimal, setelah itu dipilih nilai maksimum dari setiap probabilitas kelas yang ada. menggunakan persamaan berikut:

$$y(x_i) = \arg max\ p\big((y) \prod_{i=1}^{a} P(x_i \mid y\big) \quad (6)$$

Pembobotan atribut tiap kelas dapat meningkatkan efek prediksi dan memperhitungkan bobot atribut kelas, sehingga akurasi klasifikasi tidak hanya didasarkan pada probabilitas tetapi juga pada bobot. setiap atribut. Dengan menggunakan rumus sebagai berikut:

$$P((y, x = \mathrm{p(y)} \prod_{i=1}^{a} C\ P(x_i \mid y)^{wi} \quad (7)$$

Evaluasi

Tahap terakhir yaitu dilakukan evaluasi terhadap pengolahan dataset, tahap ini dilakukan untuk mengetahui efektifitas dan akurasi yang dihasilkan dari pengguan gain ratio terhadap naive bayes.

3. Hasil dan Pembahasan

Dalam penelitian ini, data yang menjadi fokus penelitian bersumber dari UCI Dataset, sebuah repositori machine learning yang diakses melalui situs web UCI Repository di https://archive.ics.uci.edu/dataset/45/heart+disease. Data yang berhasil diperoleh menggambarkan informasi terkait pasien yang menderita penyakit jantung, dengan melibatkan 13 atribut yang merinci berbagai aspek relevan. Penggunaan dataset ini dipilih dengan cermat untuk memberikan dasar yang kuat dalam melakukan analisis machine learning terkait dengan kondisi kesehatan jantung, sehingga mampu menghasilkan temuan yang lebih mendalam dan signifikan dalam konteks riset ini. Berikut data pasien

penyakit jantung dengan 13 atribut yang digunakan yakni:

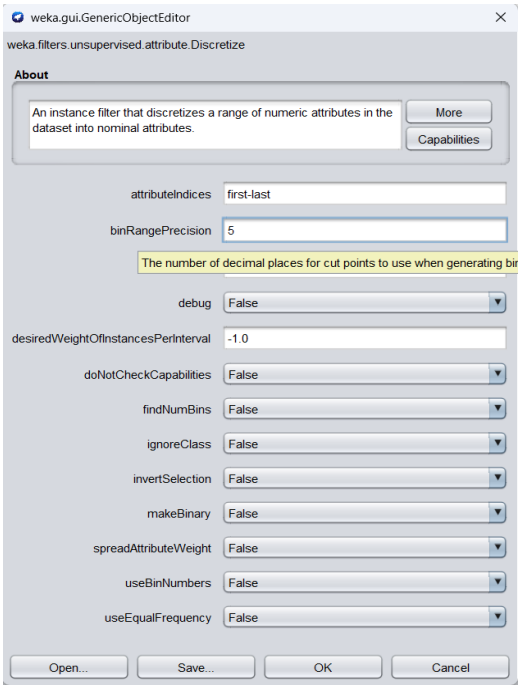
Tabel 1. Atribut Penyakit Jantung

No	Atribut	Nilai	Kategori
1	*Age*	*<40*	*Inventus*
		40-50	*Verilitas*
		55-64	*Prasenium*
		>=65	*Senium*
2	*Sex*	*1*	*Pria*
		0	*Wanita*
3	*Chest Pain Type*	*1*	*Typical Angina*
		2	*Atypical Angina*
		3	*Non Angina Pain*
		4	*Asymtomatyc*
4	*Testing Blood Pressure*	<120	Normal
		120-139	*Prehipertensi*
		140-159	*Hypertensi level 1*
		>159	*Hypertensi Level 2*
5	*Serum Cholesterol (mg/dl)*	<200	Normal
		200-239	Batas normal
		>239	Tinggi
6	*Fasting Blood Sugar >120 mg/dl*	1	Iya
		0	Tidak
7	*RestingElectrocardiographic Result*	0	Normal
		1	*Having ST-T wave abnormality*
		2	*Left Ventricular Hypertropy*
8	*Maximum heart Rate Archeived*	220 -usia	Normal
		<>220 - Usia	Tidak Normal
9	*Excerside induced angina*	0	Tidak
		1	Ya
10	*Old Peak*	<1	0
		>= 1 dan <2	1
		>=2 dan <3	2
		>= 3 dan <4	3
		>=4	4
11	*The Slope of the park excersise ST Segmen*	1	*Unsloping*
		2	*Flate*
		3	*Downsloping*
12	*Number of Major Vessel*	*1*	*1*
		2	*2*
		3	*3*
13	*Thal*	*3*	*Normal*
		6	*Fixed Defetc*
		7	*Reversible Devect*

3.1. Tahap Persiapan Data

Setelah menyelesaikan tahap pemahaman data pasien, langkah berikutnya adalah tahap pengolahan data yang menghasilkan dataset akhir yang akan digunakan dalam pemodelan data. Proses ini melibatkan serangkaian operasi dan transformasi data yang dirancang untuk memastikan kualitas dan keberlanjutan dataset yang dihasilkan. Dengan dataset yang telah disiapkan, langkah selanjutnya adalah mengimplementasikan proses data mining menggunakan aplikasi Weka. Penerapan alat ini memungkinkan analisis mendalam terhadap pola dan hubungan dalam data, memungkinkan penemuan informasi berharga yang dapat mendukung pemahaman lebih lanjut mengenai kondisi kesehatan pasien atau fenomena yang sedang dipelajari.

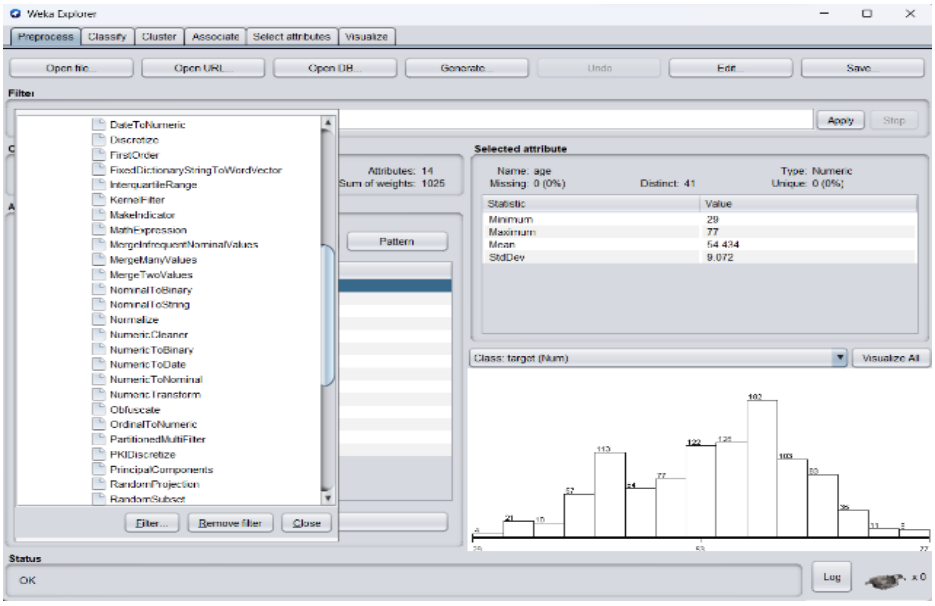
Dalam tahap awal pemrosesan dataset, dilakukan praproses data untuk mengatasi karakteristik data awal yang bersifat numerik. Adapun langkah pertama yang diambil adalah mengubah tipe data dari numerik menjadi nominal. Proses ini diperlukan agar dataset dapat dibaca dan dipahami oleh Aplikasi Weka, seperti yang terlihat pada ilustrasi Gambar 2.



Gambar 2. Konversi data Numerik Ke Nominal

Transformasi ini memungkinkan aplikasi untuk menerapkan metode analisis data mining yang optimal, sesuai dengan kebutuhan dan spesifikasi dataset yang diolah.

Langkah berikutnya dalam pengolahan data melibatkan perangkingan data untuk menangani beragam jenis ukuran yang dimiliki oleh dataset, yang berkisar dari angka 1 hingga ratusan. Pendekatan yang diambil adalah dengan membagi kelas data ke dalam lima kategori dengan proses diskritisasi.

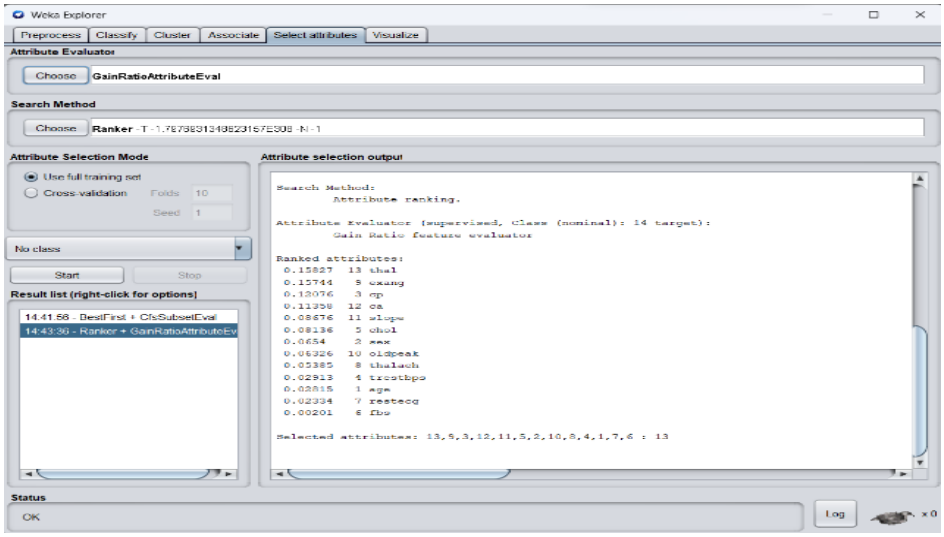


Gambar 3. Proses Diskrit data

Proses diskritisasi ini dilakukan dengan tujuan utama untuk memfasilitasi seleksi fitur pada tahap selanjutnya dari pengolahan data. Dengan membagi data ke dalam kelas-kelas yang lebih terstruktur, memungkinkan penyusunan kriteria seleksi fitur yang lebih efektif dan efisien dalam konteks analisis data yang dilakukan.

3.2. Tahap Pemodelan

Dalam langkah Pemodelan, penelitian ini secara sengaja memilih teknik data mining yang melibatkan penerapan algoritma klasifikasi yang terkait dengan metode Data Mining *Classification Rules.* Sebelum memasuki fase pengolahan data menggunakan model klasifikasi yang telah dipilih, dilakukan serangkaian langkah tambahan untuk mempersiapkan dataset. Tahap persiapan ini mencakup pengolahan data yang melibatkan perangkingan, dan proses ini diinisiasi dengan menerapkan metode gain ratio, sebagaimana tercermin pada ilustrasi yang ditampilkan dalam gambar 4.



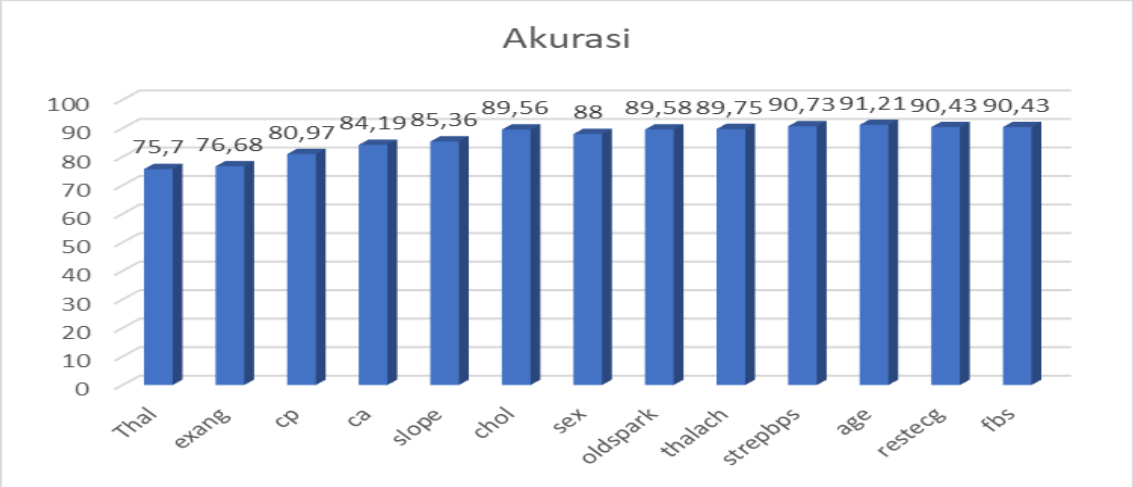
Gambar 4. Hasil Perangkingan

Proses perangkingan ini merupakan langkah strategis untuk mengorganisir data dengan lebih efektif, sehingga dapat memfasilitasi eksekusi model klasifikasi dengan lebih optimal pada tahap selanjutnya dalam penelitian ini.

Setelah melalui proses perangkingan dengan menggunakan seleksi fitur gain ratio,

atribut-atribut yang diurutkan secara menurun adalah sebagai berikut: Thal, Exang, cp, ca, slope, chol, sek, oldpeak, thalach, strepbps, age, restecg, dan fbs. Dari urutan perangkingan tersebut, terlihat bahwa tiga atribut memiliki nilai peringkat di bawah 5, yaitu age, restecg, dan fbs. Oleh karena itu, dilakukan suatu percobaan dengan menghapus fitur-fitur yang terkait untuk meningkatkan hasil akurasi pada model Naïve Bayes.

Percobaan ini bertujuan untuk mengevaluasi dampak penghapusan fitur-fitur tersebut terhadap performa model. Dengan demikian, hasil dari perangkingan dijadikan dasar untuk eksplorasi lebih lanjut dan pembuktian apakah penghapusan fitur-fitur tertentu dapat meningkatkan akurasi model Naïve Bayes. Evaluasi ini diharapkan dapat memberikan wawasan lebih lanjut terkait kontribusi masing-masing fitur terhadap kinerja model klasifikasi, sehingga dapat diambil keputusan yang lebih terinformasi terkait dengan pemilihan fitur dalam konteks penelitian ini.



Gambar 5. Hasil Pengujian Data

Berdasarkan hasil percobaan setelah melakukan seleksi dan penghapusan fitur fbs dan Restecg, diperoleh nilai akurasi tertinggi sebesar 91.21%. Hasil ini mencerminkan peningkatan performa model Naïve Bayes setelah dilakukan penyesuaian terhadap fitur-fitur yang dipilih.

Confusion matrix, yang merupakan alat evaluasi kinerja model klasifikasi, memberikan gambaran lebih rinci tentang performa model dalam mengklasifikasikan instance-data pada setiap kelas. Meskipun detail matriks konfusi tidak disediakan dalam pertanyaan Anda, matriks konfusi dapat memberikan informasi tentang nilai true positive, true negative, false positive, dan false negative untuk setiap kelas yang diamati. Hal ini dapat digunakan untuk evaluasi lebih lanjut terkait keberhasilan model dalam mengidentifikasi kelas-kelas tertentu.

Tabel 2. Tabel Confusion Matrik

		Kenyataan	
		True	**False**
Prediksi	**True**	449	50
	False	40	482

$$\frac{TP + TN}{TP + FP + FN + TN} \; x \; 100$$

$$\frac{449 + 482}{449 + 40 + 50 + 482} \; x \; 100$$

$$\frac{935}{1025} \; x \; 100 = \frac{93.500}{102.500}$$

$$= 0{,}912$$

3.3. Tahap Evaluasi

Tahap evaluasi ini merupakan langkah kritis dalam menilai apakah pemodelan data klasifikasi penyakit jantung dengan menggunakan seleksi fitur Gain Ratio telah mencapai tujuan penelitian dan menghasilkan hasil yang sesuai dengan harapan awal. Melalui perbandingan antara metode ini dengan metode Naïve Bayes tanpa seleksi fitur, beberapa kesimpulan dapat diperoleh dari hasil penghitungan yang telah dilakukan.

H hasil evaluasi menunjukkan bahwa penerapan seleksi fitur menggunakan metode Gain Ratio memberikan dampak positif terhadap performa model klasifikasi. Hal ini tercermin dari peningkatan nilai akurasi yang diperoleh, sejalan dengan tujuan utama penelitian.

Dengan mengintegrasikan metode Naïve Bayes dan seleksi fitur Gain Ratio, terlihat bahwa model dapat memberikan hasil yang lebih optimal dalam mengklasifikasikan penyakit jantung. Pemilihan fitur-fitur yang lebih relevan dan signifikan dapat meningkatkan akurasi model dan mengurangi potensi overfitting.

Perbandingan dengan metode Naïve Bayes tanpa seleksi fitur memberikan gambaran lebih jelas tentang kontribusi positif dari tahap seleksi fitur dalam meningkatkan kualitas hasil klasifikasi. Ini mengindikasikan bahwa penggunaan seleksi fitur dapat memberikan insight berharga dalam meningkatkan efisiensi dan ketepatan model.

Tahap evaluasi ini menegaskan bahwa penerapan seleksi fitur Gain Ratio pada metode Naïve Bayes memberikan hasil yang sesuai dengan harapan dan memperbaiki performa model klasifikasi penyakit jantung. Hasil ini dapat dijadikan dasar untuk rekomendasi penggunaan metode tersebut dalam analisis lebih lanjut atau aplikasi di lingkungan klinis.

Berikut hasil evaluasi perbandingan antara metode naïve bayes dengan naïve bayes menggunakan gain ratio yang dapat ilihat pada tabel 3.

Tabel 3. Evaluasi Perbandingan Naïve Bayes dan Naïve Bayes+Gain Ratio

	Naïve Bayes	**Naïve Bayes + Gain Ratio**
Acuraccy	0,904	0,912
Precision	0,910	0,918
Recall	0,892	0,899

Berdasarkan data yang terdapat dalam Tabel 3, ditemukan bahwa penggunaan metode Naïve Bayes dengan penerapan seleksi fitur menggunakan Gain Ratio pada dataset penyakit jantung menghasilkan peningkatan nilai akurasi klasifikasi. Terdapat kenaikan sebesar 1.00% jika dibandingkan dengan penggunaan metode Naïve Bayes tanpa seleksi fitur.

Selain peningkatan akurasi, terdapat peningkatan pada nilai precision dan recall meskipun tidak signifikan, namun hal ini memberikan indikasi bahwa penerapan seleksi fitur dengan Gain Ratio memiliki dampak positif terhadap kualitas prediksi model. Meskipun peningkatan ini mungkin tidak besar, tetapi hasil ini menggambarkan bahwa metode seleksi fitur dapat memberikan kontribusi yang berarti terhadap presisi dan kemampuan model untuk mengenali secara efisien kelas-kelas tertentu.

Hal ini menunjukkan bahwa menggunakan seleksi fitur dengan Gain Ratio dalam konteks klasifikasi penyakit jantung dapat memberikan perbaikan yang signifikan terhadap performa model Naïve Bayes. Oleh karena itu, hasil ini memberikan dukungan bagi penerapan strategi seleksi fitur dalam upaya meningkatkan akurasi dan keandalan model klasifikasi dalam domain kesehatan.

Berikut ini merupakan perbandingan hasil penelitian model yang diusulkan dengan penelitian sebelumnya dengan menggunakan dataset yang sama dan dengan menggunakan metode yang berbeda, hasil perbandingan hasil penelitian dapat dilihat pada table 4 berikut ini:

Tabel 4. Perbandingan Penelitian Sebelumnya

NO	**Judul Penelitian**	**Algoritma yang digunakan**	**Accuracy**
1	Komparasi Algoritma Klasifikasi Svm-Pso Dan C4.5-Pso Dalam Prediksi Penyakit Jantung	SVM + PSO C4.5 + PSO	84.41% 80.0%
2	Perbandingan Kinerja Algoritma untuk Prediksi Penyakit Jantung dengan Teknik Data Mining Derisma	Algoritma Naive Bayes, Random Forest, Neural Network	83%
3	Komparasi *Support Vector Machine, Logistic Regression* Dan *Artificial Neural Network* dalam Prediksi	Logistic Regression dengan	83%

	Penyakit Jantung		
4	Algoritme *Stacking* Untuk Klasifikasi Penyakit Jantung Pada Dataset Imbalanced Class	STACKING	81%
5	Komparasi Algoritma *Neural Network* Dan *Naïve Bayes* Untuk Memprediksi Penyakit Jantung	ANN Naïve Bayes	84,52% 79,88%
6	Metode Yang Diusulkan	Naïve Bayes + Gain Ratio	**91.2 %**

4. Kesimpulan

Penelitian ini menerapkan seleksi fitur Gain Ratio dan metode Naïve Bayes pada dataset penyakit jantung. Proses pemodelan ini melibatkan serangkaian pengujian untuk mendapatkan nilai akurasi, presisi, dan recall dengan menggunakan algoritma yang telah dipilih. Hasil dari pengujian ini menunjukkan bahwa penerapan Naïve Bayes dengan metode seleksi fitur Gain Ratio menghasilkan nilai akurasi sebesar 91.2%.

Angka akurasi ini memberikan bukti bahwa penerapan Gain Ratio pada Naïve Bayes mampu meningkatkan nilai akurasi dibandingkan dengan penggunaan Naïve Bayes tanpa seleksi fitur. Dengan demikian, dapat disimpulkan bahwa penggunaan Gain Ratio sebagai metode seleksi fitur memberikan kontribusi positif terhadap kemampuan model dalam memprediksi penyakit jantung.

Selain akurasi, evaluasi juga mencakup presisi dan recall, yang memberikan wawasan tambahan tentang performa model dalam mengidentifikasi dan mengklasifikasikan instansi-data. Dengan nilai akurasi yang mencapai 91.2%, hasil ini memberikan dukungan kuat terhadap efektivitas penerapan metode seleksi fitur Gain Ratio pada Naïve Bayes sebagai suatu pendekatan yang memperbaiki kemampuan prediksi model terkait dengan penyakit jantung.

Referensi

Alhabib, I. (2022). Komparasi Metode Deep Learning, Naïve Bayes Dan Random Forest Untuk Prediksi Penyakit Jantung. *INFORMATICS FOR EDUCATORS AND PROFESSIONAL : Journal of Informatics*, *6*(2), 176. https://doi.org/10.51211/itbi.v6i2.1881

ANGGARA, B., & Halim, R. M. N. (2020). Penerapan Data Mining Dirumah Sakit Umum Daerah Prabumulih Menggunakan Algoritma Naive Bayes Study Kasus: Penyakit Jantung. *Bina Darma …*, 209–220. http://conference.binadarma.ac.id/index.php/BDCCS/article/view/1546%0Ahttps://conference.binadarma.ac.id/index.php/BDCCS/article/download/1546/690

Aziz, M. I., Fanani, A. Z., & Affandy, A. (2023). Analisis Metode Ensemble Pada Klasifikasi Penyakit Jantung Berbasis Decision Tree. *Jurnal Media Informatika Budidarma*, *7*(1), 1–12. https://doi.org/10.30865/mib.v7i1.5169

Cahya Putri Buani, D. (2021). Penerapan Algoritma Naïve Bayes dengan Seleksi Fitur Algoritma Genetika Untuk Prediksi Gagal Jantung. *EVOLUSI : Jurnal Sains Dan Manajemen*, *9*(2), 43–48. https://doi.org/10.31294/evolusi.v9i2.11141

Fitriani, I., Basuki, S., & Minarno, A. E. (2020). Seleksi Fitur Relieff Pada Klasifikasi Malware Android Menggunakan Support Vector Machine(SVM). *Jurnal Repositor*, *2*(11), 1529. https://doi.org/10.22219/repositor.v2i11.901

Gligorijevic, N., Robajac, D., & Nedic, O. (2019). Повышенная Чувствительность Тромбоцитов К Действию Инсулиноподобного Фактора Роста 1 У Больных Сахарным Диабетом 2-Го Типа. *Биохимия*, *84*(10), 1511–1518. https://doi.org/10.1134/s0320972519100129

Haryadi, D., & Atmaja, D. M. U. (2021). Penerapan Algoritma K-Means Clustering Untuk Pengelompokan Tingkat Risiko Penyakit Jantung. *Journal of Informatics and*

Communication Technology (JICT), *3*(2), 51–66. https://doi.org/10.52661/j_ict.v3i2.85

Indrayanti, Sugianti, D., & Karomi, M. A. Al. (2017). Peningkatan Akurasi Algoritma KNN dengan Seleksi Fitur Gain Ratio untuk Klasifikasi Penyakit Diabetes Mellitus. *IC-Tech*, *7*(2), 1–6. https://ejournal.stmik-wp.ac.id/index.php/ictech/article/view/3/2

Larassati, D., Zaidiah, A., & Afrizal, S. (2022). Sistem Prediksi Penyakit Jantung Koroner Menggunakan Metode Naive Bayes. *JIPI (Jurnal Ilmiah Penelitian Dan Pembelajaran Informatika)*, *7*(2), 533–546. https://doi.org/10.29100/jipi.v7i2.2842

Majid, A.M; Dwi Raharja, M. . (2020). *Indonesian Journal of Business Intelligence*. *3*(2), 54–60.

Nurmasani, A., & Pristyanto, Y. (2021). Algoritme Stacking Untuk Klasifikasi Penyakit Jantung Pada Dataset Imbalanced Class. *Pseudocode*, *8*(1), 21–26. https://doi.org/10.33369/pseudocode.8.1.21-26

Pradana, D., Luthfi Alghifari, M., Farhan Juna, M., & Palaguna, D. (2022). Klasifikasi Penyakit Jantung Menggunakan Metode Artificial Neural Network. *Indonesian Journal of Data and Science*, *3*(2), 55–60. https://doi.org/10.56705/ijodas.v3i2.35

Pradana, M. G., Saputro, P. H., & Wijaya, D. P. (2022). Komparasi Metode Support Vector Machine Dan Naïve Bayes Dalam Klasifikasi Peluang Penyakit Serangan Jantung. *Indonesian Journal of Business Intelligence (IJUBI)*, *5*(2), 87. https://doi.org/10.21927/ijubi.v5i2.2659

Prasetyo, E., & Prasetiyo, B. (2020). Peningkatan Akurasi Klasifikasi Algoritma C 4.5 Menggunakan Teknik Bagging pada Diagnosis Penyakit Jantung. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, *7*(5), 1035–1040. https://doi.org/10.25126/jtiik.2020752379

Riani, A., Susianto, Y., & Rahman, N. (2019). Implementasi Data Mining Untuk Memprediksi Penyakit Jantung Mengunakan Metode Naive Bayes. *Journal of Innovation Information Technology and Application (JINITA)*, *1*(01), 25–34. https://doi.org/10.35970/jinita.v1i01.64

Septianingrum, F., & Irawan, A. S. Y. (2021). Metode Seleksi Fitur Untuk Klasifikasi Sentimen Menggunakan Algoritma Naive Bayes: Sebuah Literature Review. *Jurnal Media Informatika Budidarma*, *5*(3), 799. https://doi.org/10.30865/mib.v5i3.2983

Sonhaji Akbar, M., & Rochimah, S. (2018). Prediksi Cacat Perangkat Lunak Dengan Optimasi Naive Bayes Menggunakan Pemilihan Fitur Gain Ratio. *Jurnal Sistem Dan Informatika*, *11*(1), 147–155.

Wijayadhi, A., Effendi, M. M., & Rahardjo, S. B. (2023). Prediksi Penyakit Jantung Dengan Algoritma Regresi Linier. *Bulletin of Information Technology (BIT)*, *4*(1), 15–28. https://journal.fkpt.org/index.php/BIT

Wintana, D. (2020). Integrasi metode diskritisasi dan gain ratio pada prediksi cacat perangkat lunak berbasis naive bayes tesis. *Repository Nusa Mandiri*, 1–48. https://repository.bsi.ac.id/index.php/unduh/item/337645/14002220_Dede-Wintana-Full-1.pdf

Zulaikhah Hariyanti Rukmana, S., Aziz, A., & Harianto, W. (2022). Optimasi Algoritma K-Nearest Neighbor (Knn) Dengan Normalisasi Dan Seleksi Fitur Untuk Klasifikasi Penyakit Liver. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *6*(2), 439–445. https://doi.org/10.36040/jati.v6i2.4722