



Implementasi Metode *Naïve Bayes* untuk Klasifikasi Penderita Penyakit Jantung

Bowo Hirwono^{1*}, Arief Hermawan², Donny Avianto³

^{1*} Program Studi Informatika, Fakultas Teknik, Universitas Teknologi Yogyakarta, Kabupaten Sleman, Daerah Istimewa Yogyakarta, Indonesia.

^{2,3} Program Studi Magister Teknologi Informasi, Fakultas Teknik, Universitas Teknologi Yogyakarta, Kabupaten Sleman, Daerah Istimewa Yogyakarta, Indonesia.

article info

Article history:

Received 21 January 2023

Received in revised form

6 April 2023

Accepted 1 May 2023

Available online July 2023

DOI:

<https://doi.org/10.35870/jtik.v7i3.910>

Keywords:

Naïve Bayes; Heart Disease;
Classification.

Kata Kunci:

Naïve Bayes; Penyakit Jantung;
Klasifikasi.

abstract

Heart attack is a very serious heart disorder. This disorder occurs when the heart muscle does not get good blood flow. This condition will interfere with the function of the heart in flowing blood flow throughout the body. This study aims to develop a system capable of classifying people with heart disease using the Naïve Bayes method. Naïve Bayes is a method that works based on the probability that a person has a heart disease or not based on their medical record data. This algorithm is used with the aim of calculating the probability of a person suffering from heart disease based on their medical records. This data was obtained from the University of California Irvine Machine Learning website with a total of 303 datasets with 13 attributes. This research was conducted by dividing the data into 75% for training data and 25% for testing data. The results of this study indicate that the Naïve Bayes algorithm used gives a fairly high accuracy value of 86.84%.

abstract

Serangan jantung merupakan gangguan jantung yang sangat serius. Gangguan ini terjadi ketika otot jantung tidak mendapatkan aliran darah yang baik. Kondisi inilah yang akan mengganggu fungsi jantung dalam mengalirkan aliran darah ke seluruh tubuh. Penelitian ini bertujuan untuk mengembangkan sistem yang mampu mengklasifikasikan penderita penyakit jantung menggunakan metode naïve bayes. Naïve bayes adalah suatu metode yang bekerja berdasarkan probabilitas kemungkinan seseorang menderita penyakit jantung atau tidak berdasarkan data rekam medisnya. Algoritma ini digunakan dengan tujuan menghitung probabilitas kemungkinan seseorang menderita penyakit jantung berdasarkan rekam medisnya. Data ini diperoleh dari website University of California Irvine Machine learning dengan jumlah keseluruhan ada 303 dataset dengan 13 atribut, penelitian ini dilakukan dengan membagi data menjadi 75% untuk data latih dan 25% untuk data training, hasil penelitian ini menunjukkan bahwa algoritma Naïve Bayes yang digunakan memberikan nilai akurasi yang cukup tinggi sebesar 86.84%.

Corresponding Author. Email: hirwonob@gmail.com ^{1}.

1. Latar Belakang

Kesehatan adalah yang paling utama untuk menjalani kehidupan sehari-hari. Banyak hal yang harus diperhatikan dalam menjalani kehidupan yang sehat, salah satunya menjaga kesehatan jantung. Jantung yang juga merupakan sistem peredaran darah pada manusia merupakan salah satu organ vital yang penting yang ada pada tubuh manusia. Jantung adalah organ yang mempunyai peranan penting dalam kelangsungan hidup manusia karena fungsinya mendistribusikan darah dari paru-paru ke setiap kantong di tubuh yang faktanya bahwa darah ini mengandung banyak oksigen, jantung merupakan organ penting dalam kehidupan manusia karena membantu mengatur metabolisme di tubulus manusia [1]. Jantung bertugas sebagai alat pompa untuk mengedarkan darah keseluruh tubuh manusia. Serangan jantung merupakan gangguan jantung yang sangat serius. Gangguan ini terjadi ketika otot jantung tidak mendapatkan aliran darah yang baik. Kondisi inilah yang akan mengganggu fungsi jantung dalam mengalirkan aliran darah ke seluruh tubuh. Hal ini dapat berakibat fatal bagi kesehatan manusia [2].

Jika jantung mengalami suatu gangguan atau kerusakan maka akan mengakibatkan terganggunya seluruh kinerja organ didalam tubuh manusia. Berdasarkan data Badan Kesehatan Dunia (WHO) pada 2016, total kematian di Indonesia mencapai 1.863.000 jiwa, Sebanyak 73 persen di antaranya disebabkan oleh penyakit tidak menular. Adapun jantung koroner menjadi penyebab kematian nomor 1 sebanyak 35 persen. Tingkat penyakit jantung tertinggi didominasi oleh manula berusia 75 tahun ke atas yakni 4,7 persen, Namun bukan berarti usia muda bebas dari ancaman penyakit jantung [2]. Serangan Jantung adalah salah satu penyakit yang paling mematikan di dunia. dan salah satu penyakit yang banyak penderitanya adalah Penyakit Jantung [3], Penyakit jantung adalah suatu penyakit degeneratif yang berkaitan dengan gaya hidup, dan sosial ekonomi masyarakat. Penyakit ini merupakan masalah kesehatan utama di negara maju. Badan Kesehatan Dunia (WHO) mencatat lebih dari 7 juta orang meninggal akibat penyakit jantung di seluruh dunia pada tahun 2002. Angka ini diperkirakan meningkat hingga 11 juta orang pada tahun 2020 [4].

Kebanyakan para penderita penyakit jantung tidak mengetahui bahwa dia mengidap penyakit tersebut dikarenakan penyakit ini sulit untuk dideteksi, sehingga tidak sedikit para penderita yang meninggal dunia dikarenakan serangan penyakit jantung yang kambuh secara tiba-tiba. Sebenarnya penyakit jantung bisa diketahui sejak dini, namun dikarenakan kurangnya edukasi sehingga hal tersebut terlupakan dan penyakit tersebut akan dibawa hingga dia dewasa.

sejauh ini proses untuk mengetahui bahwa penderita mengalami penyakit jantung adalah secara manual yaitu melakukan pemeriksaan ataupun konsultasi langsung ke dokter ahli atau spesialis jantung dan mengikuti beragam pemeriksaan laboratorium untuk mengetahui hasil apakah orang tersebut mengidap penyakit jantung atau tidak. Namun banyak keluhan yang terjadi dikarenakan hal tersebut tentu saja memerlukan biaya yang tidak sedikit, ditambah dengan resiko yang begitu tinggi, maka dengan ini diperlukanlah suatu sistem klasifikasi yang nantinya dapat membantu untuk mendeteksi bahwa seseorang menderita penyakit jantung secara akurat, efisien dan pastinya hanya dengan biaya yang relatif murah.

Klasifikasi adalah jenis analisis data yang dapat membantu orang memprediksi label kelas sampel harus diklasifikasikan, Berbagai macam teknik klasifikasi telah diusulkan dalam bidang-bidang seperti pembelajaran mesin, sistem pakar dan statistik[5]. Perkembangan teknologi informasi dan ilmu komputasi, telah membuka peluang yang sangat besar untuk menjadi bagian dari solusi alternatif penyelesaian masalah terkait sulitnya identifikasi penyakit jantung [6]. Hal ini membuat banyak para peneliti melakukan penelitian terhadap penyakit jantung, salah satunya dengan cara metode yang berbasis komputer. Cara ini sudah banyak digunakan dengan komputasi yang begitu cerdas dan memproses data dengan jumlah yang begitu banyak, pengolahan data tersebut tentunya dilakukan dengan cara klasifikasi menggunakan beberapa metode dengan hasil yang akurat dan cepat.

Penelitian yang dilakukan [7] dalam peramalan, untuk mendapatkan hasil yang akurat dan bermanfaat, terdapat dua hal yang harus diperhatikan. Pertama Data yang dikumpulkan haruslah berupa informasi yang relevan sehingga dapat menghasilkan peramalan

yang akurat, dan yang kedua penggunaan metode peramalan yang tepat. Selanjutnya peneliti melakukan pengujian menghasilkan perbandingan kedua algoritma Neural Network dan Naïve Bayes untuk melakukan prediksi terhadap penyakit jantung dengan 14 atribut hasilnya lebih dari 75% tepat untuk mendeteksi penyakit jantung pada manusia [7]. Peneliti lain melakukan penelitian menggunakan Data Riset Kesehatan Dasar 2013, Badan Litbangkes Kementerian Kesehatan RI dan Data Penduduk Sasaran, Pusdatin Kementerian Kesehatan RI.2 (<https://m5.gs/bE42Nm>). atribut yang digunakan dengan 6 atribut seperti; Nama, Sesak nafas, Denyut nadi, Mual muntah, gelisah, gagal jantung. Proses klasifikasi ini menggunakan Tools Weka, Setelah mengeksekusi dataset dengan algoritma yang dipilih didapatlah algoritma naïve bayes dengan hasil akurasi 80% Untuk pengembangannya, peningkatan akurasi bisa dilakukan dengan menghibridkan algoritma klasifikasi ini dengan beberapa algoritma lain, sehingga tidak menutup kemungkinan bisa menghasilkan kinerja yang berbeda dari algoritma tersebut [8].

Berdasarkan uraian yang disebutkan, menunjukan bahwa permasalahan sistem klasifikasi penderita penyakit jantung menghasilkan akurasi cukup rendah, oleh sebab itu penelitian ini kali ini akan mengimplementasikan metode naïve bayes untuk klasifikasi penderita penyakit jantung menggunakan data yang diperoleh dari website University of California Irvine Machine learning data repository yang di unggah oleh David W dengan judul dataset yaitu *Heart Disease*. Dataset ini bisa digunakan sebagai informasi mengenai keadaan seseorang apakah mengidap penyakit jantung atau tidak. dengan menggunakan metode ini yang nantinya dijadikan bahan perbandingan untuk peneliti lain, sistem tersebut diharapkan bisa digunakan dalam memprediksi penyakit jantung berdasarkan gejala ataupun rekam medis yang dimiliki oleh pasien.

Machine learning adalah subdivisi yang muncul dari kecerdasan buatan yang merupakan sebuah teknologi yang dapat memproses data dalam volume besar yang digunakan oleh perusahaan untuk mengubah data mentah menjadi informasi yang berguna untuk membuat suatu keputusan bisnis yang sangat penting. Pada dasarnya Data Mining mempunyai 7

fungsi yaitu *Description, Classification, Clustering, Assosiation, Sequencing, Forecasting, dan Prediction*. Data mining memiliki tujuan sebagai Explanatory yaitu untuk menjelaskan beberapa kondisi terkait dengan suatu penelitian. Fokus utamanya adalah merancang sistem, memungkinkan mereka untuk belajar dan membuat prediksi berdasarkan pengalaman. Melatih algoritma pembelajaran mesin menggunakan kumpulan data pelatihan untuk membuat model [9].

Pengklasifikasi Naive Bayes adalah pengklasifikasi probabilistik berdasarkan teorema Bayes, yang mengasumsikan bahwa setiap fitur memberikan kontribusi yang independen dan setara ke kelas target. Mengasumsikan bahwa setiap fitur adalah independen dan tidak berinteraksi satu sama lain, sehingga setiap fitur secara independen dan sama-sama berkontribusi pada kemungkinan sampel untuk menjadi bagian dari kelas tertentu, Pengklasifikasi Naive Bayes mudah diimplementasikan dan secara komputasi cepat dan berkinerja baik pada kumpulan data besar yang memiliki dimensi tinggi. Kondusif untuk aplikasi waktu nyata dan tidak sensitif terhadap noise. Pengklasifikasi Naive Bayes memproses set data pelatihan untuk menghitung probabilitas kelas dan probabilitas bersyarat, yang menentukan frekuensi setiap nilai fitur untuk nilai kelas tertentu dibagi dengan frekuensi instance dengan nilai kelas tersebut. Pengklasifikasi Naive Bayes berkinerja terbaik ketika fitur yang berkorelasi dihapus karena fitur yang berkorelasi akan dipilih dua kali dalam model yang mengarah ke penekanan berlebihan pada pentingnya fitur yang berkorelasi [2].

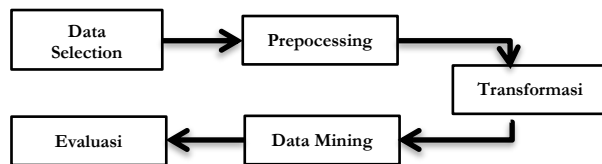
$$P(c|x) = \frac{P(x|c)P(c)}{P(x)}$$

$$P(c|X) = P(x_1|c) \times P(x_2|c) \times P(x_3|c) \times \dots \times P(x_n|c) \times P(c)$$

- 1) $P(c|x)$ adalah probabilitas posterior kelas (target) diberikan prediktor (atribut).
- 2) $P(c)$ adalah probabilitas kelas sebelumnya.
- 3) $P(x|c)$ adalah peluang yang merupakan peluang kelas yang diberikan oleh prediktor.
- 4) $P(x)$ adalah probabilitas sebelumnya dari predictor

2. Metode Penelitian

Penelitian yang dilakukan saat ini menggunakan metode eksperimen atau percobaan, dimana peneliti melakukan percobaan menggunakan data public yang telah disediakan oleh *University of California Irvine Machine learning data repository* dengan kata kunci *Heart Disease*. Dataset ini kemudian di uji menggunakan tools RapidMiner dengan metode klasifikasi Naïve Bayes. Peneliti menguji keakuratan hasil dari sistem yang tadinya sudah dilakukan dengan menghitung persentase dari hasil prediksi sistem yang telah dikerjakan. Langkah penelitian adalah proses untuk memperoleh, menerima, mendapatkan, atau mendiskusikan informasi tentang menyelesaikan masalah yang telah diidentifikasi secara ilmiah, sistematis dan logis. Langkah-langkah penelitian yang dilakukan oleh peneliti saat ini adalah sebagai berikut.



Gambar 1. Langkah-langkah penelitian

1) Data Selection

Pada tahap ini dilakukan seleksi data menggunakan parameter Filter Examples dengan tujuan untuk mereduksi dataset pasien penyakit jantung dan memilih fitur-fitur yang relevan dan yang tidak diperlukan terhadap diagnosis penyakit jantung.

2) Preprocessing

Data preprocessing dilakukan menggunakan parameter PCA untuk membersihkan data dari noise. Teknik data preprocessing dilakukan untuk mengatasi ukuran database yang terlalu besar adalah dengan cara membagi database menjadi beberapa bagian sehingga akan lebih mempercepat proses scanning data saat menerapkan algoritma data mining.

3) Transformasi

Transformasi ini dilakukan untuk mendeteksi keluaran dari prediksi seseorang dengan melihat apakah data tersebut sudah sesuai atau tidak dengan algoritma yang diterapkan. Data yang

berjenis numerikal akan dilakukan perubahan data menjadi bentuk kategorikal.

4) Klasifikasi Data Mining

Klasifikasi terhadap kumpulan dataset ini menggunakan Algoritma Naive Bayes pada Tools RapidMiner dalam penyelesaiannya.

5) Evaluasi

Tahapan Evaluasi langkah akhir yang dilakukan untuk melihat nilai akurasi yang dihasilkan dari metode yang digunakan.

Dataset yang digunakan

Dataset yang akan digunakan adalah data *Heart Disease* publik dari *University of California Irvine Machine learning* yang sudah tersebar di internet. Dataset ini berjumlah 303 data dengan 13 variabel. Untuk pengambilan data ini bisa di akses di internet menggunakan halaman website <http://archive.ics.uci.edu/ml/datasets/Heart+Disease>. Berikut adalah tabel yang berisikan kumpulan data penyakit jantung yang akan di uji.

Tabel 1. Daftar Fitur

No	Atribut	Deskripsi	Value
1	Age	Age In Years	Continuous
2	Sex	Male or Female	1 = Male 0 = Female 1 = Typical Type 1 2 = Typical Type
3	Cp	Chest Pain Type	Agina 3 = Non-Agina Pain 4 = Asymptomatic
4	Threstbps	Resting Blood Pressure	Continuous Value In Mm Hg
5	Chol	Cholesterol	Mm/Dl 0 = Norma 1 = Having ST_T
6	Restecg	Resting Electrographic Result	Wave abnormal 2 = Left Ventricular Hypertrophy
7	Fbs	Fasting Blood Sugar	1 ≥ 120 Mg/Dl 0 ≤ 120 Mg/Dl
8	Thalach	Maximum Heart	Continuous Value
9	Exang	Sugar Exercise-induced	0 = No

No	Atribut	Deskripsi	Value
			1 = Yes
10	Oldpeak	Agina ST Depression Induced Bt	Continuous Value
		Exercise Relative To Rest	
		The Slope Of The	
11	Slope	Peak Exercise ST	1 = Unsloping 2 = Flat 3 = Downsloping
		Segment Number Of	
		Major	0-3 Value
12	Ca	Vessels Colored By Florsopy	
13	Thal	Defect Type	3 = Normal 6 = Fixed 7 = Reversible defect

Berikut adalah penjelasan masing-masing fitur dari dataset penyakit jantung di atas

- 1) *Age*: Usia dalam tahun (numerik);
- 2) *Sex*: 1: Laki-laki, 0: Perempuan;
- 3) *Chest pain type* (CP): (a) typical angina (1: *angina*), (b) atypical angina (2: *abnang*), (c) nonanginal pain (3: *notang*), (d) asymptomatic (4: *asympt*). Pengertian secara medis: *Typical angina* adalah kondisi rekam medis pasien menunjukkan gejala umum nyeri dada sehingga kemungkinan memiliki penyumbatan arteri koroner yang tinggi sedangkan *Atypical angina* adalah kondisi dimana gejala pasien tidak rinci sehingga kemungkinan penyumbatan lebih rendah dan *Non-anginal pain* adalah rasa sakit yang menusuk seperti pisau atau kondisi menyakitkan yang dapat berlangsung dalam jangka waktu pendek atau panjang dan yang terakhir adalah *Asymptomatic pain* tidak menunjukkan gejala penyakit dan kemungkinan tidak akan menyebabkan atau menunjukkan gejala penyakit.
- 4) *Trestbps*: resting blood pressure pasien dalam mm Hg pada saat masuk rumah sakit (numerik);
- 5) *Chol*: Serum kolesterol dalam mg/dl (numerik);
- 6) *Fbs*: Ukuran boolean yang menunjukkan apakah fasting blood sugar lebih besar dari 120 mg/dl atau tidak: (1: True; 0: false).
- 6) *Restecg*: Hasil ECG selama istirahat : 0: normal, 1: abnormal (memiliki kelainan gelombang ST-T), 2: hipertrofi ventrikel.

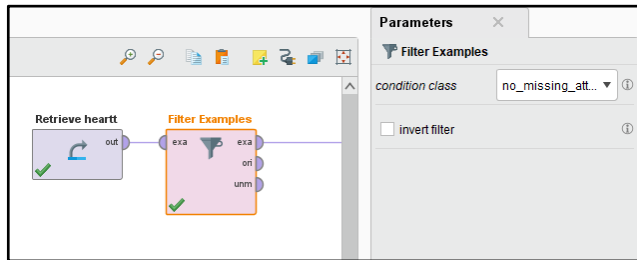
- 7) *Thalac*: detak jantung maksimum yang dicapai (numerik).
- 8) *Exang*: Ukuran boolean yang menunjukkan apakah latihan angina induksi telah terjadi (1: ya, 0: tidak).
- 9) *Oldpeak*: segmen ST yang diperoleh dari latihan relatif terhadap istirahat (numerik).
- 10) *Slope*: kemiringan segmen ST untuk latihan maksimum (puncak). Terdapat tiga jenis nilai yaitu 1: *upsloping*, 2: flat, 3: *downsloping*.
- 11) *Ca*: jumlah vessel utama yang diwarnai oleh *fluoroskopi* (numerik).
- 12) *Thal*: status jantung (3: normal, 6: cacat tetap, 7: cacat reversibel).

Deskripsi fitur diatas diperoleh dari jurnal penelitian orang lain yang memiliki kesamaan data dengan peneliti [10].

3. Hasil dan Pembahasan

Data Cleaning

Tahap awal yang harus dilakukan pada proses ini adalah Cleaning Data atau menghapus Missing Value, dataset awal yang berjumlah 303 record kemudian akan di proses apakah terdapat missing value atau tidak. Preprocessing Missing Value ini menggunakan Operator Filter Examples. Pengolahan data yang dilakukan adalah membersihkan data dengan cara mengecek apakah data yang digunakan memiliki informasi (fitur) yang hilang atau tidak atau biasa yang disebut dengan missing value (larassati, 2022). Operator ini menghilangkan record yang kosong ataupun tidak sesuai dengan atribut, setelah dilakukan proses ini maka akan menghasilkan sisa data yang lengkap dan sesuai untuk di proses selanjutnya. Tahapan pada proses missing value adalah menggunakan *Operator Filter Examples* dan kemudian memilih *No_Missing_Attribute* pada bagian *Parameter Condition*. Hasil dari proses *Filter Examples* menunjukkan bahwa tidak ada attribute yang terdapat Missing Values pada dataset penyakit jantung, hal ini menunjukkan semua atribut yang berjumlah 303 data penyakit jantung bisa digunakan tanpa harus membuangnya. Berikut adalah gambar proses Missing Value pada RapidMiner.



Gambar 2. Proses Missing Value

Row No.	target	tavg	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak
1	1	63	1	3	145	233	1	0	150	0	2.300
2	1	37	1	2	130	250	0	1	187	0	3.500
3	1	57	1	2	150	168	0	1	174	0	1.600
4	1	54	1	0	140	239	0	1	160	0	1.200
5	1	66	0	3	150	226	0	1	114	0	2.900
6	1	59	1	0	135	234	0	1	161	0	0.500
7	1	71	0	1	190	302	0	1	152	0	0.400
8	1	59	1	2	150	212	1	1	157	0	1.500
9	1	51	1	2	110	175	0	1	123	0	0.800
10	1	41	0	1	105	198	0	1	158	0	0

Gambar 5. Data Testing

chol	fbs	restecg	thalach	exang	oldpeak
233	1	0	150	0	2.300
250	0	1	187	0	3.500
204	0	0	172	0	1.600
236	0	1	178	0	1.200
354	0	1	163	1	2.900
192	0	1	148	0	0.500
294	0	0	153	0	0.400

Gambar 3. Hasil Missing Value

Pembagian Data

Pembagian data dilakukan menjadi 2 bagian yaitu 75% data training dan 25% data testing. Data training digunakan untuk melatih classifier dalam mengenali karakteristik pasien yang positif terkena jantung koroner maupun yang negatif. Data testing digunakan dalam uji coba terhadap model klasifikasi yang dihasilkan dan menentukan performa dari model klasifikasi dengan cara membandingkan hasil klasifikasi model terhadap tiap data dalam data testing dengan label sebenarnya [11]. Dataset yang digunakan untuk data tes dari jumlah total keseluruhan dataset yaitu 303 data, maka total data yang digunakan sebagai data latih adalah 228 record sedangkan untuk data testing sebanyak 75 record.

Row No.	target	tavg	sex	cp	trestbps	chol	fbs	restecg	thalach	exang	oldpeak
1	1	41	0	1	130	204	0	0	172	0	1.400
2	1	56	1	1	120	236	0	1	178	0	0.800
3	1	57	0	0	120	354	0	1	163	1	0.600
4	1	57	1	0	140	192	0	1	148	0	0.400
5	1	56	0	1	140	284	0	0	153	0	1.300
6	1	44	1	1	120	263	0	1	173	0	0
7	1	52	1	2	172	199	1	1	162	0	0.500
8	1	48	0	2	130	275	0	1	139	0	0.200
9	1	49	1	1	130	266	0	1	171	0	0.600
10	1	64	1	3	110	211	0	0	144	1	1.800
11	1	58	0	3	150	283	1	0	162	0	1
12	1	50	0	2	120	219	0	1	158	0	1.600

Gambar 4. Data Training

Pengujian

Berikut adalah persamaan dari teorema Bayes

$$P(H|X) = \frac{p(X|H) \cdot p(H)}{P(X)}$$

Keterangan :

- X : Data dengan class yang belum diketahui
- H : Hipotesis data X merupakan suatu class spesifik
- $P(H|X)$: Probabilitas hipotesis H berdasarkan kondisi X (posteriori probability)
- $P(H)$: Probabilitas hipotesis H (prior probability)
- $P(X|H)$: Probabilitas X berdasarkan kondisi pada hipotesis H
- $P(X)$: Probabilitas X

Selanjutnya adalah proses perhitungan probabilitas diagnosa penyakit jantung akan sebagai berikut.

Probability Healthy dengan Sakit (1)

$$\frac{\text{Healthy dengan hasil (1)}}{\text{Total Recor}} = \frac{124}{228} = 0,54385965$$

Probability Healthy dengan Sehat (0)

$$\frac{\text{Healthy dengan hasil (0)}}{\text{Total Recor}} = \frac{104}{228}$$

Dengan demikian maka diperoleh akurasi menggunakan Metode Naïve pada Tools RapidMiner adalah sebagai berikut:

Tabel 2. Hasil Perhitungan Naïve Bayes

Metode	Akurasi	Precision	Recall
Naïve bayes	86.84%	85.07%	89.36 %
			91.94 %
			80.77 %

Tabel 3. Perbandingan Hasil

Author	Data	Metode	Akurasi
hirwonob	UCI	Naïve Bayes	86.84%
(7)	UCI	Naïve Bayes	79.88%
(7)	UCI	Neural Network	84.52%

Tabel 4. Confusion Matrix

	True 1	True 0
Pred 1	114	20
Pred 0	10	84

Dari proses uji coba menggunakan Tools Rapid Miner diketahui bahwa hasil diagnosis dengan membandingkan hasil probability Sakit (1)=0,54385965 dan Tidak Sakit (0)=0,45614035 sehingga hasil diagnosis adalah Sakit (1) lebih besar daripada Tidak Sakit (0).

4. Kesimpulan

Penelitian yang dilakukan peneliti menggunakan data penyakit jantung sebanyak 303 record dengan 13 atribut yang diperoleh dari website UCI Machine Learning Repository kata kunci *Heart Disease* dengan menggunakan algoritma Naïve Bayes. Dengan membagi jumlah data menjadi 75:25 atau 228 data untuk training dan 75 data sebagai testing. Hasil penelitian ini mendapatkan nilai akurasi menggunakan Naïve Bayes tertinggi sebesar 86.84%. hal ini menunjukkan perbedaan yang signifikan dengan penelitian yang dilakukan oleh [7] menghasilkan akurasi sebesar 79.88%. hasil penelitian ini menunjukkan bahwa pengklasifikasian algoritma Naïve Bayes yang dilakukan oleh peneliti kali ini tampil baik dengan cukup baik.

5. Daftar Pustaka

- [1] Wibisono, A.B. and Fahrurrozi, A., 2020. Perbandingan Algoritma Klasifikasi Dalam Pengklasifikasian Data Penyakit Jantung Koroner. *jurnal ilmiah teknologi dan rekayasa*, 24(3), pp.161-170. DOI: <http://dx.doi.org/10.35760/tr.2019.v24i3.2393>.
- [2] Yudistira, N. and Putra, A.F., 2021. Algoritma Decision Tree Dan Smote Untuk Klasifikasi Serangan Jantung Miokarditis Yang Imbalance. *Jurnal Litbang Edusaintech*, 2(2), pp.112-122. DOI: <https://doi.org/10.51402/jle.v2i2.48>.
- [3] Alhamad, A., Azis, A.I., Santoso, B. and Taliki, S., 2019. Prediksi Penyakit Jantung Menggunakan Metode-Metode Machine Learning Berbasis Ensemble-Weighted Vote. *JEPIN (Jurnal Edukasi dan Penelitian Informatika)*, 5(3), pp.352-360. DOI: <http://dx.doi.org/10.26418/jp.v5i3.37188>.
- [4] Riani, A., Susianto, Y. and Rahman, N., 2019. Implementasi Data Mining Untuk Memprediksi Penyakit Jantung Menggunakan Metode Naive Bayes. *Journal of Innovation Information Technology and Application (JINITA)*, 1(01), pp.25-34.
- [5] Larassati, D., Zaidiah, A. and Afrizal, S., 2022. Sistem prediksi penyakit jantung koroner menggunakan metode naïve bayes. *JUPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, 7(2), pp.533-546. DOI: <https://doi.org/10.29100/jupi.v7i2.2842>.
- [6] Al Hafiz, D.M., Amaly, K., Jonathan, J. and Rachmatullah, M.T., 2021. Sistem Prediksi Penyakit Jantung Menggunakan Metode Naive Bayes. *Jurnal Rekayasa Elektro Srinwijaya*, 2(2), pp.151-157. DOI: <https://doi.org/10.36706/jres.v2i2.29>.

- [7] Nawawi, H.M., Purnama, J.J. and Hikmah, A.B., 2019. Komparasi Algoritma Neural Network Dan Naïve Bayes Untuk Memprediksi Penyakit Jantung. *Jurnal Pilar Nusa Mandiri*, 15(2), pp.189-194. DOI: <https://doi.org/10.33480/pilar.v15i2.669>.
- [8] 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), pp.176-185. DOI: <https://doi.org/10.51211/itbi.v6i2.1881>.
- [9] Bianto, M.A., Kusriani, K. and Sudarmawan, S., 2020. Perancangan Sistem Klasifikasi Penyakit Jantung Menggunakan Naïve Bayes. *Creative Information Technology Journal*, 6(1), pp.75-83. DOI: <https://doi.org/10.24076/citec.2019v6i1.231>.
- [10] Annisa, R., 2019. Analisis Komparasi Algoritma Klasifikasi Data Mining Untuk Prediksi Penderita Penyakit Jantung. *JTIK (Jurnal Teknik Informatika Kaputama)*, 3(1), pp.22-28.
- [11] Rasywir, E., Meisak, D., Pratama, Y. and Feranika, A., 2022. Penerapan Algoritma Naïve Bayes Untuk Mengetahui Pasien Penyakit Gagal Jantung. *Jurnal Informatika Dan Rekayasa Komputer (JAKAKOM)*, 2(2), pp.263-266. DOI: <https://doi.org/10.33998/jakakom.2022.2.2.117>.
- [12] Sahar, S., 2020. Analisis Perbandingan Metode K-Nearest Neighbor dan Naïve Bayes Classifier Pada Dataset Penderita Penyakit Jantung. *Indonesian Journal of Data and Science*, 1(3), pp.79-86. DOI: <https://doi.org/10.33096/ijodas.v1i3.20>.
- [13] Shah, D., Patel, S. and Bharti, S.K., 2020. Heart disease prediction using machine learning techniques. *SN Computer Science*, 1, pp.1-6. DOI: <https://doi.org/10.1007/s42979-020-00365-y>.
- [14] Munandar, T.A. and Munir, A.Q., 2022. Implementasi K-Nearest Neighbor Untuk Prototype Sistem Pakar Identifikasi Dini Penyakit Jantung. *Respati*, 17(2), pp.44-50. DOI: <https://doi.org/10.35842/jtir.v17i2.457>.
- [15] Partogi, Y. and Pasaribu, A., 2021. Perancangan Metode Decision Tree Terhadap Sistem Perpustakaan STMIK Kuwera. *Jurnal Sistem Informasi dan Teknologi (SINTEK)*, 1(2), pp.20-25. DOI: <https://doi.org/10.56995/sintek.v1i2.4>.