

PENERAPAN DATA MINING UNTUK KLASIFIKASI PENYAKIT JANTUNG KORONER MENGGUNAKAN ALGORITMA NAÏVE BAYES

Agus Fajar Riany^{1*)}, Gusmelia Testiana²

^{1,2}Program Studi Sistem Informasi, Fakultas Sains dan Teknologi, UIN Raden Fatah Palembang
¹agusfajarriany76@gmail.com, ²gusmeliatestiana_uin@radenfatah.ac.id

Kata kunci:

data mining; klasifikasi; naïve bayes; jantung koroner

Abstract: Coronary heart disease is one of the leading causes of death in all age groups after stroke. Coronary heart disease is a disease caused by plaque build-up in the coronary arteries that supply oxygen to the heart muscle. The importance of early detection of coronary heart disease symptoms is early prevention. Therefore, this study aims to analyze data related to coronary heart disease to determine whether the patient will suffer from coronary heart disease in the next ten years and to determine the performance of the Naïve Bayes algorithm. There are 16 attributes included in the causes of coronary heart disease including; gender, age, education, current smoker, cigs per day, BP Meds, prevalent stroke, prevalent hypertension, diabetes, total cholesterol, sys, BP, dia BP, BMI, heart rate, and glucose. The results showed that classification using the Naïve Bayes algorithm can be applied in classifying coronary heart disease data resulting in an accuracy of 79.10% in the fairly accurate category.

Abstrak: Penyakit jantung koroner merupakan salah satu penyakit penyebab utama kematian pada semua kelompok umur setelah stroke. Penyakit jantung koroner adalah penyakit yang disebabkan oleh plak yang menumpuk di arteri koroner yang memasok oksigen ke otot jantung. Pentingnya deteksi dini gejala penyakit jantung koroner merupakan pencegahan awal. Oleh karena itu, penelitian ini bertujuan untuk menganalisis data terkait penyakit jantung koroner untuk menentukan apakah pasien tersebut sepuluh tahun mendatang menderita penyakit jantung koroner dan untuk mengetahui performansi algoritma Naïve Bayes. Terdapat 16 atribut yang termasuk dalam penyebab penyakit jantung koroner diantaranya; gender, age, education, current smoker, cigs per day, BP Meds, prevalent stroke, prevalent hypertension, diabetes, total cholesterol, sys, BP, dia BP, BMI, heart rate, dan glucose. Hasil penelitian menunjukkan bahwa klasifikasi dengan menggunakan algoritma Naïve Bayes dapat diterapkan dalam mengklasifikasi data penyakit jantung koroner menghasilkan akurasi sebesar 79,10% dalam kategori cukup akurat.

Riany & Testiana. (2023). Penerapan Data Mining untuk Klasifikasi Penyakit Jantung Koroner Menggunakan Algoritma Naïve Bayes. *MDP Student Conference 2023*

PENDAHULUAN

Penyakit jantung koroner adalah penyakit jantung yang terjadi disebabkan oleh penyempitan pembuluh darah koroner. Terjadi gangguan pada arteri yang mengalirkan darah dari otot jantung sehingga jantung tidak dapat memompa darah untuk mencapai organ penting [1]. Penyakit ini termasuk dalam penyakit kardiovaskular. Penyakit kardiovaskular (CVDs) adalah sekelompok penyakit kardiovaskular, termasuk penyakit jantung koroner, penyakit serebrovaskular, penyakit jantung rematik dan penyakit lainnya, yang merupakan penyebab utama kematian di seluruh dunia [2].

Menurut Organisasi Kesehatan Dunia (WHO), tidak kurang dari 17,9 juta orang meninggal akibat penyakit kardiovaskular pada tahun 2019, atau 32% dari seluruh kematian global. 85 % dari kematian ini disebabkan oleh serangan jantung dan stroke. Kemudian, pada tahun 2019, 17 juta kematian dini (dibawah usia 70 tahun) disebabkan oleh penyakit tidak menular (PTM), 38% di antaranya disebabkan oleh penyakit kardiovaskular [2]. Penyakit jantung koroner merupakan salah satu penyakit mematikan di Indonesia. Menurut Hasil survei *Sample Registration System* (SYS) pada tahun 2014, penyakit jantung koroner sebesar 12,9% merupakan penyakit yang menyebabkan kematian tertinggi setelah stroke pada semua umur [3].

Penyakit jantung koroner merupakan masalah kesehatan dapat memberikan dampak pada faktor sosial-ekonomi, karena perawatan dan pengobatan yang membutuhkan biaya yang mahal sehingga diperlukan upaya pencegahan dan pengendalian sejak dini sangat diperlukan. Beberapa faktor yang menyebabkan penyakit jantung koroner yaitu: usia, jenis kelamin, hipertensi, obesitas, diabetes mellitus, stroke, merokok, penyakit arteri perifer, letak geografis, makanan tinggi lemak dan alkohol, hipertensi dan dislipidemia [4]. Masalah stroke di Indoneasia memerlukan perhatian serius karena jumlah kasus terus meningkat dan angka kematian yang tinggi.

Alternatif solusi yang dilakukan pada penelitian ini adalah memanfaatkan data penyakit jantung koroner untuk memeriksa informasi baru, atau informasi yang dikandungnya, sehingga dapat digunakan untuk membuat model baru untuk mengklasifikasikan pasien untuk memprediksi apakah pasien menderita penyakit jantung koroner sepuluh tahun mendatang. Mengekstraksi informasi baru dari kumpulan data disebut data mining. Metode data mining yang digunakan dalam penelitian ini adalah klasifikasi. Metode klasifikasi adalah metode data mining dengan mengklasifikasikan kelas target secara akurat. Tujuan dari metode ini adalah untuk mengklasifikasikan nilai suatu variabel yang tidak diketahui dari variabel lain yang telah ditentukan sebelumnya [5].

Beberapa algoritma dapat digunakan untuk klasifikasi seperti *Decision Tree*, *Naïve Bayes*, *K-Nearest Neighbor*, dan lain-lain. Penelitian ini menggunakan algoritma *Naïve Bayes* untuk mengklasifikasikan penyakit jantung koroner. Algoritma *Naïve Bayes* adalah klasifikasi menggunakan probabilitas sederhana dan metode statistik berdasarkan teorema Bayes dengan asumsi independensi yang kuat [6]. *Naïve Bayes* telah terbukti memiliki keakurasian dan kecepatan yang tinggi ketika diterapkan pada *database* dengan jumlah data yang besar [7].

Terdapat penelitian terdahulu yang berkaitan dengan penerapan algoritma *Naïve Bayes* dan pengklasifikasian penyakit jantung koroner dengan menggunakan data mining. Sebuah penelitian yang bertujuan untuk menentukan apakah algoritma *Naïve Bayes* dapat memprediksi penyakit jantung dan mendapatkan wawasan tentang kinerja data uji menggunakan algoritma *Naïve Bayes* menunjukkan bahwa dari pembagian data 80:20 memberikan akurasi sebesar 70,00% [8]. Penelitian selanjutnya untuk membuat sistem yang mampu memberikan informasi dan mengklasifikasikan informasi atau gejala awal serangan jantung menunjukkan kinerja algoritma *Naïve Bayes* menghasilkan akurasi sebesar 90,61 [9]. Penelitian selanjutnya yang bertujuan untuk mendeteksi penyakit jantung menggunakan algoritma *Naïve Bayes* dan PSO menemukan akurasi sebesar 82,14% untuk algoritma *Naïve Bayes*, sedangkan algoritma PSO memiliki akurasi sebesar 92,86% [10].

Berdasarkan gambaran latar belakang masalah di atas, penelitian ini bertujuan untuk mengklasifikasikan pasien, apakah pasien tersebut sepuluh tahun mendatang menderita penyakit jantung koroner atau tidak menggunakan algoritma *Naïve Bayes* dengan menghitung probabilitas masing-masing atribut. Tujuan selanjutnya adalah untuk mengetahui akurasi dari algoritma *Naïve Bayes*.

METODE

Metode penelitian adalah metode yang digunakan untuk mengumpulkan data guna mencapai suatu tujuan. Metode penelitian kuantitatif adalah metode penelitian yang menerapkan filosofi positivisme, dimana populasi atau sampel tertentu dipelajari dengan cara mengumpulkan data dengan alat penelitian, analisis datanya bersifat kuantitatif dengan bertujuan untuk menguji hipotesis yang telah ditetapkan [11]. Tujuan

metode penelitian kuantitatif adalah untuk mengidentifikasi masalah yang ditetapkan sebagai tujuan penelitian. Pendekatan deskriptif merupakan salah satu cara untuk mengetahui keberadaan variabel bebas. Penelitian ini menggunakan metode kuantitatif dengan menggunakan data maning dan perhitungan algoritma *Naïve Bayes*.

Metode Pengumpulan Data

Pengumpulan data dilakukan untuk mendukung penelitian ini. Dalam penelitian ini metode pengumpulan data untuk memperoleh sumber data adalah dengan menggunakan metode data sekunder. Data Sekunder adalah informasi berdasarkan penelitian sebelumnya, yaitu peneliti tidak harus mengambil bahan ke lapangan, tetapi sudah dikumpulkan dan diolah oleh peneliti lain [12]. Sumber data yang digunakan dalam penelitian ini yaitu data dari *Kaggle*. Jumlah sampel yang digunakan dalam penelitian ini adalah sebanyak 4.133 data. Data ini terdiri dari 16 atribut, dimana 15 merupakan atribut dan 1 merupakan label atau target.

Algoritma Naïve Bayes

Salah satu algoritma yang termasuk dalam metode klasifikasi adalah algoritma Naïve Bayes [13]. Algoritma *Naïve Bayes* adalah *classifier* yang merepresentasikan setiap kelas objek berdasarkan infrensi atau ringkasan probabilistik dan menemukan kelas yang paling mungkin sesuai untuk setiap objek yang kelasnya ditentukan dari objek uji yang ada berdasarkan atribut atau variabel yang nilainya diketahui [14]. Naïve Bayes merupakan algoritma klasifikasi sederhana dimana setiap atribut bersifat independen dan dapat mempengaruhi keputusan akhir [15]. Algoritma *Naïve Bayes* dapat digunakan untuk memprediksi probabilitas milik kelas. Algoritma Naïve Bayes sangat cocok untuk klasifikasi data bertipe nominal [16]. *Naïve bayes* telah terbukti memiliki keakurasian dan kecepatan yang tinggi ketika diterapkan pada *database* dengan jumlah data yang besar [7]. Teorema Bayes memiliki bentuk umum sebagai berikut:

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)} \quad (1)$$

Dalam hal ini, X adalah data dari kelas yang tidak diketahui, H adalah hipotesis bahwa data X adalah kelas tertentu, P(H|X) adalah probabilitas dari hipotesis H berdasarkan kondisi X (*posteriori probability*), P(H) adalah probabilitas hipotesis H (*piror probability*), P(H|X) adalah probabilitas X berdasarkan kondisi hipotesis H dan P(X) adalah probabilitas dari X.

HASIL DAN PEMBAHASAN

Pengolahan data mining merupakan hasil utama dari penelitian ini yang hasilnya dibahas pada bagian hasil dan pembahasan. Proses pengolahan data mining dilakukan sesuai langkah-langkah *Cross Industry Standard Process for Data Mining* (CRISP-DM). Berikut adalah penjelasan dari setiap langkah pemrosesan data mining yang dilakukan.

Pemahaman Penelitian (Research Understanding)

Proses data mining dimuali dengan pemahaman penelitian (*research understanding*) yang tujuannya adalah untuk memahami masalah dan menentukan tujuan dari data mining. Seperti permasalahan yang telah dijelaskan pada latar belakang, penerapan data mining pada penelitian ini bertujuan untuk mengetahui hasil klasifikasi penyakit jantung koroner, untuk mengetahui apakah pasien tersebut sepuluh tahun mendatang menderita jantung koroner atau tidak menggunakan algoritma *Naïve Bayes* dengan melakukan perhitungan probabilitas dan untuk mengetahui keakurasian dari model yang dilakukan.

Pemahaman Data (Data Understanding)

Tahapan selanjutnya setelah pemahaman penelitian yaitu pemahaman data (*data understanding*). Tahap ini bertujuan untuk memahami karakteristik data yang diproses. Sumber data yang digunakan dalam

penelitian ini yaitu data dari *Kaggle*. Data penyakit jantung koroner dari *kaggle* yang digunakan sebanyak 4.133 data. Data ini terdiri dari 16 atribut, 15 diantaranya merupakan atribut dan 1 merupakan label atau target.

Persiapan Data (Data Preparation)

Dataset penyakit jantung koroner memiliki beberapa atribut yang teridentifikasi sebagai *missing value*, sehingga data tersebut harus di proses terlebih dahulu dengan menghilangkan *missing value*. Proses membersihkan data disebut dengan *preprocessing data*. Data yang diperoleh dari dataset *Kaggle* terdiri dari 4.133 data, dari data tersebut diantaranya 18 data teridentifikasi sebagai data *missing value* pada atribut BMI. Kemudian data *missing value* dihilangkan untuk memvalidasi data sehingga data yang diproses tanpa *missing value* adalah 4.115 data.

Setelah data *missing value* dibersihkan, selanjutnya dilakukan proses transformasi pada dataset awal. Proses transformasi dilakukan untuk mengubah nilai atribut dari tipe nominal menjadi numerik. Seluruh atribut ditransformasikan, atribut-atribut tersebut antara lain *gender, age, education, current smoker, cigs per day, BP Meds, prevalent stroke, prevalent hypertension, diabetes, total cholesterol, sys, BP, dia BP, BMI, heart rate, glucose* dan *ten year CHD*.

Tabel 1. Dataset Hasil Preparation Data

<i>Gender</i>	<i>Age</i>	<i>Education</i>	<i>Current Smoker</i>	<i>Cigs Per Day</i>	...	<i>Heart Rate</i>	<i>Glucose</i>	<i>Stroke</i>
<i>Male</i>	Dewasa Akhir	<i>College Degree and Higher</i>	<i>No</i>	Tidak Merokok	...	Normal	Normal	<i>No</i>
<i>Female</i>	Lansia Awal	<i>Less than High School and High School degrees</i>	<i>No</i>	Tidak Merokok	...	Normal	Normal	<i>No</i>
<i>Male</i>	Lansia Awal	<i>Less than High School and High School degrees</i>	<i>Yes</i>	Perokok Berat	...	Bradikarbi	Normal	<i>No</i>
<i>Female</i>	Lansia Akhir	<i>College Degree and Higher</i>	<i>Yes</i>	Perokok Berat	...	Bradikarbi	Prediabetes	<i>Yes</i>
<i>Female</i>	Lansia Awal	<i>Less than High School and High School degrees</i>	<i>Yes</i>	Perokok Berat	...	Normal	Normal	<i>No</i>

Selanjutnya, setelah proses transformasi, dataset dibagi menjadi dua bagian, yaitu data latih dan data uji. Proses pembagian data dilakukan dengan *Split Train Test*, yaitu membagi data menjadi data latih (*training*) dan data uji (*testing*). Persentase pembagian data adalah 80% data latih dan 20% data uji. Membagi kumpulan data dengan rasio 80:20 dapat memberikan kinerja model terbaik [17]. Jadi, pada penelitian ini dengan dataset sebanyak 4.115, terdapat 3.292 data latih dan 823 data uji.

Pemodelan (Modeling)

Tahap ini merupakan tahapan implementasi dari teknologi dan algoritma yang dipilih. Pada penelitian ini klasifikasi diterapkan menggunakan algoritma terpilih yaitu *Naïve Bayes*. Jumlah sampel pasien penyakit jantung koroner yang dijadikan sampel untuk proses perhitungan algoritma *Naïve Bayes* untuk mengetahui pasien tersebut terklasifikasi menderita penyakit jantung koroner atau tidak yang dilihat berdasarkan probabilitas dari setiap kelas dan probabilitas setiap kejadian perkelas dan keakurasian yang dihasilkan.

Menghitung Probabilitas Setiap Kelas

Dari data yang digunakan terdapat 2 class yaitu *Yes* dan *No*. *Class Yes* (Penyakit Jantung Koroner) berjumlah 490 data, sedangkan jumlah *class No* (Tidak Menderita Jantung Koroner) berjumlah 3.291 data. Untuk menghitung probabilitas tiap kelas, dilakukan perhitungan dengan membagi jumlah tiap kelas dibagi dengan seluruh data.

Tabel 2. Probabilitas Setiap Kelas

Kelas	Probabilitas Kelas P(H)
<i>Yes</i>	0,149
<i>No</i>	0,851

Menghitung Probabilitas Setiap Kejadian Perkelas

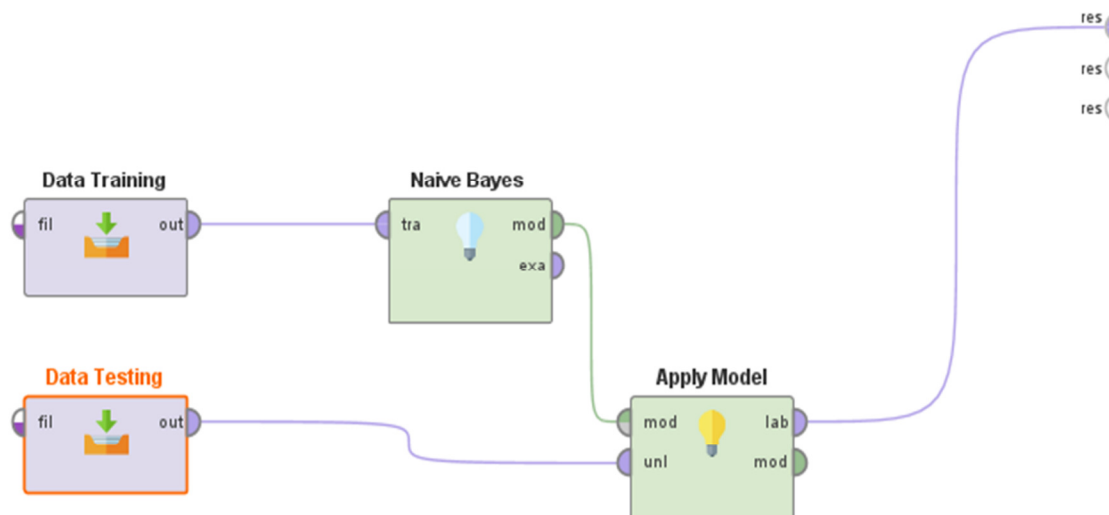
Hitung nilai probabilitas dari setiap atribut dengan cara menghitung jumlah kejadian atau atribut dalam suatu kategori dibagi dengan kategori yang ada. Nilai Probabilitas untuk setiap atribut adalah sebagai berikut:

Tabel 3. Probabilitas Setiap Kejadian Perkelas

No	Atribut	Sub Atribut	Probabilitas P(XIH)	
			YES	NO
1	<i>Gender</i>	<i>Male</i>	0,524	0,404
		<i>Female</i>	0,476	0,596
2	<i>Age</i>	<i>Dewasa Awal</i>	0,004	0,017
		<i>Dewasa Akhir</i>	0,178	0,392
		<i>Lansia Awal</i>	0,353	0,349
		<i>Lansia Akhir</i>	0,402	0,221
		<i>Manula</i>	0,063	0,020
3	<i>Education</i>	<i>Less than High School and High School</i>	0,739	0,714
		<i>Collage Degree and Higher</i>	0,261	0,286
4	<i>Current Smoker</i>	<i>Yes</i>	0,520	0,493
		<i>No</i>	0,480	0,507
5	<i>Cigs Per Day</i>	<i>Tidak Merokok</i>	0,486	0,524
		<i>Perokok Ringan</i>	0,104	0,140
		<i>Perokok Sedang</i>	0,063	0,057
		<i>Perokok Berat</i>	0,347	0,279
6	<i>BP Meds</i>	<i>Yes</i>	0,069	0,025
		<i>No</i>	0,931	0,975
7	<i>Prevalent Stroke</i>	<i>Yes</i>	0,014	0,003
		<i>No</i>	0,986	0,997
8	<i>Prevalent Hypertension</i>	<i>Yes</i>	0,506	0,270
		<i>No</i>	0,494	0,730
9	<i>Diabetes</i>	<i>Yes</i>	0,061	0,016
		<i>No</i>	0,939	0,984
10	<i>Total Cholesterol</i>	<i>Normal</i>	0,141	0,209
		<i>Batas Tinggi</i>	0,359	0,363
		<i>Tinggi</i>	0,500	0,428
11	<i>Sys BP</i>	<i>Normal</i>	0,357	0,562

		<i>High Normal</i>	0,153	0,178
		<i>Mild Hypertension</i>	0,265	0,168
		<i>Moderate Hypertension</i>	0,131	0,061
		<i>Hypertensive Emergence</i>	0,094	0,031
12	<i>Dia BP</i>	<i>Normal</i>	0,465	0,622
		<i>High Normal</i>	0,143	0,256
		<i>Mild Hypertension</i>	0,218	0,247
		<i>Moderate Hypertension</i>	0,100	0,052
		<i>Hypertensive Emergence</i>	0,073	0,023
13	BMI	<i>Underweight</i>	0,014	0,015
		<i>Normal</i>	0,367	0,465
		<i>Overweight</i>	0,467	0,403
		<i>Obese</i>	0,114	0,091
		<i>Extremely Obese</i>	0,037	0,026
14	<i>Heart Rate</i>	Bradikarbi	0,618	0,634
		Normal	0,353	0,343
		Takikarbi	0,029	0,024
15	<i>Glucose</i>	Normal	0,865	0,929
		Prediabetes	0,071	0,058
		Diabetes	0,063	0,013

Selanjutnya, dari hasil nilai probabilitas dari setiap atribut di atas dapat digunakan untuk mengklasifikasi atau menentukan label data *testing*. Pengujian dilakukan dengan menggunakan *tools* Rapid Miner. Berikut ini pemodelan di Rapid Miner untuk mempresentasikan hasil klasifikasi pada data *testing* yaitu:



Gambar 1. Pemodelan di Rapid Miner

Berdasarkan model yang dibentuk pada Rapid Miner, model di atas mampu menghasilkan klasifikasi dari semua data uji (*testing*) yang mengklasifikasi apakah seorang pasien akan menderita penyakit jantung koroner dalam sepuluh tahun mendatang atau tidak. Berikut adalah hasil klasifikasi yang dihasilkan:

Open in  Turbo Prep  Auto Model Filter (823 / 823 examples):

Row No.	Ten Year CHD	prediction(T...	confidence(...	confidence(...	Gender	Age	Education	Current Sm...	Cigs Per Day	BP Meds	Prevalent St...
1	?	No	0.954	0.046	Female	Lansia Awal	Less than Hi...	No	Tidak Merokok	No	No
2	?	No	0.900	0.100	Male	Lansia Awal	Less than Hi...	Yes	Perokok Berat	No	No
3	?	Yes	0.413	0.587	Male	Lansia Awal	Less than Hi...	Yes	Perokok Sed...	No	No
4	?	No	0.968	0.032	Male	Dewasa Akhir	Less than Hi...	No	Tidak Merokok	No	No
5	?	No	0.988	0.012	Female	Dewasa Akhir	Less than Hi...	Yes	Perokok Berat	No	No
6	?	No	0.930	0.070	Male	Lansia Awal	Less than Hi...	Yes	Perokok Berat	No	No
7	?	No	0.686	0.314	Male	Lansia Awal	Less than Hi...	Yes	Perokok Ring...	No	No
8	?	No	0.867	0.133	Male	Lansia Akhir	Less than Hi...	No	Tidak Merokok	No	No
9	?	No	0.591	0.409	Female	Lansia Akhir	Less than Hi...	No	Tidak Merokok	No	No
10	?	No	0.956	0.044	Female	Lansia Awal	Less than Hi...	Yes	Perokok Berat	No	No
11	?	No	0.985	0.015	Male	Dewasa Akhir	College Degr...	Yes	Perokok Ring...	No	No
12	?	No	0.964	0.036	Female	Dewasa Akhir	Less than Hi...	No	Tidak Merokok	No	No
13	?	No	0.788	0.212	Female	Dewasa Akhir	Less than Hi...	No	Tidak Merokok	No	No
14	?	Yes	0.307	0.693	Male	Lansia Akhir	Less than Hi...	Yes	Perokok Berat	No	No

Gambar 2. Prediksi Klasifikasi Setiap Kelompok Penyakit Jantung Koroner

Evaluasi (Evaluation)

Tahap evaluasi ialah tahap yang dilakukan untuk memberi penilaian terhadap kinerja prosedur pemecahan masalah klasifikasi yang dipergunakan pada penelitian [18]. Proses evaluasi model yang digunakan pada penelitian ini menggunakan fungsi *confusion matrix* dengan bantuan *software* rapid miner. Gambar 3 berikut ini adalah hasil evaluasi model yang telah dilakukan yaitu:

accuracy: 79.10%

	true No	true Yes	class precision
pred. No	609	86	87.63%
pred. Yes	86	42	32.81%
class recall	87.63%	32.81%	

Gambar 3. Hasil Evaluasi dari Algoritma Naïve Bayes

Hasil evaluasi menunjukkan bahwa akurasi dari algoritma *Naïve Bayes* sebesar 79,10%. Artinya akurasi algoritma *Naïve bayes* dalam menentukan atau mengklasifikasi pasien penyakit jantung koroner untuk mengetahui apakah sepuluh tahun mendatang akan menderita jantung koroner atau tidak, pada kategori Cukup Akurat.

SIMPULAN

Berdasarkan hasil rangkaian tahapan penelitian yang dilakukan, dapat disimpulkan bahwa proses data mining menggunakan algoritma *Naïve Bayes* dengan teknik klasifikasi dapat diterapkan dalam pengelompokan pasien untuk mengetahui apakah pasien tersebut sepuluh tahun mendatang akan menderita penyakit jantung koroner atau tidak, dimana hasil akurasi yang diperoleh sebesar 79,10% yang berada dalam kategori cukup akurat.

DAFTAR PUSTAKA

- [1] J. J. Pangaribuan, C. Tedja, and S. Wibowo, "Perbandingan Metode Algoritma C4.5 dan Extreme Learning Machine Untuk Mendiagnosis Penyakit Jantung Koroner," *Informatics Eng. Res. Technol.*, Vol. 1, No. 1, pp. 1–7, 2019.
- [2] W. H. O. (WHO), "Cardiovascular Diseases," 2019.
- [3] Kemenkes RI, "Penyakit Jantung Penyebab Kematian Tertinggi, Kemenkes Ingatkan CERDIK," 2017.
- [4] L. Ghani, M. D. Susilawati, and H. Novriani, "Faktor Risiko Dominan Penyakit Jantung Koroner di Indonesia," *Bul. Penelit. Kesehat.*, Vol. 44, No. 3, pp. 153–164, 2016, doi: 10.22435/bpk.v44i3.5436.153-164.
- [5] F. F. Nugraha, I. Sunandar, and C. Juliane, "Penerapan Data Mining Dengan Metode Klasifikasi Menggunakan Algoritma C4.5," *J. Teknol. Inform. dan Sist. Inf.*, Vol. 9, No. 4, pp. 2862–2869, 2022, [Online]. Available: <https://jurnal.mdp.ac.id/index.php/jatisi/article/view/2518/1006>.
- [6] E. Prasetyo, *Data Mining: Konsep dan Aplikasi Menggunakan Mtlab*. Yogyakarta: Andi Offset, 2014.
- [7] E. T. Luthfi and Kusriani, *Algoritma Data Mining*. Yogyakarta: Andi Offset, 2009.
- [8] D. Maulana and R. Yahya, "Implementasi Algoritma Naive Bayes untuk Klasifikasi Penderita Penyakit Jantung di Indonesia Menggunakan Rapid Miner," *J. Teknol. Pelita Bangsa*, Vol. 10, No. 2, pp. 2407–3903, 2019, doi: 10.1134/s0320972519100129.
- [9] M. A. Bianto, K. Kusriani, and S. Sudarmawan, "Perancangan Sistem Klasifikasi Penyakit Jantung Menggunakan Naive Bayes," *Creat. Inf. Technol. J.*, Vol. 6, No. 1, p. 75, 2020, doi: 10.24076/citec.2019v6i1.231.
- [10] N. A. Widiastuti, S. Santosa, and C. Supriyanto, "Algoritma Klasifikasi Data Mining Naive Bayes Berbasis Particle Swarm Optimization untuk Deteksi Penyakit Jantung," *Nat. Pseudocode*, Vol. 1, No. 1, pp. 2355–5920, 2014, doi: 10.1038/nmeth.f.284.
- [11] Sugiyono, *Metode Penelitian Kuantitatif, Kualitatif dan R&D*. Bandung: Alfabeta, 2017.
- [12] Suryani and Hendryadi, *Metode Riset Kuantitatif: Teori dan Aplikasi pada Penelitian Bidang Manajemen dan Ekonomi Islam*, 2nd ed. Jakarta: Prenadamedia Group, 2018.

- [13] J. C. Aponno, “Penerapan Algoritma Sentimen Analysis dan Naïve Bayes Terhadap Opini Pengunjung di Tempat Wisata Pantai Pintu Kota, Kota Ambon,” *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, Vol. 9, No. 4, pp. 3180–3188, 2022, doi: 10.35957/jatisi.v9i4.2697.
- [14] M. Arhami and M. Nasir, *Data Mining: Algoritma dan Implementasi*. Yogyakarta: Andi Offset, 2020.
- [15] G. Testiana, “Perancangan Model Prediksi Kelulusan Mahasiswa Tepat Waktu pada UIN Raden Fatah,” *JUSIFO (Jurnal Sist. Informasi)*, Vol. 4, No. 1, pp. 49–62, 2018, [Online]. Available: <http://jurnal.radenfatah.ac.id/index.php/jusifo/article/view/1932>.
- [16] J. Suntoro, *Data Mining: Algoritma dan Implementasi Dengan Pemograman PHP*. Jakarta: PT ELex Media Komputindo, 2019.
- [17] H. Bugis, “Metode Naïve Bayes Untuk Memprediksi Penyakit Stroke,” *J. SISKOM-KB (Sistem Komput. dan Kecerdasan ...)*, 2022, [Online]. Available: <https://doi.org/10.47970/siskom-kb.v6i1.317>.
- [18] A. I. Tanggraeni and M. N. N. Sitokdana, “Analisis Sentimen Aplikasi E-Government pada Google Play Menggunakan Algoritma Naïve Bayes,” *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, Vol. 9, No. 2, pp. 785–795, 2022, doi: 10.35957/jatisi.v9i2.1835.