
PENERAPAN METODE NAÏVE BAYES DALAM MEMPREDIKSI PENYAKIT JANTUNG

Atthohiroh^{1*}, Rama Ayu², Sri Maharani³

^{1,2,3}Sistem Informasi, STMIK Royal Kisaran

email: *atthohiroh6@gmail.com

Abstract: *The heart is one of the human organs that has an important function to circulate blood throughout the body. In caring for the human heart, one must know how to take care of the heart and check the health conditions of the heart regularly and do it early in order to avoid dangerous diseases that attack the heart. In this study, a heart disease diagnostic examination was carried out using machine learning with the Naïve Bayes method. The criteria used in predicting heart disease with Naïve Bayes in this study were gender, BP (pneumonia), Cholesterol (cholesterol), and Maxhr. The purpose of using Naive Bayes in this study is to predict using simple probabilities based on the application of Bayes' rules. In this study, Naive Bayes produced a Precision value for heart disease of 87% and for those without 75%, a recall value for those with heart disease 81% and those without heart disease 82%, support values for those with heart disease 84% and those who do not. 78%.*

Keywords: *machine learning, naïve bayes, jantung*

Abstrak: Jantung merupakan salah satu organ manusia yang memiliki fungsi penting untuk mengalirkan darah keseluruh tubuh. Dalam merawat jantung manusia harus mengetahui cara menjaga jantung dan memeriksa kondisi kesehatan jantung secara berkala dan dilakukan sejak dini guna menghindari penyakit berbahaya menyerang jantung. Pada penelitian ini dilakukan pemeriksaan diagnosa penyakit jantung menggunakan *machine learning* dengan metode *Naïve Bayes*. Kriteria yang digunakan dalam memprediksi penyakit jantung dengan *Naïve Bayes* dalam penelitian ini yaitu gender(jenis kelamin), BP(radang paru-paru), Cholesterol(kolestrol), dan Maxhr. Tujuannya penggunaan Naive bayes dalam penelitian ini untuk memprediksi menggunakan probabilitas sederhana berdasarkan pada penerapan aturan bayes. Pada penelitian ini *naive Bayes* menghasilkan nilai *Precision* untuk penyakit jantung 87% dan untuk yang tidak 75%, nilai *recall* untuk yang terkena penyakit jantung 81% dan yang tidak terkena penyakit jantung 82%, nilai *support* untuk yang terkena penyakit jantung 84% dan yang tidak 78% .

Kata Kunci: machine learning, naïve bayes, jantung

PENDAHULUAN

Kemajuan teknologi saat ini sangat berkembang pesat, kemajuan teknologi akan berjalan dengan kemajuan ilmu pengetahuan yang ada. Teknologi memberikan banyak kemudahan untuk segala aktivitas masyarakat terkhusus dalam bidang kesehatan teknologi sudah sangat maju dan berkembang. Salah satu penerapan teknologi yakni untuk mengatasi penyakit jantung.

Jantung merupakan salah satu organ manusia yang sangat penting yang berfungsi untuk mengalirkan darah keseluruh tubuh. Data Badan Kesehatan Dunia (WHO) pada 2016, total kematian di Indonesia mencapai 1.863.000 jiwa. Sebanyak 73 persen diantaranya disebabkan penyakit tidak menular. Prevalensi penyakit jantung tertinggi berusia 75 tahun ke atas yaitu 4,7 persen. Data Rinkesdas 2018, angka prevalensi penyakit jantung di Indonesia dimulai umur dari 1 tahun dan tersebar disemua kalangan. Semakin tinggi usia seseorang, maka resiko terkena penyakit jantung akan semakin tinggi.

Dalam kasus ini banyak kasus kematian yang diakibatkan oleh penyakit jantung dimana penyebabnya pola hidup yang tidak sehat serta makan berkolestrol. Penyakit jantung ini lebih diderita oleh laki-laki akibat penggunaan alkohol, tembakau dan lainnya. Pasien yang terkena penyakit banyak mengalami penyakit jantung. Besarnya jumlah data penyakit jantung ini, tentunya para dokter akan memiliki banyak waktu untuk memeriksa pasien yang terkena penyakit jantung.

Dalam hal ini masyarakat masih kurang pengetahuan terhadap gejala atau tanda-tanda saat seseorang mengidap penyakit jantung. Apalagi

para masyarakat banyak mengkomsumsi makanan berkolestrol.

Naive Bayes merupakan pengklasifikasian dengan metode probabilitas yang ditemukan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya sehingga dikenal sebagai teorema Bayes [1]. Teorema tersebut dikombinasikan dengan naive di mana diasumsikan kondisi antar petunjuk (atribut) saling bebas. Naive bayes adalah teknik memprediksi berbasis probabilitas sederhana yang berdasarkan pada penerapan aturan bayes dengan asumsi ketidaktergantungan yang kuat. Selain itu, naive bayes juga dapat menganalisis variabel-variabel yang paling mempengaruhinya dalam bentuk peluang [2].

Salah satu pengaplikasian dari Naive Bayes yaitu pada bidang kesehatan. Kesehatan adalah keadaan sejahtera dari badan, jiwa, dan sosial yang memungkinkan setiap orang hidup produktif secara sosial dan ekonomi. Keadaan tubuh dalam kondisi sakit membuat apapun yang ingin dilakukan menjadi tidak maksimal atau tidak nyaman. Untuk itu menjaga kesehatan adalah salah satu hal yang penting untuk selalu diupayakan.

Oleh karena itu, penelitian ini bertujuan untuk mendapatkan hasil prediksi dan mengetahui pengukuran manakah yang akan menghasilkan nilai yang lebih akurat dengan menggunakan metode Naive Bayes. Nantinya hasil yang paling bagus akan diintegrasikan dengan sebuah *deploy* sehingga pengguna dapat lebih memahami dalam menggunakannya.

METODE

Metode riset memuat hal penting terkait pelaksanaan penelitian dan tahapan yang akan ditempuh dalam penelitian untuk mencakup hasil yang diinginkan. Berikut ini hal yang mencakup ke dalam metode riset antara lain :

1. Peralatan Riset

Peralatan yang digunakan dalam penelitian ini adalah berupa komputer dengan spesifikasi yang mendukung, *software* yang akan digunakan dalam proses pengolahan berupa *software Jupyter Notebook* dalam *Anaconda Navigator* dimana bahasa pemrograman *Python*, *Visual Studio Code* sebagai *teks editor* dan menggunakan *HTML* sebagai bahasa *markup* yang digunakan untuk membangun *deploy* aplikasinya.

2. Variabel yang Terdapat dalam Riset

Dalam *machine learning*, penelitian ini dikategorikan sebagai penelitian *supervised learning* yaitu sebuah model dimana data yang akan diproses memiliki label/target/class dengan tujuan untuk mengetahui hubungan kualitas antara variabel bebas dan variabel yang menjadi target/labelnya [3]. Oleh sebab itu, variabel yang terdapat dalam penelitian ini terdiri dari variabel independen atau variabel X meliputi *Gender* (jenis kelamin), *BP* (radang paru-paru), *Cholestrol* (kolestrol), *Max HR* (detak jantung). Serta yang menjadi variabel Y adalah *Heart Disease* (penyakit jantung).

3. Tahapan – Tahapan Riset

Tahapan yang akan ditempuh dalam penelitian ini adalah sebagai berikut ini [4]:

Data Preprocessing yakni dalam proses tersebut terdapat *data cleaning* untuk melakukan *handling missing* pada data.

Data Selection, untuk menyeleksi data berdasarkan kolom yang akan diteliti
Data Visualization yaitu untuk menyajikan representasi dari data yang digunakan.

Modelling yaitu tahapan pengolahan data dengan metode *Naïve Bayes*.

Data Testing, merupakan tahapan pengujian dari model yang digunakan hal tersebut bertujuan untuk menguji akurasi yang dihasilkan metode *Naïve Bayes*.

4. Data Riset

Data riset yang akan digunakan adalah dataset yang berasal dari *kaggle.com* pada tahun 2022. Data tersebut berjumlah 300 baris data. Terdapat 4 kolom (*Gender*, *BP*, *Cholesterol*, *MaxHR*) yang akan digunakan untuk pengujian dan 1 kolom (*Heart Disease*) sebagai kolom target/label. Sampel dataset yang akan digambarkan dalam penelitian ini menggunakan 10 data teratas seperti yang terdapat dalam Tabel (1) berikut ini :

	Gender	BP	Cholesterol	MaxHR	HeartDisease
0	1	130	322	109	Presence
1	0	115	564	160	Absence
2	1	124	261	141	Presence
3	1	128	263	105	Absence
4	0	120	269	121	Absence
5	1	120	177	140	Absence
6	1	130	256	142	Presence
7	1	110	239	142	Presence
8	1	140	293	170	Presence
9	0	150	407	154	Presence

Gambar 1. Dataset Penyakit Jantung

HASIL DAN PEMBAHASAN

5. Pembagian Data

Pembagian data dalam penelitian ini menggunakan *data training* dan *data testing* dengan 4 rasio. Adapun rasio tersebut diantaranya rasio 90:10, 80:20, 70:20, dan 60:40 yang digunakan dalam penelitian ini sehingga rasio pembagian data lebih bervariasi dimana tujuan yang diharapkan adalah memperoleh model Naive Bayes pada rasio pembagian data manakah yang memiliki hasil prediksi terbaik.

6. Prediksi dengan Naive Bayes

Naive bayes merupakan sebuah metode dengan pendekatan 2 penafsiran yang sedikit berbeda. Menurut penjelasan bayes, mode ini menjelaskan tingkat derajat kepercayaan pada suatu subjektif harus dapat merubah secara rasional ketika adanya suatu tujuan baru [7] [8] [9]. Adapun rumus dalam naive bayes adalah sebagai berikut [5]:

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

Keterangan :

X merupakan data dengan class yang belum diketahui.

H : hipotesis data X merupakan suatu class spesifik

P(H|X) : Probabilitas hipotesis H berdasarkan kondisi X

P(H) : Probabilitas hipotesis H

P(X|H) : Probabilitas X berdasarkan kondisi pada hipotesis H

P(X) : Probabilitas X

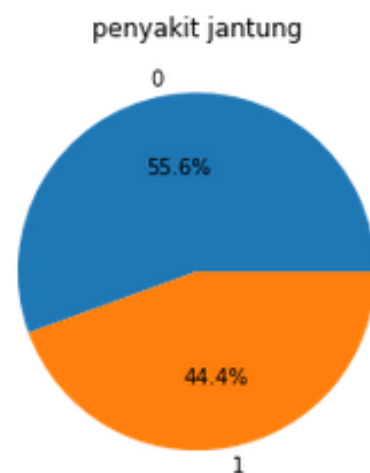
7. Pengukuran Tingkat Akurasi

Data yang diolah menggunakan model *Naive Bayes* kemudian diukur nilai akurasinya. Pengolahan data yang bersifat prediksi yang masih mengandung unsur ketidakpastian maka diperlukan pengujian.

Pada pembahasan dalam penelitian ini akan dilakukan visualisasi data agar data dapat lebih mudah untuk diamati serta dapat mengetahui setiap hubungan antar variabel.

a. Visualisasi Data

Visualisasi data bertujuan untuk memudahkan seseorang dalam memahami dan mengamati sebuah data [6]. Terdapat berbagai model yang dapat digunakan untuk menyajikan sebuah data ke dalam bentuk visualnya.



Gambar 2. *pie chart* penyakit jantung

Pada visualisasi gambar diatas mendeskripsikan besaran data yang digunakan dalam penelitian ini. Pada diagram warna biru dengan kode 0 yang artinya besaran data penderita jantung terdapat data sebanyak 55,6%. Kemudian diagram berwarna jingga dengan kode 1 menggambarkan besaran yang bukan penyakit jantung sebanyak 44,4%. Pada data tersebut dapat diketahui bahwa penderita penyakit jantung lebih besar daripada non penderita penyakit jantung sehingga perlu pencegahan lebih lanjut.

b. Pengujian Data

Pada penelitian ini menggunakan metode naive bayes dimana dilakukan perbandingan menggunakan naive bayes agar diperoleh hasil akurasi terbaik digunakan dalam prediksi penyakit jantung. Penelitian ini menggunakan pembagian data untuk menguji data sebanyak 3 bentuk rasio pembagian data. Berikut ini pembahasan penelitian untuk menentukan nilai naive bayes terbaik kedalam rincian sebagai berikut :

Tabel 1. Pengujian Data

No	Rasio	Akurasi
1	80 : 20	76%
2	70 : 30	81%
3	60 : 40	66%

Berdasarkan tabel diatas, pada pengujian rasio 80:20, maka diperoleh akurasi 76%, pada pengujian rasio 70:30 diperoleh nilai akurasi 81% dan pengujian rasio 60:40 diperoleh nilai akurasi 66%. Sebuah performance pada naive bayes dapat dihitung nilai rata-rata sebesar 74,3%. Perbandingan dari pengujian metode *Naive Bayes* diatas, diperoleh rasio 70:30 adalah akurasi yang terbaik.

Perolehan confusion matrix untuk metode *Naive Bayes* adalah :

	precision	recall	f1-score	support
0	0.87	0.81	0.84	48
1	0.75	0.82	0.78	33
accuracy			0.81	81
macro avg	0.81	0.82	0.81	81
weighted avg	0.82	0.81	0.82	81

Gambar 3. Nilai Akurasi

Berdasarkan hasil perolehan dari pengujian model *Naive Bayes* maka di dapat hasil nilai *Naive Bayes* menghasilkan nilai *precision* untuk penyakit jantung 87% dan untuk yang tidak 75%, nilai *recall* untuk yang terkena penyakit jantung 81% dan yang tidak terkena penyakit jantung 82%, nilai *support* untuk yang terkena penyakit jantung 84% dan yang tidak 78%. Sehingga diperoleh hasil bahwa metode *Naive Bayes* dapat di implementasikan untuk memprediksi penyakit jantung.

KESIMPULAN

Berdasarkan hasil pengolahan data menggunakan metode naive bayes dengan *software jupyter notebook*, *visual studio* dan *python* sebagai bahasa pemrograman serta pengujian menggunakan 3 rasio pengukuran nilai yaitu 80:20, 70:30, dan 60:40 maka diperoleh dari 3 metode terbaik yang menghasilkan akurasi yang baik yaitu 80:20 dengan nilai akurasinya 81%. Dengan demikian metode *Naive Bayes* akan diimplementasikan dalam pembuatan *Deploy* aplikasi untuk prediksi penerapan penyakit jantung.

DAFTAR PUSTAKA

- [1] Anggara, Bayu, and R M Nasrul Halim D, 'Penerapan Data Mining Rumah Sakit Umum Prabumulih Menggunakan Algoritma Naive Bayes (Study Kasus : Penyakit Jantung)', 209–20
- [2] Depari, Deo Haganta, Yuni Widiastiwi, Mayanda Mega Santoni, Fakultas Ilmu Komputer, Universitas Pembangunan, Nasional Veteran, and others, 'Perbandingan Model Decision Tree , Naive Bayes Dan Random Forest Untuk Prediksi Klasifikasi Penyakit Jantung', 4221 (2022), 239–48
- [3] Hafiz, Dimsyiar M Al, Khoirul Amaly, Javen Jonathan, and M Teranggono Rachmatullah, 'Sistem Prediksi Penyakit Jantung Menggunakan Metode Naive Bayes', 2.2, 151–57
- [4] Informatika, Jurnal, Dan Rekayasa, Komputer Jakakom, Rdm Fiqriansyah, Farchan Akbar, Virginia Casanova Andiko, and others, 'Penerapan Algoritma Naive Bayes Untuk Mengetahui Pasien Penyakit Gagal Jantung Jurnal Informatika Dan Rekayasa Komputer (JAKAKOM)', 2.September (2022)
- [5] Journal, Citec, Mufti Ari Bianto, Magister Teknik Informatika, and Universitas Amikom Yogyakarta, 'Perancangan Sistem Klasifikasi Penyakit Jantung Menggunakan Naive Bayes', 6.1 (2019)
- [6] Nasional, Konferensi, Ilmu Sosial, Tri Retnasari, and Eva Rahmawati, 'Diagnosa Prediksi Penyakit Jantung Dengan Model Algoritma Naive Bayes Dan Algoritma C4.5', 2017, 7–12
- [7] Science, Computer, 'Prediksi Penyakit Jantung Dengan Algoritma Klasifikasi', 5.1 (2019), 978–79
- [8] Sulaksono, Juli, Diaagnosa Penyakit, and Jantung Menggunakan, 'Sistem Pakar Penentuan Penyakit Gagal Jantung', 2015, 6–8
- [9] Utomo, Dito Putro, 'Analisis Komparasi Metode Klasifikasi Data Mining Dan Reduksi Atribut Pada Data Set Penyakit Jantung', 4.April (2020), 437–44
<<https://doi.org/10.30865/mib.v4i2.2080>>