

Implementasi Analisa Penyakit Jantung Menggunakan Naïve Bayes

¹Lailatul Risqia, ²Zaehol Fatah

^{1,2}Sistem Informasi, Sains dan Teknologi Universitas Ibrahimy
email: lailatulrisqia123@gmail.com

Abstrak

Penyakit jantung adalah suatu penyebab utama kematian di dunia. Deteksi dini dapat membantu mengurangi angka mortalitas yang tinggi akibat penyakit ini. Dalam penelitian ini, algoritma Naive Bayes digunakan untuk menganalisis data medis dan memprediksi kemungkinan penyakit jantung. Dataset yang digunakan adalah Cleveland Heart Disease Dataset, yang berisi fitur-fitur medis penting. Hasil menunjukkan bahwa model Naive Bayes mampu mencapai akurasi prediksi hingga 81%, menjadi metode yang efektif untuk analisa penyakit jantung.

Kata Kunci: Penyakit Jantung, Naive Bayes, Data Mining, Prediksi, analisa Data

Abstract

Heart disease is a leading cause of death in the world. Early detection can help reduce the high mortality rate due to this disease. In this study, the Naive Bayes algorithm was used to analyze medical data and predict the likelihood of heart disease. The dataset used is the Cleveland Heart Disease Dataset, which contains important medical features. The results show that the Naive Bayes model is able to achieve prediction accuracy of up to 81%, becoming an effective method for analyzing heart disease.

Keywords: Heart Disease, Naïve Bayes, Data Mining, Prediction, Data Analysis

PENDAHULUAN

Jantung sebuah rongga organ berotot yang memompa darah melalui pembuluh darah dengan kontraksi berirama yang terus berulang merupakan salah satu organ manusia yang berperan dalam sistem peredaran darah. Darah kemudian menyuplai oksigen dan nutrisi pada tubuh, dimana darah juga membantu menghilangkan sisa-sisa dari metabolisme. Letak Jantung berada di rongga dada sekitar sebelah kiri. Jantung sebagai salah organ terpenting dalam tubuh memiliki resiko kematian jika ada kelainan yang terjadi pada jantung. Beberapa masalah pada jantung dibagi menjadi dua yaitu penyakit jantung dan serangan jantung. WHO berdasarkan data menyatakan bahwa ada sebanyak 7,3 juta penduduk di dunia yang meninggal dikarenakan penyakit jantung. Tipe penyakit jantung terjadi dikarenakan jantung tidak dapat

melaksanakan tugasnya dengan baik, seperti Otot jantung yang lemah atau adanya celah antara serambi kanan dan serambi kiri (Tempola, 2020).

Banyak faktor predisposisi seperti kebiasaan pribadi serta predisposisi genetik menyebabkan penyakit jantung. Berbagai faktor risiko kebiasaan seperti merokok, terlalu sering menggunakan alkohol dan kafein, stres, dan aktivitas fisik bersama dengan faktor fisiologis lainnya seperti obesitas, hipertensi, kolesterol, darah tinggi dan juga kondisi jantung yang sudah ada sebelumnya merupakan faktor predisposisi penyakit jantung. Diagnosis medis penyakit jantung yang efisien dan akurat serta dini memainkan peran penting dalam mengambil tindakan pencegahan untuk mencegah kematian (Atthohiroh, Ayu and Maharani, 2023).

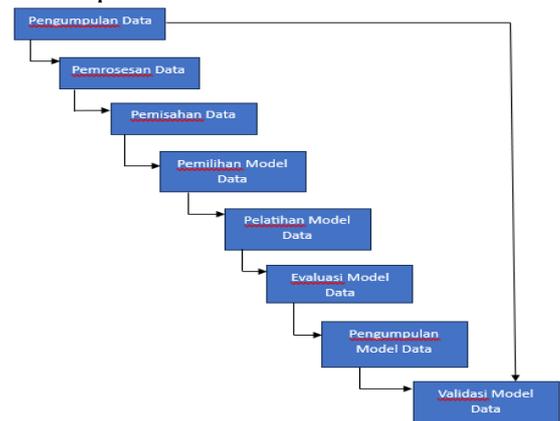
Naïve Bayes merupakan pengklasifikasi dengan metode probabilitas dengan metode probabilitas yang ditemukan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya sehingga dikenal sebagai teorema Bayes. Teorema tersebut dikombinasikan dengan naive dimana diasumsi kondisi antara petunjuk (atribut) saling bebas. Naive Bayes adalah teknik memprediksi berbasis probabilitas pada penerapan aturan bayes dengan asumsi ketidak ketergantungan yang kuat. 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 Naïve Bayes dalam implementasi penyakit jantung sehingga dapat lebih memahami dalam menggunakannya (Atthohiroh, Ayu and Maharani, 2023).

Metode Dan Tahap Penelitian

Penelitian ini menggunakan metode Naïve Bayes (NB) untuk pengelolaan data terkait penyakit jantung, dengan tujuan mengidentifikasi faktor risiko dan memprediksi penyakit tersebut. Data dikumpulkan dari hasil pencarian di internet dan diolah melalui beberapa tahap, termasuk pemrosesan data untuk membangun model Naïve Bayes, pemisahan data berdasarkan nilai untuk memudahkan seleksi model, pelatihan model untuk mengenali pola data, serta evaluasi model dengan membaca data

atribut dari file CSV. Setelah pelatihan, model digunakan untuk melakukan prediksi pada data uji guna memastikan akurasi klasifikasi. Validasi dilakukan untuk menyempurnakan metrik evaluasi, sehingga metode ini mampu menghasilkan prediksi yang akurat dan dapat membantu dalam analisis penyakit jantung. Berikut adalah tahapan desain.



Gambar 1. Langkah-langkah dalam metode Naïve Bayes

Terdapat tahapan dalam menggunakan metode naïve bayes berikut penjelasan dari setiap Langkah yang di berikan:

1. Pengumpulan Data
 - Tahap ini mencakup pencarian dan pengumpulan dataset yang relevan. Dataset tersebut akan berisi fitur atau variabel yang diperlukan untuk analisis.
2. Pemrosesan Data
 - Pada tahap ini, data yang telah dikumpulkan diolah untuk menghilangkan kesalahan atau inkonsistensi, seperti data duplikat, data kosong (missing values), atau data yang tidak sesuai.
3. Pemilihan Data
 - Data yang relevan dipisahkan dari data yang tidak dibutuhkan. Fitur yang paling signifikan dipilih untuk meningkatkan efisiensi model.
4. Pelatihan Model Data
 - Model dilatih menggunakan algoritma tertentu, dalam hal ini Naive Bayes. Proses ini melibatkan pembelajaran pola dari data pelatihan yang telah dipilih.

5. Evaluasi Model Data
 - Model diuji menggunakan data uji untuk mengukur akurasi, presisi, recall, atau metrik evaluasi lainnya.
6. Pengumpulan Model Data
 - Data tambahan atau hasil model sebelumnya dikumpulkan untuk memperbaiki model dan meningkatkan performa.
7. Validasi Model Data
 - Model yang telah diuji divalidasi untuk memastikan hasilnya dapat diterapkan pada data baru. Validasi ini dilakukan untuk memastikan keakuratan dan kemampuan generalisasi model.

Jenis penelitian

Untuk penelitian ini penulis menggunakan jenis Penelitian Deskriptif Kuantitatif serta *Library Online Research*. Penelitian deskriptif kuantitatif dipilih agar dapat menggambarkan, mendiskripsikan serta menjelaskan sesuatu secara objektif serta dapat menarik kesimpulan dari data berupa angka-angka yang telah disajikan, dan dengan cara ini peneliti menghimpun

mengelola data serta menganalisa data-data yang terkait dengan klasifikasi penyakit jantung dengan metode Naive Bayes dengan rujukan seperti jurnal, perpustakaan serta buku yang kemudian dibaca dan di catat untuk selanjutnya mengolah bahan penelitian (Yogianto, Homaidi and Fatah, 2024).

Metode pengumpulan Data

Pengumpulan data dilakukan untuk mendukung penelitian ini. Dalam penelitian ini metode pengumpulan data yang dilakukan untuk mengumpulkan data terkait dengan penelitian yang sedang dilakukan, metode pengumpulan data ini merupakan suatu cara independen untuk melakukan analisa data atau bisa dijadikan alat utama dalam menganalisa data. Untuk mendapatkan data-data yang terkait dengan penyakit jantung, penulis mengumpulkan data yang terkait dengan penelitian dengan cara Study Literature dengan jenis data sekunder yang didapatkan dari Website Kaggle dengan judul Implementasi Penyakit Jantung.

Tabel 1. Dataset

Row No.	HeartDisea...	confidence(...	confidence(...	prediction(...	Age	Gender	ChestPainT...	RestingBP	Cholesterol	FastingBS	RestingECG	MaxHR	ExerciseAn...	Oldr
1	1	1.000	0.000	1.0	63	Male	Typical	145.0	233.0	1.0	Normal	150.0	?	2.3
2	0	0.000	1.000	0.0	37	Male	Non-anginal	130.0	250.0	0.0	ST	187.0	?	3.5
3	0	0.000	1.000	0.0	41	Female	Atypical	130.0	204.0	0.0	Normal	172.0	?	1.4
4	1	0.992	0.008	1.0	56	Female	Asymptomat	120.0	236.0	0.0	LVH	178.0	?	0.8
5	1	1.000	0.000	1.0	57	Male	Typical	140.0	354.0	0.0	Normal	163.0	?	0.6
6	1	1.000	0.000	1.0	62	Male	Atypical	135.0	354.0	0.0	ST	150.0	?	1.0
7	0	0.000	1.000	0.0	54	Female	Non-anginal	120.0	354.0	0.0	LVH	178.0	?	2.8
8	0	0.000	1.000	0.0	48	Female	Typical	115.0	354.0	0.0	Normal	168.0	?	1.5
9	1	1.000	0.000	1.0	58	Male	Asymptomat	150.0	354.0	0.0	ST	140.0	?	2.0
10	1	1.000	0.000	1.0	60	Male	Typical	140.0	354.0	0.0	LVH	155.0	?	2.4

Keterangan:

1. HeartDisease : Indikator apakah pasien memiliki penyakit jantung (1: ya, 0: tidak)
2. Condifindence : Nilai keyakinan model untuk prediksi, biasanya menunjukkan Probibalitas kelas
3. Prediction : Prediksi model terhadap penyakit jantung (1: ada penyakit, 0: tidak ada penyakit)
4. Age : Usia Pasien
5. Gender : Jenis Kelamin
6. ChestPainType : Tipe nyeri dada yang dirasakan

7. RestingBp : Tekanan darah saat istirahat
8. Cholestrol : Tingkatan kolestrol
9. FastingBS : Kadar gula darah setelah puasa
10. MaxHR : Denyut jantung maksimum yang dicapai
11. ExenrciseAngina : Nyeri dada akibat olahraga
12. Oldpeak : Depresi St yag diindusikan oleh olahraga relatif terhadap istirahat (mengindikasikan tingkat keparahan iskemia jantung)

Data mining

Pada penelitian kali ini, peran dari *data mining* menjadi sangat penting, karna *data mining* merupakan suatu teknik yang diperhunakan dalam menemukan informasi dalam suatu *database* (Ayub). *Data mining* dalam prakteknya mengimolementasikan berbagai teknik diantaranya, Kecerdasan buatan, Pemebelajaran mesin, dan teknik statis yang digunakan untuk mengekstrasikan data dan mengindentifikasi data untuk kebutuhan informasi dari sebuah data base yang besar(Dodi Nofri Yoliad, 2023). *Data mining* mengacu pada proses untuk menambang (mining) pengetahuan dari sekumpulan data yang sangat besar. *Data mining* ialah menuju langkah *Knowledge Discovery In Database* (KDD).

Naive Bayes (NB)

Naive Bayes merupakan algoritma klasifikasi sederhana dimana setiap atribut bersifat independen dan dapat mempengaruhi keputusan akhir. Algoritma Naive Bayes dapat digunakan untuk memprediksi probabilitas milik kelas. Algoritma Naive Bayes sangat cocok untuk klasifikasi data bertipe nominal. Naive bayes telah terbukti memiliki keakurasian dan kecepatan yang tinggi ketika diterapkan pada database dengan jumlah data yang besar.(Hirwono, Hermawan and Avianto, 2023). Algoritma Naive 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(Riany and Testiana, 2023).

Berikut ini adalah langkah-langkah dan rumus dalam pengelolaan data penyakit jantung dengan metode Naive Bayes[7].

$$P(C_k|x_1, x_2, \dots, x_n) =$$

$$\frac{P(C_k) \times P(x_1|C_k) \times P(x_2|C_k) \dots \times P(x_n|C_k)}{P(x_1) \times P(x_2) \dots \times P(x_n)}$$

Keterangan:

$P(C_k|x_1, x_2, \dots, x_n)$ merupakan probalitas kelas C_k yang mempunyai fitur x_1, x_2, \dots, x_n
 $P(C_k)$ merupakan probalitas prior dari kelas C_k

$P(x_i|C_k)$ merupakan probalitas like hood fitur X_i ke kelas C_k

$P(x_1), P(x_2), \dots, P(x_n)$ merupakan probalitas margin dari x_1, x_2, \dots, x_n , yang bisa dihitung dengan metode apapun.

Impelementasi

Implementasi adalah suatu kegiatan atau suatu tindakan dari sebuah rencana yang dibuat secara terperinci untuk mencapai suatu tujuan. Implementasi mulai dilakukan apabila seluruh perencanaan sudah dianggap sempurna. Dan implementasi adalah suatu perluas aktivitas yang saling menyesuaikan dalam tindakan yang dilakukan baik individu maupun kelompok(Egziabher and Edwards, 2017).

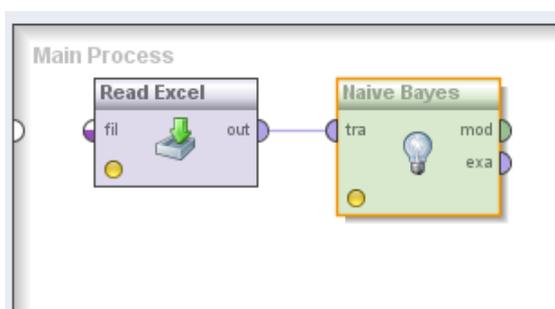
Penyakit Jantung

Jika terdapat pasien yang menderita penyakit jantung memerlukan pegangan yang tepat dan baik, karena penyakit ini termasuk salah satu penyakit yang paling mematikan meskipun tidak menular, penyakit jantung merupakan penyebab utama dari kematian di seluruh dunia. Setiap tahun sekitar 50% penduduk meninggal akibat penyakit jantung. Pada tahun 2012, Badan Kesehatan dunia

menyebutkan bahwa terdapat 17,5 juta orang meninggal di dunia disebabkan oleh penyakit kardiovaskular dan ikut menyumbang 31% dari total 56,5 juta kasus kematian di dunia. Hal ini menyebabkan terjadinya gangguan pada jantung dan memerlukan perhatian khusus. Hal perlu kiranya untuk melakukan penelitian yang dapat mengklasifikasikan terhadap pasien yang memiliki penyakit jantung atau tidak, hal ini juga dapat membantu tenaga medis didalam pengambilan keputusan dan tindakan pencegahan terhadap pasien penderita penyakit jantung. (Gobel and Mahkota, 2006)

HASIL DAN PEMBAHASAN

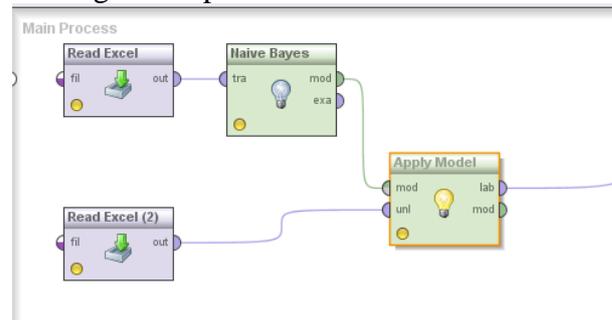
Pengembangan model klasifikasi dengan akurasi model 81% pada pengujian data validasi, menunjukkan kemampuan prediksi yang cukup baik dalam implementasi penyakit jantung, model berhasil mengidentifikasi faktor-faktor risiko utama seperti usia, tekanan darah, kolesterol, dan jenis kelamin yang secara signifikan memengaruhi kemungkinan seseorang menderita penyakit jantung. Dataset yang digunakan terdiri dari berbagai fitur medis penting seperti tipe nyeri dada, denyut jantung maksimum, dan depresi ST yang relevan dalam analisis penyakit jantung.



Gambar 2. Transformasi Data
Transformasi Data

Transformasi data ini merupakan suatu cara yang dilakukan untuk inisialisasi kepada *dataset* dan mempersiapkan data

agar sesuai dengan asumsi algoritma dan meningkatkan performa model.



Gambar 3. Model Metode Naïve Bayes (KB)

Implementasi *Data Mining* dengan algoritma Naïve Bayes. Gambar tersebut secara khusus menampilkan sebuah *process* atau alur kerja dalam RapidMiner. Alur kerja ini menggambarkan serangkaian langkah yang akan dilakukan untuk mencapai suatu tujuan analisis data.

KESIMPULAN

Penelitian ini membuktikan bahwa algoritma Naïve Bayes merupakan metode yang efektif dalam menganalisis dan memprediksi kemungkinan penyakit jantung dengan tingkat akurasi mencapai 81%. Dengan menggunakan Cleveland Heart Disease Dataset, penelitian ini berhasil mengidentifikasi faktor-faktor risiko utama seperti usia, tekanan darah, kadar kolesterol, jenis kelamin, dan tipe nyeri dada yang berpengaruh signifikan terhadap kemungkinan seseorang menderita penyakit jantung.

Proses data mining yang dilakukan memberikan wawasan penting melalui pengolahan dan analisis data yang sistematis, sehingga menghasilkan model yang dapat diandalkan. Implementasi metode ini tidak hanya relevan dalam mendukung diagnosa dini, tetapi juga berpotensi menjadi alat yang membantu tenaga medis dalam pengambilan keputusan yang lebih baik serta tindakan pencegahan yang lebih tepat sasaran terhadap pasien dengan risiko tinggi penyakit jantung.

UCAPAN TERIMA KASIH

Ungkapan terima kasih kami haturkan untuk segala macam dukungan, kerja sama serta bantuan pemikiran yang telah dibeikan dalam proses penelian ini. Tanpa sumbangsih dari anda semua, pencapaian serta kesuksesan yang kami raih akan sulit terwujud.

DAFTAR PUSTAKA

- Atthohiroh, A., Ayu, R. and Maharani, S. (2023) 'Penerapan Metode Naive Bayes Untuk Memprediksi Penyakit Jantung', *Jurnal Teknisi*, 3(1), p. 8. Available at: <https://doi.org/10.54314/teknisi.v3i1.1252>.
- Ayub, M. (no date) 'Proses Data Mining dalam Sistem Pembelajaran Berbantuan Komputer', pp. 21–30.
- Dodi Nofri Yoliad (2023) 'Data mining Dalam Analisis Tingkat Penjualan Barang Elektronik Menggunakan Algoritma K-means', *Insearch (Information System Research) Journal*, 3(1).
- Egziabher, T.B.G. and Edwards, S. (2017) 'The Policy Implementation', *Africa's Potential for the Ecological iIensification of Agriculture*, 53(9), pp. 1689–1699.
- Gobel, F.A. and Mahkota, R. (2006) 'Faktor-faktor yang Mempengaruhi Kematian Pasien Penyakit Jantung Koroner di Pusat Jantung Nasional Harapan Kita Tahun 2004', *Kesmas: National Public Health Journal*, 1(3), p. 99. Available at: <https://doi.org/10.21109/kesmas.v1i3.303>.
- Hirwono, B., Hermawan, A. and Avianto, D. (2023) 'Implementasi Metode Naive Bayes untuk Klasifikasi Penderita Penyakit Jantung', *Jurnal JTIK (Jurnal Teknologi Informasi dan Komunikasi)*, 7(3), pp. 450–457. Available at: <https://doi.org/10.35870/jtik.v7i3.910>.
- Riany, A.F. and Testiana, G. (2023) 'Penerapan Data Mining untuk Klasifikasi Penyakit Jantung Koroner Menggunakan Algoritma Naive Bayes', *MDP Student Conference*, 2(1), pp. 297–305. Available at: <https://doi.org/10.35957/mdp-sc.v2i1.4388>.
- Tempola, firman (2020) 'Implementasi Metode Naive Bayes Untuk Memprediksi Resiko Penyakit Jantung', *Patria Artha Technological Journal*, 4(2), pp. 66–70. Available at: <https://doi.org/10.33857/patj.v4i2.351>.
- Yogianto, A., Homaidi, A. and Fatah, Z. (2024) 'Implementasi Metode K-Nearest Neighbors (KNN) untuk Klasifikasi Penyakit Jantung', *G-Tech: Jurnal Teknologi Terapan*, 8(3), pp. 1720–1728. Available at: <https://doi.org/10.33379/gtech.v8i3.4495>.