

Pengelolaan Data Penyakit Jantung Dengan Menggunakan Metode *Naive Bayes*

Oskar^{1*}, Fransiska Ria², Rupina³, Noviyanti. P⁴

^{1,2,3,4}Program Studi Teknologi Informasi, Institut Shanti Bhuana Bengkayang

E-mail: oskar2112@shantibhuana.ac.id¹, fransiska2106@shantibhuana.ac.id²,
rupina2115@shantibhuana.ac.id³, noviyanti@shantibhuana.ac.id⁴

Abstrak

Penyakit jantung merupakan penyakit dengan angka kematian yang tinggi di seluruh dunia. Oleh karena itu, penelitian ini bertujuan untuk mengelola data penyakit jantung dengan menggunakan metode data mining Naive Bayes. Hasil analisis menunjukkan bahwa metode Naive Bayes dapat mengklasifikasikan data penyakit jantung secara efektif dengan akurasi 79,92% dibandingkan hasil diagnosis medis. Namun metode Naive Bayes masih memiliki kelemahan yaitu tidak dapat menghitung keakuratan klasifikasi. Dengan menggunakan teknik analisis data tingkat lanjut, hasil penelitian ini diharapkan dapat memberikan pedoman yang lebih baik dalam upaya penanganan penyakit jantung.

Abstract

Heart disease is a disease with a high mortality rate worldwide. Therefore, this study aims to manage heart disease data using the Naive Bayes data mining method. The analysis results show that the Naive Bayes method can classify heart disease data effectively with 79.92% accuracy compared to medical diagnosis results. However, the Naive Bayes method still has the disadvantage that it cannot calculate the accuracy of classification. By using advanced data analysis techniques, the results of this study are expected to provide better guidelines in efforts to treat heart disease.

Info Naskah:

Naskah masuk: 13 Juni 2024

Direvisi: 11 Juli 2024

Diterima: 15 Juli 2024

Keywords:

Data management

heart diseases

Method Naive Bayes

...

*Penulis korespondensi:

Oskar, Fransiska Ria, dan Rupina

E-mail: oskar2112@shantibhuana.ac.id, fransiska2106@shantibhuana.ac.id, rupina2115@shantibhuana.ac.id

1. Pendahuluan

Organ jantung adalah salah satu yang terpenting dalam tubuh manusia dan terdiri dari otot dan empat ruang. Serangan jantung adalah suatu kondisi di mana pembuluh darah tersumbat akibat timbunan lemak. Merupakan penyakit yang menimbulkan berbagai gejala seperti sesak napas dan nyeri dada[1]. Tingkat kematian yang tinggi terjadi akibat penyakit serangan jantung[2]. Saat tubuh sakit, fungsi tubuh tidak akan berjalan maksimal dan tidak nyaman untuk tidur, duduk, ataupun makan. Oleh karena itu, menjaga Kesehatan tubuh merupakan hal yang sangat penting yang harus selalu diperhatikan. Salah satu penyakit paling mematikan di dunia adalah penyakit jantung[3]. Dalam hal ini banyak kematian yang disebabkan oleh penyakit jantung akibat pola hidup tidak sehat dan konsumsi kolesterol tinggi[4]. Karena risiko kematian sangat tinggi, maka perlu dilakukan pengumpulan dan analisis data yang dapat mendeteksi penyakit jantung pada pasien secara akurat. Data yang diperoleh diolah dengan data mining menggunakan metode Naive Bayes[5].

Algoritma Naive Bayes sangat cenderung dalam mendiagnosis penyakit jantung karena nilai akurasi yang didapat sangat tinggi. Kelebihannya adalah cepat dalam perhitungan dan algoritmanya juga sederhana sehingga sangat akurat dalam menentukan diagnosa[6]. Nilai posteriori tertinggi dipilih sebagai hasil klasifikasi[7]. Penelitian ini dilakukan untuk mengatasi masalah tersebut dengan menggunakan data klasifikasi untuk memprediksi penyakit jantung. Diperlukan suatu metode atau teknik yang dapat mengolah data yang ada.[8]. Data yang dianalisis diharapkan dapat memberikan informasi dan solusi terbaik bagi pengobatan penyakit jantung. [9]. Namun metode Naive Bayes sendiri mempunyai kelemahan yaitu tidak dapat menghitung seberapa akurat klasifikasi yang dilakukan[10]. Maka dari itu dengan adanya penelitian ini dapat dimanfaatkan secara maksimal[11]. Dan juga dapat terlaksana dengan baik demi mencapai nilai yang lebih akurat[12]. Mengumpulkan informasi sebagai tolak ukur untuk melakukan penelitian. [13]. Ini termasuk penelitian Deo Haganta Depar tahun 2022, "Perbandingan Model Pohon Keputusan, Naive Bayes, dan Random Forest untuk Memprediksi Klasifikasi Penyakit Jantung.[14].Kajian Derisma tahun 2020 berjudul "Perbandingan Kinerja Algoritma Prediksi Penyakit Jantung Menggunakan Teknik Data Mining."[15]. Oleh karena itu, dua publikasi ilmiah yang telah dibahas sebelumnya antara lain adalah perbandingan pohon keputusan, Naive Bayes, dan model random forest untuk memprediksi klasifikasi penyakit jantung, serta perbandingan kinerja algoritma prediksi penyakit jantung dengan menggunakan teknik data mining. Dua penelitian sebelumnya dapat dibandingkan dengan penelitian yang sedang disusun oleh penulis berjudul

"Mengelola Data Penyakit Jantung Menggunakan Metode Naïve Bayes" menangani pengelolaan data penyakit jantung dengan akurasi sangat tinggi menggunakan metode Naive Bayes.

2. Metode Penelitian

Metode Naïve Bayes adalah alat yang berguna untuk pengelolaan data penyakit jantung, yang dapat membantu dalam mengidentifikasi faktor-faktor resiko dan mengetahui penyakit. Dengan mengikuti Langkah-langkah di atas, penulis dapat mengembangkan model, yang dapat memberikan wawasan berharga kepada para profesional medis dan peneliti dalam upaya pencegahan dan pengelolaan data penyakit jantung. Berikut ini adalah langkah-langkah dan rumus dalam pengelolaan data penyakit jantung dengan Metode Naive Bayes.

2.1 Rumus Naive Bayes

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

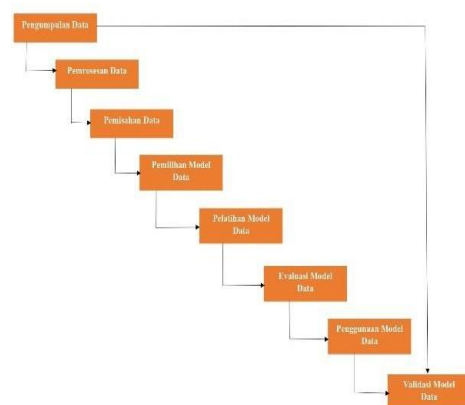
Keterangan:

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

$P(C_k)$ merupakan probabilitas prior dari kelas C_k

$P(x_i|C_k)$ merupakan probabilitas like hood dari fitur x_i ke kelas C_k

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



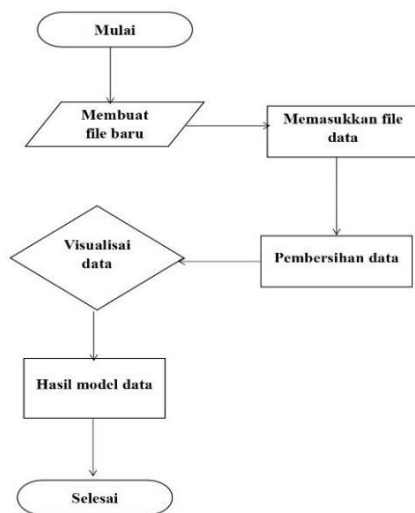
Gambar 2.1 Langkah-langkah dalam Metode Naïve Bayes

Pada gambar 2.1 terdapat tahapan dalam menggunakan metode Naive Bayes berikut penjelasan dari setiap Langkah yang di berikan: Pengumpulan Data adalah data diperoleh dari hasil pencarian di internet, dan dikumpulkan untuk memudahkan dalam pengelolaan data menggunakan metode Naïve Bayes, Pemrosesan Data adalah untuk membuat model Naïve Bayes agar dapat mengetahui sejauh mana data yang diperoleh menghasilkan akurasi dari data yang sudah

dikumpulkan. Pemisahan data untuk memudahkan dalam pemilihan data model, berdasarkan nilai yang dihasilkan,

Pemilihan model data mendefinisikan model data di mana data tersebut dilatih dan diproses untuk menghasilkan data yang akurat, Pelatihan Model Data adalah untuk mengajar komputer atau algoritma bagaimana memahami pola dan informasi yang ada dalam data sehingga dapat diketahui atau klasifikasi pada data yang belum pernah dilihat sebelumnya, Evaluasi Model Data adalah yang berfungsi untuk membaca data dari file CSV data yang terdiri dari atribut, Penggunaan Model Data adalah untuk melatih Naïve Bayes dengan data training dan kemudian melakukan prediksi kelas untuk data uji, dan Validasi Model Data adalah untuk menyesuaikan metrik evaluasi sesuai dengan kebutuhan kita.

2.2 Flowchart



Gambar 2.2 Flowchart

Pada gambar berikut di jelaskan alur sistem yang akan di buat yang di mulai dari membuat file baru, menemukan file data, pembersihan data, visualisasi data, dan yang terakhir hasil model pendataan.

3. Hasil dan Pembahasan

Secara keseluruhan, penggunaan perangkat lunak Python untuk mengelola data penyakit jantung menggunakan metode Naive Bayes merupakan langkah penting dalam memahami dan mengobati penyakit jantung. Di bawah ini adalah contoh data yang digunakan untuk penelitian ini.

1	age	gender	impluse	pressure	pressure	glucose	kcm	troponin	class
2	64	1	66	160	83	160	01.08	0.012	negative
3	21	1	94	98	46	296	0,302083	01.06	positive
4	55	1	64	160	77	270	0,110417	0.003	negative
5	64	1	70	120	55	270	0,602083	0,084722	positive
6	55	1	64	112	65	300	01.08	0.003	negative
7	58	0	61	112	58	87	0,099306	0.004	negative
8	32	0	40	179	68	102	0,049306	0.003	negative
9	63	1	60	214	82	87	300	02.37	positive
10	44	0	60	154	81	135	02.35	0.004	negative
11	67	1	61	160	95	100	0,141667	0.011	negative
12	44	0	60	166	90	102	02.39	0.006	negative
13	63	0	60	150	83	198	02.39	0.013	negative
14	64	1	60	199	99	92	03.43	05.37	positive
15	54	0	94	122	67	97	01.42	0.012	negative
16	47	1	76	120	70	319	02.57	0.003	negative
17	61	1	81	118	66	134	01.49	0.017	positive
18	86	0	73	114	68	87	01.11	0,538889	positive
19	45	0	70	100	68	96	0,420833	0.004	negative
20	37	0	72	107	86	274	0,145139	0.003	negative

Gambar 3.1 Data Penyakit Jantung

Untuk menggambarkan data penyakit jantung, angka 1 mewakili jenis kelamin laki-laki, angka 0 mewakili jenis kelamin perempuan, dan sampel data terdiri dari 20 kolom dan 9 baris.

3.2 Halaman Utama Python

Disini penulis menggunakan perangkat lunak Python, langkah awal untuk mendapatkan hasil, didalam perangkat lunak Python penulis juga sudah menyiapkan sebuah file agar perangkat lunak Python dapat digunakan dan berjalan dengan lancar, karena tanpa sebuah file perangkat lunak Python tidak bisa berjalan atau digunakan untuk pengelolaan hasil data. Berikut adalah tampilan utama perangkat lunak Python sebelum membuat file.



Gambar 3.2 Tampilan Sebelum Membuat File

Berdasarkan Gambar 3.2 Tampilan Sebelum Membuat File, ini adalah tampilan sebelum membuat file ke dalam Python dan Python tidak dapat dijalankan.

3.3 Tampilan sesudah membuat file



Gambar 3.2 Tampilan Sesudah Membuat File

Berdasarkan Gambar 3.2 Tampilan Sesudah Membuat File, dan ini adalah tampilan sesudah membuat file di Python, dan sistem dapat berjalan.

3.4 Membaca Dataset ke Dalam Python

Membaca Dataset ke Dalam Python adalah untuk mengambil data yang tersimpan di dalam sebuah file seperti CSV, Excel, dan Format file lainnya. Berikut ini adalah contoh cara membaca dataset dalam format

CSV menggunakan pandas ke dalam perangkat lunak Python.

[illegible]

Gambar 3.3 Memasukkan Nama File Data kedalam Python untuk membaca dataset menggunakan kode Pandas

Berdasarkan Gambar 3.3 adalah untuk memasukkan nama file yang akan kita Kelola dalam perangkat lunak Python fungsinya agar data mudah didapatkan dan dataset suatu data mudah dikelola dengan baik.

3.5 Pembersihan Data

Pembersihan Data adalah proses mengidentifikasi penangan untuk memperbaiki masalah dalam suatu data yang dapat mempengaruhi kualitas pemodelan data.

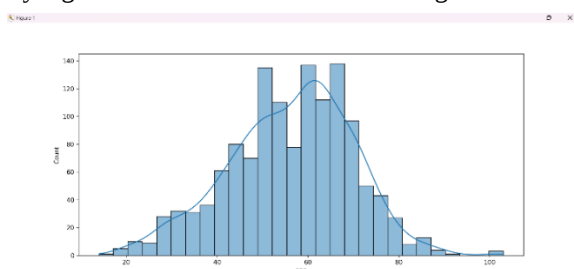
[illegible]

Gambar 3.3 Hasil Pembersihan Data

Berdasarkan Gambar 3.3 Hasil Pembersihan Data adalah fungsinya untuk membersihkan data-data yang kosong dan data yang kurang jelas maka data akan dihapus berdasarkan data yang akan diperlukan dalam pengelolaan.

3.6 Visualisasi Data

Visualisasi Data adalah membuat data untuk lebih mudah dipahami dan memungkinkan analisis yang baik untuk membantu dalam mengidentifikasi



pola. Gambar 3.4 Hasil Visualisasi Data dalam bentuk Histogram

Berdasarkan Gambar 3.6 Hasil Visualisasi Data dalam bentuk Histogram adalah untuk memberikan hasil data yang jelas dan dapat ditandai berdasarkan data yang diinginkan, seperti pada gambar diatas dapat dilihat bahwa variabel data yang ditandai yaitu Age/Umur, dengan adanya data

berbentuk histogram ini dapat lebih mempermudah dalam pengelolaan data.

3.7 Pemodelan dan Evaluasi

Pemodelan dan evaluasi merupakan tahap pengelolaan data dan pengembangan model statistik atau pembelajaran untuk memprediksi informasi berdasarkan data yang ada.

[illegible]

Gambar 3.7 Hasil Pemodelan dan Evaluasi Data

Kajian hasil analisis program komputer yang mengimplementasikan algoritma klasifikasi data. Program ini menggunakan data penyakit jantung untuk melatih model. Model ini digunakan untuk memprediksi apakah seorang pasien menderita penyakit jantung. Hasil analisis menunjukkan akurasi model ini sebesar 79,92%. Akurasi ini dicapai dengan membandingkan prediksi model dengan diagnosis dokter. Kedua, hasil analisis menunjukkan bahwa model tersebut lebih akurat dalam memprediksi pasien tanpa penyakit jantung (76%) dibandingkan pasien dengan penyakit jantung (82%). Hal ini menunjukkan bahwa model tersebut lebih baik dalam mengidentifikasi pasien tanpa penyakit jantung dibandingkan dengan pasien dengan penyakit jantung. Dalam hal ini, model tersebut memperkirakan bahwa 70 pasien yang sebenarnya menderita penyakit jantung akan menderita penyakit jantung. Model tersebut memperkirakan penyakit jantung pada 31 pasien yang sebenarnya tidak menderita penyakit jantung.

3.8 Simpan Model

Simpan Model adalah untuk menyimpan data yang sudah di Kelola dan siap untuk disimpan.

```

C:\Users\user>python3 -c "import sys; sys.path.append('C:\Users\user\AppData\Local\Programs\Python\Python310\Scripts'); import httpx; print('httpx version:', httpx.__version__)"
httpx version: 0.24.0

```

Gambar 3.8 Menyimpan Hasil Pemodelan

Berdasarkan Gambar 3.8 Menyimpan Hasil Pemodelan, Dimana data akan disimpan setelah semua data dibersihkan dan di evaluasi untuk pengelolaan data lebih lanjut.

4 Kesimpulan

Dalam penelitian ini, pengelolaan data penyakit jantung dengan menggunakan metode Naive Bayes telah menghasilkan temuan penting. Hasil analisis menunjukkan bahwa metode ini dapat dengan efektif mengklasifikasikan data penyakit jantung, memberikan kontribusi signifikan terhadap pemahaman dan deteksi dini kondisi kesehatan tersebut. Kesimpulan yang dapat ditarik adalah bahwa hasil model data

analisis mengenai nilai akurasi mencapai 79,92% dari membandingkan hasil prediksi diagnosa dokter.

Ucapan Terimakasih

Ucapan terima kasih yang tulus ingin saya sampaikan kepada dosen pengampu dalam proses penelitian ini. Terima kasih atas bimbingan, arahan, dan dukungan yang luar biasa selama penelitian berlangsung. Keberhasilan penelitian ini tidak terlepas dari dedikasi dan pemahaman Ibu dalam memberikan panduan yang sangat berharga.

Saya juga ingin menyampaikan apresiasi yang tinggi atas dukungan finansial yang diberikan, yang telah memberikan kontribusi besar terhadap kelancaran dan kelengkapan penelitian ini. Semua bantuan ini menjadi pilar penting dalam mencapai hasil yang memuaskan.

Semoga ucapan terima kasih ini dapat menjadi ungkapan kecil penghargaan saya terhadap Ibu Christian Cahyaningtyas, S.Kom, M.Kom, atas dedikasi dan kontribusinya dalam kesuksesan penelitian ini. Terima kasih banyak.

Daftar Pustaka

- [1] D. Sitanggang, N. Nicholas, V. Wilson, A. R. A. Sinaga, and A. D. Simanjuntak, "Implementasi Data Mining Untuk Memprediksi Penyakit Jantung Menggunakan Metode K-Nearest Neighbor Dan Logistic Regression," *Jurnal Teknik Informasi dan Komputer (Tekinkom)*, vol. 5, no. 2, p. 493, 2022, doi: 10.37600/tekinkom.v5i2.698.
- [2] E. Nurlia and U. Enri, "Penerapan Fitur Seleksi Forward Selection Untuk Menentukan Kematian Akibat Gagal Jantung Menggunakan Algoritma C4.5," *Jurnal Teknik Informatika Musirawas (Elin Nurlia)*, vol. 6, no. 1, p. 42, 2021.
- [3] B. ANGGARA and R. M. N. Halim, "Penerapan Data Mining Dirumah Sakit Umum Daerah Prabumulih Menggunakan Algoritma Naive Bayes Study Kasus: Penyakit Jantung," *Bina Darma ...*, pp. 209–220, 2020, [Online]. Available: <http://conference.binadarma.ac.id/index.php/BDCCS/article/view/1546%0Ahttps://conference.binadarma.ac.id/index.php/BDCCS/article/download/1546/690>.
- [4] R. Ayu, S. Maharani, S. Informasi, and S. Royal Kisaran, "Penerapan Metode Naïve Bayes Dalam Memprediksi Penyakit Jantung," *Jurnal Teknologi Komputer dan Sistem Informasi Februari*, vol. 2023, no. 1, pp. 8–13, 2023, [Online]. Available: <http://jurnal.goretanpena.com/index.php/teknisi>.
- [5] A. Samosir, M. S. Hasibuan, W. E. Justino, and T. Hariyono, "Komparasi Algoritma Random Forest, Naïve Bayes dan K- Nearest Neighbor Dalam klasifikasi Data Penyakit Jantung," *Prosiding Seminar Nasional Darmajaya*, vol. 1, no. 0, pp. 214–222, 2021, [Online]. Available: <https://jurnal.darmajaya.ac.id/index.php/PSND/article/view/2955>.
- [6] D. Larassati, A. Zaidiah, and S. Afrizal, "Sistem Prediksi Penyakit Jantung Koroner Menggunakan Metode Naive Bayes," *JUPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 7, no. 2, pp. 533–546, 2022, doi: 10.29100/jupi.v7i2.2842.
- [7] E. Harianto, "Implementasi Metode Naive Bayes Pada Sistem Pakar Diagnosa Penyakit Arteriosklerosis," *Jurnal Majalah Ilmiah Informasi dan Teknologi Ilmiah (INTI)*, vol. 7, no. 2, pp. 165–169, 2020, [Online]. Available: <https://ejurnal.stmik-budidarma.ac.id/index.php/inti/article/view/2380/1762>.
- [8] Julia Triani, Yovi Pratama, and E. Yanti, "Komparasi Dalam Prediksi Gagal Jantung Dengan Menggunakan Metode C4.5 dan Naïve Bayes," *Jurnal Informatika Dan Rekayasa Komputer (JAKAKOM)*, vol. 3, no. 1, pp. 394–402, 2023, doi: 10.33998/jakakom.2023.3.1.759.
- [9] S. Muharni and S. Andriyanto, "Sistem Diagnosa Penyakit Jantung Berbasis Case Based Reasoning (CBR)," *Prosiding Seminar Nasional Darmajaya*, vol. 1, pp. 1–11, 2021, [Online]. Available: <https://jurnal.darmajaya.ac.id/index.php/PSND/article/view/2910>.
- [10] T. Arifin and D. Ariesta, "Prediksi Penyakit Ginjal Kronis Menggunakan Algoritma Naive Bayes Classifier Berbasis Particle Swarm Optimization," *Jurnal Tekno Insentif*, vol. 13, no. 1, pp. 26–30, 2019, doi: 10.36787/jti.v13i1.97.
- [11] K. Abdul Khalim, U. Hayati, and A. Bahtiar, "Perbandingan Prediksi Penyakit Hipertensi Menggunakan Metode Random Forest Dan Naïve Bayes," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 7, no. 1, pp. 498–504, 2023, doi: 10.36040/jati.v7i1.6376.
- [12] S. Sahar, "Analisis Perbandingan Metode K-Nearest Neighbor dan Naïve Bayes Classifier Pada Dataset Penyakit Jantung," *Indonesian Journal of Data and Science*, vol. 1, no. 3, pp. 79–86, 2020, doi: 10.33096/ijodas.v1i3.20.
- [13] A. Putranto, N. L. Azizah, I. Ratna, I. Astutik, F. Sains, and D. Teknologi, "Sistem Prediksi Penyakit Jantung Berbasis Web Menggunakan Metode SVM dan Framework Streamlit," *Jurnal Penerapan Sistem Informasi (Komputer & Manajemen)*, vol. 4, no. 2, pp. 442–452, 2023, [Online]. Available: <https://archive.ics.uci.edu/ml/datasets/heart+disease>.
- [14] D. H. Depari, Y. Widiastiw, and M. M. Santoni, "Perbandingan Model Decision Tree, Naive Bayes dan Random Forest untuk Prediksi Klasifikasi Penyakit Jantung," *Informatik : Jurnal Ilmu Komputer*, vol. 18, no. 3, p. 239, 2022, doi: 10.52958/iftk.v18i3.4694.
- [15] D. Derisma, "Perbandingan Kinerja Algoritma untuk Prediksi Penyakit Jantung dengan Teknik Data Mining," *Journal of Applied Informatics and*

Computing, vol. 4, no. 1, pp. 84–88, 2020, doi:
10.30871/jaic.v4i1.2152.