

Implementasi Data Mining Menggunakan Metode Naïve Bayes Untuk Klasifikasi Penyakit Jantung

^{1*} Arinda Sekar Bilbina S, ² Anita Fatmawati, ³Erna Daniati

¹ Sistem Informasi, Universitas Nusantara PGRI Kediri

E-mail: ^{1*} sarindabilbina@gmail.com, ² itsita.all22@gmail.com, ³ ernadaniati@unp.ac.id

Penulis Korespondens : Erna Daniati

Abstrak—Fakta bahwa penyakit jantung menjadi salah satu sebab kematian tertinggi di seluruh dunia termasuk di Indonesia sehingga deteksi dini sangat penting dilakukan. Peneliti menggunakan metode algoritma Naïve Bayes untuk melakukan klasifikasi kemungkinan penyakit jantung berdasarkan 303 data Kaggle dengan 14 fitur. Preprocessing, pelatihan model menggunakan Python di Google Colab, dan seleksi fitur (SelectKBest) adalah komponen proses. Hasil menunjukkan bahwa model ini efektif untuk membantu diagnosis awal dengan akurasi sebesar 83% dan recall sebesar 97% pada pasien sakit.

Kata Kunci: Penyakit jantung, klasifikasi, Naïve bayes, data mining, diagnosis dini

Abstract—The fact that heart disease is one of the highest causes of death worldwide, including in Indonesia, makes early detection very important. We used Naïve Bayes algorithm method to classify the possibility of heart disease based on 303 Kaggle data with 14 features. Preprocessing, model training using Python on Google Colab, and feature selection (SelectKBest) are the process components. The results show that the model is effective to help early diagnosis with 83% accuracy and 97% recall in sick patients.

Keywords: Heart disease, classification, Naïve bayes, data mining, early diagnosis

This is an open access article under the CC BY-SA License.



I. PENDAHULUAN

Organisasi Kesehatan Dunia atau yang biasa disebut WHO melaporkan jika penyakit jantung menjadi salah satu penyakit yang mematikan diseluruh dunia. Penyakit jantung mencakup berbagai kondisi yang melibatkan jantung dan pembuluh darah. Jantung merupakan organ vital yang bertanggung jawab atas sistem peredaran darah, dengan fungsi utama memompa darah kaya oksigen ke seluruh tubuh untuk mendistribusikan nutrisi [1].Penyakit jantung terjadi ketika fungsi jantung atau sistem peredaran darah tidak berjalan sebagaimana mestinya. Menurut [2], beberapa faktor risiko utama yang menyebabkan penyakit jantung meliputi obesitas, diabetes, hipertensi, kolesterol tinggi, dan riwayat keluarga dengan penyakit serupa. Penyakit jantung sendiri sebanarnya bisa diketahui sejak kini, tapi karena kurangnya edukasi yang tersampaikan pada masyarakat Indonesia, hal seperti itu jadi terlupakan dan penyakit tersebut akan dibawa terus hingga dewasa [3].

Deteksi dini terhadap risiko penyakit jantung masih belum berjalan secara optimal, terutama yang berbasis teknologi. Banyak penderita baru mengetahui kondisi kesehatannya ketika penyakit sudah dalam tahap lanjut. Selain itu, sistem berbasis data yang dapat memberikan dukungan keputusan secara cepat dan tepat masih terbatas. Oleh karena itu, diperlukan pendekatan prediktif yang mampu mengolah data kesehatan untuk mengantisipasi potensi penyakit jantung sebelum gejalanya berkembang serius [4]. Peneliti memperoleh dataset dari kaggle dengan judul Klasifikasi Penyakit Jantung yang dibuat oleh Dirga

Halim Susilo, terdapat 14 kolom yang ada dalam data set ini yaitu umur, jenis kelamin, chest pain, tekanan darah, kolesterol, gula darah, hasil EKG, detak jantung, angina, depresi ST, segmen ST, jumlah pembuluh darah utama, thalassemia, dan label [5]. Jumlah kolom yang ada pada dataset sangat banyak sehingga membutuhkan sistem untuk melakukan klasifikasi sehingga dapat hasil yang akurat. Untuk mengelola dataset tersebut, kami menggunakan metode Data mining.

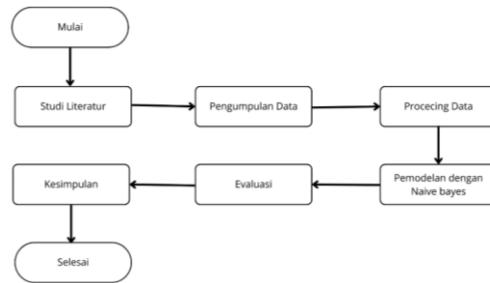
Menurut penelitian terdahulu dalam [6], Data mining adalah proses yang digunakan untuk mendukung pengambilan keputusan serta mengidentifikasi hubungan tersembunyi antar data yang sebelumnya tidak diketahui, dengan pendekatan yang mudah dipahami oleh pengguna. Data mining yang kami ambil metodenya adalah klasifikasi. Klasifikasi merupakan metode untuk menentukan sebuah anggota kedalam suatu kelas tertentu yang telah ditentukan sebelumnya, anggota tersebut dimasukan kedalam kelas tertentu berdasarkan persamaan karakter dari data tersebut [7]. Proses klasifikasi yang kami gunakan yaitu dengan Naïve Bayes. Pada [8] mengatakan, jika metode Naive Bayes juga dinilai berpotensi baik dalam mengklasifikasi dokumen dibandingkan metode pengklasifikasian yang lain dalam hal akurasi dan efisiensi komputasi

Penelitian yang dilakukan oleh Nur Aeni dan yang lainnya, bicara bahwa saat eksperimen awal yang dihasilkan dengan algoritma Naïve Bayes sebesar 82.14% dengan nilai AUC 0.686 dengan kategori poor classification. Pada eksperimen kedua dengan menggunakan metode Naïve Bayes juga perbasis PSO menjadi 92.86% dari nilai AUC 0.839 dengan kategori good classification. Terbukti pada eksperimen kedua dengan penambahan opsi dapat meningkatkan nilai akurasi [9].

Dengan demikian, penelitian ini ditujukan untuk membangun sistem klasifikasi penyakit jantung menggunakan algoritma Naïve Bayes, dengan harapan dapat menghasilkan tingkat akurasi yang optimal serta mendukung proses deteksi dini penyakit jantung.

II. METODE PENELITIAN

Untuk menyusun penelitian ini peneliti memerlukan metode-metode penelitian untuk memproses data yang telah kami peroleh. Tahapan-tahapan yang akan kami gunakan pada penelitian ini terdiri atas studi literatur, pengumpulan data, procecing data, pemodelan, evaluasi, dan kesimpulan. Untuk mempermudah pemahaman mengenai tahapan penelitian secara keseluruhan, berikut ditampilkan diagram alur dari proses penelitian ini.



Gambar 2.1 Alur Penelitian

A. Pengumpulan Data

Data yang kami gunakan pada penelitian ini diperoleh dari situs website Kaggle [5], yang merupakan platform kumpulan data set dengan kode sumber untuk keperluan analisis data. Dataset dengan judul

Klasifikasi Penyakit Jantung, yang berisi data riwayat medis dengan beberapa parameter pasien seperti usia, jenis kelamin, darah tinggi, diabetes, dan lain-lain.

B. Prosesing data

Langkah pertama yang harus dilakukan dalam pengecekan terhadap dataset pasien penyakit jantung yaitu melakukan tahapan Preprocessing, yang mana data preprocessing adalah proses membersihkan data dari noise[10]. Untuk mengatasi data yang mungkin memiliki ukuran database yang cukup besar dapat dibagi menjadi beberapa bagian atau dihapus pada bagian-bagian yang kemungkinan tidak digunakan untuk proses penelitian, sehingga dapat mempercepat scanning data saat algoritma data mining diterapkan [11].

C. Pemodelan dengan metode Naïve Bayes

Pada tahapan ini dilakukan pelatihan model klasifikasi dengan menggunakan metode Naïve Bates untuk mengetahui hasil memprediksi penyakit jantung [4]. Dataset penyakit jantung dijadikan sampel, untuk mulai perhitungan untuk mengetahui pasien tersebut terkласifikasi menderita penyakit jantung atau tidak yang dilihat berdasarkan probabilitas dari setiap kelas [12]. Dalam algoritma Naive Bayes, yang didasarkan pada teori Bayes, tiap fitur dianggap independen satu sama lain dan memberikan kontribusi yang seimbang untuk penentuan kelas [13]. Pengklasifikasi Naive Bayes mudah diimplementasikan dan secara komputasi cepat dan berkinerja baik pada kumpulan data besar yang memiliki dimensi tinggi. Selain itu juga kondusif untuk aplikasi waktu nyata dan tidak sensitif terhadap noise [13]. Dalam [14] mengatakan ada beberapa algoritma klasifikasi yang banyak digunakan secara luas, yaitu Decision/classification tree, Bayesian classifiers/ Naïve Bayes classifiers, Neural networks, Analisa Statistik, Algoritma Genetika, Rough sets, k-nearest neighbor, Metode Rule Based, Memory based reasoning, dan Support vector machines (SVM). Berikut rumus naïve bayes yang akan digunakan[15]:

$$P(C|X) = \frac{P(X|C_i) \cdot P(C_i)}{P(X)} \quad (1)$$

Keterangan:

X: data dengan class yang belum diketahui

Ci: Hipotesis data X yang merupakan suatu class spesifik

P(Ci|X): probabilitas hipotesis Ci berdasarkan kondisi X (posteriori probability)

P(Ci): probabilitas hipotesis Ci (priorprobability)

P(X|Ci): probabilitas X berdasarkan kondisi pada hipotesis Ci

P(X): probabilitas dari X

Untuk menilai kinerja metode klasifikasi yang telah dilakukan, proses berikutnya adalah evaluasi metrik akurasi. Untuk mengetahui tingkat keakurasaian model tersebut, maka perhitungan akurasi akan dilakukan dengan perumusan sebagai berikut[4]:

$$Akurasi = \frac{\sum \text{data benar}}{n} \times 100 \quad (2)$$

Keterangan:

Akurasi: Akurasi sistem

\sum Data Benar: Jumlah benar dalam sampel uji

n: Banyaknya jumlah sampel uji

III. HASIL DAN PEMBAHASAN

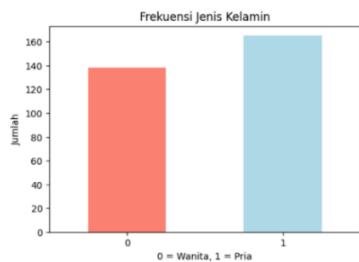
Dalam penelitian ini menggunakan algoritma Naive Bayes digunakan untuk mengklasifikasi penyakit jantung. Bahasa pemrograman yang digunakan adalah *python* dengan menggunakan Google Colab sebagai media eksperimen. Sumber data eksperimen berasal dari platform kaggle dengan nama Klasifikasi Penyakit Jantung yang memiliki 303 data pasien dengan 14 fitur. Masing - masing fitur akan dijelaskan dalam table 3.1 dibawah ini:

Table 3.1 Atribut Dataset

Atribut	Penjelasan
age	Usia pasien
sex	Jenis kelamin — 1 = laki - laki, 0 = perempuan.
cp	Tipe nyeri dada — 0 = typical angina, 1 = atypical angina, 2 = non-anginal pain, 3 = asymptomatic.
trestbps	Tekanan darah saat istirahat (mm Hg) — nilai > 130 - 140 dianggap tinggi
chol	Kadar kolesterol serum dalam mg/dl — nilai > 200 dianggap tinggi.
fbs	Gula darah puasa > 120 mg/dl (1 = true, 0 = false) — nilai > 126 mg/dl mengindikasi diabetes.
restecg	Hasil EKG saat istirahat — 0 = normal, 1 = ST-T abnormality, 2 = left ventricular hypertrophy.
thalach	Denyut jantung maksimum yang tercapai.
exang	Angina yang dipicu oleh latihan (1 = ya, 0 = tidak).
oldpeak	Indikator stress jantung.
slope	Kemiringan segmen ST latihan puncak — 0 = upsloping, 1 = flat, 2 = downsloping.
ca	Jumlah pembulu darah utama yang terlihat melalui fluoroskopi (0-3).
thal	Hasil stress thalium — 1,3 = normal, 6 = fixed defect, 7 = reversible defect.
target	Diagnosis penyakit jantung — 0 = tidak sakit, 1 = mengindap penyakit jantung.

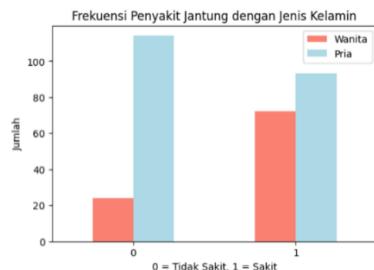
Sebelum memulai pelatihan model, langkah awal yang dilakukan adalah memuat sejumlah library seperti pandas dan numpy untuk keperluan manipulasi data, matplotlib dan seaborn untuk visualisasi, serta scikit-learn yang digunakan dalam tahap praproses dan pemodelan. Data diimpor dalam format CSV, kemudian diperiksa untuk mendeteksi adanya duplikasi maupun nilai hilang. Hasil pengecekan menunjukkan bahwa tidak diperlukan proses imputasi ataupun penghapusan data. Setelah data dinyatakan bersih, dilakukan analisis eksploratif untuk memahami distribusi tiap fitur, hubungan antar atribut, serta pola-pola yang berpengaruh terhadap proses klasifikasi.

A. Visualisasi Data Awal



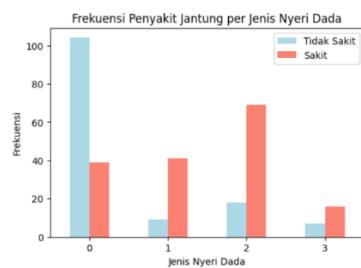
Gambar 1 Frekuensi Jenis Kelamin

Gambar 1 memperlihatkan distribusi pasien berdasarkan jenis kelamin. Terlihat bahwa jumlah pasien laki-laki (dengan kode 1) lebih tinggi dibandingkan pasien perempuan (kode 0). Ketidakseimbangan distribusi gender ini perlu diperhatikan, karena dapat memengaruhi performa model yang dibangun, khususnya jika variabel jenis kelamin memiliki keterkaitan dengan kondisi penyakit jantung.



Gambar 2 Frekuensi Penyakit Jantung dengan Jenis Kelamin

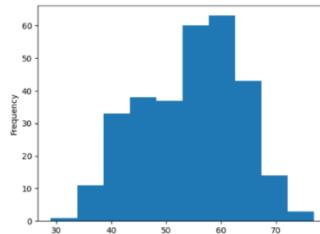
Gambar 2 memperlihatkan distribusi kasus penyakit jantung berdasarkan jenis kelamin, di mana kategori 0 merepresentasikan individu yang tidak menderita penyakit jantung dan kategori 1 menunjukkan individu yang menderita. Dari grafik terlihat bahwa jumlah pria lebih dominan pada kedua kategori. Meskipun pada kategori tidak sakit jumlah pria dan wanita relatif seimbang, dominasi jumlah pria tetap terlihat pada kategori sakit. Temuan ini mengindikasikan bahwa kasus penyakit jantung juga cukup signifikan terjadi pada wanita, namun memperkuat indikasi bahwa pria cenderung memiliki risiko lebih tinggi terhadap penyakit jantung.



Gambar 3 Frekuensi penyakit jantung per jenis nyeri dada

Gambar 3 menyajikan distribusi kasus penyakit jantung berdasarkan empat tipe nyeri dada yang diklasifikasikan dalam kode 0 hingga 3. Warna biru menunjukkan pasien yang tidak menderita penyakit jantung, sedangkan warna pink menandakan pasien yang terdiagnosa. Tipe nyeri dada 0 (typical angina) paling sering dialami oleh pasien tanpa riwayat penyakit jantung. Sebaliknya, tipe 2 (non-anginal pain) paling umum dijumpai pada pasien yang memiliki penyakit jantung. Adapun tipe 1 (atypical angina) dan

tipe 3 (asymptomatic) juga menunjukkan kecenderungan yang sama, di mana jumlah penderita penyakit jantung lebih tinggi dibandingkan yang tidak terdiagnosa.



Gambar .4 frekuensi usia pasien

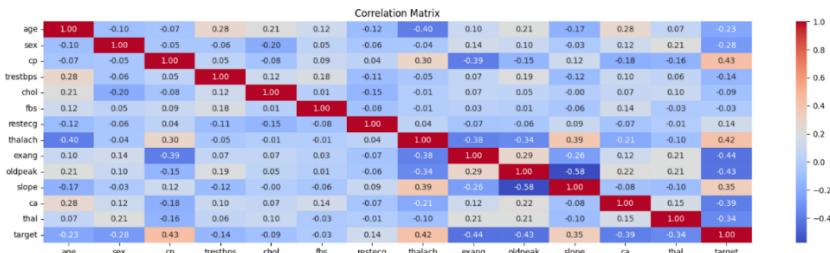
Gambar 4 menunjukkan frekuensi pasien berdasarkan kelompok usia. Sumbu horizontal merepresentasikan usia pasien, sedangkan sumbu vertikal menunjukkan jumlah pasien pada setiap usia. Dari grafik terlihat bahwa kasus penyakit jantung paling banyak terjadi pada individu berusia antara 50 hingga 62 tahun. Temuan ini mengindikasikan bahwa orang dengan usia paruh baya hingga lanjut usia memiliki risiko yang lebih tinggi terhadap penyakit jantung.



Gambar 5 Penyakit Jantung dalam Usia dan Denyut Jantung Maksimal

Gambar 5 memperlihatkan hubungan antara usia dan denyut jantung maksimal pada pasien, yang dikategorikan berdasarkan status penyakit jantung. Titik-titik merah merepresentasikan pasien dengan penyakit jantung, sedangkan titik-titik biru menunjukkan pasien yang sehat. Berdasarkan grafik, terlihat bahwa pasien penderita penyakit jantung memiliki denyut jantung maksimal yang tinggi dan berasal dari rentang usia yang bervariasi, khususnya antara 40 hingga 60 tahun. Sebaliknya, pasien tanpa penyakit jantung menunjukkan pola usia dan denyut jantung yang lebih seimbang, dengan nilai denyut jantung maksimal yang cenderung lebih rendah. Temuan ini menunjukkan bahwa meskipun usia dan denyut jantung tidak dapat dijadikan indikator tunggal untuk mendeteksi penyakit jantung, keduanya tetap bisa menjadi sinyal awal yang bermanfaat bila dikombinasikan dengan fitur lainnya.

B. Analisis Korelasi Fitur



Gambar 6 Korelasi Matrix

Gambar 6 menunjukkan bahwa beberapa variabel memiliki korelasi kuat terhadap target. Variabel cp, thalach, dan slope menunjukkan korelasi positif masing-masing sebesar 0.43, 0.42, dan 0.35, yang mengindikasikan peningkatan nilai variabel tersebut berkaitan dengan risiko penyakit jantung. Sebaliknya, exang, oldpeak, ca, dan thal memiliki korelasi negatif antara -0.34 hingga -0.44. Sementara itu, age, sex, dan chol menunjukkan hubungan yang lemah terhadap target.

Setelah fitur-fitur yang relevan dianalisis, data disiapkan untuk pelatihan model dengan membagi dataset ke dalam dua variabel: X sebagai fitur (input) dan Y sebagai target (output). Pemisahan ini bertujuan agar model dapat mempelajari hubungan antara input dan output tanpa terjadi kebocoran data. Dataset kemudian dibagi menjadi 80% data latih dan 20% data uji untuk menghindari bias saat evaluasi. Selanjutnya, fitur dinormalisasi menggunakan MinMaxScaler ke rentang 0–1 agar seluruh fitur memiliki skala yang setara dan tidak ada yang mendominasi proses pelatihan.

Setelah data disiapkan, proses seleksi fitur menggunakan metode SelectKBest dimulai. Ini digunakan untuk memilih fitur yang berpengaruh terhadap target sambil mengurangi fitur yang tidak penting agar model tidak menjadi lebih kompleks. Ada delapan fitur yang terpilih, yaitu c, thalach, exang, oldpeak, slope, ca, thal, dan sex.

Setelah data dipersiapkan, fitur dipilih menggunakan metode SelectKBest untuk menyaring delapan variabel paling berpengaruh: cp, thalach, exang, oldpeak, slope, ca, thal, dan sex. Tujuan seleksi ini adalah menyederhanakan model dan meningkatkan kinerjanya. Model Naïve Bayes kemudian dilatih dengan fitur tersebut menggunakan model.fit(X_train_selected, y_train), yang menghitung probabilitas awal tiap kelas dan peluang bersyarat tiap fitur. Proses klasifikasi didasarkan pada Teorema Bayes, untuk memprediksi kelas dengan kemungkinan tertinggi pada data uji.

$$P(C_i|X) = \frac{P(X|C_i) \cdot P(C_i)}{P(X)} \quad (3)$$

Dalam Teorema Bayes, probabilitas kelas berdasarkan data dihitung dari peluang awal kelas dan kecocokan data terhadap kelas tersebut. Setelah model dilatih, prediksi dilakukan pada data uji dan dibandingkan dengan label asli untuk mengukur akurasi.

```
Confusion Matrix:
[[19  9]
 [ 1 32]]
```

Gambar 3.2.2 Confusion Matrix

Untuk mengevaluasi performa model, digunakan confusion matrix setelah proses prediksi pada data uji. Hasilnya menunjukkan 32 prediksi benar untuk pasien dengan penyakit (true positive) dan 19 untuk yang tidak sakit (true negative). Meski begitu, masih ditemukan 9 kesalahan klasifikasi positif dan 1 negatif, yang menandakan model belum sepenuhnya akurat

Classification Report:					
	precision	recall	f1-score	support	
0	0.95	0.68	0.79	28	
1	0.78	0.97	0.86	33	
accuracy			0.84	61	
macro avg	0.87	0.82	0.83	61	
weighted avg	0.86	0.84	0.83	61	

Gambar 3.3.3 Kalsifikasi Report

Model mencapai akurasi 83% dengan precision tinggi pada kelas 0 (95%), artinya prediksi untuk pasien sehat sebagian besar tepat. Namun, recall-nya hanya 68%, menunjukkan model belum cukup sensitif

terhadap pasien sehat. Sebaliknya, recall pada pasien sakit (kelas 1) sangat tinggi, yaitu 97%, dengan F1-score sebesar 0.86, menandakan keseimbangan baik antara presisi dan sensitivitas. Artinya, model cukup andal dalam mengenali pasien yang benar-benar sakit, meskipun ada kecenderungan mengklasifikasikan sebagian pasien sehat sebagai sakit. Dalam konteks medis, kepekaan tinggi lebih diutamakan agar tidak ada kasus serius yang terlewat.

IV. KESIMPULAN

Model klasifikasi untuk penyakit jantung berhasil dikembangkan dalam penelitian ini melalui algoritma Naive Bayes, berdasarkan dataset Kaggle yang mencakup 303 pasien dan 14 atribut medis. Model yang dikembangkan menunjukkan akurasi sebesar 83% melalui proses preprocessing, pemodelan, dan evaluasi. Dengan tingkat akurasi yang tinggi, yaitu 97% pada kelas pasien yang sakit, model ini sangat efektif dalam mendeteksi pasien yang benar-benar menderita penyakit jantung. Meskipun ada beberapa kesalahan dalam klasifikasi pasien sehat, hasil ini tetap dianggap baik, terutama dalam lingkungan medis di mana lebih baik salah mengklasifikasikan pasien sehat sebagai sakit daripada melewatkannya pasien yang benar-benar sakit. Penggunaan Naïve Bayes terbukti efisien, sederhana, dan cepat dalam memproses data, juga cocok untuk pengklasifikasian awal dalam sistem deteksi penyakit jantung.

DAFTAR PUSTAKA

- [1] A. A. Surya and Y. Yamasari, “Penerapan Algoritma Naïve Bayes (NB) untuk Klasifikasi Penyakit Jantung,” *J. Informatics Comput. Sci.*, vol. 5, no. 03, pp. 447–455, 2024, doi: 10.26740/jinacs.v5n03.p447-455.
- [2] A. Yogianto, A. Homiadi, and Z. Fatah, “Implementasi Metode K-Nearest Neighbors (KNN) untuk Klasifikasi Penyakit Jantung,” *G-Tech J. Teknol. Terap.*, vol. 8, no. 3, pp. 1720–1728, 2024, doi: 10.33379/gtech.v8i3.4495.
- [3] B. Hirwono, A. Hermawan, and D. Avianto, “Implementasi Metode Naïve Bayes untuk Klasifikasi Penderita Penyakit Jantung,” *J. JTIK (Jurnal Teknol. Inf. dan Komunikasi)*, vol. 7, no. 3, pp. 450–457, 2023, doi: 10.35870/jtik.v7i3.910.
- [4] D. Larassati, A. Zaidiah, and S. Afrizal, “Sistem Prediksi Penyakit Jantung Koroner Menggunakan Metode Naive Bayes,” *JIPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.)*, vol. 7, no. 2, pp. 533–546, 2022, doi: 10.29100/jipi.v7i2.2842.
- [5] D. H. SUSILO, “Klasifikasi Penyakit Jantung.” Accessed: Apr. 29, 2025. [Online]. Available: <https://www.kaggle.com/code/dirgahalimsusilo/klasifikasi-penyakit-jantung/notebook>
- [6] Q. A'yuniyah *et al.*, “Implementasi Algoritma Naïve Bayes Classifier (NBC) untuk Klasifikasi Penyakit Ginjal Kronik,” *J. Sist. Komput. dan Inform.*, vol. 4, no. 1, p. 72, 2022, doi: 10.30865/json.v4i1.4781.
- [7] M. A. Bianto, K. Kusrini, and S. Sudarmawan, “Perancangan Sistem Klasifikasi Penyakit Jantung

Menggunakan Naïve Bayes,” *Creat. Inf. Technol. J.*, vol. 6, no. 1, p. 75, 2020, doi: 10.24076/citec.2019v6i1.231.

- [8] G. P. Kawani, “Implementasi Naive Bayes,” *J. Informatics, Inf. Syst. Softw. Eng. Appl.*, vol. 1, no. 2, pp. 73–81, 2019, doi: 10.20895/inista.v1i2.73.
- [9] C. S. Nur Aeni Widiastuti, Stefanus Santosa, “ALGORITMA KLASIFIKASI DATA MININGNAÏVE BAYES BERBASIS PARTICLE SWARM OPTIMIZATION UNTUK DETEKSI PENYAKIT JANTUNG.pdf,” *Nat. Methods*, vol. 1, no. 1, p. 4, 2014, doi: <https://doi.org/10.33369/pseudocode.1.1.11-14>.
- [10] A. Riani, Y. Susianto, and N. Rahman, “Implementasi Data Mining Untuk Memprediksi Penyakit Jantung Menggunakan Metode Naive Bayes,” *J. Innov. Inf. Technol. Appl.*, vol. 1, no. 01, pp. 25–34, 2019, doi: 10.35970/jinita.v1i01.64.
- [11] D. Gunawan, “Evaluasi Performa Pemecahan Database dengan Metode Klasifikasi Pada Data Preprocessing Data mining,” *Khazanah Inform. J. Ilmu Komput. dan Inform.*, vol. 2, no. 1, pp. 10–13, 2016, doi: 10.23917/khif.v2i1.1749.
- [12] A. F. Riany and G. Testiana, “Penerapan Data Mining untuk Klasifikasi Penyakit Jantung Koroner Menggunakan Algoritma Naïve Bayes,” *MDP Student Conf.*, vol. 2, no. 1, pp. 297–305, 2023, doi: 10.35957/mdp-sc.v2i1.4388.
- [13] Dimsyiar M Al Hafiz, Khoirul Amaly, Javen Jonathan, M Teranggono Rachmatullah, and Rosidi, “Sistem Prediksi Penyakit Jantung Menggunakan Metode Naive Bayes,” *J. Rekayasa Elektro Sriwij.*, vol. 2, no. 2, pp. 151–157, 2021, doi: 10.36706/jres.v2i2.29.
- [14] H. Annur, “Klasifikasi Masyarakat Miskin Menggunakan Metode Naive Bayes,” *Ilk. J. Ilm.*, vol. 10, no. 2, pp. 160–165, 2018, doi: 10.33096/ilkom.v10i2.303.160-165.
- [15] H. F. Putro, R. T. Vulandari, and W. L. Y. Saptomo, “Penerapan Metode Naive Bayes Untuk Klasifikasi Pelanggan,” *J. Teknol. Inf. dan Komun.*, vol. 8, no. 2, 2020, doi: 10.30646/tikomsin.v8i2.500.