

# Studi Literatur Review Penerapan Data Mining Untuk Prediksi Penyakit Jantung Menggunakan *Naïve Bayes*

Anisa Rizki Septia\*, Hetty Rohayani

Informatika, Sains dan Teknologi, Universitas Muhammadiyah Jambi, Kota Jambi, Indonesia

Jl.Kapt.Pattimura Simpang IV Sipin, Kota Jambi, Provinsi Jambi, Indonesia

Email: <sup>1,\*</sup>aseptia597@gmail.com, <sup>2</sup>hettyrohayani@gmail.com

Email Penulis Korespondensi: aseptia597@gmail.com

**Abstrak**— Penyakit jantung merupakan salah satu penyebab utama kematian global yang sering kali sulit terdeteksi sejak dini akibat manifestasi gejala klinis yang tidak spesifik. Untuk mengatasi masalah keterbatasan diagnosis manual tersebut, penerapan teknik data mining memanfaatkan algoritma *Naïve Bayes* hadir sebagai solusi komputasi yang efisien dan akurat. Penelitian ini bertujuan untuk menganalisis dan memetakan efektivitas implementasi *Naïve Bayes* dalam memprediksi penyakit jantung melalui pendekatan Systematic Literature Review (SLR). Kontribusi dari penelitian ini adalah menyajikan panduan taksonomi yang komprehensif mengenai pengaruh geometri data, teknik preprocessing, serta integrasi metode seleksi fitur terhadap optimalisasi performa model probabilistik. Hasil penelaahan literatur menunjukkan bahwa tingkat akurasi model bervariasi antara 58% hingga 91,80%, dengan mayoritas performa stabil pada kisaran 79%-91% yang dipengaruhi secara deterministik oleh kualitas reduksi dimensi data. Secara keseluruhan, proses penambangan data berbasis *Naïve Bayes* terbukti memiliki potensi besar sebagai sistem pendukung keputusan klinis dalam mendukung tindakan preventif medis secara dini.

**Kata Kunci:** Penyakit Jantung; Data Mining; *Naïve Bayes*; Prediksi Penyakit; Deteksi Dini

**Abstract**— Heart disease is one of the leading causes of global death, often difficult to detect early due to non-specific clinical symptoms. To overcome the limitations of manual diagnosis, the application of data mining techniques utilizing the *Naïve Bayes* algorithm presents an efficient and accurate computational solution. This study aims to analyze and map the effectiveness of *Naïve Bayes* implementation in predicting heart disease through a Systematic Literature Review (SLR) approach. The contribution of this study is to provide a comprehensive taxonomic guide regarding the influence of data geometry, preprocessing techniques, and the integration of feature selection methods on optimizing the performance of probabilistic models. The results of the literature review indicate that the model accuracy level varies between 58% and 91.80%, with the majority of performance stable in the range of 79%-91% which is deterministically influenced by the quality of data dimensionality reduction. Overall, the *Naïve Bayes*-based data mining process has proven to have great potential as a clinical decision support system in supporting early medical preventive measures.

**Keywords:** Heart Disease; Data Mining; *Naïve Bayes*; Disease Predictions; Early Detection

## 1. PENDAHULUAN

Penyakit jantung sampai saat ini masih menduduki peringkat atas sebagai salah satu penyebab angka kematian (mortalitas) tertinggi di dunia, tidak terkecuali di Indonesia. Kompleksitas gangguan kardiovaskular sering kali menjadi ancaman serius karena manifestasi klinis atau gejala awalnya yang sangat samar dan tidak spesifik. Gejala umum seperti rasa lelah yang berkepanjangan, nyeri dada intensitas ringan, atau sesak napas pendek sering kali diabaikan oleh masyarakat awam. Akibatnya, sebagian besar pasien baru melakukan pemeriksaan medis ketika kondisi patologis jantungnya telah memasuki fase kritis atau stadium lanjut, yang secara drastis menurunkan probabilitas keberhasilan terapi medis dan meningkatkan risiko kematian mendadak. Oleh karena itu, urgensi mengenai pengembangan sistem deteksi dini (*early warning system*) yang cepat, murah, dan akurat menjadi hal yang sangat krusial bagi dunia kesehatan modern saat ini demi meminimalkan tingkat fatalitas penyakit tersebut.

Di antara berbagai algoritma klasifikasi yang tersedia dalam ilmu kecerdasan buatan, algoritma *Naïve Bayes* menjadi salah satu metode yang paling populer dan sering diimplementasikan oleh para peneliti untuk skenario prediksi medis. Algoritma ini bekerja menggunakan pendekatan probabilistik yang mengacu pada Teorema Bayes, disertai dengan sebuah asumsi khas yang "naif", yaitu menganggap bahwa setiap atribut data atau gejala klinis bersifat independen dan tidak saling memengaruhi satu sama lain. Alasan mendasar di balik tingginya frekuensi penggunaan *Naïve Bayes* adalah kesederhanaan strukturnya, efisiensi waktu komputasi yang sangat cepat karena hanya membutuhkan satu kali pemindaian data (*single-pass learning*), serta performa akurasinya yang dinilai cukup tangguh meskipun dihadapkan pada dataset dengan dimensi atribut yang kompleks.

Dengan demikian, dalam ranah riset empiris, penerapan *Naïve Bayes* standar tidak luput dari kendala akademik. Kelemahan mutlak dari asumsi independensi antar-fitur sering kali bertolak belakang dengan kondisi riil kedokteran, di mana gejala klinis pasien pada kenyataannya saling berkaitan erat. Permasalahan ini memicu terjadinya fluktuasi nilai akurasi yang cukup lebar pada hasil-hasil penelitian terdahulu. Selain itu, belum meratanya standarisasi maupun pemetaan yang komprehensif mengenai metode optimasi apa yang paling efektif untuk menutupi kelemahan *Naïve Bayes* pada data penyakit jantung, memunculkan kebutuhan mendesak akan sebuah kajian literatur terstruktur.

Untuk memahami posisi penelitian ini, diperlukan penelaahan terhadap penelitian sejenis serta metode-metode perbandingan yang lazim digunakan oleh peneliti terdahulu. Sebagai contoh, Sahar(2020) melakukan studi komparatif secara langsung antara metode *K-Nearest Neighbor* (KNN) dan *Naïve Bayes Classifier* (NBC) menggunakan dataset 7.000 sampel dari *UCI Machine Learning Repository*. Hasil riset tersebut menunjukkan bahwa metode perbandingan KNN mengungguli NBC dengan akurasi 67% berbanding 58%, yang membuktikan bahwa *Naïve Bayes* tanpa pemrosesan

lanjutan akan mengalami penurunan performa drastis ketika menangani dataset besar yang memiliki banyak *noise* [1]. Di sisi lain, penelitian oleh Khodijah dan Sriyanto (2023) menggunakan algoritma pohon keputusan C4.5 dan *Random Forest* sebagai pembanding untuk memprediksi gagal jantung pada 918 observasi data Kaggle, di mana *Naïve Bayes* mampu bersaing dengan menorehkan akurasi sebesar 87,25%[2]. Evaluasi komparatif juga dilakukan oleh Putri Anasia Sihotang dan Delima Sitanggang (2024), yang membandingkan algoritma C4.5 dengan *Naïve Bayes*. Hasilnya menunjukkan bahwa *Naïve Bayes* justru lebih unggul dengan akurasi 84% dan nilai *Area Under the Curve* (AUC) mencapai 93%, mengindikasikan sensitivitas klasifikasi medis yang sangat baik [3].

Beberapa peneliti terdahulu juga mencoba melakukan modifikasi struktural maupun integrasi metode penapis untuk meningkatkan performa *Naïve Bayes*. Stevani M. Meilisa dkk. (2023) memperkenalkan variasi berupa pendekatan *Correlated Naïve Bayes* pada dataset Kaggle guna mengatasi kelemahan independensi fitur, yang berhasil memperoleh akurasi stabil sebesar 83,15% serta sukses memetakan empat atribut klinis paling berpengaruh [4]. Selanjutnya, Bowo Hirwono dkk. (2023) menitikberatkan risetnya pada proses seleksi dan pembersihan data mentah sebelum divalidasi dengan *Naïve Bayes* standar, menghasilkan tingkat akurasi 86,84% [5]. Modifikasi mutakhir ditunjukkan oleh Siti Roziana Azizah dkk. (2023) yang menggabungkan *Naïve Bayes* dengan teknik seleksi fitur *hybrid* berbasis *Filter* dan *Wrapper*, sehingga berhasil mencatatkan rekor akurasi tertinggi sebesar 91,80% [6]. Dampak positif dari reduksi dimensi data juga diamini oleh Alifa Lutfia dkk. (2024) yang menyandingkan seleksi fitur *Gain Ratio* (GR) dengan *Naïve Bayes*, sehingga mendongkrak performa klasifikasi hingga mencapai akurasi 91,2% [7].

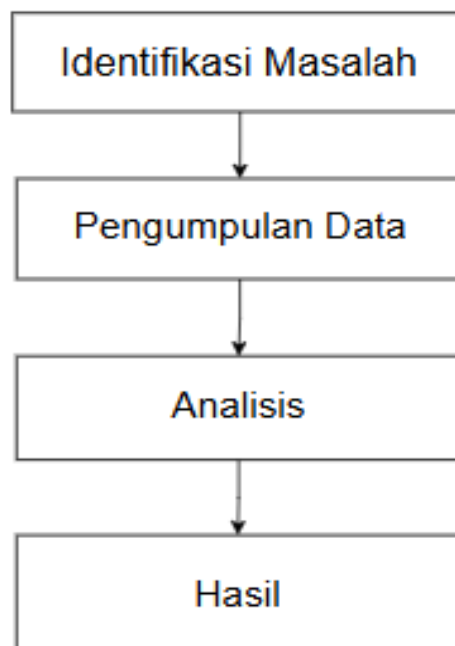
Meskipun terdapat banyak publikasi ilmiah mengenai topik ini, terdapat kesenjangan penelitian yang nyata. Mayoritas penelitian terdahulu hanya berfokus pada pengujian eksperimental satu atau dua model optimasi secara lokal pada satu dataset tertentu saja. Belum ada kajian meta-analisis berskala *Systematic Literature Review* (SLR) yang secara khusus merangkum, membandingkan, dan menguji benang merah teoritis mengenai korelasi antara porsi pembagian data (*data splitting*), pengaruh teknik reduksi dimensi fitur, serta dampak algoritma metaheuristik terhadap naik-turunnya nilai akurasi *Naïve Bayes* lintas-dataset medis. Ketiadaan standardisasi komparatif ini menyulitkan para praktisi sistem pakar kedokteran untuk menentukan arsitektur algoritma yang paling ideal.

Oleh karena itu, penelitian ini dilaksanakan dengan tujuan utama untuk melakukan analisis mendalam dan merangkum secara sistematis berbagai temuan ilmiah terdahulu mengenai implementasi algoritma *Naïve Bayes* khusus pada kasus prediksi penyakit jantung. Lebih lanjut, penelitian ini bertujuan untuk memetakan efektivitas taksonomi metode seleksi fitur, pengaruh rasio pembagian data latihan-uji, serta merumuskan rekomendasi arsitektur hibrida *Naïve Bayes* yang paling optimal sebagai landasan pengembangan sistem pendukung keputusan klinis (*clinical decision support system*) di masa mendatang.

## 2. METODOLOGI PENELITIAN

### 2.1 Tahapan Penelitian

Penelitian ini dirancang menggunakan metode *Systematic Literature Review* (SLR) yang mengkaji secara komprehensif literatur ilmiah terkait aplikasi data mining pada deteksi penyakit jantung. Tahapan pelaksanaan penelitian disusun secara terstruktur sebagai berikut:



Gambar 1. Diagram Penelitian

a. **Identifikasi Masalah**

Tahap awal untuk merumuskan masalah mengenai tingginya risiko fatal penyakit jantung akibat keterlambatan diagnosis dini. Di sini ditentukan urgensi untuk menganalisis sejauh mana efektivitas algoritma klasifikasi *Naive Bayes* dalam membantu ketepatan prediksi medis.

b. **Pengumpulan Data**

Data dikumpulkan dari artikel ilmiah bereputasi yang diterbitkan dalam rentang waktu 5 tahun terakhir (2020-2025). Proses pencarian dilakukan secara digital melalui database Google Scholar menggunakan kombinasi kata kunci: "*Naive Bayes*", "*Data Mining*", dan "*Prediksi Penyakit Jantung*". relevansi isi dengan topik penelitian, serta kelengkapan informasi seperti metode, data yang digunakan, dan hasil evaluasi model.[8]

c. **Analisis**

Artikel yang telah dikumpulkan kemudian disaring ketat melalui kriteria inklusi (artikel wajib membahas penerapan *Naive Bayes* langsung pada objek penyakit jantung). Informasi penting seperti variasi dataset, jumlah atribut, teknik pemrosesan awal (*preprocessing*), kombinasi seleksi fitur, hingga metrik akurasi diekstraksi untuk dianalisis.

d. **Hasil**

Melakukan komparasi mendalam terhadap hasil ekstraksi data untuk memetakan kekuatan, kelemahan, serta potensi optimalisasi algoritma *Naive Bayes* guna penyusunan rekomendasi sistem pendukung keputusan medis di masa depan.[9]

### 3. HASIL DAN PEMBAHASAN

Pada bagian ini, kami menyajikan hasil studi literatur yang telah dilakukan serta pembahasan terkait Penerapan Data Mining Untuk Prediksi Penyakit Jantung Menggunakan *Naive Bayes*. Studi ini dilakukan dengan pendekatan Systematic Literature Review (SLR), yang bertujuan untuk mengidentifikasi, mengevaluasi, dan mensintesis penelitian-penelitian sebelumnya yang relevan. Hasil kajian difokuskan pada efektivitas, kelebihan, serta tantangan penggunaan algoritma *Naive Bayes* dalam Prediksi Penyakit Jantung. Selain itu, bagian ini juga dilengkapi dengan penyajian data dalam bentuk tabel, penjelasan, serta ilustrasi yang mendukung pemahaman hasil studi.[8]

**Tabel 1.** Penggunaan Algoritma *Naive Bayes* Dalam Prediksi Penyakit Jantung

Nama	Judul Jurnal	Data	Metode	Hasil
Sahar (2020)	Analisis Perbandingan Metode <i>K-Nearest Neighbor</i> dan <i>Naive Bayes Classifier</i> Pada Data Set Penyakit Jantung[1]	7.000 sampel pasien, 11 atribut klinis (UCI Repository)	Komparasi langsung: <i>KNN</i> vs <i>Naive Bayes Classifier</i> (NBC)	Akurasi KNN 67%; NBC 58%. Performa NBC rendah tanpa adanya manipulasi seleksi fitur pada data berskala besar.
Stevani M Sagarung, Meilisa S Eko Hari Parmadi (2023)	Pendekatan <i>Correlated Naive Bayes</i> Pada Klasifikasi Potensi Penyakit Jantung[4]	918 records, 11 atribut, 1 label target (Kaggle)	Pendekatan <i>Correlated Naive Bayes</i> untuk mereduksi kelemahan independensi fitur.	Akurasi tertinggi 83,15% pada rasio pembagian data latih-uji 80:20.
Bowo Hermawan, Hirwono, Arief Donny Avianto (2023)	Implementasi Metode <i>Naive Bayes</i> Untuk Klasifikasi Penderita Penyakit Jantung[5]	303 records, 13 variabel klinis (UCI Public Dataset)	Penguatan pada fase <i>preprocessing</i> data dan standarisasi <i>splitting</i> .	Nilai akurasi mencapai 86,84%. Algoritma bawaan ( <i>standard</i> ) mampu tampil sangat baik pasca pembersihan data.
Sultan Rahma Putri Dina Fernanda, Mardika (2025)	Implementasi Data Mining Untuk Mendiagnosa Penyakit Jantung Dengan Algoritma <i>Naive Bayes</i> Di RS TK II MOH.RIDWAN MAURAKSA[10]	25 data uji klinis riil, diintegrasikan dengan data basis Kaggle	Implementasi algoritma berbasis teori Bayes konvensional di RS TK II Moh. Ridwan Mauraksa	Akurasi sebesar 80% pada pengujian sampel data uji medis skala kecil.

Nama	Judul Jurnal	Data	Metode	Hasil
Siti Roziana Azizah, Rudy Hertanto, Andi Farmadi, Dwi Kartini, Irwan Budiman (2023)	Kombinasi Seleksi Fitur Berbasis Filter Dan Wrapper Menggunakan <i>Naïve Bayes</i> Pada Klasifikasi Penyakit Jantung[6]	Rekam medis penyakit jantung (UCI Machine Learning Repository)	Kombinasi seleksi fitur berbasis filter ( <i>Information Gain &amp; Gain Ratio</i> ) dan <i>Wrapper</i> .	Memecahkan rekor akurasi tertinggi sebesar 91,80% melalui teknik integrasi fitur ( <i>union</i> ).
Khodijah, Sriyanto (2023)	Perbandingan kinerja Algoritma C4.5, <i>Naïve Bayes</i> Dan Random Forest Dalam Prediksi Penyakit jantung[2]	918 observasi data profil gagal jantung, 12 atribut (Kaggle)	Analisis kinerja komparatif multi-algoritma: C4.5, <i>Naïve Bayes</i> , dan <i>Random Forest</i> .	<i>Naïve Bayes</i> mencatatkan nilai akurasi yang kompetitif sebesar 87,25%.
Agung Suko Wijoyo, Arief Jananto (2023)	Rancang Bangun Sistem Informasi Prediksi Penyakit Jantung Berbasis Algoritma <i>Naïve Bayes</i> [11]	Dataset publik Kaggle, 11 atribut numerik dan biner	Rancang bangun sistem informasi prediksi berbasis <i>Naïve Bayes Classifier</i> (NBC)	Akurasi terbaik diraih pada rasio data latih 90% dengan perolehan hasil sebesar 84%.
Atthohiroh, Rama Ayu, Sri Maharani (2023)	Penerapan Metode <i>Naïve Bayes</i> Dalam Memprediksi Penyakit Jantung[12]	10 sampel data klinis uji coba (Kaggle)	Diagnosis probabilistik berbantuan <i>Machine Learning</i>	Rasio pengujian data 80:20 menghasilkan performa akurasi terbaik sebesar 81%.
Mahdiawan Nurkholifah, Andri Nofiar.Am, Fenty Kurnia Oktorina (2023)	Analisa Penyakit Jantung Menggunakan Algoritma <i>Naïve Bayes</i> [13]	Dataset pasien penyakit jantung terkait stroke (Kaggle Open Dataset), 14 atribut	Klasifikasi probabilistik <i>Naïve Bayes</i> terintegrasi.	Diperoleh akurasi prediksi sebesar 83% untuk penanganan kasus komorbiditas stroke dengan rasio 80:20.
Romdan Muhamad Ubaidillah, Toto Andri Puspito (2025)	Optimasi Prediksi Penyakit Jantung Dengan <i>Naïve Bayes</i> Dan <i>Particle Swarm Optimization(PSO)</i> [14]	<i>Heart failure clinical records dataset</i> (UCI Repository)	Optimasi hibrida: <i>Naïve Bayes</i> + <i>Particle Swarm Optimization</i> (PSO)	Akurasi meningkat secara signifikan dari 79,12% (standar) menjadi 86,37% pasca optimasi fitur swarm.
Putri Anasia Sihotang, Delima Sitanggang (2024)	Perbandingan algoritma C4.5 Dengan <i>Naive Bayes</i> Untuk Memprediksi Penyakit Jantung[3]	Dataset klinis penyakit kardiovaskular (UCI Repository)	Analisis perbandingan kinerja model: Algoritma C4.5 vs <i>Naïve Bayes</i>	<i>Naïve Bayes</i> terbukti lebih unggul dengan raih akurasi 84% dan kekuatan diskriminasi kelas (AUC) sebesar 93%.
Alifa Lutfia, Gunawan, Ramdhan Saepul Rohman, A Gunawan (2024)	Penerapan Seleksi Fitur Gain Ratio Pada Prediksi Penyakit Jantung Berbasis <i>Naïve Bayes</i> [7]	Dataset publik rekam medis jantung (UCI Repository)	Penerapan algoritma seleksi fitur berbasis <i>Gain Ratio</i> (GR) sebelum klasifikasi.	Mampu mendongkrak performa klasifikasi secara signifikan hingga mencapai akurasi sebesar 91,2%.

Nama	Judul Jurnal	Data	Metode	Hasil
Ivana Alhabib,Ahmad Faqih, Fatihanursari Dikananda (2022)	Komparasi Metode <i>Deep Learning,Naïve Bayes</i> Dan Random Forest Untuk Prediksi Penyakit Jantung[15]	Dataset analisis dan prediksi serangan jantung (Kaggle)	Komparasi multi-model: <i>Deep Learning, Naïve Bayes, dan Random Forest</i>	Algoritma pembandingan <i>Deep Learning</i> memimpin dengan akurasi 83,49%, sedikit di atas <i>Naïve Bayes standard</i> .
Oskar,Frisiska Ria,Rupina, Noviyanti (2024)	Pengelolaan Data Penyakit Jantung Dengan Menggunakan Metode <i>Naïve Bayes</i> [16]	Dataset rekam medis jantung terstandardisasi.	Pengelolaan dan pengujian sistem klasifikasi data penyakit jantung.	Menunjukkan bahwa <i>Naïve Bayes</i> mampu melakukan pengelompokan secara efektif dengan capaian akurasi stabil di angka 79%.

Berdasarkan data ringkasan literatur yang telah dipetakan, tampak jelas bahwa ukuran geometri dataset memberikan pengaruh yang sangat signifikan terhadap kinerja prediksi algoritma *Naïve Bayes*. Ketika diuji pada dataset dengan ukuran sampel kecil hingga menengah, algoritma *Naïve Bayes* standar mampu bekerja secara optimal dan menghasilkan akurasi yang cukup tinggi. Hal ini dibuktikan melalui penelitian Bowo Hirwono dkk. yang menggunakan sampel sebanyak 303 rekam medis serta riset Khodijah dkk. dengan 918 sampel data, di mana masing-masing berhasil mencatatkan nilai akurasi sebesar 86,84% dan 87,25%. Namun, performa model standar ini mengalami penurunan yang sangat drastis ketika dihadapkan langsung pada dataset skala besar tanpa adanya proses seleksi fitur atau pembersihan data yang mendalam. Fenomena tersebut terlihat jelas pada studi Sahar (2020) yang menguji 7.000 sampel pasien, di mana tingkat akurasi *Naïve Bayes* jatuh hingga menyentuh angka 58%. Penurunan akurasi pada data bervolume besar ini terjadi karena penumpukan atribut mentah memicu akumulasi *noise* atau gangguan data yang mengacaukan kalkulasi nilai probabilitas akhir model.

Faktor mendasar lain yang membatasi performa *Naïve Bayes* konvensional di sektor medis adalah adanya asumsi independensi bersyarat yang menganggap seluruh gejala klinis pasien berdiri sendiri dan tidak saling memengaruhi. Pada realitas patofisiologi kedokteran, asumsi tersebut kurang realistis karena gejala klinis penyakit jantung pada kenyataannya saling berkaitan erat, seperti hubungan biologis antara detak jantung maksimal dengan karakteristik tipe nyeri dada yang dirasakan pasien. Ketika model memaksa menghitung peluang dari setiap indikator medis tersebut secara terpisah, model rentan kehilangan informasi kontekstual penting yang berujung pada kekeliruan prediksi. Sebagai solusi teoretis untuk mengatasi kelemahan asumsi ini, implementasi pendekatan *Correlated Naïve Bayes* seperti yang diterapkan oleh Stevani M. Meilisa dkk. terbukti efektif karena menghitung nilai korelasi antar-atribut terlebih dahulu sebelum probabilitas akhir dikalikan, sehingga mampu menjaga stabilitas hasil klasifikasi pada tingkat akurasi sebesar 83%.

Intervensi komputasi yang terbukti paling ampuh dalam mendongkrak performa klasifikasi *Naïve Bayes* hingga mampu melampaui angka 90% adalah integrasi teknik seleksi fitur sebelum data dimasukkan ke dalam algoritma utama. Rekam medis penyakit jantung pada umumnya melibatkan dimensi atribut yang sangat kompleks dan sering kali mengandung variabel yang redundan atau tidak terlalu relevan. Penerapan seleksi fitur berfungsi untuk menyaring dan mengeliminasi variabel-variabel pengganggu tersebut sehingga ruang data menjadi lebih ringkas dan bersih. Dampak positif dari reduksi dimensi ini ditunjukkan secara nyata dalam riset Siti Roziana Azizah dkk. yang menerapkan kombinasi seleksi fitur hibrida *Filter-Wrapper* dan sukses memecahkan rekor akurasi tertinggi sebesar 91,80%. Selaras dengan temuan tersebut, Alifa Lutfia dkk. yang mengintegrasikan metode seleksi fitur berbasis *Gain Ratio* juga berhasil meningkatkan ketepatan prediksi model hingga mencapai nilai akurasi sebesar 91,2%.

Selain kualitas fitur, penentuan proporsi pembagian data untuk fase pelatihan (*training*) dan pengujian (*testing*) juga memegang peranan penting dalam mengukur objektivitas performa model. Tren eksperimen pada mayoritas studi terdahulu menunjukkan bahwa penggunaan rasio pembagian data sebesar 80% : 20% merupakan standardisasi terbaik yang mampu menghasilkan performa klasifikasi paling stabil sekaligus aman dari risiko *overfitting*. Skenario pembagian data latih dan data uji dengan porsi 80:20 ini telah diadopsi secara sukses oleh sebagian besar peneliti dalam daftar literatur, termasuk dalam pengujian model terkorelasi oleh Meilisa yang menghasilkan akurasi 83%, Athhohiroh dengan capaian 81%, serta Nurkholifah yang memperoleh nilai akurasi sebesar 83% pada klasifikasi penyakit jantung terkait risiko stroke.

#### 4. KESIMPULAN

Berdasarkan hasil penelaahan terstruktur terhadap seluruh rangkaian draf penelitian lintas-studi terdahulu, dapat disimpulkan secara konkret bahwa efisiensi algoritma *Naïve Bayes* sebagai arsitektur pendukung keputusan klinis sangat bergantung pada perlakuan dimensi geometri data serta teknik pemrosesan yang digunakan. Melalui pemetaan matematis

teorema peluang ini, tingkat akurasi yang dihasilkan terbukti sangat bervariasi dengan rentang nilai antara 58% hingga 91,80%, di mana sebagian besar hasil pengujian secara konsisten berada dalam kisaran 79% hingga 91%. Karakteristik rekam medis yang berjumlah kecil hingga menengah mampu diklasifikasikan secara instan dengan tingkat ketepatan yang tinggi, sehingga metode ini dinilai sangat andal dan memiliki potensi besar untuk diterapkan pada sistem pendukung keputusan di sektor kesehatan, khususnya dalam membantu tenaga medis melakukan deteksi dini penyakit jantung. Namun, tantangan utama berupa penurunan akurasi pada data berskala besar mengharuskan adanya intervensi komputasi tambahan berupa optimasi pembagian porsi data latih serta pemilihan fitur yang ketat guna mengeliminasi variabel bising. Secara keseluruhan, integrasi yang sistematis antara pembersihan rekam medis, penentuan proporsi sampel uji yang berimbang, serta penyaringan atribut klinis yang tepat merupakan satu kesatuan pipa data yang valid untuk menghasilkan model prediksi otomatis yang presisi, sehingga *Naïve Bayes* dapat menjadi salah satu pilihan terbaik dalam mendukung keputusan medis secara cepat dan efisien.

## REFERENCES

- [1] S. Sahar, "Analisis Perbandingan Metode K-Nearest Neighbor dan Naïve Bayes Classifier Pada Dataset Penyakit Jantung," *Indones. J. Data Sci.*, vol. 1, no. 3, pp. 79–86, 2020, doi: 10.33096/ijodas.v1i3.20.
- [2] Khodijah and Sriyanto, "Perbandingan Kinerja Algoritma C4.5. Naive Bayes Dan Random Forest Dalam Prediksi Penyakit Jantung," *J. Tek.*, vol. 17, no. 2, pp. 419–426, 2023.
- [3] P. A. Sihotang and D. Sitanggang, "Penerapan Metode Algoritma C4.5 Dan Naive Bayes Untuk Prediksi Penyakit Jantung," *J. Tek. Inf. dan Komput.*, vol. 7, no. 2, p. 899, 2024, doi: 10.37600/tekinkom.v7i2.1535.
- [4] S. M. Mellisa, S. Sagarung, and H. Parmadi, "Pendekatan correlated naïve bayes pada klasifikasi potensi penyakit jantung," *Pros. Semin. Nas. Penelit. dan Pengabd. Kpd. Masy. LPPM Univ. 'Aisyiyah Yogyakarta*, vol. 1, pp. 22–2023, 2023, [Online]. Available: <https://www.kaggle.com/datasets/firdaus9914/penyakit-jantung>.
- [5] 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.
- [6] S. R. Azizah, R. Herteno, A. Farmadi, D. Kartini, and I. Budiman, "Kombinasi Seleksi Fitur Berbasis Filter dan Wrapper Menggunakan Naive Bayes pada Klasifikasi Penyakit Jantung," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 10, no. 6, pp. 1361–1368, 2023, doi: 10.25126/jtiik.2023107467.
- [7] A. Lutfia, G. Gunawan, R. S. Rohman, and A. Gunawan, "Penerapan Seleksi Fitur Gain Ratio Pada Prediksi Penyakit Jantung Berbasis Naïve Bayes," *J. Responsif Ris. Sains dan Inform.*, vol. 6, no. 1, pp. 1–10, 2024, doi: 10.51977/jti.v6i1.1396.
- [8] A. Nurfazila and H. Rohayani, "Literature Review : Implementation of the Naive Bayes Algorithm for Classification in Various Fields of Data Mining," vol. 3, no. 1, pp. 240–253, 2026, doi: <https://doi.org/10.61098/ijiretm.v3i1.269>
- [9] H. Rohayani, S. N. Alam, M. Fauzi, and R. Rico, "Prediksi Penyebaran Virus COVID-19 Dari Hasil PCR Menggunakan Metode Naïve Bayes," *J. Comput. Syst. Informatics*, vol. 4, no. 1, pp. 109–115, 2022, doi: 10.47065/josyc.v4i1.2577.
- [10] S. R. Fernanda and P. D. Mardika, "Implementasi Data Mining untuk Mendiagnosa Penyakit Jantung dengan Algoritma Naïve Bayes di RS TK II Moh. Ridwan Mauraksa," *Semnas Ristek (Seminar Nas. Ris. dan Inov. Teknol.)*, vol. 9, no. 1, pp. 114–118, 2025, doi: 10.30998/semnasristek.v9i1.7779.
- [11] A. Suko Wijoyo and A. Jananto, "Rancang Bangun Sistem Informasi Prediksi Penyakit Jantung Berbasis Algoritma Naive Bayes," *J. Tek. Inform. Unika ST. Thomas*, vol. 8, no. 2, pp. 181–190, 2023.
- [12] A. Atthohiroh, R. Ayu, and S. Maharani, "Penerapan Metode Naive Bayes Untuk Memprediksi Penyakit Jantung," *J. Tek.*, vol. 3, no. 1, p. 8, 2023, doi: 10.54314/teknisi.v3i1.1252.
- [13] A. N. Am, M. Nurkholifah, and F. K. Oktorina, "Analisa Penyakit Jantung Menggunakan Algoritma Naïve Bayes," *J. Syst. Comput. Eng.*, vol. 4, no. 1, pp. 26–36, 2023, doi: 10.47650/jsce.v4i1.671.
- [14] R. M. Ubaidillah and T. A. Puspito, "Optimasi Prediksi Penyakit Jantung Dengan Naïve Bayes Dan Particle Swarm Optimization (Pso)," *J. Inf. dan Komput.*, vol. 13, no. 1, p. 2025, 2025, [Online]. Available: [www.kaggle.com](http://www.kaggle.com)
- [15] I. Alhabib, "Komparasi Metode Deep Learning, Naïve Bayes Dan Random Forest Untuk Prediksi Penyakit Jantung," *INFORMATICS Educ. Prof. J. Informatics*, vol. 6, no. 2, p. 176, 2022, doi: 10.51211/itbi.v6i2.1881.
- [16] Oskar, Fransiska Ria, Rupina, and N. P., "Pengelolaan Data Penyakit Jantung Dengan Menggunakan Metode Naive Bayes," *J. Sains Dan Komput.*, vol. 8, no. 02, pp. 49–54, 2024, doi: 10.61179/jurnalinfact.v8i02.531.