Prosiding Seminar Nasional Bisnis, Teknologi dan Kesehatan
Universitas Muhammadiyah Muara Bungo
Volume 3, No 1, 2026
e-ISSN 3089-3763
Hal:-181-190

PERBANDINGAN ALGORITMA K-NEAREST NEIGHBOR DAN NAIVE BAYES UNTUK MENENTUKAN KELANGSUNGAN HIDUP PASIEN DENGAN PENYAKIT JANTUNG

Hali Mukid[1], Ahmad Homaidi[2] Zaehol Fatah[3]
Universitas Ibrahimy
halimukid79@gmail.com

ABSTRAK

Penyakit kardiovaskular merupakan penyebab utama kematian di dunia, termasuk di Indonesia. Penelitian ini bertujuan untuk membandingkan efektivitas dua algoritma machine learning, yaitu K-Nearest Neighbors (K-NN) dan Naive Bayes, dalam mengklasifikasikan risiko kematian atau kelangsungan hidup pasien penyakit jantung. Penelitian menggunakan pendekatan kuantitatif komparatif dengan dataset sekunder yang diunduh dari Kaggle.com, mencakup 500 baris data dengan fitur klinis seperti usia, tekanan darah, dan kadar kreatinin. Implementasi algoritma dilakukan menggunakan bahasa pemrograman PHP dengan tahapan preprocessing (penanganan missing value dengan mean, normalisasi Min-Max), pembagian data (80% latih, 20% uji), serta evaluasi menggunakan confusion matrix. Hasil penelitian menunjukkan bahwa algoritma K-NN (K=3) mencapai akurasi 95%, presisi 100%, recall 50%, dan F1-score 66,67%, sedangkan Naive Bayes Gaussian memperoleh akurasi 94%, presisi 64,29%, recall 90%, dan F1-score 75%. K-NN unggul dalam akurasi dan presisi tanpa false positive, sehingga direkomendasikan untuk skenario klinis yang menghindari intervensi tidak perlu. Naive Bayes unggul dalam recall, cocok untuk skenario yang menuntut sensitivitas tinggi. Sistem yang dibangun telah terbukti fungsional dari halaman login hingga visualisasi evaluasi. Keterbatasan penelitian meliputi ukuran dataset yang kecil dan ketidakseimbangan kelas, sehingga penelitian mendatang disarankan menggunakan teknik SMOTE, algoritma ensemble, serta integrasi dengan rekam medis elektronik..

Kata Kunci: K-Nearest Neighbors, Naive Bayes, Penyakit jantung, klasifikasi, komparasi algoritma, PHP.

ABSTRACT

Cardiovascular disease is a leading cause of death globally, including in Indonesia. This study aims to compare the effectiveness of two machine learning algorithms—K-Nearest Neighbors (K-NN) and Naive Bayes—in classifying the risk of death or survival among heart disease patients. A comparative quantitative approach was employed using a secondary dataset from Kaggle.com, comprising 500 data entries with clinical features such as age, blood pressure, and creatinine levels. The algorithms were implemented using PHP, involving preprocessing stages (handling missing values via mean imputation and Min-Max normalization), data splitting (80% training, 20% testing), and evaluation using a confusion matrix. The results indicate that the K-NN algorithm (K=3) achieved 95% accuracy, 100% precision, 50% recall, and an F1-score of 66.67%, whereas Gaussian Naive Bayes achieved 94% accuracy, 64.29% precision, 90% recall, and an F1-score of 75%. K-NN demonstrated superior performance in terms of accuracy and precision with zero false positives, making it suitable for clinical scenarios that prioritize avoiding unnecessary interventions. Conversely, Naive Bayes excelled in recall, making it appropriate for scenarios requiring high sensitivity. The developed system functioned effectively across all components, from the login page to the evaluation visualization. Study limitations include the small dataset size and class imbalance; therefore, future research is recommended to utilize techniques such as SMOTE and ensemble algorithms, as well as integration with electronic medical records.

***Keywords**: K-Nearest Neighbors, Naive Bayes, heart disease, classification, algorithm comparison, PHP.*

Prosiding Seminar Nasional Bisnis, Teknologi dan Kesehatan
Universitas Muhammadiyah Muara Bungo
Volume 3, No 1, 2026
e-ISSN 3089-3763
Hal:-181-190

Pendahuluan

Penyakit kardiovaskular (CVD), yang mencakup penyakit jantung, terus menjadi alasan utama kematian di seluruh dunia: menurut laporan WHO, diperkirakan pada tahun 2022, 19,8 juta individu kehilangan nyawa akibat CVD, yang berkontribusi sekitar 32% dari total kematian global dari total ini, sekitar 85% disebabkan oleh serangan jantung dan stroke [1].

Perkembangan teknik pembelajaran mesin (ML) dalam beberapa tahun terakhir menunjukkan kemampuan yang luar biasa untuk menentukan hasil klinis pada pasien dengan masalah kardiovaskular. Model-model ML yang lebih rumit seperti pohon ansambel dan peningkatan berdasarkan gradien telah menunjukkan hasil yang positif dalam meramalkan angka kematian dan hasil dalam jangka menengah atau panjang pada sejumlah kohort besar [2]. Namun, algoritma yang lebih sederhana dan mudah dipahami masih tetap menarik untuk digunakan di lingkungan klinis berkat kemudahan dalam penerapan dan menjelaskan hasil kepada tim medis [3].

Data epidemiologi terkini menguatkan pandangan mengenai beban global tersebut. Sebuah kajian internasional yang diterbitkan pada tahun 2024 menyatakan bahwa beban penyakit kardiovaskular-termasuk kejadian, angka prevalensi, dan tingkat kematian - masih tetap tinggi dan didominasi oleh negara-negara dengan pendapatan rendah hingga menengah, tempat di mana lebih dari 75 persen dari beban global penyakit kardiovaskular berada [4].

Selain itu, studi yang dilakukan di berbagai wilayah menunjukkan bahwa beban penyakit jantung dan pembuluh darah di negara-negara berkembang cenderung tetap atau bahkan bertambah dalam tiga puluh tahun terakhir [5]. Misalnya, sebuah penelitian yang dilakukan pada populasi di Indonesia dan menganalisis data dari tahun 1990 hingga 2019 menemukan bahwa meskipun beberapa indikator seperti Disability-Adjusted Life Years (DALYs) mengalami penurunan di beberapa daerah, secara keseluruhan di tingkat nasional, beban penyakit kardiovaskular masih belum memperlihatkan penurunan yang signifikan hal ini menunjukkan bahwa angka kematian dan sakit akibat penyakit jantung masih menjadi masalah kesehatan masyarakat yang penting di Indonesia [6].

Dua jenis algoritma yang umum digunakan dan sering dibandingkan dalam penelitian kesehatan adalah K-Nearest Neighbors (K-NN) dan Naive Bayes (NB). K-NN beroperasi dengan mengandalkan kesamaan antara fitur pasien (berbasis instansi), sedangkan Naive Bayes merupakan pendekatan probabilistik yang beranggapan bahwa fitur-fitur bersifat independen dengan syarat terhadap kategori [7]. Kedua algoritma ini cukup sederhana untuk diterapkan, responsif pada kumpulan data berukuran sedang, dan sering kali dijadikan tolok ukur dalam analisis penentuan kesehatan. Meskipun begitu, sifat data klinis yakni perbedaan kelas yang tidak seimbang, data yang hilang, dan hubungan antara fitur dapat berdampak secara berbeda terhadap efektivitas masing-masing algoritma [8]. Oleh karenanya, penting untuk melakukan sistematisasi perbandingan K-Nearest Neighbor dan Naive Bayes dalam konteks penentuan keberlangsungan hidup pasien penderita penyakit jantung yang relevan baik dari perspektif akademis maupun praktis [8].

Menjadi penting untuk menentukan keberlangsungan hidup pasien penderita penyakit jantung menggunakan kumpulan data yang ada sehingga dapat memberikan rekomendasi antisipasi dan tindakan menggunakan metode K-Nearest Neighbor dan Naïve Bayes.

Materi dan Metode

1. Jenis Penelitian

Penelitian ini menggunakan jenis penelitian kuantitatif dengan pendekatan komparatif, yang berfokus pada pengolahan dan analisis data numerik dari rekam medis pasien untuk menghasilkan model perbandingan yang objektif dan terukur, terutama untuk membandingkan efektivitas algoritma K-

Prosiding Seminar Nasional Bisnis, Teknologi dan Kesehatan
Universitas Muhammadiyah Muara Bungo
Volume 3, No 1, 2026
e-ISSN 3089-3763
Hal:-181-190

Nearest Neighbors (K-NN) dan Naive Bayes dalam mengklasifikasikan risiko kematian atau kelangsungan hidup pasien penyakit jantung. [9]. Implementasi kedua algoritma dilakukan dengan bahasa pemrograman PHP menggunakan dataset sekunder yang memuat variabel klinis seperti usia, kadar serum kreatinin, dan tekanan darah, serta diuji dengan pembagian data latih dan data uji yang identik agar hasil komparasi valid, objektif, dan konsisten [10]. Teknik pengumpulan data yang digunakan meliputi dokumentasi, yaitu mengunduh dataset pasien penyakit jantung dari Kaggle.com, dan studi pustaka untuk mempelajari konsep serta teori terkait penyakit jantung, klasifikasi data mining, algoritma K-NN, dan Naïve Bayes guna membangun landasan teoritis yang kuat [11]. Dataset yang digunakan merupakan data sekunder yang memuat variabel klinis penting, seperti usia, kadar serum kreatinin, dan tekanan darah. Untuk memastikan hasil komparasi bersifat valid, objektif, dan konsisten, kedua algoritma diuji menggunakan dataset yang sama melalui pembagian data latih (training) dan data uji (testing) yang identic.

2. Metode Pengembangan Sistem

Penelitian ini menggunakan metode CRISP-DM (Cross-Industry Standard Process for Data Mining) sebagai pendekatan dalam pengembangan sistem. Metode ini merupakan standar yang banyak digunakan dalam proyek data mining dan analisis data, khususnya dalam mendukung pengambilan keputusan berbasis machine learning. CRISP-DM dipilih karena fleksibel dan dapat diterapkan di berbagai bidang, termasuk bidang kesehatan, di mana hasil analisis dapat membantu tenaga medis dalam menentukan tindakan berdasarkan hasil komparasi model [12].

Tahapan dalam metode CRISP-DM pada penelitian ini dijelaskan sebagai berikut:

a. Business Understanding Tahap awal berfokus pada pemahaman tujuan penelitian, yaitu menentukan kelangsungan hidup pasien penderita penyakit jantung. Permasalahan utama dalam penelitian ini adalah meminimalkan tingkat kesalahan prediksi, khususnya False Negative, karena keterlambatan diagnosis dapat berakibat fatal. Oleh karena itu, penelitian ini bertujuan untuk membandingkan performa algoritma K-Nearest Neighbor (KNN) dan Naive Bayes berdasarkan metrik akurasi, presisi, dan recall guna menentukan model yang paling andal dalam mendukung keputusan medis.

b. Data Understanding
Pada tahap ini dilakukan pengumpulan data sekunder yang diperoleh dari situs Kaggle. Dataset yang digunakan mencakup berbagai fitur klinis, seperti usia, jenis kelamin, tekanan darah, kadar kolesterol, dan status kelangsungan hidup pasien. Selanjutnya, dilakukan eksplorasi data untuk memahami distribusi kelas serta hubungan antar variabel, sehingga dapat diidentifikasi pola awal yang relevan dengan prediksi kelangsungan hidup pasien.

c. Data Preparation
Data mentah diproses terlebih dahulu agar siap digunakan dalam tahap pemodelan. Proses ini mencakup penanganan data yang hilang (missing values), deteksi serta penanganan outlier, dan normalisasi data untuk memastikan konsistensi skala antar variabel. Selain itu, dataset kemudian dibagi menjadi training set guna memastikan bahwa proses evaluasi model dilakukan secara objektif dan tidak bias.

d. Modeling
Pada tahap ini, dua algoritma diterapkan dan dibandingkan, yaitu K-Nearest Neighbor (KNN) dan Naive Bayes. KNN menggunakan pendekatan berbasis jarak dengan perhitungan jarak Euclidean untuk menentukan kedekatan antar data pasien, sehingga data baru diklasifikasikan berdasarkan kemiripan dengan data historis pada tetangga terdekat. Sementara itu,

Prosiding Seminar Nasional Bisnis, Teknologi dan Kesehatan
Universitas Muhammadiyah Muara Bungo
Volume 3, No 1, 2026
e-ISSN 3089-3763
Hal:-181-190

Naive Bayes menggunakan pendekatan probabilitas berdasarkan Teorema Bayes dengan asumsi independensi antar fitur, di mana setiap variabel klinis dianggap memberikan kontribusi secara mandiri terhadap probabilitas kelangsungan hidup pasien.

e. Evaluation
Evaluasi model dilakukan menggunakan Confusion Matrix untuk mengukur performa masing-masing algoritma. Metrik yang digunakan dalam penilaian meliputi akurasi, sensitivity (recall), dan F1-Score. Berdasarkan hasil evaluasi tersebut, model dengan tingkat kesalahan paling rendah serta performa terbaik akan dipilih sebagai model yang direkomendasikan.

f. Deployment
Sistem prediksi penyakit jantung ini, yang mengandalkan model pembelajaran mesin (machine learning), dirancang agar mudah digunakan oleh para profesional kesehatan untuk diagnosis dini dan pengambilan keputusan. Sistem ini terdiri dari elemen-elemen berikut:

Modul Input Data: Pengguna dapat memasukkan data pasien secara manual atau mengunggah file CSV. Modul ini memverifikasi akurasi dan kelengkapan data masukan melalui pemeriksaan validasi data.

Modul Pra-pemrosesan (Preprocessing): Menjalankan tugas-tugas pra-pemrosesan, seperti mengelola data yang hilang, mendeteksi dan memperbaiki pencilan (outliers), serta melakukan normalisasi data. Modul ini membersihkan dan menyiapkan data untuk dianalisis.

Modul Pelatihan: Modul ini melatih algoritma KNN dan Naive Bayes menggunakan data yang telah diproses sebelumnya. Modul ini dilengkapi dengan kemampuan pengaturan parameter (parameter tuning) dan validasi silang (cross-validation) untuk mengoptimalkan performa model.

Modul Prediksi: Menerapkan model yang telah dipelajari untuk menentukan risiko penyakit jantung pada pasien baru. Modul ini menyajikan prediksi secara waktu nyata (real-time) beserta akurasi, presisi, recall, dan skor F1 sebagai metrik performa utama.

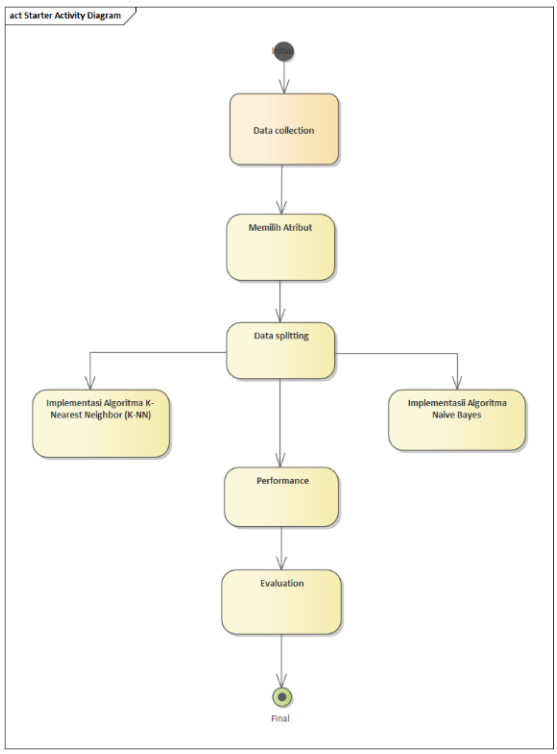
Modul Visualisasi: Menghasilkan kurva Receiver Operating Characteristic (ROC), matriks kebingungan (confusion matrices), dan plot kepentingan fitur (feature importance) dari performa model. Pengguna dapat menilai pengambilan keputusan dan keandalan model melalui visualisasi.

3. Perancangan Sistem

Komparasi Algoritma K-Nearest Neighbor Dan Naive Bayes Untuk Penentuan Keberlangsungan Hidup Pasien Penderita Penyakit Jantung dilakukan menggunakan system informasi. Dalam perancangan ini perancangan berfungsi sebagai cetak biru (blueprint) yang memberikan gambaran menyeluruh dan memastikan alur proses berjalan logis serta terstruktur, mulai dari pengumpulan data, pemilihan atribut, pembagian data, implementasi kedua algoritma (K-NN dan Naive Bayes), evaluasi kinerja, hingga tahap akhir. Perancangan juga berperan dalam menentukan komponen dan teknologi yang digunakan, seperti bahasa pemrograman PHP, format data CSV, metode Euclidean distance untuk K-NN, serta Gaussian Naive Bayes. Selain itu, perancangan menjamin perlakuan yang adil (fair comparison) dengan memastikan pembagian data latih dan uji yang identik serta metrik evaluasi yang sama bagi kedua algoritma, sehingga hasil komparasi menjadi objektif dan valid. Fungsi lainnya adalah mengidentifikasi kebutuhan data dan persiapan awal, menjadi dasar untuk pengujian dan validasi sistem, mempermudah pemeliharaan serta pengembangan selanjutnya (misalnya menambah algoritma lain), serta mengelola risiko teknis seperti normalisasi fitur atau keterbatasan ukuran dataset. Dengan demikian, perancangan sistem bukan sekadar

Prosiding Seminar Nasional Bisnis, Teknologi dan Kesehatan
Universitas Muhammadiyah Muara Bungo
Volume 3, No 1, 2026
e-ISSN 3089-3763
Hal:-181-190

gambar alur, melainkan alat komunikasi, panduan implementasi, dan jaminan kualitas yang memastikan sistem yang dibangun terstruktur, mudah diuji, dan hasilnya dapat dipertanggungjawabkan secara ilmiah [13].



Gambar 1. Alur Perancangan Sistem

Hasil dan Pembahasan

Hasil

Penelitian ini telah berhasil menentukan proyeksi keberlangsungan hidup pasien penderita penyakit jantung dengan membandingkan efektivitas metode K-Nearest Neighbor (KNN) dan Naïve Bayes.

1. Analisis Performa

Kedua klasifikasi, baik KNN maupun Naive Bayes, diuji menggunakan dataset Penyakit Jantung yang tersedia di website Kaggle.com. Akurasi KNN (95%) melampaui akurasi Naive Bayes (94%). Model KNN yang diusulkan mengungguli klasifikasi lainnya dalam hal akurasi. Metrik terperinci seperti presisi, recall, dan F1-score digunakan untuk memperoleh wawasan komprehensif mengenai performa klasifikasi.

Dalam hal ini, 100% dari kasus yang dilabeli KNN sebagai penderita penyakit jantung merupakan kasus positif yang sebenarnya. Presisi KNN dalam mengidentifikasi kejadian penyakit jantung adalah sebesar 100%. Sementara itu, presisi Naive Bayes adalah 64,29%, dengan sedikit peningkatan pada hasil positif palsu (false positives).

Recall mengukur proporsi kasus positif aktual yang berhasil diidentifikasi dengan benar oleh model. Model KNN memiliki nilai recall sebesar 50%, yang menunjukkan bahwa model tersebut berhasil mengidentifikasi dengan benar 90% dari seluruh kasus penyakit jantung dalam dataset. Di sisi lain, 90% dari kasus positif aktual berhasil diidentifikasi menggunakan metode Naive Bayes, yang berarti masih terdapat jumlah kasus yang tidak teridentifikasi dalam jumlah yang lebih tinggi. Skor F1-score sebesar 66,67% untuk KNN terbukti lebih kecil dibandingkan dengan metode Naive Bayes model (75%, karena keseimbangan yang lebih baik antara presisi dan recall (lihat pada Tabel 2).

Tabel 1. Tabel 1. Akurasi model pada dataset penyakit jantung.

Algoritma	Akurasi
KNN	95%
NB	94%

Table 2. Model Precision, Recall, And F1 *Scores For The Heart Deases Dataset.*

Algoritma	Presisi	Recall	F1Score
KNN	100%	50%	66,67%
NB	64,29%	90%	75%

2. Sistem yang di Usulkan

Komparasi algoritma k-nearest neighbor dan naive bayes untuk penentuan keberlangsungan hidup pasien penderita penyakit jantung dengan akurasi, presisi, recall, dan skor F1 yang tinggi. Sistem yang diusulkan menyederhanakan klasifikasi penyakit jantung melalui antarmuka yang intuitif dan ramah pengguna, sehingga menghasilkan hasil yang cepat dan andal.

Halaman login menjamin keamanan sistem. Pengguna harus memasukkan nama pengguna

Prosiding Seminar Nasional Bisnis, Teknologi dan Kesehatan
Universitas Muhammadiyah Muara Bungo
Volume 3, No 1, 2026
e-ISSN 3089-3763
Hal:-181-190

dan kata sandi yang valid untuk mengautentikasi informasi medis guna memastikan keamanan dan privasi informasi medis. Kepatuhan terhadap peraturan perlindungan data kesehatan memerlukan penerapan langkah keamanan ini (Gambar 2).

Tampilan dashboard pada sistem tersebut menampilkan judul utama "Sistem Prediksi Keberlangsungan Hidup Pasien Jantung" serta sambutan yang menjelaskan bahwa sistem ini digunakan untuk membandingkan performa algoritma K-Nearest Neighbors (K-NN) dan Naive Bayes dalam menentukan keberlangsungan hidup pasien penyakit jantung. Dashboard menyediakan dua metode utama untuk memasukkan data, yaitu melalui unggah file CSV dengan tombol "Mulai Upload" untuk memproses dataset pasien secara kolektif, serta melalui input data manual dengan tombol "Input Manual" untuk memasukkan data pasien satu per satu. Dengan antarmuka yang sederhana dan informatif, halaman ini menjadi titik akses awal bagi pengguna sebelum melakukan proses klasifikasi dan evaluasi kedua algoritma (Gambar 3).

Halaman “Upload CSV” pada sistem prediksi keberlangsungan hidup pasien jantung menampilkan antarmuka untuk mengunggah file dataset pasien dalam format CSV. Pada halaman ini, pengguna disuguhi tombol Choose file dengan indikator “No file chosen” yang menandakan belum ada file yang dipilih, serta informasi batasan bahwa maksimal 500 baris data diperbolehkan agar komputasi tetap cepat. Tersedia pula dua tombol aksi, yaitu Upload untuk mengirimkan file yang telah dipilih ke sistem, dan Batal untuk membatalkan proses unggahan. Dengan demikian, halaman ini berfungsi sebagai gerbang input data yang membatasi jumlah baris demi menjaga kinerja algoritma K-NN dan Naive Bayes (Gambar 4).

Setelah pengguna mengunggah file CSV menampilkan daftar lengkap kolom yang tersedia dalam dataset, seperti patient_id, gender, age, systolic_blood_pressure, hingga heart_attack, serta menyediakan opsi bagi pengguna untuk memilih satu kolom sebagai target prediksi. Selain itu, halaman ini juga menyajikan tabel pratinjau (preview) dari 10 baris pertama data sehingga pengguna dapat memeriksa kualitas dan struktur dataset sebelum melanjutkan proses. Di bagian bawah, tersedia dua tombol aksi, yaitu “Gunakan target ini” untuk mengonfirmasi pilihan kolom target dan melanjutkan ke tahap berikutnya, serta “Upload ulang” jika pengguna ingin mengganti file CSV yang diunggah (Gambar 5).

Halaman Preprocessing Data pada sistem ini menampilkan konfigurasi praproses yang akan diterapkan secara otomatis, yaitu penanganan missing value dengan metode Mean (mengisi nilai kosong dengan rata-rata), tanpa penanganan outlier, serta normalisasi Min-Max untuk menyamakan skala fitur numerik. Halaman ini juga menyajikan pratinjau (preview) 10 baris pertama data yang telah melalui proses tersebut, di mana kolom kategorikal seperti patient_id, gender, dan smoker telah diubah menjadi nilai numerik (0/1), sementara kolom age dan body_mass_index mempertahankan nilai aslinya atau diisi 0 jika data hilang. Dengan demikian, pengguna dapat memverifikasi hasil praproses sebelum melanjutkan ke tahap pembagian data dan implementasi algoritma, serta tersedia tombol “Kembali” untuk mengubah konfigurasi jika diperlukan (Gambar 6).

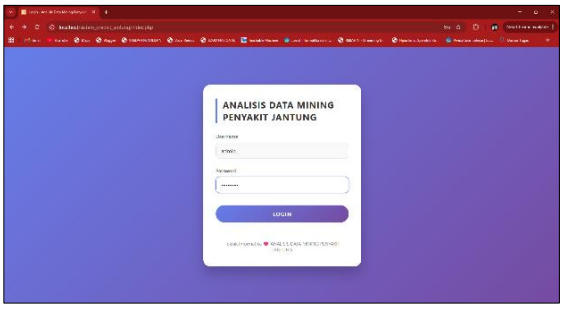
Halaman Pembagian Data dalam sistem ini menampilkan informasi bahwa total data yang tersedia setelah preprocessing adalah 500 baris. Pengguna dapat menentukan proporsi data uji (test size) yaitu sebesar 0.2 (atau 20%), yang berarti 20% dari data akan digunakan sebagai data uji, sedangkan sisanya 80% sebagai data latih. Tersedia tombol Split untuk menjalankan proses pembagian data secara acak namun konsisten (agar perbandingan algoritma K-NN dan Naive Bayes adil), serta tombol Kembali untuk mengubah konfigurasi preprocessing atau unggahan file. Halaman ini menjadi tahap

Prosiding Seminar Nasional Bisnis, Teknologi dan Kesehatan
Universitas Muhammadiyah Muara Bungo
Volume 3, No 1, 2026
e-ISSN 3089-3763
Hal:-181-190

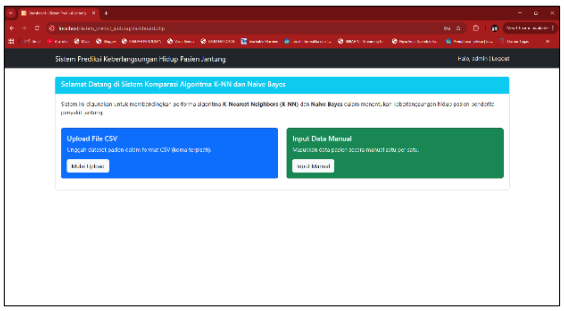
penting sebelum implementasi kedua algoritma klasifikasi (Gambar 7).

Halaman Komparasi Performa Algoritma yang membandingkan hasil klasifikasi antara K-Nearest Neighbors (K-NN) dengan K=3 dan Naive Bayes (Gaussian). Untuk K-NN, sistem menampilkan akurasi sebesar 95%, presisi 100%, recall 50%, dan F1-Score 66,67%, dengan rincian confusion matrix: TP (True Positive) = 5, TN (True Negative) = 90, FP (False Positive) = 0, dan FN (False Negative) = 5. Sementara itu, Naive Bayes memperoleh akurasi 94%, presisi 64,29%, recall 90%, dan F1-Score 75%, dengan TP = 9, TN = 85, FP = 5, dan FN = 1. Dari data tersebut terlihat bahwa K-NN unggul dalam akurasi dan presisi (tanpa FP), tetapi memiliki recall yang lebih rendah dibanding Naive Bayes, sehingga Naive Bayes lebih baik dalam menangkap kasus positif (pasien berisiko), sedangkan K-NN lebih unggul dalam menghindari kesalahan positif. Halaman ini juga menyediakan tombol Lanjut ke Visualisasi dan Dashboard untuk eksplorasi lebih lanjut (Gambar 8).

Halaman Visualisasi Evaluasi yang menyajikan dua diagram confusion matrix untuk membandingkan kinerja algoritma K-Nearest Neighbors (K-NN) dan Naive Bayes (Gaussian). Pada confusion matrix K-NN, sumbu X menampilkan empat metrik: TP (True Positive), FP (False Positive), FN (False Negative), dan TN (True Negative), sementara sumbu Y menunjukkan nilai dari 0 hingga 90 (mewakili jumlah kasus). Legenda pada confusion matrix K-NN menandakan batang berwarna hijau untuk K-NN dan biru untuk Naive Bayes (meskipun ini agak tidak biasa karena masing-masing seharusnya hanya menampilkan satu algoritma). Sementara itu, confusion matrix untuk Naive Bayes juga memiliki sumbu X dan Y yang sama, dengan legenda batang berwarna hijau untuk Naive Bayes. Dengan tampilan ini, pengguna dapat secara visual membandingkan distribusi kesalahan dan kebenaran prediksi dari kedua algoritma secara langsung (Gambar 9).

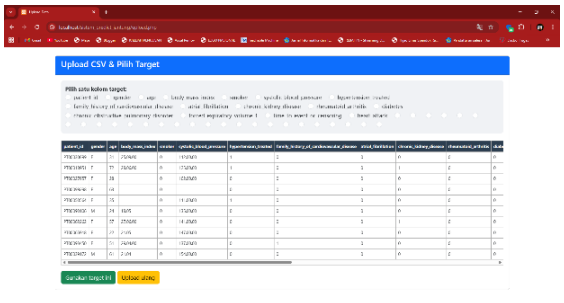


Gambar 2. Halaman Login



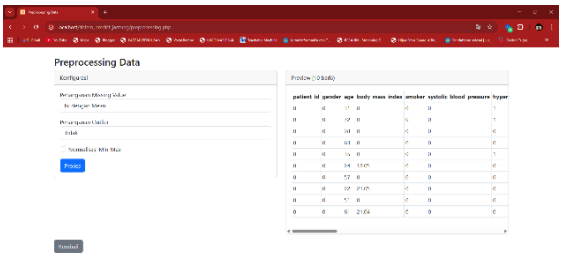
Gambar 3. Halaman Dashboard

Gambar 4. Halaman Upload Data CSV

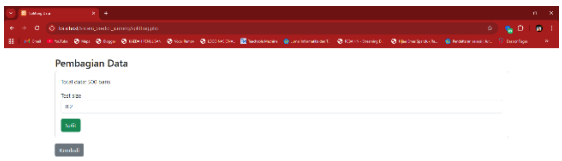


Gambar 5. Halaman Tampilan Data

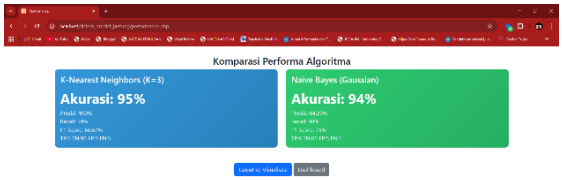
Prosiding Seminar Nasional Bisnis, Teknologi dan Kesehatan
Universitas Muhammadiyah Muara Bungo
Volume 3, No 1, 2026
e-ISSN 3089-3763
Hal:-181-190



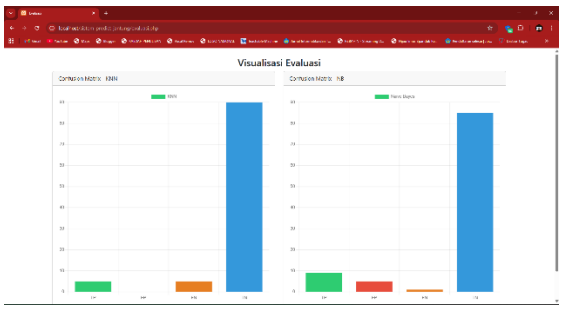
Gambar 6. Halaman Preprosessing Data



Gambar 6. Halaman Pembagian Data



Gambar 8. Halaman Komparasi Performa



Gambar 9. Halaman Visualisasi Evaluasi

Pembahasan

Pengujian sistem dimulai dari halaman Login yang berfungsi mengautentikasi pengguna dengan username dan password untuk menjamin keamanan data medis, dan sistem berhasil memverifikasi akses. Selanjutnya, Dashboard Utama menampilkan dua metode input data (unggah CSV dan input manual) dengan tombol yang berfungsi baik. Pada halaman Upload CSV, sistem berhasil menerima file CSV dengan batasan maksimal 500 baris, menyediakan tombol Upload dan Batal. Setelah file diunggah, halaman Pilih Target menampilkan seluruh 16 kolom dataset, pratinjau 10 baris pertama, serta opsi memilih kolom target, dan tombol "Gunakan target ini" berfungsi dengan baik. Kemudian, halaman Preprocessing Data menerapkan konfigurasi penanganan missing value dengan Mean, tanpa outlier, dan normalisasi Min-Max, disertai preview data yang telah diubah ke numerik, serta tombol Kembali. Pada halaman Pembagian Data, sistem menampilkan total 500 baris data dan test size 0,2 (20% data uji, 80% data latih), dengan tombol Split yang berhasil menjalankan pembagian secara konsisten. Halaman Komparasi Performa Algoritma menunjukkan hasil K-NN (K=3): akurasi 95%, presisi 100%, recall 50%, F1-score 66,67% dengan TP=5, TN=90, FP=0, FN=5; sedangkan Naive Bayes Gaussian: akurasi 94%, presisi 64,29%, recall 90%, F1-score 75% dengan TP=9, TN=85, FP=5, FN=1. Terakhir, halaman Visualisasi Evaluasi menyajikan dua diagram confusion matrix dalam bentuk batang untuk membandingkan TP, FP, FN, TN dari kedua algoritma. Seluruh fitur dari login hingga visualisasi berfungsi dengan baik dan sistem siap digunakan.

Kesimpulan

Penelitian ini berhasil membandingkan efektivitas algoritma K-Nearest Neighbors (K-NN) dan Naive Bayes dalam mengklasifikasikan risiko kematian atau kelangsungan hidup pasien penyakit jantung. Dari hasil pengujian pada dataset dengan total 500 baris data dan pembagian 80% latih serta 20% uji, diperoleh bahwa algoritma K-NN dengan nilai K=3 menunjukkan akurasi (95%) lebih tinggi dibandingkan Naive Bayes (94%), serta presisi sempurna (100%) tanpa menghasilkan false positive. Sementara itu, Naive Bayes unggul dalam recall (90%) dan F1-score (75%), yang berarti lebih baik dalam

menangkap kasus positif (pasien berisiko meninggal) meskipun dengan konsekuensi lebih banyak false positive.

Secara keseluruhan, K-NN lebih direkomendasikan untuk skenario klinis yang memprioritaskan penghindaran kesalahan positif (misalnya mencegah intervensi yang tidak perlu), sedangkan Naive Bayes lebih cocok untuk skenario yang menuntut sensitivitas tinggi agar tidak ada pasien berisiko yang terlewat. Sistem yang dibangun telah terbukti fungsional mulai dari login, unggah CSV, preprocessing, pembagian data, hingga visualisasi evaluasi. Meskipun demikian, keterbatasan dataset yang kecil dan tidak seimbang perlu diatasi pada penelitian mendatang dengan teknik SMOTE, penambahan algoritma ensemble, serta integrasi ke sistem rekam medis elektronik untuk prediksi waktu nyata.

Ucapan Terima Kasih

Penulis mengucapkan terima kasih kepada Rumah Jurnal Ilmiah Universitas Muhammadiyah Muara Bungo atas dukungan dan kesempatan yang diberikan sehingga kegiatan penulis bisa berjalan dengan baik. Penulis juga menyampaikan apresiasi kepada semua yang terlibat dalam penelitian ini sehingga sampai artikel ini diterbitkan.

Daftar Pustaka

[1] Lv, H., Bi, X., Shang, S., Wei, M., Zhou, X., Wang, K., Tang, B., & Lu, Y. (2025). Machine learning model for predicting in-hospital cardiac mortality among atrial fibrillation patients. 1–10.

[2] Lin, C., Lai, Y., Chuang, C., Yu, C., Lo, C., Chen, M., & Shia, B. (2025). Machine Learning-Based Prediction of Three-Year Heart Failure and Mortality After Premature Ventricular Contraction Ablation. 1–18.

[3] Rizky, F., Chaq, M., Chaq, E., Multazam, Z., Mustofa, A., & Socha, W. (2024). The 30 Years of Shifting in The Indonesian Cardiovascular Burden — Analysis of The Global Burden of Disease Study Global Burden of Disease. Journal of Epidemiology and Global Health, 193–212. https://doi.org/10.1007/s44197-024-00187-8

[4] Li, J., Fan, H., Yang, Y., Huang, Z., Yuan, Y., & Liang, B. (2024). Global burden of myocarditis from 1990 to 2021 : findings from the Global Burden of Disease Study 2021.

[5] Kumar, R., Garg, S., Kaur, R., & Lozanović, J. (2025). A comprehensive review of machine learning for heart disease prediction : challenges , trends , ethical considerations , and future directions. May, 1–31. https://doi.org/10.3389/frai.2025.1583459

[6] Pahlevi, R., Negara, E. S., Sutabri, T., & Herdiansyah, M. I. (2023). Penerapan Metode Naive Bayes untuk Menentukan Klasifikasi Kelayakan Penerimaan Bantuan Rehabilitasi dan Pembangunan Sekolah pada Dinas Pendidikan dan Kebudayaan Kabupaten Banyuasin Magister Teknik Informatika , Universitas Bina Darma Provinsi Sumatera Selatan . karakteristik klasifikasi untuk melengkapi proses kelayakan bantuan pemugaran gedung Salah satu penelitian yang menjadikan landasan peneliti melakukan penelitian ini adalah Klasifikasi Penerima Bantuan Program Rehabilitasi Rumah Tidak Layak Huni Menggunakan Algoritme K-Nearest Neighbor . Data yang digunakan dalam penelitian ini. 9(2), 1176–1188.

[7] Dewi, S. C., Putra, C. E., Nugraheni, A. G., Bangsa, U. P., Bangsa, U. P., & Bangsa, U. P. (2024). Implementasi Metode K-Nearest Neighbors (KNN) dan Naive Bayes untuk Klasifikasi Penyakit Jantung. 3(2).

[8] Putra, H. S., Atina, V., & Hartanti, D. (2024). Application of Naive Bayes Algorithm for Sentiment Analysis of Service and Facility Satisfaction (Case Study : PKU Muhammadiyah Sukoharjo General Hospital). 6(1), 61–72.

[9] Tingkat, K., & Kuantitas, D. A. N. (2024). PERANCANGAN WEBSITE UNTUK MEDIA PEMBELAJARAN. 12(2).

[10] Wilders, C. (2025). The Journal of Academic Librarianship Predicting the Role of Library Bookshelves in 2025. The Journal of Academic Librarianship, 43(5), 384–391. https://doi.org/10.1016/j.acalib.2017.06.019

[11] Harahap, M. R., & Albina, M. (2025). Model Penelitian Campuran : Kajian Literatur atas Jenis , Langkah , dan Manfaat Mixed Method dalam Studi Ilmiah. 85–96.

[12] Pengaruh, A., Terhadap, F., & Badan, T. (2025). Analisis Pengaruh Fitur Terhadap Tinggi Badan Anak menggunakan SHAP. 11(2), 271–276.

[13] Permana, A. A., Engineering, I., Tangerang, U. M., & Perintis, J. (2024). Enhancing Early Diagnosis of Heart Disease : A Comparative Study of K-NN and Naive Bayes Classifiers Using the UCI Heart Disease Dataset. 5(1), 35–42. https://doi.org/10.26714/jichi.v5i1.11251