POTERS
PROCEEDINGS OF TECHNOLOGY, ENGINEERING AND COMPUTERS

POTERS: Proceedings of Technology, Engineering and Computers

Vol. 1, No. 2, pp. 64 - 71, 2025
p-ISSN. xxxx - xxxx, e-ISSN. 3090 - 3602

PERBANDINGAN K-NN DAN NAIVE BAYES UNTUK PREDIKSI KELANGSUNGAN HIDUP PASIEN GAGAL JANTUNG

Adisty Maharani[1], Desiyanna Lasut[2*]
[1, 2] Program Studi Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Buddhi Dharma
*Corresponding Author, email: desiyanna.lasut@ubd.ac.id

ABSTRAK

Gagal jantung merupakan salah satu penyebab utama kematian global, dengan WHO mencatat 17,9 juta kematian akibat penyakit kardiovaskular pada 2019. Dalam konteks ini, deteksi dini menjadi krusial untuk meningkatkan harapan hidup pasien, untuk mengatasi tantangan ini dunia medis mulai mengadopsi teknologi guna memprediksi peluang hidup pasien secara lebih akurat. Penelitian ini bertujuan untuk membandingkan efektivitas dua algoritma machine learning, yaitu K-Nearest Neighbors (K-NN) dan Naive Bayes, dalam memprediksi keberlangsungan hidup pasien gagal jantung menggunakan data medis. Metode yang digunakan mengikuti pendekatan CRISP-DM, dimulai dari pemahaman bisnis dan data, persiapan data, hingga pemodelan dan evaluasi. Dataset yang digunakan terdiri dari 299 pasien berusia 40 tahun ke atas dengan 12 atribut dan 1 variabel target, diambil dari UCI Machine Learning Repository. Hasil pengujian menunjukkan bahwa algoritma K-NN memiliki akurasi pelatihan sebesar 82,85%, lebih tinggi dibandingkan Naive Bayes yang mencapai 82,42%. Meskipun perbedaannya kecil, K-NN menunjukkan performa yang lebih baik dalam akurasi prediksi, sementara Naive Bayes unggul dalam efisiensi waktu proses. Kesimpulan dari penelitian ini menyatakan bahwa kedua algoritma memiliki keunggulan masing-masing, sehingga pilihan algoritma dapat disesuaikan dengan prioritas pengguna, apakah lebih mengutamakan akurasi atau kecepatan prediksi. Penelitian ini diharapkan dapat menjadi referensi bagi peneliti dan praktisi kesehatan dalam pengembangan sistem prediksi medis yang lebih efektif.

Kata kunci: Data Mining, Gagal Jantung, *K-Nearest Neighbor*, *Naïve Bayes.*

I. PENDAHULUAN

Gagal jantung merupakan salah satu penyebab utama kematian global. Pada 2019, WHO mencatat 17,9 juta kematian akibat penyakit kardiovaskular, 85% di antaranya disebabkan oleh serangan jantung dan stroke. Gagal jantung terjadi ketika kemampuan jantung memompa darah menurun drastis, memicu gejala seperti sesak napas, kelelahan, dan pembengkakan (Effendy et al., 2023). Untuk mengatasi tantangan ini, dunia medis mulai mengadopsi teknologi data mining guna memprediksi peluang hidup pasien secara lebih akurat. Data mining digunakan untuk menemukan pola tersembunyi guna meningkatkan efisiensi layanan dan mendukung pengambilan keputusan medis (Amna et al., 2023). Dalam praktiknya,

Published by Universitas Buddhi Dharma
Available online at https://jurnal.ubd.ac.id/index.php/poters/index
Kota Tangerang, 9 Juli 2025.

Vol. 1, No. 2, pp. 64 - 71, 2025
p-ISSN. xxxx - xxxx, e-ISSN. 3090 - 3602

digunakan algoritma untuk membentuk model dari data sesuai karakteristik masalah (Kartika Dewi & Yahfizham, 2023).

Machine Learning adalah cabang dari AI yang memungkinkan komputer memproses dan mengklasifikasikan data baru secara otomatis menggunakan algoritma berbasis data sebelumnya (Bansal et al., 2022). Salah satu bahasa pemprograman yang digunakan untuk mengimplementasikan *data mining* adalah python. Python dikenal dengan sintaks yang sederhana, mudah dipelajari, open source, dan portable (Cutting & Stephen, 2021). Selain itu, Python memiliki pustaka standar yang luas untuk berbagai keperluan seperti analisis data, web scraping, pemrosesan gambar, machine learning, dan pengolahan teks (Gholizadeh, 2022). Prediksi digunakan untuk memperkirakan hasil di masa depan berdasarkan pola dan bukti dari data, dengan pendekatan matematika dan statistik (Muharrom, 2023). Sementara itu, perbandingan algoritma data mining bertujuan untuk mengevaluasi efektivitas, efisiensi, dan akurasi dalam menyelesaikan masalah. Naive Bayes adalah algoritma klasifikasi berbasis Teorema Bayes dengan asumsi independensi antar atribut. Algoritma ini cepat, efisien, dan cocok untuk dataset besar meskipun pendekatannya sederhana (Rino, 2021). Sementara itu, K-Nearest Neighbor (K-NN) didefinisikan algoritma lazy learning yang mengklasifikasikan data baru berdasarkan jarak terdekat ke data pelatihan, tanpa proses pelatihan kompleks. Klasifikasi ditentukan oleh mayoritas dari k tetangga terdekat (Sinaga et al., 2020). K-Nearest Neighbor (kNN) merupakan algoritma nonparametrik yang dikenal karena kesederhanaan, efektivitas, dan kemampuannya menangani data kompleks tanpa asumsi distribusi tertentu (Pan et al., 2020).

Penelitian ini bertujuan membangun dan membandingkan model prediksi gagal jantung menggunakan algoritma Naïve Bayes dan K-Nearest Neighbor (K-NN) berdasarkan data medis dari Institut Kardiologi Faisalabad, Pakistan. Tujuannya adalah mengetahui algoritma yang lebih akurat dan efisien dalam memprediksi keberlangsungan hidup pasien, serta mengembangkan aplikasi prediksi medis. Penelitian ini bermanfaat bagi peneliti dan praktisi kesehatan dalam memilih algoritma yang tepat, sekaligus menghasilkan aplikasi pendukung diagnosis. Ruang lingkup penelitian dibatasi pada dua algoritma klasifikasi dengan

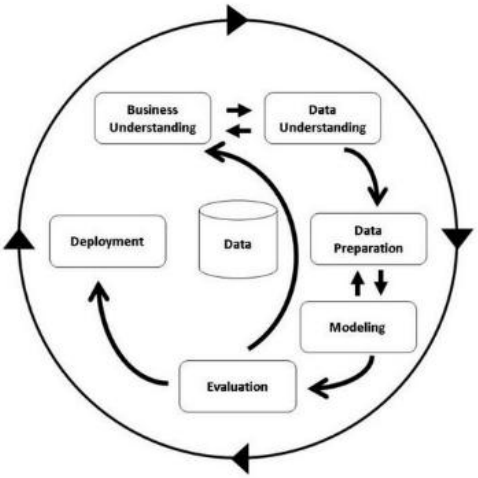
Published by Universitas Buddhi Dharma
Available online at https://jurnal.ubd.ac.id/index.php/poters/index
Kota Tangerang, 9 Juli 2025.

POTERS
PROCEEDINGS OF TECHNOLOGY, ENGINEERING AND COMPUTERS
POTERS: Proceedings of Technology, Engineering and Computers
Vol. 1, No. 2, pp. 64 - 71, 2025
p-ISSN. xxxx - xxxx, e-ISSN. 3090 - 3602

data sekunder berjumlah 299 pasien berusia 40 tahun ke atas, terdiri dari 13 atribut dan 1 kelas target, yang diambil dari UCI Machine Learning Repository.

II. METODOLOGI

Proses perancangan model dalam penelitian ini mengikuti pendekatan CRISP-DM. CRISP-DM (Cross-Industry Standard Process for Data Mining) adalah pendekatan standar lintas industri untuk proyek data mining yang dirancang untuk memandu proses analitik secara sistematis dan terstruktur (Azeroual et al., 2025).



Gambar 1. Tahapan CRISP-DM

Sejumlah penelitian telah dilakukan untuk mengevaluasi performa algoritma Naïve Bayes dan K-Nearest Neighbor (KNN). (Timotius & Fenriana, 2024) membandingkan Naïve Bayes dengan Regresi Linear menggunakan dataset berisi 918 data, dan hasilnya Naïve Bayes unggul dengan akurasi 86,26%. Sementara itu, penelitian (Kumar, 2020) dengan dataset besar berisi 70.000 data menunjukkan bahwa KNN memberikan akurasi tertinggi sebesar 72,28%, mengungguli Logistic Regression dan Naïve Bayes, membuktikan bahwa KNN dapat unggul tergantung konteks dan karakteristik data. Di sisi lain, (Hansen & Hariyanto, 2023) menunjukkan dominasi Naïve Bayes dalam klasifikasi diabetes dibanding C4.5, dengan akurasi 85% dan AUC 0.936. Dari berbagai studi ini dapat disimpulkan bahwa baik Naïve Bayes maupun KNN memiliki keunggulan masing-masing, tergantung pada kompleksitas data, jumlah atribut, dan kebutuhan sistem, sehingga pemilihan algoritma perlu disesuaikan dengan konteks permasalahan yang dihadapi.

Published by Universitas Buddhi Dharma
Available online at https://jurnal.ubd.ac.id/index.php/poters/index
Kota Tangerang, 9 Juli 2025.

POTERS
PROCEEDINGS OF TECHNOLOGY, ENGINEERING AND COMPUTERS
POTERS: Proceedings of Technology, Engineering and Computers
Vol. 1, No. 2, pp. 64 - 71, 2025
p-ISSN. xxxx - xxxx, e-ISSN. 3090 - 3602

III. HASIL DAN PEMBAHASAN

3.1 Business Understanding

Penelitian ini membandingkan algoritma K-Nearest Neighbor (K-NN) dan Naive Bayes untuk memprediksi kelangsungan hidup pasien gagal jantung berdasarkan data medis 299 pasien. Tujuannya adalah membantu deteksi dini pasien berisiko tinggi. Hasilnya digunakan untuk menentukan algoritma terbaik serta membangun aplikasi prediktif sebagai alat bantu pengambilan keputusan medis.

3.2 Data Understanding

Data yang digunakan dalam penelitian ini diambil dari situs https://archive.ics.uci.edu/, uraian masing atribut disajikan Tabel 1.

Tabel 1. Keterangan Dataset

Nama Variabel	Tipe	Deskripsi	Satuan
Age	*Integer*	usia pasien	Tahun
Anaemia	*Binary*	anemia atau penurunan jumlah sel darah merah/hemoglobin	
creatinine_phosphokinase	*Integer*	kadar enzim CPK dalam darah	mcg/L
diabetes	*Binary*	apakah pasien menderita diabetes	
ejection_fraction	*Integer*	persentase darah yang dipompa keluar dari jantung setiap kontraksi	%
high_blood_pressure	*Binary*	apakah pasien mengalami tekanan darah tinggi atau hipertensi	
platelets	*Continuous*	jumlah trombosit dalam darah	kiloplatelets/mL
serum_creatinine	*Continuous*	kadar kreatinin serum dalam darah	mg/dL
serum_sodium	*Integer*	kadar natrium serum dalam darah	mEq/L
sex	*Binary*	jenis kelamin pasien, perempuan atau laki-laki	
smoking	*Binary*	apakah pasien merokok atau tidak	
time	*Integer*	durasi masa tindak lanjut	days
death_event	*Binary*	apakah pasien meninggal dunia selama masa tindak lanjut	

Published by Universitas Buddhi Dharma
Available online at https://jurnal.ubd.ac.id/index.php/poters/index
Kota Tangerang, 9 Juli 2025.

POTERS
PROCEEDINGS OF TECHNOLOGY, ENGINEERING AND COMPUTERS
POTERS: Proceedings of Technology, Engineering and Computers
Vol. 1, No. 2, pp. 64 - 71, 2025
p-ISSN. xxxx - xxxx, e-ISSN. 3090 - 3602

3.3 Data Preparation

Tabel 2. Tahap Data Preparation

```
X = data.drop(columns=["DEATH_EVENT"])
y = data["DEATH_EVENT"]
X_train, X_test, y_train, y_test = train_test_split(
X, y, test_size=0.2, random_state=42, stratify=y
)
# Standardisasi fitur
scaler = StandardScaler()
X_train_scaled = scaler.fit_transform(X_train)
X_test_scaled = scaler.transform(X_test)
```

Pada tahap ini, dataset terlebih dahulu dipisahkan antara fitur dan target. Target yang digunakan adalah DEATH_EVENT, yaitu indikator apakah pasien mengalami kematian akibat gagal jantung. Selanjutnya, data dibagi menjadi dua bagian: 80% untuk pelatihan *(training)* dan 20% untuk pengujian *(testing)*. Berdasarkan (Joseph, 2022) rasio 80:20 merupakan rasio yang paling sering digunakan serta sering dipilih karena merujuk pada prinsip Pareto. Pembagian ini dilakukan menggunakan fungsi *train_test_split* dengan parameter *stratify=y* untuk memastikan distribusi kelas pada data pelatihan dan pengujian tetap seimbang. Setelah pembagian, dilakukan proses standardisasi fitur menggunakan StandardScaler, terutama untuk model K-Nearest Neighbors (K-NN) yang sangat bergantung pada perhitungan jarak antar data.

3.4 Modelling

Tabel 3. Tahap Modelling

```
# Model K-NN
knn = KNeighborsClassifier(n_neighbors=5)
knn.fit(X_train_scaled, y_train)
# Model Naive Bayes
nb = GaussianNB()
nb.fit(X_train, y_train)
```

Proses selanjutnya adalah *modelling*, yaitu pelatihan model menggunakan data latih. Model K-NN dilatih menggunakan data yang telah distandarisasi karena algoritma ini menentukan hasil prediksi berdasarkan jarak terdekat antar data, sehingga skala fitur sangat mempengaruhi akurasi model. Sementara itu, algoritma Naive Bayes dilatih menggunakan data asli tanpa proses scaling, karena metode ini

Published by Universitas Buddhi Dharma
Available online at https://jurnal.ubd.ac.id/index.php/poters/index
Kota Tangerang, 9 Juli 2025.

POTERS
PROCEEDINGS OF TECHNOLOGY, ENGINEERING AND COMPUTERS
POTERS: Proceedings of Technology, Engineering and Computers
Vol. 1, No. 2, pp. 64 - 71, 2025
p-ISSN. xxxx - xxxx, e-ISSN. 3090 - 3602

berbasis probabilistik dan lebih fokus pada distribusi nilai tiap fitur dibandingkan pada jarak antar data. Masing-masing model di-fit atau dilatih menggunakan *fit()* dengan input data latih dan label, untuk mempelajari hubungan antara fitur dan kemungkinan terjadinya kematian akibat gagal jantung.

3.5 Evaluation

Tabel 4. Tahap Evaluation

```
# Evaluasi akurasi model pada data latih
knn_train_acc = accuracy_score(y_train, knn.predict(X_train_scaled))
nb_train_acc = accuracy_score(y_train, nb.predict(X_train))
```

Evaluasi merupakan proses untuk mengukur kinerja model setelah pelatihan, dengan menghitung akurasi prediksi terhadap data latih menggunakan fungsi accuracy_score. Model K-NN dievaluasi pada data yang telah distandarisasi, sedangkan Naive Bayes menggunakan data asli. Hasil akurasi menunjukkan sejauh mana model memahami pola data, dan disimpan dalam database sebagai bagian dari riwayat pelatihan yang mencakup nama dataset, waktu, durasi, dan akurasi model. Evaluasi ini penting untuk memastikan model siap digunakan dalam prediksi nyata melalui aplikasi.

3.6 Development

Dengan menerapkan algoritma Naive Bayes dan k-Nearest Neighbor (k-NN), penelitian ini membandingkan akurasi kedua model dalam memprediksi keberlangsungan hidup pasien. Hasil prediksi diharapkan dapat meningkatkan kewaspadaan individu terhadap kondisi kesehatannya dan mendorong pemeriksaan medis lebih awal guna memungkinkan penanganan awal.

IV. SIMPULAN

Sistem prediksi gagal jantung telah berhasil dikembangkan menggunakan algoritma K-Nearest Neighbors (K-NN) dan Naive Bayes. Aplikasi ini dibangun dengan bahasa pemrograman Python, antarmuka interaktif berbasis Streamlit, serta MySQL sebagai basis data untuk menyimpan informasi pengguna dan histori pelatihan. Berdasarkan hasil pengujian yang tersimpan dalam tabel *train_history*, algoritma K-NN *consistently* menunjukkan akurasi pelatihan lebih tinggi, yaitu

Vol. 1, No. 2, pp. 64 - 71, 2025
p-ISSN. xxxx - xxxx, e-ISSN. 3090 - 3602

sekitar 82.85%, dibandingkan dengan Naive Bayes yang memiliki akurasi sekitar 82.42%. Meski perbedaannya kecil, K-NN terbukti lebih unggul dalam akurasi prediksi. Di sisi lain, Naive Bayes unggul dalam kecepatan proses karena algoritmanya lebih sederhana dan efisien. Dengan demikian, kedua algoritma memiliki keunggulan masing-masing dan dapat dipilih berdasarkan prioritas pengguna: apakah lebih mengutamakan akurasi atau efisiensi waktu prediksi.

DAFTAR PUSTAKA

Amna, Gede Iwan Sudipa, I., Andi Putra, T. E., Jurnaidi Wahidin, A., Alfa Syukrilla, W., Khrisna Wardhani, A., Heryana, N., Indriyani, T., Willyanto Santoso, L., & S, W. (2023). *Data Mining* (D. Ediana & Yanto Ari, Eds.; 1st ed.). PT.GLOBALEKSEKUTIFTEKNOLOGI. www.globaleksekutifteknologi.co.id

Azeroual, O., Nacheva, R., Nikiforova, A., & Störl, U. (2025). A CRISP-DM and Predictive Analytics Framework for Enhanced Decision-Making in Research Information Management Systems. *Informatica*, *49*(18), 67–86. https://doi.org/10.31449/inf.v49i18.5613

Bansal, M., Goyal, A., & Choudhary, A. (2022). A Comparative Analysis of K-Nearest Neighbor, Genetic, Support Vector Machine, Decision Tree, and Long Short Term Memory Algorithms in Machine Learning. *Decision Analytics Journal*, 100071. https://doi.org/10.1016/j.dajour.2022.100071

Cutting, V., & Stephen, N. (2021). A Review on using Python as a Preferred Programming Language for Beginners. *International Research Journal of Engineering and Technology (IRJET)*, *08*(08). https://www.researchgate.net/publication/359379004

Effendy, E., Siregar, E. A., Fitri, P. C., & Damanik, I. A. S. (2023). Mengenal Sistem Informasi Manajemen Dakwah (Pengertian Sistem, Karakteristik Sistem). *Jurnal Pendidikan Dan Konseling*, *5*, 4343–4349. https://doi.org/10.31004/jpdk.v5i2.14061

Gholizadeh, S. (2022). Top Popular Python Libraries in Research. *Journal of Robotics and Automation Research*, *3*(2), 142–145. https://doi.org/10.22541/au.164580055.55493761/v1

Published by Universitas Buddhi Dharma
Available online at https://jurnal.ubd.ac.id/index.php/poters/index
Kota Tangerang, 9 Juli 2025.

Vol. 1, No. 2, pp. 64 - 71, 2025
p-ISSN. xxxx - xxxx, e-ISSN. 3090 - 3602

Hansen, & Hariyanto, S. (2023). Perbandingan Algoritma Data Mining Dalam Mengklasifiksi Penyakit Diabetes Menggunakan Model C4.5 Dan Naive Bayes. *JURNAL ALGOR*, *4*(2). https://jurnal.buddhidharma.ac.id/index.php/algor/index

Joseph, V. R. (2022). Optimal ratio for data splitting. *Statistical Analysis and Data Mining*, *15*(4), 531–538. https://doi.org/10.1002/sam.11583

Kartika Dewi, N., & Yahfizham. (2023). Pengenalan Dasar Algoritma Pemograman Bagi Mahasiswa. *JournalOfInformaticsAndBusisnes*, *01*(03), 156–161.

Kumar, D. (2020). Cardiovascular Disease Prediction Using Machine Learning. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, *6*(5), 46–54. https://doi.org/10.32628/CSEIT20659

Muharrom, M. (2023). Analisis Komparasi Algoritma Data Mining Naive Bayes, K-Nearest Neighbors dan Regresi Linier Dalam Prediksi Harga Emas. *BULLETIN OF INFORMATION TECHNOLOGY (BIT)*, *4*(4), 430–438. https://doi.org/10.47065/bit.v3i1

Pan, Z., Wang, Y., & Pan, Y. (2020). A New Locally Adaptive K-Nearest Neighbor Algorithm Based On Discrimination Class. *Knowledge-Based Systems*, *204*. https://doi.org/10.1016/j.knosys.2020.106185

Rino. (2021). Comparison of Data Mining Methods Using C4.5 Algorithm and Naive Bayes in Predicting Heart Disease. *JournalTECH-E*, *4*(2). http://bsti.ubd.ac.id/e-jurnal

Sinaga, L. M., Sawaluddin, & Suwilo, S. (2020). Analysis of Classification and Naïve Bayes Algorithm K-Nearest Neighbor in Data Mining. *IOP Conference Series: Materials Science and Engineering*, *725*, 1–5. https://doi.org/10.1088/1757-899X/725/1/012106

Timotius, A., & Fenriana, I. (2024). Perancangan Aplikasi Prediksi Penyakit Jantung Menggunakan Metode Naive Bayes. *Jurnal Algor*, *5*(2). https://jurnal.buddhidharma.ac.id/index.php/algor/index

Published by Universitas Buddhi Dharma
Available online at https://jurnal.ubd.ac.id/index.php/poters/index
Kota Tangerang, 9 Juli 2025.