

Perbandingan Kedua Algoritma Decision Tree dan Naive Bayes untuk Prediksi Gagal Jantung

Aji Setiawan¹, Alvin Nugraha Saputra², Zahra Alifia Dewi³, Yamin Nuryamin, M.Kom⁴, Ade Priyatna, M.Kom⁵

¹ajiisetiawan09@gmail.com, ²alvinoegraha@gmail.com, ³zahralfia18@gmail.com, ⁴yamin.yny@bsi.ac.id,

⁵ade.aeq@bsi.ac.id

^{1,2,3} Mahasiswa, Teknologi Informasi Universitas Bina Sarana Informatika,

^{4,5} Dosen, Teknologi Informasi Universitas Bina Sarana Informatika.

Abstract— Heart disease remains a leading global cause of mortality; however, early diagnosis is hampered by limited trained medical personnel and diagnostic resources. This research compares two classification algorithms—Decision Tree C4.5 and Naive Bayes—to predict heart failure risk. Using a preprocessed dataset split into 80% training and 20% testing data, we evaluated both approaches' effectiveness in detecting complex disease patterns. Results demonstrate Naive Bayes achieved superior accuracy of 86.41%, outperforming Decision Tree C4.5 at 80.43%. Despite relying on feature independence assumptions, the probabilistic approach of Naive Bayes proved more adaptive to the unique characteristics of cardiac datasets than the tree-based method. These findings establish Naive Bayes as a promising reference for early diagnostic systems that can accelerate clinical diagnosis and enhance predictive accuracy, ultimately contributing to improved patient outcomes and lives saved.

Intisari— Penyakit jantung tetap menjadi penyebab utama kematian global, namun diagnosis dini masih terhambat oleh keterbatasan tenaga medis terlatih dan sumber daya diagnostik. Penelitian ini membandingkan dua algoritma klasifikasi—Decision Tree C4.5 dan Naive Bayes—untuk memprediksi risiko gagal jantung. Menggunakan dataset yang telah dipreprocessing dan dibagi 80% pelatihan serta 20% pengujian, kami mengevaluasi efektivitas kedua pendekatan dalam mendeteksi pola kompleks penyakit jantung. Hasil menunjukkan Naive Bayes mencapai akurasi 86,41%, melampaui Decision Tree C4.5 yang memperoleh 80,43%. Meskipun berbasis asumsi independensi fitur, pendekatan probabilistik Naive Bayes terbukti lebih adaptif terhadap karakteristik unik dataset jantung dibanding pendekatan berbasis pohon keputusan. Temuan ini menunjukkan potensi Naive Bayes sebagai referensi untuk sistem diagnosa dini yang dapat mempercepat diagnosis klinis dan meningkatkan akurasi prediksi risiko, berkontribusi signifikan dalam penyelamatan nyawa.

Kata Kunci— Gagal Jantung, perbandingan algoritma, decision tree (C4.5), naive bayes.

I. PENDAHULUAN

Penyakit kardiovaskular atau gagal jantung, merupakan penyebab kematian nomor satu di seluruh dunia [1]. Data dari *World Health Organization* (WHO) menunjukkan bahwa penyakit jantung merenggut jutaan nyawa setiap tahunnya dan terus meningkat, termasuk di negara Indonesia. Gagal jantung merupakan kondisi kronis di mana jantung tidak mampu memiliki kekuatan memompa darah secara efisien guna memenuhi kebutuhan tubuh, yang mengakibatkan tingginya angka mortalitas dan morbiditas, serta berdampak signifikan pada kualitas hidup pasien [2].

Kompleksitas diagnosis dan kebutuhan penanganan segera menjadikan prediksi dini sebagai tantangan besar di dunia medis. Keterbatasan tenaga ahli dan sumber daya seringkali menghambat proses diagnosis yang cepat dan tepat [3]. Dalam konteks ini, *data mining* dan *machine learning* hadir sebagai pendekatan inovatif. Dengan menganalisis data riwayat medis pasien dalam jumlah besar, teknik ini mampu menemukan pola tersembunyi dan membangun model prediksi untuk mengidentifikasi individu berisiko tinggi [4].

Di antara banyaknya algoritma klasifikasi, Decision Tree (C4.5) dan Naive Bayes adalah kedua metode yang cukup banyak digunakan karena efektivitas dan kemudahannya untuk diinterpretasikan[5]. Beberapa penelitian telah melaporkan kinerja yang baik dari kedua algoritma ini dalam berbagai kasus medis[6]. Namun, perbandingan langsung untuk kasus prediksi gagal jantung masih perlu dikaji lebih dalam untuk memberikan panduan yang jelas bagi para praktisi dan peneliti.

Dalam pengembangan model prediktif yang efektif, pemilihan fitur yang tepat memiliki peran krusial. Prinsip pemilihan fitur ini juga berlaku pada algoritma klasifikasi untuk diagnosis medis, di mana fitur-fitur dengan korelasi kuat terhadap target (seperti MaxHR, Oldpeak, dan ST_Slope pada prediksi penyakit jantung) akan menjadi predictor yang lebih powerful dalam model klasifikasi [7].

Karena itu, tujuan dari penelitian ini adalah untuk menganalisis dan membandingkan secara komprehensif kinerja dari kedua algoritma tersebut dalam memprediksi risiko penyakit jantung.

II. TINJAUAN PUSTAKA

Penelitian sebelumnya di bidang *data mining* telah banyak mengeksplorasi penggunaan algoritma klasifikasi untuk mendukung diagnosis dan prediksi penyakit jantung. Tinjauan ini secara khusus berfokus pada studi-studi yang menerapkan dan membandingkan algoritma Decision Tree (C4.5) dan Naive Bayes dalam konteks diagnosis dan prediksi penyakit jantung. Sejumlah penelitian melakukan perbandingan langsung antara keduanya untuk menentukan model mana yang akurat dalam memprediksi penyakit jantung.

Gagal jantung merupakan kondisi kronis di mana jantung tidak mampu untuk memompa darah secara efisien guna memenuhi kebutuhan tubuh, sehingga mengakibatkan tingginya angka mortalitas dan morbiditas, serta berdampak signifikan pada kualitas hidup pasien[2].

faktor risiko utama penyakit jantung koroner dan serangan berulang mencakup perilaku merokok, hipertensi, aktivitas fisik, pola makan (terutama mengonsumsi karbohidrat dan lemak), serta kepatuhan minum obat[2].

Sebuah studi komparatif oleh Khodijah & Sriyanto [3] menguji tiga algoritma populer: C4.5, Naive Bayes, dan Random Forest. Menunjukkan bahwa algoritma C4.5 memiliki kinerja yang sangat baik dengan akurasi mencapai 88,34%. Algoritma Naive Bayes juga menunjukkan performa kompetitif dengan akurasi 87,25%.

Hasanah et al. [6] menganalisis penggunaan varian Gaussian Naive Bayes (GNB) khusus untuk data pasien gagal jantung yang bersifat kontinu. Hasil validasi menggunakan 10-fold menunjukkan akurasi tertinggi yang dicapai GNB adalah 69%. Meskipun akurasinya lebih rendah dibandingkan studi lain, GNB menunjukkan nilai *recall* tertinggi sebesar 95,91%, yang mengindikasikan kemampuan model yang sangat baik dan unggul dalam mengidentifikasi pasien penderita gagal jantung.

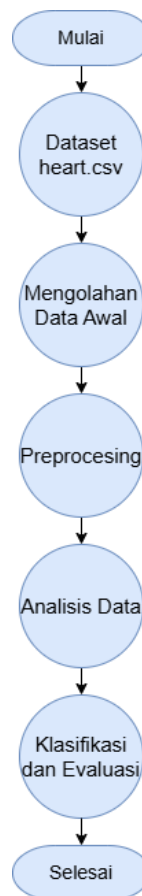
Penelitian terdahulu telah mengeksplorasi berbagai algoritma untuk prediksi penyakit jantung. Subarkah et al. [8] melakukan perbandingan komprehensif antara algoritma Correlated Naive Bayes Classifier (C-NBC) dan Naive Bayes Classifier (NBC) pada dataset gagal jantung dari UCI Repository. Hasil penelitian menunjukkan bahwa algoritma C-NBC mencapai akurasi 80.6%, melampaui Naive Bayes Classifier standar yang hanya mencapai 76.5%. Penelitian ini membuktikan bahwa dengan mempertimbangkan korelasi fitur, kinerja algoritma dapat ditingkatkan secara signifikan.

Arifudin & Kristinawati [9] menemukan bahwa masalah psikologis seperti kecemasan juga depresi berdampak pada hidup pasien gagal jantung, yang merupakan aspek krusial dalam perawatan jangka panjang.

Nurdianto & Prihati [10] secara khusus menganalisis performansi C4.5. Mereka menemukan bahwa nilai akurasi terbaik sebesar 88,35% diperoleh perbandingan data latih dan data uji sebesar 90%:10%.

III. METODE PENELITIAN

Pada penelitian ini, proses perancangan telah direncanakan dan disusun secara sistematis agar alur penelitian sesuai dengan tujuan. Pendekatan yang digunakan adalah metode kuantitatif, di mana dilakukan pada populasi atau sampel tertentu dengan bantuan instrumen pengumpulan data yang sudah dirancang sebelumnya. Sehingga data yang diperoleh lalu diolah dan dianalisis secara statistik untuk menguji hasil hipotesis. Metode ini disebut discovery karena melalui proses ini, peneliti menemukan informasi baru guna mengembangkan pengetahuan yang sudah ada menjadi lebih kaya dan bermakna.



Gbr. 1 Diagram Metode Alur Penelitian

A. Pengumpulan Data

Berbagai penelitian sebelumnya telah mencoba berbagai cara dalam memprediksi penyakit jantung, namun sayangnya belum terdapat kesimpulan pasti terkait algoritma mana yang paling akurat. Oleh karena itu, penelitian ini bertujuan untuk membandingkan kinerja dari kedua algoritma tersebut, guna mengetahui metode mana yang memiliki hasil prediksi terbaik pada kasus penyakit jantung.

Dataset yang digunakan merupakan data publik dari Kaggle: <https://www.kaggle.com/datasets/fedesoriano/heart-failure-prediction>. Dataset terdiri dari 918 data pasien dengan 12 atribut, yang mencakup berbagai informasi kondisi kesehatan terkait risiko gagal jantung. Dataset tersebut dipilih karena bersifat terbuka, banyak digunakan dalam penelitian sejenis, dan memiliki variasi informasi yang cukup luas untuk proses klasifikasi.

B. Praproses

Setelah dataset dipastikan bersih dan siap digunakan, tahap preprocessing dilakukan untuk mempersiapkan data bagi algoritma machine learning. Pada tahap ini, variabel target yaitu HeartDisease dipisahkan dari fitur input. Atribut kategorikal seperti Sex, ChestPainType, RestingECG, Exercise Angina, dan ST_Slope kemudian dikonversi menjadi data numerik menggunakan teknik One-Hot Encoding, agar seluruh fitur dapat dibaca oleh model secara optimal. Selain itu, atribut numerik seperti usia, tekanan darah, kadar kolesterol, detak jantung maksimum, dan nilai Oldpeak distandarisasi menggunakan StandardScaler. Proses penyetaraan skala ini dilakukan untuk menghindari perbedaan rentang nilai antar fitur yang dapat mempengaruhi performa sebagian algoritma, terutama Naive Bayes yang lebih sensitif terhadap skala data.

Setelah seluruh proses konversi dan normalisasi selesai, Kami membagi dataset menjadi 80% data pelatihan (734 sampel) dan 20% data pengujian (184 sampel). Pembagian ini memungkinkan model belajar dari sebagian besar data dan menguji kemampuannya pada data yang belum pernah dilihat sebelumnya.

C. Analisis Data

Tahap analisis data (Exploratory Data Analysis/EDA) dilakukan untuk memahami karakteristik dan pola dataset sebelum membangun model. Pada tahap ini, kami mengevaluasi kelengkapan data, distribusi variabel, keseimbangan kelas, korelasi antar fitur, dan kehadiran outliers. Analisis mendalam ini memastikan dataset siap digunakan dan mengidentifikasi fitur-fitur yang paling berpengaruh terhadap prediksi penyakit jantung.

a. Pemeriksaan Kelengkapan Data

Dataset terdiri dari 918 sampel dengan 12 atribut. Hasil pemeriksaan menunjukkan tidak ada missing values (0%) dan data duplikat, sehingga semua sampel dapat digunakan tanpa perlu imputasi atau penghapusan data.

b. Statistik Deskriptif Variabel Numerik

Analisis statistik menunjukkan karakteristik data sebagai berikut:

1. Age: 28-77 tahun, rata-rata 53.51 tahun
2. RestingBP: 0-200 mmHg, rata-rata 131.62 mmHg (tergolong tinggi)
3. Cholesterol: 0-603 mg/dL, rata-rata 246.26 mg/dL
4. MaxHR: 60-202 bpm, rata-rata 136.81 bpm
5. Oldpeak: 0-6.2, rata-rata 1.04 (mayoritas < 2.0)

Data menunjukkan sebaran yang beragam dan mencerminkan populasi pasien dengan risiko penyakit jantung.

c. Analisis Distribusi Kelas

Hasil pemeriksaan distribusi target variable:

1. Kelas 0 (Tanpa penyakit): 508 sampel (55.34%)
2. Kelas 1 (Dengan penyakit): 410 sampel (44.66%)
3. Rasio: 1.24:1 (Seimbang)

Keseimbangan kelas yang baik ini memastikan model tidak akan bias terhadap satu kelompok dan metrik akurasi akan reliabel.

d. Analisis Korelasi Fitur dengan Target

Hasil korelasi Pearson menunjukkan variabel-variabel dengan hubungan terkuat terhadap penyakit jantung:

1. MaxHR: -0.422 (korelasi negatif kuat) - Detak jantung maksimum rendah = risiko tinggi
2. Oldpeak: +0.405 (korelasi positif kuat) - Depresi ST tinggi = risiko tinggi
3. ST_Slope: -0.380 (korelasi negatif sedang) - Pola EKG abnormal = risiko tinggi
4. ExerciseAngina: +0.325, Age: +0.285, Sex: +0.286 (korelasi sedang)
5. RestingBP: +0.101, Cholesterol: +0.089 (korelasi lemah)

Fitur-fitur dengan korelasi kuat (MaxHR, Oldpeak, ST_Slope) akan menjadi predictor yang lebih powerful dalam model klasifikasi.

e. Analisis Multicollinearity

Pemeriksaan korelasi antar fitur menunjukkan bahwa semua pasangan fitur memiliki korelasi < 0.5, dengan hubungan terkuat adalah:

1. Age vs MaxHR: -0.398 (masuk akal medis)
2. MaxHR vs Oldpeak: -0.305
3. Semua pasangan lainnya: < 0.15 (sangat lemah)

Hasil ini menunjukkan tidak ada multicollinearity yang problematic, sehingga fitur-fitur cukup independen untuk digunakan dalam algoritma Naive Bayes.

f. Identifikasi Outliers

Analisis outliers menggunakan IQR method menunjukkan:

1. Age: 0.87% outliers (sangat sedikit)
2. RestingBP & Cholesterol: 2.4% - 3.05% outliers
3. MaxHR: 0% outliers (sempurna)
4. Oldpeak: 14.16% outliers (paling banyak, namun penting secara medis)

Semua outliers dipertahankan karena dalam konteks medis nilai-nilai ekstrem merepresentasikan kondisi pasien yang nyata dan memiliki informasi klinis penting.

g. Perbandingan Fitur Antar Kelas

Perbandingan rata-rata fitur antara pasien sehat dan sakit:

1. Oldpeak: 0.68 (sehat) vs 1.54 (sakit) = +126.5%
2. MaxHR: 142.29 bpm (sehat) vs 129.33 bpm (sakit) = -9.1%
3. Age: 52.13 (sehat) vs 55.57 (sakit) = +6.6%
4. RestingBP: 129.40 (sehat) vs 134.93 (sakit) = +4.3%
5. Cholesterol: 246.02 (sehat) vs 246.59 (sakit) = +0.2% (tidak berbeda)

Hasil menunjukkan Oldpeak dan MaxHR adalah pembeda paling jelas antara pasien sehat dan sakit.

D. Pengujian Algoritma

a. Algoritma Decision Tree

Algoritma ini memiliki kemampuan memecah data ke dalam beberapa cabang keputusan berdasarkan nilai fitur. Pemilihan atribut terbaik pada setiap percabangan dilakukan dengan menghitung impurity atau ketidakmurnian data. Salah satu metode umum yang digunakan adalah Entropy dan Information Gain.

b. Algoritma Naive Bayes

Algoritma yang menggunakan prinsip Teorema Bayes dengan asumsi fitur yang saling independen. Tujuan utamanya adalah menghitung probabilitas kelas berdasarkan data masukan.

IV. HASIL & PEMBAHASAN

A. Pengumpulan Data

Penelitian ini menggunakan dataset publik dari Kaggle bernama Heart Failure Prediction yang terdiri dari 918 sampel pasien dengan 12 atribut yang mencakup berbagai informasi kondisi kesehatan terkait risiko penyakit jantung. Dataset dipilih karena bersifat terbuka, telah banyak digunakan dalam penelitian sejenis, dan memiliki variasi informasi yang cukup luas untuk proses klasifikasi. Setelah dataset dikonfirmasi bersih dan siap digunakan, tahap preprocessing dilakukan sesuai prosedur yang telah dijelaskan pada bagian Metode Penelitian (III.B). Tabel 1 menunjukkan sampel data mentah sebelum preprocessing, dengan kolom-kolom yang mencakup informasi demografis (Age, Sex), hasil pemeriksaan klinis (RestingBP, Cholesterol, MaxHR, Oldpeak), dan tipe kondisi (ChestPainType, RestingECG, ExerciseAngina, ST_Slope). Setelah tahap preprocessing yang meliputi One-Hot Encoding untuk variabel kategorikal dan StandardScaler untuk normalisasi variabel numerik, data siap untuk tahap pelatihan dan pengujian model.

TABEL 1 Struktur dan Sampel Data yang Digunakan

No	Age	Sex	CP Type	RestBP	Chol	MaxHR	Oldpeak	ST Slope	HeartDisease
1	40	M	ATA	140	289	172	0.0	Up	0
2	49	F	NAP	160	180	156	1.0	Flat	1
3	37	M	ATA	130	283	98	0.0	Up	0
4	48	F	ASY	138	214	108	1.5	Flat	1
5	54	M	NAP	150	195	122	0.0	Up	0

B. Praproses

Tahap analisis eksplorasi data (EDA) dilakukan untuk memahami karakteristik dan pola dalam dataset sebelum pelatihan model. Hasil analisis menunjukkan distribusi kelas yang tergolong seimbang, dengan 410 sampel pasien yang memiliki penyakit jantung (45%) dan 508 sampel pasien tanpa penyakit jantung (55%). Distribusi yang seimbang ini memastikan bahwa model tidak akan bias terhadap salah satu kelas, sehingga tidak diperlukan teknik balancing khusus seperti oversampling atau undersampling. Analisis korelasi fitur dengan target menunjukkan bahwa beberapa fitur memiliki hubungan yang lebih kuat dengan penyakit jantung dibandingkan fitur lainnya. Fitur-fitur seperti Age (usia), MaxHR (detak jantung maksimum), dan Oldpeak (depresi segmen ST) menunjukkan korelasi signifikan dengan risiko penyakit jantung, mengindikasikan bahwa faktor-faktor tersebut berpotensi menjadi predictor yang kuat dalam model klasifikasi. Gbr. 3 menampilkan implementasi kode preprocessing yang menunjukkan proses One-Hot Encoding untuk variabel kategorikal dan StandardScaler untuk normalisasi. Setelah preprocessing selesai, dataset dibagi menjadi 80% data pelatihan (734 sampel) dan 20% data pengujian (184 sampel) untuk memastikan evaluasi model yang objektif dan reliabel.

```
# 9 Pisahkan fitur dan target
X = data.drop('HeartDisease', axis=1)
y = data['HeartDisease']

# 10 Tentukan kolom kategorikal dan numerik
categorical = ['Sex', 'ChestPainType', 'RestingECG', 'ExerciseAngina', 'ST_Slope']
numeric = X.select_dtypes(include=np.number).columns.tolist()

# 1.1 Preprocessing
preprocessor = ColumnTransformer(
    transformers=[
        ('onehot', OneHotEncoder(handle_unknown='ignore'), categorical),
        ('scaler', StandardScaler(), numeric)
    ]
)

# 1.2 Split Train-Test
X_train, X_test, y_train, y_test = train_test_split(
    X, y, test_size=0.2, random_state=42
)
```

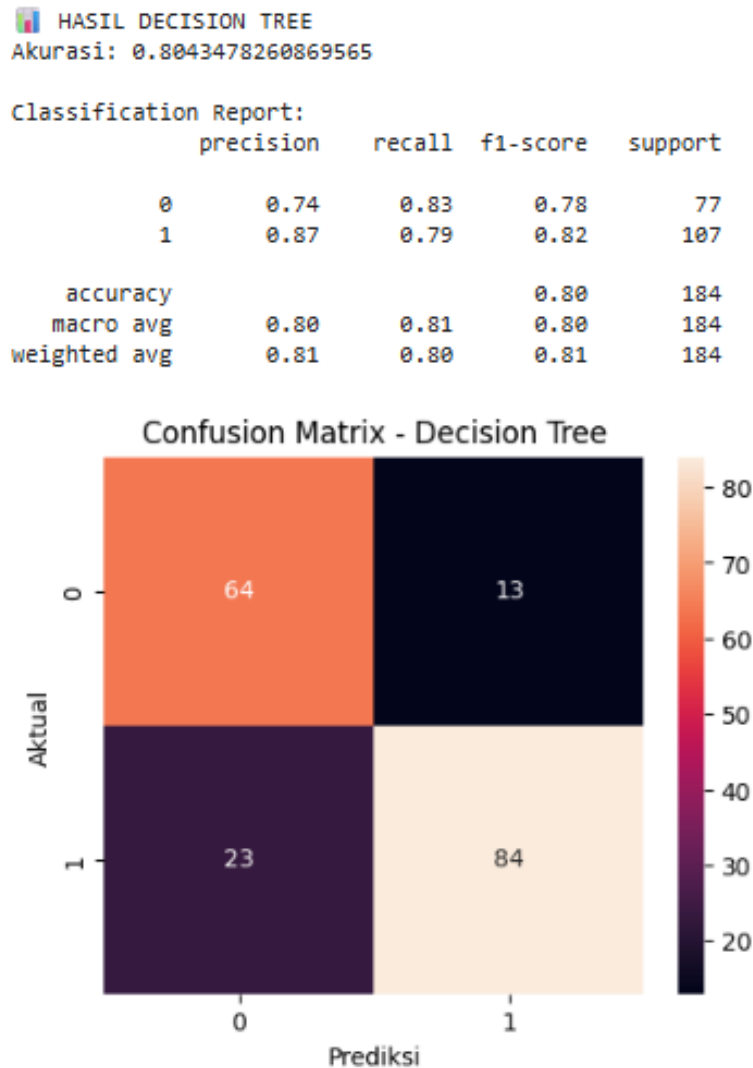
Gbr. 2 Kode Praproses

C. Hasil Algoritma

Setelah preprocessing dan pelatihan model selesai, kedua algoritma diuji pada dataset pengujian untuk mengevaluasi kinerja mereka. Berikut adalah hasil evaluasi untuk masing-masing algoritma.

a. Decision Tree

Pengujian mencakup gambaran mengenai kemampuan model dalam membedakan pasien dengan dan tanpa penyakit jantung. Secara keseluruhan, model mencatat tingkat akurasi sebesar 80.43%. Detail performanya dapat dilihat pada tabel berikut:

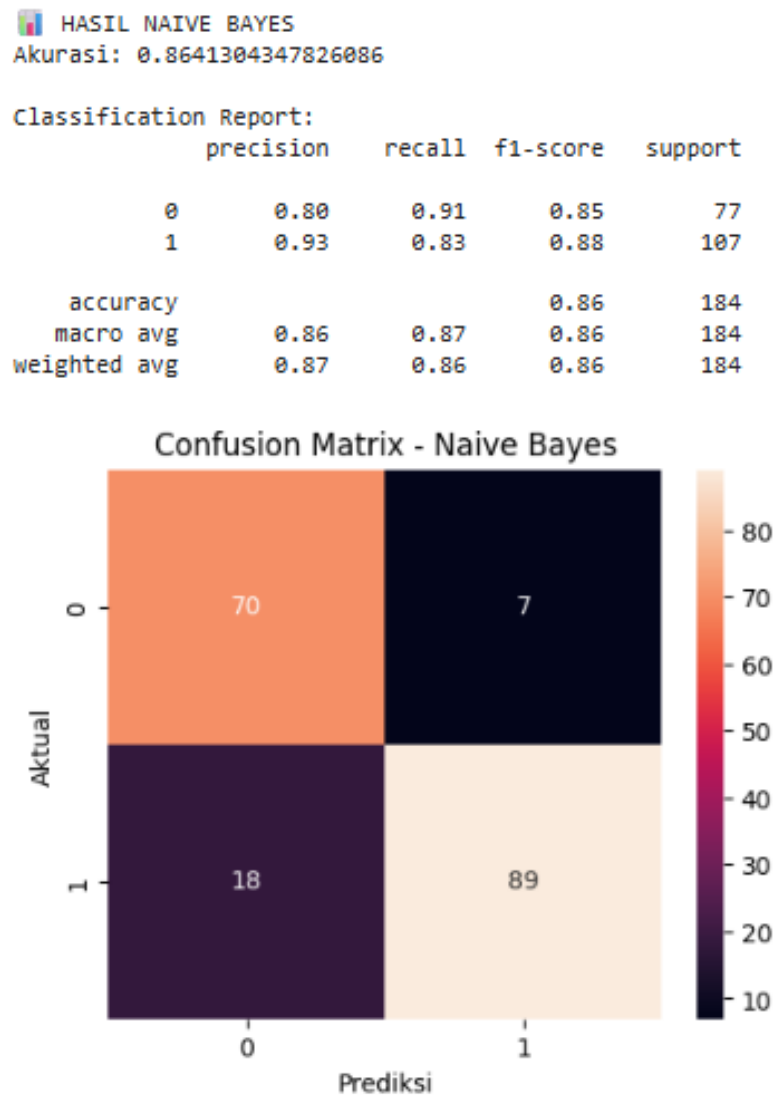


Gbr. 3 Hasil Algoritma Decision Tree

Model Decision Tree terlihat cukup baik dalam mengidentifikasi pasien yang tidak memiliki penyakit jantung (kelas 0), ditunjukkan oleh nilai recall 0.83. Artinya, sebagian besar individu sehat berhasil dikenali dengan benar oleh model. Meski begitu, kemampuan model dalam mengenali pasien yang berpotensi mengalami penyakit jantung masih dapat ditingkatkan. Nilai recall pada kelas pasien berisiko berada di angka 0.79, yang menunjukkan masih terdapat beberapa kasus positif yang belum terdeteksi secara optimal. Secara keseluruhan, performa ini menunjukkan Decision Tree mampu memberikan hasil yang cukup solid, namun masih ada ruang perbaikan—terutama dalam mendeteksi pasien dengan risiko penyakit jantung lebih tinggi.

b. Naive Bayes

Pada bagian pengujian yang dilakukan ini melihat sejauh mana model mampu memprediksi risiko penyakit jantung. Secara keseluruhan, algoritma ini memiliki performa yang cukup kuat, dengan akurasi 86.41%, lebih tinggi dibandingkan model Decision Tree. Rincian hasil evaluasi ditampilkan pada tabel berikut:

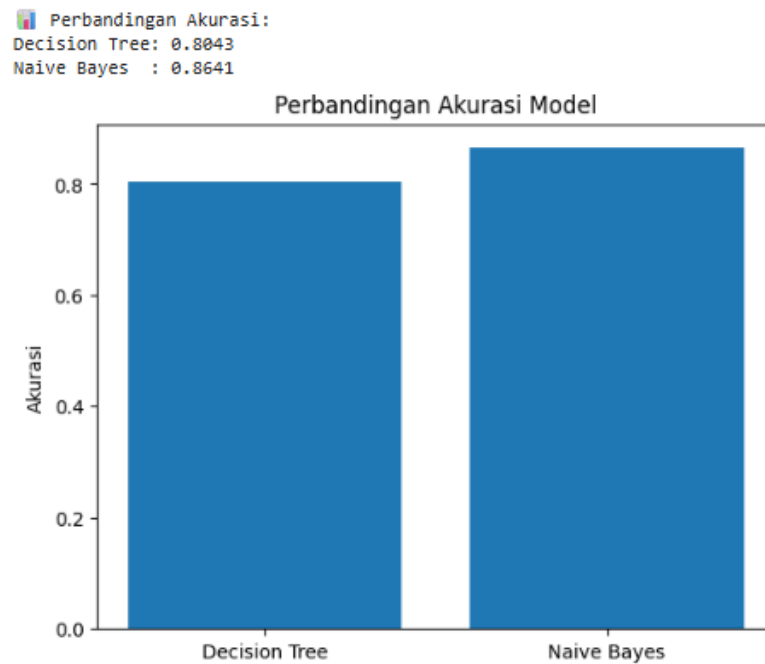


Gbr. 4 Hasil Algoritma Naive Bayes

Naive Bayes menunjukkan kinerja yang lebih baik, terutama dalam mencari tahu kondisi pasien yang memiliki penyakit jantung. Terlihat dari nilai precision kelas 1 yang mencapai 0.93, yang berarti sebagian besar prediksi pasien berisiko benar adanya. Selain itu, nilai recall pada kelas pasien sehat kelas 0 juga cukup tinggi, yaitu 0.91, menandakan model mampu mengenali mayoritas individu tanpa penyakit jantung dengan akurat. Secara keseluruhan, model Naive Bayes memberikan hasil yang konsisten dan seimbang pada kedua kelas, serta menunjukkan kinerja yang lebih unggul dan baik dibandingkan Decision Tree dalam penelitian ini. Hasil akhir ini mengindikasikan bahwa pendekatan probabilistik yang digunakan cukup baik dalam menangani karakteristik data pada kasus prediksi penyakit jantung.

V. KESIMPULAN

Penelitian ini membandingkan kinerja Decision Tree (C4.5) dan Naive Bayes dalam memprediksi risiko penyakit jantung menggunakan 918 sampel pasien dengan 12 atribut dari Kaggle. Semua proses, dari pengolahan data hingga evaluasi model, dilakukan menggunakan Python di platform Google Colab. Tahap preprocessing meliputi: (1) pemisahan fitur dan label target, (2) One-Hot Encoding untuk variabel kategorikal, (3) standarisasi numerik dengan StandardScaler, dan (4) pembagian data menjadi 80% pelatihan dan 20% pengujian..



Gbr. 5 Hasil Perbandingan Algoritma Decision Tree & Naive Bayes

Pengujian menggunakan pembagian data 80% pelatihan dan 20% pengujian menunjukkan bahwa kedua algoritma mampu memberikan performa yang baik, namun Naive Bayes lebih superior dibandingkan Decision Tree. Algoritma Naive Bayes mencapai akurasi sebesar 86.41%, sementara Decision Tree menghasilkan akurasi 80.43%.

UCAPAN TERIMA KASIH

Terima kasih kepada Bapak Yamin Nuryamin, M.Kom dan Ade Priyatna, M.Kom yang memberikan kesempatan dan membimbing dalam proses membuat perbandingan algoritma yang terbaik untuk diagnosa penyakit gagal jantung.

REFERENSI

- [1] W. Hanifah, O. S. Wanda, and H. Nisa, "Penelitian gizi dan makanan.," *Nutr. Food Res.*, vol. 43, no. 1, pp. 29–40, 2020, [Online]. Available: <https://www.neliti.com/publications/223576/hubungan-asupan-energi-lemak-dan-serat-dengan-rasio-kadar-kolesterol-total-hdl>
- [2] V. Leutualy *et al.*, "Faktor Risiko yang Berhubungan dengan Kejadian Serangan Berulang Pasien Penyakit Jantung Koroner; Literature Review," *JUSTE (Journal Sci. Technol.*, vol. 3, no. 1, pp. 68–79, 2022, doi: 10.51135/justevol3issue1page68-79.
- [3] 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.
- [4] E. Nurlia and U. Enri, "Penerapan Fitur Seleksi Forward Selection Untuk Menentukan Kematian Akibat Gagal Jantung Menggunakan Algoritma C4.5," *J. Tek. Inform. Musirawas) Elin Nurlia*, vol. 6, no. 1, p. 42, 2021.
- [5] L. Saroinsong, E. L. Jim, and S. H. Rampengan, "Diagnosis dan Tatalaksana Terkini Gagal Jantung Akut," *e-CliniC*, vol. 9, no. 1, pp. 60–67, 2021, doi: 10.35790/ecl.v9i1.31857.
- [6] Q. Hasanah, H. Oktavianto, and Y. D. Rahayu, "Analisis Algoritma Gaussian Naive Bayes Terhadap Klasifikasi Data Pasien Penderita Gagal Jantung," *J. Smart Teknol.*, vol. 3, no. 4, pp. 382–389, 2022, [Online]. Available: <http://jurnal.unmuhjember.ac.id/index.php/JST/article/view/7597>
- [7] Sekar Setyaningtyas, B. Indarmawan Nugroho, and Z. Arif, "Tinjauan Pustaka Sistematis Pada Data Minin: Studi Kasus Algoritma K-Means Clustering," *J. Teknoif Tek. Inform. Inst. Teknol. Padang*, vol. 10, no. 2, pp. 52–61, 2022.
- [8] P. Subarkah, W. R. Damayanti, and R. A. Permana, "Comparison of Correlated Algorithm Accuracy Naive Bayes Classifier and Naive Bayes Classifier for Classification of heart failure," *Ilk. J. Ilm.*, vol. 14, no. 2, pp. 120–125, 2022, doi: 10.33096/ilkom.v14i2.1148.120-125.
- [9] N. F. Arifudin and B. Kristinawati, "The Impact of Psychological Problems on the Quality of Life of Heart Failure Patients," *Heal. Inf. J. Penelit.*, vol. 15, p. e796, 2023, [Online]. Available: <https://myjurnal.poltekkes-kdi.ac.id/index.php/hijp>
- [10] P. K. Nurdianto, "Jurnal Cakrawala Informasi," *Cakrawala Inf.*, vol. 1, no. 1, pp. 1–14, 2021.