

## **Prediksi Penyakit Gagal Jantung Menggunakan Algoritma Naive Bayes** *The Prediction Of Brain Failure Using The Naive Bayes Algorithm*

**Moch. Rafli Febrin<sup>1)</sup>, Ilham Saifudin<sup>2)</sup>, Wiwik Suharso<sup>3)</sup>**

<sup>1</sup>Mahasiswa Fakultas Teknik, Universitas Muhammadiyah Jember  
email: febrinrafli26@gmail.com

<sup>2</sup>Dosen Fakultas Teknik, Universitas Muhammadiyah Jember  
email: ilhamsaifudin@unmuhjember.ac.id

<sup>3</sup>Dosen Fakultas Teknik Universitas Muhammadiyah Jember  
email: wwiksuharso@unmuhjember.ac.id

### **Abstrak**

Gagal jantung adalah kondisi medis serius di mana jantung tidak dapat memompa darah dengan baik, sering disebabkan oleh hipertensi, diabetes, dan penyakit jantung koroner. Penyakit jantung adalah salah satu penyakit paling mematikan di dunia, dengan lebih dari 17,7 juta kematian setiap tahun menurut WHO. Mengingat angka kematian yang tinggi, diagnosis dini meningkatkan peluang bertahan hidup. Studi ini membuat model prediksi gagal jantung menggunakan algoritma Naive Bayes, yang populer karena cepat dan mudah digunakan. Tujuan model ini adalah membantu tenaga medis menemukan pasien berisiko tinggi, memungkinkan intervensi dini yang lebih baik. Algoritma Naive Bayes dipilih karena kelebihanannya dalam kesederhanaan dan kecepatan proses, penting dalam situasi medis yang membutuhkan keputusan cepat. Pengujian model menggunakan dataset relevan menunjukkan bahwa algoritma Naive Bayes memiliki tingkat akurasi memadai dalam memprediksi risiko gagal jantung. Dengan demikian, model ini dapat diintegrasikan dalam sistem kesehatan untuk meningkatkan efektivitas diagnosis dan perawatan, meningkatkan kualitas hidup pasien, dan mengelola sumber daya medis lebih efisien. Selain itu, model ini dapat dikembangkan lebih lanjut untuk memperhitungkan berbagai variabel klinis tambahan, meningkatkan akurasi dan kegunaannya dalam skenario medis yang lebih luas.

**Keywords:** *Gagal Jantung, Naive Bayes, Prediksi.*

### **Abstract**

*Heart failure is a serious medical condition in which the heart cannot pump blood properly, often caused by hypertension, diabetes, and coronary heart disease. Heart disease is one of the deadliest diseases in the world, with more than 17.7 million deaths each year according to the WHO. Given the high mortality rate, early diagnosis increases the chances of survival. The study created a model for predicting heart failure using the Naive Bayes algorithm, which is popular because of its speed and ease of use. The purpose of this model is to help medical personnel find high-risk patients, enabling better early intervention. The Naive Bayes algorithm was chosen because of its advantages in the simplicity and speed of the process, essential in medical situations that require quick decisions. Testing models using relevant datasets showed that the Naive bayes algorithm has an adequate degree of accuracy in predicting the risk of heart failure. Thus, these models can be integrated into the health system to improve the effectiveness of diagnosis and treatment, improve the quality of life of patients, and manage medical resources more efficiently. In addition, the model could be further developed to take into account various additional clinical variables, improving its accuracy and usefulness in wider medical scenarios.*

**Keywords:** *Heart Failure, Naive Bayes, Prediction.*

## 1. PENDAHULUAN

Gagal jantung adalah kondisi medis serius di mana jantung tidak dapat memompa darah dengan cukup efisien untuk memenuhi kebutuhan tubuh. Kondisi ini sering berkembang secara bertahap dan bisa disebabkan oleh berbagai faktor seperti penyakit jantung koroner, tekanan darah tinggi, dan diabetes. Menurut Organisasi Kesehatan Dunia (WHO), lebih dari 17,7 juta orang diperkirakan meninggal karena penyakit jantung, yang menyebabkan 31% dari seluruh kematian secara global. Dengan kata lain bahwa penyakit kardiovaskular khususnya penyakit jantung adalah salah satu penyakit paling mematikan baik di negara maju maupun berkembang, perhatian terhadap penyakit tersebut sangatlah penting dan sangat diperlukan.

Terlepas dari fakta bahwa angka kematian akibat penyakit jantung tinggi, peluang untuk bertahan hidup lebih tinggi jika diagnosis dilakukan sejak awal. Oleh karena itu, peneliti membuat model prediktor untuk menemukan pasien yang memiliki tingkat risiko yang tinggi. Pendekatan pembelajaran mesin (ML) menjadi lebih populer baru-baru ini untuk membangun model untuk diagnosis awal penyakit jantung.

Upaya dalam bidang medis untuk memanfaatkan teknologi pembelajaran mesin adalah prediksi penyakit jantung menggunakan algoritma Naive Bayes. Dokter dapat dengan efektif menentukan risiko penyakit jantung pada pasien dengan algoritma ini karena mudah digunakan dan cepat. Meskipun memiliki beberapa kekurangan, Naive Bayes tetap menjadi pilihan yang baik untuk aplikasi medis karena mampu memberikan diagnosis awal yang akurat dengan pengumpulan dan pra-pemrosesan data yang tepat.

Penelitian ini bertujuan untuk mengembangkan sistem prediksi gagal jantung yang menggunakan algoritma Naive-Bayes. Tujuan utama dari penelitian ini adalah untuk membantu masyarakat dalam mengurangi risiko penyakit gagal jantung dengan memberikan prediksi dini yang akurat. Dengan menggunakan algoritma Naive-Bayes, diharapkan dapat memperoleh nilai akurasi

yang tinggi, yang akan memungkinkan prediksi risiko dilakukan dengan lebih efektif.

## 2. TINJAUAN PUSTAKA

### A. Penyakit Jantung

Penyakit jantung mengacu pada berbagai penyakit yang memengaruhi pembuluh darah dan kesehatan jantung. Secara umum, genetika, gaya hidup tidak sehat, tekanan darah tinggi, diabetes, obesitas, dan tingkat stres yang tinggi adalah beberapa penyebab penyakit jantung. Ini dapat menyebabkan berbagai masalah kesehatan yang melibatkan jantung, mulai dari penyakit arteri koroner hingga gangguan irama jantung. Penyakit jantung koroner (PJK) adalah salah satu kondisi utama dalam spektrum penyakit jantung, di mana pembuluh darah yang mengirimkan darah ke jantung menyempit atau tersumbat, yang mengurangi aliran darah dan oksigen ke jantung. Ini dapat menyebabkan gejala seperti nyeri dada atau angina, serta serangan jantung yang hebat. Penyakit jantung juga dapat mencakup kerusakan pada katup jantung, yang bertanggung jawab untuk menjalankan fungsi jantung.

### B. Prediksi

Prediksi adalah kemampuan untuk menggunakan data historis dan algoritma untuk membuat estimasi atau penilaian tentang nilai masa depan atau status tidak diketahui dari variabel tertentu. Teori di balik prediksi dalam machine learning mencakup beberapa konsep dasar, termasuk pembelajaran dari data, model pembelajaran, dan evaluasi kinerja model.

### C. Machine Learning

Machine learning (ML) adalah cabang kecerdasan buatan (AI) yang memungkinkan sistem komputer belajar dan meningkatkan kinerjanya dari pengalaman tanpa pemrograman eksplisit. Secara umum, ML menggunakan algoritma dan model matematika untuk menganalisis data, mengidentifikasi pola, dan membuat prediksi atau keputusan tanpa intervensi manusia.

ML mencoba meniru bagaimana makhluk cerdas belajar dan menggeneralisasi, dengan dua aplikasi utama: klasifikasi dan prediksi. Biasanya digunakan untuk menggantikan atau meniru perilaku manusia dalam otomatisasi atau penyelesaian masalah (Hasibuan et al., 2022).

ML adalah cabang AI yang dikembangkan agar mesin dapat belajar secara otomatis, memungkinkan analisis data untuk menentukan hubungan antara input dan output yang diinginkan (Wahyu Setiyo Aji et al., 2023).

#### D. Scikit-Learn

Scikit-learn (sklearn) adalah perpustakaan populer dalam bahasa pemrograman Python yang digunakan untuk machine learning dan data mining.

Scikit-learn menyediakan berbagai macam algoritma machine learning dengan interface yang berorientasi pada tugas, memudahkan perbandingan metode (Reza Fahlepi et al., 2019).

Meskipun algoritma CART hanya dapat mengolah data numerikal, Scikit-learn memiliki fitur untuk menangani data kategorikal (Latifah et al., 2019).

Scikit-learn juga menyediakan alat untuk preprocessing data seperti normalisasi, penskalaan, pengkodean kategori, dan pemilihan fitur. Selain itu, Scikit-learn memiliki fungsi evaluasi yang kuat untuk mengukur kinerja model seperti akurasi, presisi, recall, F1-score, dan kurva ROC (Ramadhan et al., 2023).

#### E. Naive Bayes

Naive Bayes adalah teknik dalam statistika dan machine learning untuk klasifikasi, menggambarkan probabilitas suatu peristiwa berdasarkan pengetahuan sebelumnya tentang faktor terkait. Naive Bayes Classifier dapat memprediksi kemungkinan keanggotaan suatu kelas dengan cepat dan akurat, terutama dalam database besar (Wahyudin et al., 2023).

Algoritma ini menggunakan teorema Bayes dan berasumsi bahwa nilai antar variabel independen pada suatu nilai output (Saputro & Sari, 2020). Naive Bayes mengklasifikasikan kejadian berdasarkan asumsi bahwa atribut kondisional independen diberikan kelas (Byna & Basit, 2020). Algoritma ini, ditemukan oleh Thomas Bayes, memprediksi peluang di masa depan berdasarkan pengalaman sebelumnya (Sabransyah et al., 2017),

Algoritma naïve bayes adalah algoritma pembelajaran mesin untuk masalah klasifikasi yang terutama digunakan untuk klasifikasi teks yang melibatkan kumpulan data pelatihan berdimensi tinggi (Ridwan, 2020).

Berikut formula umum Naive Bayes :

$$P(A|B) = (P(B|A) * P(A)) / P(B)$$

Dimana :

$P(A/B)$  = Probabilitas dari kejadian A terjadi jika B terjadi (probabilitas posterior)

$P(B/A)$  = Probabilitas dari kejadian B terjadi jika A terjadi (probabilitas likelihood)

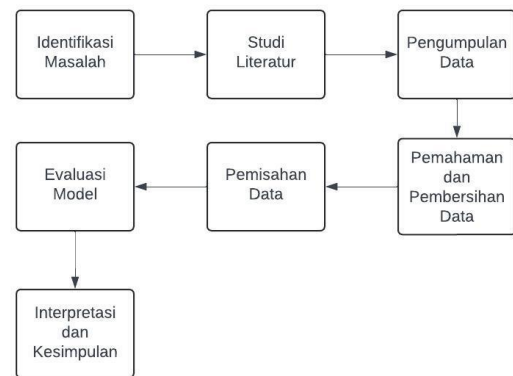
$P(A)$  = Probabilitas dari kejadian A terjadi secara independen (probabilitas prior)

$P(B)$  = Probabilitas awal atau tanpa kondisi dari peristiwa B terjadi (probabilitas marginal atau bukti).

### 3. METODOLOGI PENELITIAN

#### A. Alur Penelitian

Berikut adalah alur penelitian yang dilakukan oleh peneliti



Gambar 1. Alur Penelitian

Sumber: Penulis 2024

#### B. Identifikasi Masalah

Permasalahan pada penelitian ini adalah tingginya angka kematian yang disebabkan penyakit jantung, tantangan utama dalam penanganan penyakit jantung adalah diagnosis yang terlambat, penyakit jantung dapat disebabkan oleh berbagai faktor dan diperlukannya deteksi dini untuk meningkatkan peluang bertahan hidup pasien.

#### C. Studi Literatur

Meneliti studi-studi sebelumnya yang menggunakan berbagai algoritma pembelajaran mesin untuk prediksi penyakit jantung, dan memahami kelebihan serta kelemahan dari masing-masing metode.

#### D. Pengumpulan Data

Data diperoleh dari platform kaggle, sebuah platform yang menyediakan dataset untuk keperluan penelitian dan pengembangan di berbagai bidang, termasuk kesehatan ([www.kaggle.com](http://www.kaggle.com)). Peneliti mengambil dataset yang diupload oleh Manu Siddhartha pada tahun 2020 dengan judul “Heart Disease Dataset (Comprehensive)”. Dataset yang terdiri dari 1190 data observasi dengan 12 variabel.

Tabel 1. Variabel

Atribut	Arti	Keterangan
Age	Usia	-
Sex	Jenis kelamin	0= Male 1= Female
Chest Pain Type	Jenis nyeri dada	1=Typical 2=TypicalAngina 3=NonAngina 4=Asymtomatic
Resting Blood Pressure/Trestbps	Tekanan darah	-
Cholesterol	Kolesterol	-
Fasting Blood Sugar/Fbs > 120	Gula darah	0=False 1=True
Resting Electrocardiographic/Restecg	Hasil elektrodigrafi	0=Normal 1=Abnormal
Max Heart Rate (Thalach)	Detak jantung maksimum	-
Exercise Angina (Exang)	Nyeri dada apabila olahraga	0=No 1=Yes
Oldpeak	Segmen ST yang didapatkan berdasarkan latihan relatif pada istirahat	-
ST Slope	Kemiringan segmen ST dalam latihan maksimum	1=Upsloping 2=Flat 3=Downsloping
Target	Kelas dari fitur	0=Tidak resiko 1=Resiko

Sumber: Penulis 2024

#### E. Pemahaman dan Pembersihan Data

Proses ini mencakup pemeriksaan awal data, selanjutnya dilakukan pembersihan data.

##### 1. Pemeriksaan awal data

Dengan mengidentifikasi jenis bentuk file dataset, identifikasi tipe data.

##### 2. Pembersihan data

Identifikasi Missing Values, jika terdapat nilai yang hilang maka dapat menggunakan teknik imputasi mean (nilai rata-rata), median (nilai tengah), modus (nilai yang sering muncul).

#### F. Pemisahan Data

Dataset dibagi menjadi dua bagian, yaitu data training dan data test, pemisahan ini bertujuan untuk melatih model prediksi dan menguji performanya. Untuk memastikan model dibangun memiliki generalisasi yang baik dan tidak overfitting digunakan metode k-fold cross-validation.

#### G. Sample yang diuji

Dataset sample yang diuji menggunakan algoritma naive bayes terdiri dari 70 data yang mencakup variabel kategori seperti sex, chest pain type, resting ecg, fasting blood sugar, exercise angina.

##### 1. Menentukan Probabilitas Prior

Probabilitas prior adalah probabilitas awal dari setiap kelas target sebelum memperhitungkan data observasi.

Mencari nilai Target (Probabilitas)

Tidak Resiko =  $47/70 = 0,6714$

Resiko =  $23/70 = 0,3285$

##### 2. Menghitung Probabilitas Kondisional

Probabilitas kondisional adalah probabilitas dari suatu fitur yang muncul dalam kelas target tertentu.

##### I. Menghitung Probabilitas untuk Variabel “Sex”

$P(\text{Male } 1 | \text{Tidak Resiko}) = 26/47 = 0,55$

$P(\text{Male } 1 | \text{Resiko}) = 19/23 = 0,82$

$P(\text{Female } 0 | \text{Tidak Resiko}) = 21/47 = 0,44$

$P(\text{Female } 0 | \text{Resiko}) = 4/23 = 0,17$

##### II. Menghitung Probabilitas untuk Variabel “Chest Pain Type (CPT)”

$P(\text{Typical } | \text{Tidak Resiko}) = 1/47 = 0,021$

$P(\text{Typical Angina } | \text{Tidak Resiko}) = 27/47 = 0,57$

$P(\text{Non Angina } | \text{Tidak Resiko}) = 11/47 = 0,234$

$P(\text{Asymtomatic } | \text{Tidak Resiko}) = 8/47 = 0,17$

$P(\text{Typical } | \text{Resiko}) = 0/23 = 0$

$P(\text{Typical Angina } | \text{Resiko}) = 3/23 = 0,13$

$P(\text{Non Angina } | \text{Resiko}) = 4/23 = 0,17$

$P(\text{Asymtomatic } | \text{Resiko}) = 16/23 = 0,69$

III. Menghitung Probabilitas untuk Variabel “Fasting Blood > 120”

$$P(\text{True} | \text{Tidak Resiko}) = 2/47 = 0,042$$

$$P(\text{True} | \text{Resiko}) = 1/23 = 0,043$$

$$P(\text{False} | \text{Tidak Resiko}) = 45/47 = 0,95$$

$$P(\text{False} | \text{Resiko}) = 22/23 = 0,95$$

IV. Menghitung Probabilitas untuk Variabel “Resting Ecg”

$$P(\text{Normal} | \text{Tidak Resiko}) = 40/47 = 0,85$$

$$P(\text{Abnormal} | \text{Tidak Resiko}) = 7/47 = 0,14$$

$$P(\text{Normal} | \text{Resiko}) = 17/23 = 0,73$$

$$P(\text{Abnormal} | \text{Resiko}) = 6/23 = 0,26$$

V. Menghitung Probabilitas untuk Variabel “Exercise Angina”

$$P(\text{No} | \text{Tidak Resiko}) = 43/47 = 0,91$$

$$P(\text{Yes} | \text{Tidak Resiko}) = 4/47 = 0,08$$

$$P(\text{No} | \text{Resiko}) = 9/23 = 0,39$$

$$P(\text{Yes} | \text{Resiko}) = 14/23 = 0,60$$

3. Menghitung Likelihood

Likelihood adalah probabilitas data observasi tertentu muncul dalam kelas target tertentu

$$P(X|C) = P(x^1|C) \dots P(x_n|C)$$

$P(X|C)$ : probabilitas data observasi X yang termasuk dalam kelas C.

$P(x^1|C) \dots P(x^2|C) \dots P(x_n|C)$  : probabilitas kondisional dari setiap fitur  $x^i$  diberikan kelas C.

Sex : Female (1)  
 Chest Pain Type : Typical Angina (2)  
 Fasting Blood Sugar : False (0)  
 Resting Ecg : Normal (0)  
 Exercise Angina : (0)

I. Menghitung Likelihood untuk target “Tidak Resiko”

$$P(X|\text{Tidak Resiko}) = P(\text{Sex} = \text{Female} | \text{Tidak Resiko}) \times P(\text{CPT} = \text{Typical Angina} | \text{Tidak Resiko}) \times P(\text{FBS} = \text{False} | \text{Tidak Resiko}) \times P(\text{Resting ECG} = \text{Normal} | \text{Tidak Resiko}) \times P(\text{Exercise Angina} = \text{No} | \text{Tidak Resiko})$$

$$P(X|\text{Tidak Resiko}) = 0,44 \times 0,57 \times 0,95 \times 0,85 \times 0,91 = 0,184$$

II. Menghitung Likelihood untuk target “Resiko”

$$P(X|\text{Resiko}) = P(\text{Sex} = \text{Female} | \text{Resiko}) \times P(\text{CPT} = \text{Typical Angina} | \text{Resiko}) \times P(\text{FBS} = \text{False} | \text{Resiko}) \times P(\text{Resting ECG} = \text{Normal} | \text{Resiko}) \times P(\text{Exercise Angina} = \text{No} | \text{Resiko})$$

$$P(X|\text{Resiko}) = 0,17 \times 0,13 \times 0,95 \times 0,73 \times 0,39 = 0,005$$

4. Menghitung Posterior Probability :

Posterior probability adalah probabilitas target kelas tertentu diberikan data observasi :

$$P(C|X) = P(X|C_i) * P(C_i)$$

I. Menghitung Posterior untuk “Tidak Resiko”

$$P(\text{Tidak Resiko} | X) = P(X | \text{Tidak Resiko}) \times P(\text{Tidak Resiko}) = 0,184 \times 0,6714 = 0,123$$

II. Menghitung Posterior untuk “Resiko”

$$P(\text{Resiko} | X) = P(X | \text{Resiko}) \times P(\text{Resiko}) = 0,005 \times 0,3285 = 0,001$$

5. Menghitung Normalisasi

Normalisasi memastikan bahwa total probabilitas dari semua kelas target adalah 1.

$$P(C_i|X) = (P(X|C_i) * P(C_i)) / (P(X))$$

$$P(\text{Tidak Resiko} | X) = 0,123 / (0,123 + 0,001) = 0,123 / 0,124 = 0,99$$

$$P(\text{Resiko} | X) = 0,001 / (0,123 + 0,001) = 0,001 / 0,124 = 0,008$$

Dari hasil perhitungan diatas diperoleh :

Probabilitas pengguna berada dalam kategori “Tidak Resiko” diberikan fitur X adalah 0,99 atau 99%.

Probabilitas pengguna berada dalam kategori “Resiko” diberikan fitur X adalah 0,008 jika dibulatkan 0,01 atau 1%

Berdasarkan data fitur X, model naive bayes sangat yakin bahwa pengguna berada dalam kategori “Tidak Resiko”.

H. Skenario Uji

1. Mengumpulkan dan membersihkan nilai-nilai yang hilang.
2. Data dibagi menjadi 2 data latih dan data uji dengan metode K-fold cross-validation data latih digunakan untuk melatih model

naive bayes dan data uji untuk menguji kinerja model yang telah dilatih.

3. Fitur-fitur numerik dinormalisasi menggunakan StandardScaler agar memiliki distribusi yang seragam.
4. Model Naive Bayes dilatih menggunakan data latih yang telah dinormalisasi.
5. Prediksi data uji melibatkan perhitungan probabilitas posterior untuk setiap kelas dan pemilihan kelas dengan probabilitas tertinggi.
6. Hasil prediksi dibandingkan dengan label sebenarnya untuk menghitung metrik evaluasi seperti akurasi, presisi, recall, dan F1-score.

#### 4. HASIL DAN PEMBAHASAN

##### A. Pengumpulan Data

Data yang digunakan dalam penelitian ini diperoleh dari situs web Kaggle, sebuah platform yang menyediakan berbagai dataset untuk keperluan analisis data dan pembelajaran mesin. Dataset yang dipilih adalah dataset penyakit jantung yang berjumlah 1190 data terdiri dari 12 variabel, yaitu: age, sex, tipe chest pain type, resting blood pressure, kolesterol, fasting blood sugar >120, resting electrocardiographic, max heart rate, exercise angina, oldpeak, st slope dan target. Data ini digunakan sebagai dasar untuk membangun model prediksi penyakit gagal jantung menggunakan algoritma Naive Bayes. Dataset tersebut terdiri dari sejumlah entri yang mencerminkan berbagai kondisi pasien, dan setiap variabel memberikan informasi penting yang berkontribusi dalam proses prediksi.

##### B. Tahap Preprocessing

Preprocessing merupakan proses penting untuk membersihkan dan mempersiapkan data agar siap digunakan dalam model prediksi. Tahapan preprocessing terbagi menjadi beberapa tahapan yaitu pemeriksaan data yang hilang, pengkodean variabel kategorikal (categorical encoding), memisahkan fitur dan target.

1. Pemeriksaan Data yang Hilang.  
Pemeriksaan data yang hilang Merupakan langkah kritis dalam preprocessing data karena nilai yang hilang dapat mempengaruhi hasil

analisis dan performa model prediksi, berikut mengidentifikasi keberadaan nilai yang hilang.

```
missing_values = dt.isnull().sum()
print(missing_values)

✓ 00s
age          0
sex          0
chest pain type  0
resting bp s  0
cholesterol  0
fasting blood sugar  0
resting ecg  0
max heart rate  0
exercise angina  0
oldpeak      0
ST slope     0
target       0
dtype: int64
```

Gambar 2. Missing Value

Sumber: Penulis 2024

2. Pengkodean variabel kategorikal (categorical encoding).

Proses mengubah variabel kategorikal menjadi format numerik yang dapat diproses oleh algoritma pada penelitian ini terdapat 2 proses pengkodean variabel kategorikal.

3. Inisialisasi LabelEncoder

LabelEncoder digunakan untuk mengubah data kategorikal menjadi bentuk numerik. Variabel sex, chest pain type, fasting blood sugar, resting ecg, exercise angina merupakan variabel kategori yang perlu diproses menjadi format numerik.

```
# Inisialisasi LabelEncoder
le_sex = LabelEncoder()
le_chest_pain_type = LabelEncoder()
le_fasting_blood_sugar = LabelEncoder()
le_resting_ecg = LabelEncoder()
le_exercise_angina = LabelEncoder()
```

Gambar 3. Inisialisasi LabelEncoder

Sumber: Penulis 2024

4. Inisialisasi OneHotEncoder

Teknik untuk mengonversi variabel kategori menjadi format yang dapat digunakan oleh algoritma machine learning. Ini dilakukan dengan mengonversi setiap kategori unik dalam kolom menjadi kolom biner (0 atau 1) terpisah. Berfungsi untuk Menghindari Urutan Kategori yang Salah pada penelitian ini dilakukan pada variabel chest pain type karena terdapat 4 kategori.

### C. Pembagian dan pengujian dataset

Pembagian dan pengujian dataset menggunakan K-fold cross-validation, dataset dibagi menjadi k subset yang sama besar. Proses ini untuk mendapatkan estimasi kinerja model yang lebih stabil dengan jumlah data 1190 akan dibagi menjadi 5 bagian yang seimbang, dengan data 952 data latih (training) dan 238 data uji (testing). Hasil pembagian dan pengujian dataset menggunakan K-fold cross-validation.

Dari tabel di atas, dapat dilihat bahwa akurasi, presisi, recall, f1-score model bervariasi pada setiap uji coba, dengan nilai tertinggi sebesar 85,29% dan nilai terendah sebesar 78,14%. Nilai rata-rata akurasi, presisi, recall, f1-score dari kelima uji coba ini adalah 81,42%, 81,38%, 81,28%, 81,30%.

Selain akurasi, hasil pengujian juga mencakup metrik evaluasi lain seperti presisi, recall, dan F1 score, yang dirata-rata dari lima uji coba Hasil ini menunjukkan bahwa model memiliki performa yang konsisten dan dapat diandalkan dalam memprediksi risiko penyakit jantung. Metrik evaluasi seperti presisi, recall, dan F1 score yang semuanya mendekati nilai rata-rata akurasi mengindikasikan bahwa model tidak hanya mampu membuat prediksi yang benar, tetapi juga mempertahankan keseimbangan antara mengidentifikasi positif yang benar (true positives) dan menghindari positif yang salah (false positives).

Selain akurasi, hasil pengujian juga mencakup metrik evaluasi lain seperti presisi, recall, dan F1 score, yang dirata-rata dari lima uji coba Hasil ini menunjukkan bahwa model memiliki performa yang konsisten dan dapat diandalkan dalam memprediksi risiko penyakit jantung. Metrik evaluasi seperti presisi, recall, dan F1 score yang semuanya mendekati nilai rata-rata akurasi mengindikasikan bahwa model tidak hanya mampu membuat prediksi yang benar, tetapi juga mempertahankan keseimbangan antara mengidentifikasi positif yang benar (true positives) dan menghindari positif yang salah (false positives).

### D. Tampilan Website

Tampilan website ini dirancang untuk memberikan pengguna kemudahan dalam menginput data dan mendapatkan hasil prediksi

dengan cepat dan akurat. Berikut adalah beberapa elemen yang biasanya ditemui dalam tampilan website prediksi penyakit gagal jantung.

#### 1. Halaman Utama

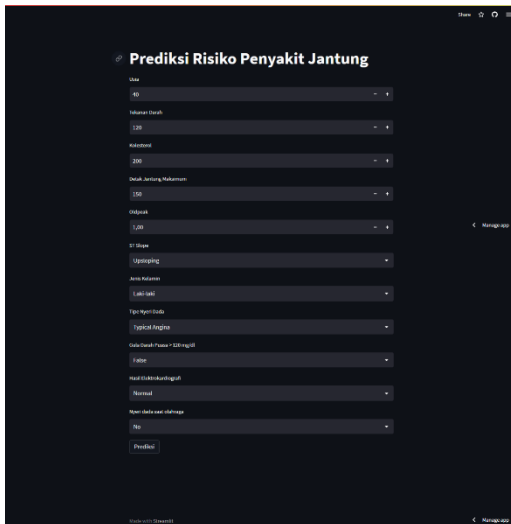
Halaman utama website ini didesain dengan fokus utama pada form input data, yang memungkinkan pengguna untuk langsung memasukkan informasi yang diperlukan dan mendapatkan hasil prediksi secara cepat.

#### 2. Form Input Data

Form Input Data: Form input data menjadi elemen utama dari halaman ini. Form ini berisi berbagai kolom input yang harus diisi oleh pengguna dengan informasi pribadi yang relevan untuk prediksi risiko penyakit gagal jantung. Beberapa kolom input yang biasanya ada termasuk:

- I. Usia (age): Kolom untuk memasukkan usia pengguna.
- II. Tekanan Darah (resting bp s): Input untuk memasukkan nilai tekanan darah sistolik pengguna saat istirahat.
- III. Kolesterol (cholesterol): Kolom untuk memasukkan kadar kolesterol dalam darah.
- IV. Detak Jantung Maksimum (max heart rate): Kolom untuk memasukkan nilai detak jantung maksimum yang dicapai pengguna.
- V. Oldpeak : Segmen ST yang didapatkan berdasarkan latihan relatif pada istirahat.
- VI. ST Slope : Kemiringan segmen ST dalam latihan maksimum.
- VII. Jenis Kelamin (sex): Pilihan untuk memilih jenis kelamin, disediakan dalam bentuk dropdown.
- VIII. Tipe nyeri dada (chest pain type): Jenis nyeri dada.
- IX. Gula Darah (fasting blood sugar) > 120: Pilihan untuk menentukan apakah gula darah puasa lebih dari 120 mg/dl.
- X. Hasil Elektrokardiografi (resting ecg): Input untuk memasukkan hasil pemeriksaan elektrokardiografi (ECG).

- XI. Nyeri Dada Saat Olahraga (exercise angina):  
Pilihan untuk menentukan apakah pengguna mengalami nyeri dada saat olahraga.

The image shows a mobile application interface for 'Prediksi Risiko Penyakit Jantung'. It features a dark theme with white text. The top section is the title 'Prediksi Risiko Penyakit Jantung'. Below it is a form with various input fields, each with a dropdown arrow. The fields include: 'Nama', 'No', 'Tanggal Masuk', 'Kategori', 'Dokter Jantung Mahkota', 'Jenis', 'Jenis', 'SI Sifat', 'Upting/PHC', 'Area Risiko', 'Laki/Laki', 'Tipe Nyeri Dada', 'Tipe Nyeri Jantung', 'Gula Darah Puasa <math>F\_{1,2}</math> (mg/dl)', 'Jenis', 'Rasio Kolesterol Total', 'Rasio', 'Nyeri Dada saat Olahraga', 'Rp', and 'Prediksi'. At the bottom right of the form is a 'Prediksi' button.

Gambar 5. Halaman Utama dan Form Input Data  
Sumber: Penulis 2024

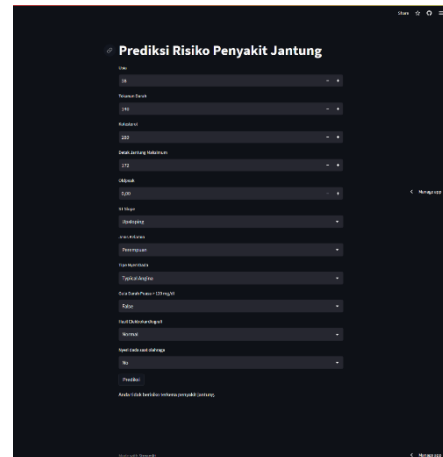
### 3. Tombol Prediksi

Setelah pengguna memasukkan data yang diperlukan ke dalam form, terdapat tombol "Prediksi" yang dapat ditekan untuk memulai proses prediksi. Tombol ini mengirimkan data input ke model prediksi yang sudah dilatih untuk menghasilkan hasil prediksi.

### 4. Hasil Prediksi

Setelah tombol "Prediksi" ditekan, hasil prediksi akan muncul di bawah form input atau pada bagian yang telah disediakan. Hasil ini biasanya dalam bentuk teks yang menyatakan risiko pengguna terkena penyakit gagal jantung, seperti "Anda tidak berisiko terkena penyakit jantung" atau "Anda berisiko terkena penyakit jantung".

Dengan tampilan yang fokus pada form input data, halaman utama website prediksi penyakit gagal jantung ini memastikan bahwa pengguna dapat dengan mudah dan cepat memasukkan informasi yang diperlukan serta mendapatkan hasil prediksi yang diinginkan. Desain yang sederhana dan langsung pada tujuan ini diharapkan dapat meningkatkan kenyamanan dan kepuasan pengguna dalam menggunakan layanan prediksi ini.

The image shows the same application interface as Gambar 5, but with the prediction result displayed. The 'Prediksi' field at the bottom of the form now contains the text: 'Anda tidak berisiko terkena penyakit jantung.' The rest of the form and the application header are the same as in Gambar 5.

Gambar 6. Hasil Prediksi tidak berisiko terkena penyakit jantung

Sumber: Penulis 2024

## 5. KESIMPULAN DAN SARAN

### A. Kesimpulan

Dengan menggunakan algoritma Naive Bayes pada prediksi penyakit gagal jantung dengan data yang tersedia, hasil evaluasi dengan metode Cross-Validation menunjukkan tingkat akurasi yang baik dengan hasil akurasi tiap fold (82,77 %, 85,29 %, 80,67 %, 80,25 %, 78,15 %) dan rata-rata akurasi sebesar 81,42%. Hasil dari evaluasi yang lain presisi, recall, F1-score memperoleh masing-masing skor 81,38%, 81,28%, dan 81,30%. Hasil ini menunjukkan bahwa metode Naive Bayes dapat memberikan prediksi yang relatif konsisten.

### B. Saran

1. Pada penelitian ini hanya berupa dataset yang tersedia di kaggle, diharapkan pada penelitian selanjutnya dapat memperoleh data yang ada di fasilitas kesehatan disekitar dan ilmu pakar yang lebih mendalam mengenai penyakit gagal jantung sebagai pembanding pada algoritma yang digunakan.
2. Untuk penelitian selanjutnya dapat mengembangkan website prediksi penyakit gagal jantung yang lebih efektif dan efisien agar dapat diakses user/pengguna



## 6. REFERENSI

### A. Artikel

- Arifin, T., & Ariesta, D. (2019). Prediksi Penyakit Ginjal Kronis Menggunakan Algoritma Naive Bayes Classifier Berbasis Particle Swarm Optimization. *Jurnal Tekno Insentif*, 13(1), 26–30. <https://doi.org/10.36787/jti.v13i1.97>
- Byna, A., & Basit, M. (2020). Penerapan Metode Adaboost Untuk Mengoptimasi Prediksi Penyakit Stroke Dengan Algoritma Naive Bayes. *Jurnal Sisfokom (Sistem Informasi Dan Komputer)*, 9(3), 407–411. <https://doi.org/10.32736/sisfokom.v9i3.1023>
- Dwi Septiani, W. (2017). Komparasi Metode Klasifikasi Data Mining Algoritma C4.5 Dan Naive Bayes untuk Prediksi Penyakit Hepatitis. *Jurnal Pilar Nusa Mandiri*, 13(1), 76–84. <http://archive.ics.uci.edu/ml/>
- Hasibuan, E., Informasi, S., Ilmu, F., Informasi, T., Gunadarma, U., Margonda, J., No, R., Cina, P., & Jawa, D. (2022). Implementasi Machine Learning untuk Prediksi Harga Mobil Bekas dengan Algoritma Regresi Linear berbasis Web. *Jurnal Ilmiah Komputasi*, 21(4), 595–602. <https://doi.org/10.32409/jikstik.21.4.3327>
- Latifah, R., Wulandari, E. S., & Kreshna, P. E. (2019). Model Decision Tree untuk Prediksi Jadwal Kerja menggunakan Scikit-Learn. *Seminar Nasional Sains Dan Teknologi*, 16, 1–6.
- Ramadhan, B., Firdaus, D., & Adiningrum, N. T. R. (2023). Analisis Data Pegawai Untuk Memprediksi Gaji Berdasarkan Faktor-Faktor Spesifik Dengan Pendekatan Machine Learning. *Naratif: Jurnal Nasional Riset, Aplikasi Dan Teknik Informatika*, 5(2), 131–139. <https://doi.org/10.53580/naratif.v5i2.205>
- Reza Fahlepi, M., Widjaja, A., & Surya Sumantri No, J. (2019). Penerapan Metode Multiple Linear Regression Untuk Prediksi Harga Sewa Kamar Kost. *Jurnal STRATEGI - Jurnal Maranatha*, 1(2), 615–629. <https://www.strategi.it.maranatha.edu/index.php/strategi/article/view/140>
- Ridwan, A. (2020). Penerapan Algoritma Naive Bayes Untuk Klasifikasi Penyakit Diabetes Mellitus. *Jurnal SISKOM-KB (Sistem Komputer Dan Kecerdasan Buatan)*, 4(1), 15–21. <https://doi.org/10.47970/siskom-kb.v4i1.169>
- Sabransyah, M., Nasution, Y. N., & Tisna, D. (2017). Aplikasi Metode Naive Bayes dalam Prediksi Risiko Penyakit Jantung. *Jurnal EKSPONENSIAL*, 8(2), 111–118. <http://jurnal.fmipa.unmul.ac.id/index.php/exponensial/article/view/31>
- Saifudin, I., & Suharso, W. (2020). Pembelajaran e-learning, pembelajaran ideal masa kini dan masa depan pada mahasiswa berkebutuhan khusus. *JP (Jurnal Pendidikan): Teori dan Praktik*, 5(2), 30-35.
- Saifudin, I. (2017). Pengenalan dan Pelatihan Software Maple guna Meningkatkan Pemahaman Geometri untuk Siswa SMK. *Jurnal Pengabdian Masyarakat Ipteks*, 3(1).
- Saifudin, I., & Umilasari, R. (2021). Automatic Aircraft Navigation Using Star Metric Dimension Theory in Fire Protected Forest Areas. *JTAM (Jurnal Teori dan Aplikasi Matematika)*, 5(2), 294-304.
- Saifudin, I., & Mubaroq, S. (2021). Pemanfaatan Aplikasi Camtasia dalam Meningkatkan Kebutuhan Multimedia Pada Video Pembelajaran Daring Bagi Guru di SMP Muhammadiyah Bondowoso. *Suluah Bendang: Jurnal Ilmiah Pengabdian Kepada Masyarakat*, 21(2), 140-147.
- Saifudin, I., & Nurhalimah, N. (2019). Screen Printing the Glassware as Souvenir to Increase Selling Value and Income on Handicraft Group in Karangrejo, Jember, East Java. *Kontribusi: Research Dissemination for Community Development*, 2(1), 24-30.
- Saputro, I. W., & Sari, B. W. (2020). Uji Performa Algoritma Naive Bayes untuk Prediksi Masa Studi Mahasiswa. *Creative Information Technology Journal*, 6(1), 1.

<https://doi.org/10.24076/citec.2019v6i1.178>

Wahyu Setiyo Aji, P., Dijaya, R., & Sains dan Teknologi, F. (2023). KESATRIA: Jurnal Penerapan Sistem Informasi (Komputer & Manajemen) Prediksi Penyakit Stroke Menggunakan Metode Random Forest. *Jurnal Penerapan Sistem Informasi*, 4(4), 916–924.

Wahyudin, W. C., Hana, F. M., Prihandono, A., Kudus, U. M., No, J. G., Email, I., Semarang, P. K., Classifier, N. B., Naive, A., Classifier, B., Classifier, N. B., & Classifier, N. B. (2023). P Rediksi S Tunting P Ada B Alita D I R Umah S Akit K Ota. 2019, 32–36.