Jurnal Smart Teknologi
Vol. 6, No. 2, Januari 2025, Halaman 264 – 275
ISSN: 2774-1702, http://jurnal.unmuhjember.ac.id/index.php/JST

Klasifikasi Penyakit Gagal Jantung Menggunakan Algoritma Naive Bayes

The Prediction Of Brain Failure Using The Naive Bayes Algorithm

Moch. Rafli Febrian[1)], Ilham Saifudin[2)], Wiwik Suharso[3)]
[1]Mahasiswa Fakultas Teknik, Universitas Muhammadiyah Jember
email: febrianrafli26@gmail.com
[2]Dosen Fakultas Teknik, Universitas Muhammadiyah Jember
email: ilhamsaifudin@unmuhjember.ac.id
[3]Dosen Fakultas Teknik Universitas Muhammadiyah Jember
email: wiwiksuharso@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 klasifikasi 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 kelebihannya 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 memklasifikasi 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, Klasifikasi.*

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 the classification of 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 algoritm has an adequate degree of accuracy in classifying 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, Clasification.*

Jurnal Smart Teknologi
Vol. 6, No. 2, Januari 2025, Halaman 264 – 275
ISSN: 2774-1702, http://jurnal.unmuhjember.ac.id/index.php/JST

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 (Alizadehsani dkk., 2019).

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 klasifikasi 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 model klasifikasi gagal jantung yang menggunakan algoritma *Naive Bayes*. Tujuan utama dari penelitian ini adalah mengidentifikasi faktor resiko penyakit gagal jantung, mengembangkan model klasifikasi berbasis algoritma *Naive Bayes*, penelitian ini diharapkan dapat memberikan kontribusi untuk mengurangi angka kematian akibat penyakit jantung.

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. Klasifikasi

Klasifikasi dalam machine learning adalah proses yang penting dan luas digunakan untuk memklasifikasi kategori atau label dari data yang diberikan. Dengan menggunakan algoritma pembelajaran mesin yang tepat, klasifikasi dapat membantu dalam berbagai aplikasi praktis.

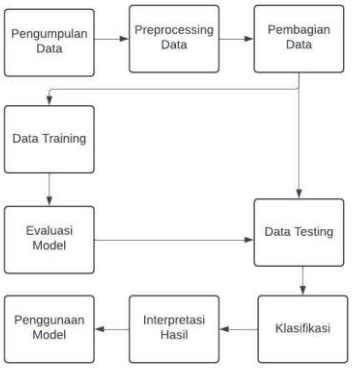
Klasifikasi merupakan proses pengolahan dataset dengan *naïve bayes*, dataset yang digunakan akan dibagi menjadi dua masing-masing untuk data train dan data test Klasifikasi merupakan proses pengolahan dataset dengan *naïve bayes*, dataset yang digunakan akan dibagi menjadi dua masing- masing untuk data train dan data test (Hayami dkk., 2022).

Klasifikasi merupakan suatu pekerjaan menilai objek data untuk memasukkannya ke dalam kelas tertentu dari sejumlah kelas yang tersedia (Putro dkk., 2020).

klasifikasi adalah kumpulan model yang dapat melakukan ilustrasi atau menggambarkan dan membedakan kelas data atau konsep, dengan tujuan mampu menggunakan model untuk melakukan klasifikasi kelas dari objek

Jurnal Smart Teknologi
Vol. 6, No. 2, Januari 2025, Halaman 264 – 275
ISSN: 2774-1702, http://jurnal.unmuhjember.ac.id/index.php/JST

yang label kelasnya tidak diketahui.(Mustofa & Mahfudh, 2019).



Ganbar 1. Klasifikasi

Sumber : Penulis 2024

C. Machine Learning

Machine learning adalah cabang dari kecerdasan buatan (*Artificial Intelligence, AI*) yang fokus pada pengembangan algoritma dan teknik yang memungkinkan komputer untuk belajar dari dan membuat klasifikasi atau keputusan berdasarkan data.

D. Naive Bayes

Naive Bayes adalah teknik peramalan probabilistik sederhana sesuai dengan pelaksanaan teorema Bayes (hukum bayes) menggunakan perkiraan independensi yang kuat.(Pebdika dkk., 2023).

Naïve Bayes merupakan satu dari beberapa algoritma klasifikasi yang paling terkenal. Efisiensinya berasal dari asumsi independensi atribut, meskipun ini mungkin dilanggar di banyak kumpulan data dunia nyata (Setiawan & Triayudi, 2022).

Model *Naive Bayes* menyajikan cara untuk menyatukan peluang dahulu berdasarkan syarat atas kemungkinan menjadi suatu formula yang bisa dipakai untuk menghitung suatu peluang dari setiap kemungkinan yang akan terjadi (Veronica Agustin & Voutama, 2023).

$$P(A \mid B)=(P(B \mid A)*P(A))/P(B)$$

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)

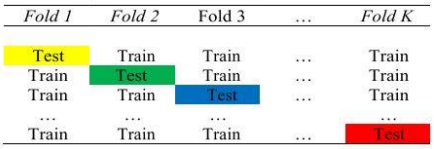
P(B) = Probabilitas awal atau tanpa kondisi dari peristiwa B terjadi (probabilitas marginal atau bukti)

E. K-Fold Cross Validation

K-Fold Validation dan Cross Validation adalah teknik evaluasi model yang digunakan dalam pembelajaran mesin untuk menilai kinerja model dan memastikan bahwa model tersebut memiliki generalisasi yang baik terhadap data baru.

K-fold Cross Validation adalah metode statistik yang digunakan untuk membandingkan dan mengevaluasi kinerja model *Machine Learning* (Tjengharwidjaja dkk., 2024).

K-fold cross validation merupakan salah satu dari teknik yang difungsikan untuk memilah data menjadi train data serta test data. Teknik ini banyak diterapkan peneliti karena didapati mengurangi bias yang didapatkan didalam pengambilan sebuah sampel (Ridwansyah, 2022).



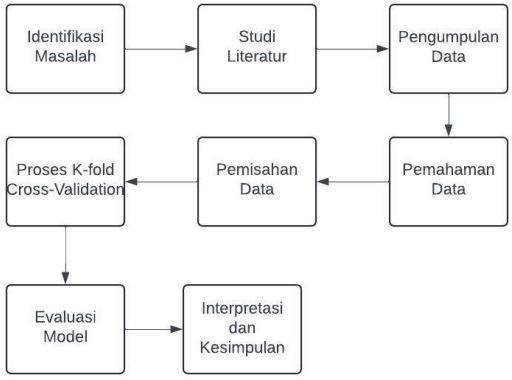
Gambar 2. Ilustrasi K-Fold

Sumber : Ridwansyah, 2022

3. METODOLOGI PENELITIAN

A. Alur Penelitian

Berikut adalah alur penelitian yang dilakukan oleh peneliti



Gambar 3. Alur Penelitian

Sumber: Penulis 2024

Jurnal Smart Teknologi
Vol. 6, No. 2, Januari 2025, Halaman 264 – 275
ISSN: 2774-1702, http://jurnal.unmuhjember.ac.id/index.php/JST

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 klasifikasi 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). Peneliti mengambil dataset yang diupload oleh Manu Siddhartha pada tahun 2020 dengan judul *"Heart Disease Dataset (Comperhensive)*. 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 Data

Proses pemeriksaan awal data merupakan langkah krusial dalam analisis data, yang bertujuan untuk memahami karakteristik dan kualitas dataset sebelum melakukan langkah-langkah analisis yang lebih mendalam.

1. Pemeriksaan awal data

Pemeriksaan awal data adalah langkah yang penting dalam setiap proyek analisis data. Tujuannya adalah untuk memahami karakteristik dan kualitas data yang akan dianalisis. Dalam penelitian ini, peneliti akan memeriksa dataset yang berkaitan dengan klasifikasi penyakit gagal jantung, yang terdiri dari berbagai variabel yang tersedia.

F. Pemisahan Data

Dataset dibagi menjadi dua bagian, yaitu data training dan data test. Pemisahan ini bertujuan untuk melatih model klasifikasi dan menguji performanya. Data training digunakan untuk membangun model, sementara data test digunakan untuk mengevaluasi kinerja model pada data yang belum pernah dilihat sebelumnya dengan menggunakan *K-fold Validation.*

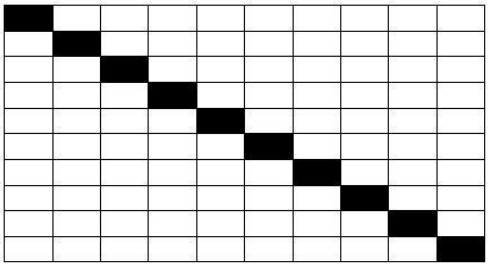
G. Proses K-Fold dan Cross Validation

K-Fold Cross-Validation adalah teknik validasi model yang bertujuan untuk mengukur kemampuan generalisasi dari model yang dibangun. Proses *k-fold cross-validation* dengan k=10 dibagi menjadi langkah-langkah berikut :

1. Dataset dibagi menjadi 10 fold yang sama besar
2. Model dilatih dan diuji sebanyak 10 kali, dengan setiap kali menggunakan 9 fold sebagai data training dan 1 fold sebagai data test.

Jurnal Smart Teknologi
Vol. 6, No. 2, Januari 2025, Halaman 264 – 275
ISSN: 2774-1702, http://jurnal.unmuhjember.ac.id/index.php/JST

3. Setiap fold secara bergantian digunakan sebagai data test, sementara fold lainnya digunakan sebagai data training.
4. Setiap data training dan testing berjumlah 1190 dengan pembagiannya berjumlah 119 data testing yang terdiri dari 56 data pasien tidak resiko, 63 data pasien resiko dan 1071 data training yang terdiri dari 505 data pasien tidak resiko, 566 data pasien resiko.



Gambar 4. Proses K-Fold dan Cross-Validation
Sumber : Penulis 2024

Pada gambar 4 kotak yang berwarna hitam merupakan data testing dan kotak yang berwarna putih merupakan data training, Setiap baris akan berada dalam data pelatihan sebanyak 9 kali dan dalam data pengujian sebanyak 1 kali selama 10 iterasi. Dengan menggunakan *k-fold cross-validation*, peneliti dapat lebih yakin bahwa model yang dibangun memiliki kemampuan generalisasi yang baik terhadap data training. Teknik ini juga membantu dalam memanfaatkan seluruh dataset secara lebih efektif, terutama ketika jumlah data terbatas.

H. Evaluasi Model

Model diuji pada data uji untuk menilai kemampuannya dalam memklasifikasi resiko gagal jantung dengan menggunakan metrik evaluasi seperti akurasi, presisi, recall, dan F1-score berikut penjelasannya :

1. Akurasi (*Accuracy*) :

Akurasi mengukur seberapa sering model membuat klasifikasi benar :

$$\text{Akurasi}=\frac{(TP+TN)}{(TP+TN+FP+FN)}$$

Dimana :
TP:*True Positive*
TN:*True Negative*
FP: *False Positive*
FN: *False Negative*

2. Presisi (*Precision*)

Presisi mengukur proporsi klasifikasi positif yang benar :

$$\text{Presisi}=\frac{TP}{TP+FP}$$

3. Recall (Sensitivitas)

Recall mengukur proporsi kasus positif yang benar-benar terdeteksi oleh model :

$$\text{Recall}=\frac{TP}{TP+FN}$$

4. F1-Score

F1-Score adalah rata-rata harmonis dari presisi dan recall, memberikan gambaran keseimbangan antara keduanya, ditunjukkan pada persamaan :

$$\text{F1-Score}=2\ .\ \frac{(Presisi\ .\ Recall)}{(Presisi+Recall)}$$

I. Interpretasi dan Kesimpulan

Peneliti menganalisis hasil dari penilitiannya yang mencakup seberapa baik algoritma *Naive Bayes* digunakan berdasarkan metrik evaluasi. Menyimpulkan hasil utama dari penelitian dalam bentuk yang ringkas dan jelas.

J. Skenario Uji

1. Pengumpulan data yang berisi fitur-fitur.
2. Pra-pemrosesan data yang meliputi transformasi fitur numerik dan fitur kategori
3. Pemisahan Data, data dibagi menjadi dua bagian yaitu data training dan data test, data training digunakan untuk melatih model, sedangkan data test digunakan untuk menguji performa model pada data yang belum pernah dilihat sebelumnya dengan menggunakan *K-Fold Cross Validation*
4. Pelatihan dan Evalasi Model, model *Naive Bayes* dilatih menggunakan data training, setelah pelatihan, model di evaluasi menggunakan data test dan juga menggunakan *K-Fold Cross-Validation* untuk mendapatkan kinerja yang baik.
5. Analisis Hasil, Metrik kinerja seperti akurasi, presisi, recall, dan F1 score akan dicari hasil yang terbaik untuk dijadikan model yang baik.

K. Sample Data 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

Jurnal Smart Teknologi
Vol. 6, No. 2, Januari 2025, Halaman 264 – 275
ISSN: 2774-1702, http://jurnal.unmuhjember.ac.id/index.php/JST

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

$$P(C) = \frac{(Jumlah\ observasi\ dalam\ kelas\ C)}{(Total\ jumlah\ observasi}$$

Perhitungan Probabilitas "Tidak Resiko": Menggunakan jumlah observasi dalam kelas "Tidak Resiko" yaitu 47 :

$$P\ (\text{Tidak Resiko}) = \frac{47}{70} = 0{,}6714$$

Perhitungan Probabilitas "Resiko": Menggunakan jumlah observasi dalam kelas "Resiko" yaitu 23 :

$$P\ (\text{Resiko}) = \frac{23}{70} = 0{,}3285$$

2. Menghitung Probabilitas Kondisional

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

$$P(Xi|C) = \frac{(Jumlah\ kemunculan\ Xi\ dalam\ kelas\ C)}{Jumlah\ total\ kelas\ C}$$

Menghitung Probabilitas untuk Variabel "Sex"
P(Male 1| Tidak Resiko) = $\frac{26}{47}$ = 0,55
P(Male 1| Resiko) = $\frac{19}{23}$ = 0,82
P(Female 0| Tidak Resiko) = $\frac{21}{47}$ = 0,44
P(Female 0| Resiko) = $\frac{4}{23}$ = 0,17

Gambar 5. Menghitung Probabilitas "Sex"
Sumber : Penulis 2024

Pada gambar 5 peneliti menghitung probabilitas variabel "*Sex*", probabilitas seorang pria tergolong dalam kelas "Tidak Resiko" adalah 55% P(Male 1|Tidak Resiko) = 0,55, sedangkan probabilitas seorang pria tergolong dalam kelas "Resiko" adalah 82% P(Male 1|Resiko) = 0,82. Untuk wanita, probabilitas dalam kelas "Tidak Resiko" adalah 44% P(Female 0|Tidak Resiko) = 0,44 dan dalam kelas "Resiko" adalah 17% P(Female 0|Resiko) = 0,17.

Menghitung Probabilitas untuk Variabel "Chest Pain Type (CPT)"
P(Typical | Tidak Resiko) = $\frac{1}{47}$ = 0,021
P(Typical Angina | Tidak Resiko) = $\frac{27}{47}$ = 0,57
P(Non Angina | Tidak Resiko) = $\frac{11}{47}$ = 0,234
P(Asymtomatic | Tidak Resiko) = $\frac{8}{47}$ = 0,17
P(Typical | Resiko) = $\frac{0}{23}$ = 0
P(Typical Angina | Resiko) = $\frac{3}{23}$ = 0,13
P(Non Angina | Resiko) = $\frac{4}{23}$ = 0,17
P(Asymtomatic | Resiko) = $\frac{16}{23}$ = 0,69

Gambar 6. Menghitung Probabilitas "CPT"
Sumber : Penulis 2024

Pada gambar 6 peneliti menghitung probabilitas variabel "CPT", probabilitas jika seorang pasien memiliki *"Typical Angina"*, ada kemungkinan 57% bahwa mereka berada dalam kelas "Tidak Resiko" dan 13% bahwa mereka berada dalam kelas "Resiko." Sebaliknya, jika seorang pasien memiliki "*Asymptomatic*" *chest pain*, ada kemungkinan 17% bahwa mereka berada dalam kelas "Tidak Resiko" dan 69% bahwa mereka berada dalam kelas "Resiko."

Menghitung Probabilitas untuk Variabel "Fasting Blood > 120"
P (True | Tidak Resiko) = $\frac{2}{47}$ = 0,042
P (True | Resiko) = $\frac{1}{23}$ = 0,043
P (False | Tidak Resiko) = $\frac{45}{47}$ = 0,95
P (False | Resiko) = $\frac{22}{23}$ = 0,95

Gambar 6. Menghitung Probabilitas "FBS > 120"
Sumber : Penulis 2024

Pada gambar 6 peneliti menghitung probabilitas variabel "FBS > 120", Probabilitas bahwa seseorang dengan gula darah puasa > 120 berada dalam kelas "Tidak Resiko" adalah 4,2%. Artinya, dari 47 orang yang tidak berisiko, hanya 2 orang memiliki gula darah puasa > 120 P(True|Tidak Resiko) = 0,042. Sebaliknya, probabilitas bahwa seseorang dengan gula darah puasa ≤ 120 berada dalam kelas "Tidak Resiko" adalah 95%. Artinya, dari 47 orang yang tidak berisiko, 45 orang memiliki gula darah puasa ≤ 120 P(False|Tidak Resiko) = 0,95. Probabilitas bahwa seseorang dengan gula darah puasa > 120 berada dalam kelas "Resiko" adalah 4,3%. Artinya, dari 23 orang yang berisiko, hanya 1 orang memiliki gula darah puasa > 120 P(True|Resiko) = 0,043. Sebaliknya, probabilitas bahwa seseorang dengan gula darah puasa ≤ 120 berada dalam kelas "Resiko" adalah 95%. Artinya, dari 23

Jurnal Smart Teknologi
Vol. 6, No. 2, Januari 2025, Halaman 264 – 275
ISSN: 2774-1702, http://jurnal.unmuhjember.ac.id/index.php/JST

orang yang berisiko, 22 orang memiliki gula darah puasa ≤ 120 P(False|Resiko) = 0,95.

Menghitung Probabilitas untuk Variabel "Resting Ecg"
P (Normal | Tidak Resiko) = $\frac{40}{47}$ = 0,85
P (Abnormal | Tidak Resiko) = $\frac{7}{47}$ = 0,14
P (Normal | Resiko) = $\frac{17}{23}$ = 0,73
P (Abnormal | Resiko) = $\frac{6}{23}$ = 0,26

Gambar 7. Menghitung Probabilitas "Resting ECG"
Sumber : Penulis 2024

Pada gambar 7 peneliti menghitung probabilitas variabel *"Resting ECG"*, Probabilitas bahwa seseorang dengan hasil "Normal" pada tes *"Resting ECG"* berada dalam kelas "Tidak Resiko" adalah 85%. Artinya, dari 47 orang yang tidak berisiko, 40 orang memiliki hasil tes *"Resting ECG"* yang normal P(Normal|Tidak Resiko) = 0,85 Probabilitas bahwa seseorang dengan hasil "Abnormal" pada tes *"Resting ECG"* berada dalam kelas "Tidak Resiko" adalah 14%. Artinya, dari 47 orang yang tidak berisiko, hanya 7 orang memiliki hasil tes *"Resting ECG"* yang abnormal P(Abnormal|Tidak Resiko) = 0,14. Probabilitas bahwa seseorang dengan hasil "Normal" pada tes *"Resting ECG"* berada dalam kelas "Resiko" adalah 73%. Artinya, dari 23 orang yang berisiko, 17 orang memiliki hasil tes *"Resting ECG"* yang normal P(Normal|Resiko)=0,73. Probabilitas bahwa seseorang dengan hasil "Abnormal" pada tes *Resting ECG* berada dalam kelas "Resiko" adalah 26%. Artinya, dari 23 orang yang berisiko, 6 orang memiliki hasil tes "Resting ECG" yang abnormal P(Abnormal|Resiko) = 0,26

Menghitung Probabilitas untuk Variabel "Exercise Angina"
P (No | Tidak Resiko) = $\frac{43}{47}$ = 0,91
P (Yes | Tidak Resiko) = $\frac{4}{47}$ = 0,08
P (No | Resiko) = $\frac{9}{23}$ = 0,39
P (Yes | Resiko) = $\frac{14}{23}$ = 0,60

Gambar 8. Menghitung Probabilitas "Exercise Angina"
Sumber : Penulis 2024

Pada gambar 8 peneliti menghitung probabilitas variabel *"Exercise Angina"*, Probabilitas bahwa seseorang tidak mengalami angina saat berolahraga berada dalam kelas "Tidak Resiko" adalah 91%, Artinya, dari 47 orang yang tidak berisiko, 43 orang tidak mengalami angina saat berolahraga P(No|Tidak Resiko)=0,91. Sebaliknya, probabilitas bahwa seseorang mengalami angina saat berolahraga berada dalam kelas "Tidak Resiko" adalah 8%. Artinya, dari 47 orang yang tidak berisiko, hanya 4 orang yang mengalami *angina* saat berolahraga P(Yes|Tidak Resiko) = 0,08. Probabilitas bahwa seseorang tidak mengalami angina saat berolahraga berada dalam kelas "Resiko" adalah 39%. Artinya, dari 23 orang yang berisiko, 9 orang tidak mengalami *angina* saat berolahraga P(No|Resiko) = 0,39
Sebaliknya, probabilitas bahwa seseorang mengalami *angina* saat berolahraga berada dalam kelas "Resiko" adalah 60%. Artinya, dari 23 orang yang berisiko, 14 orang mengalami angina saat berolahraga P(Yes|Resiko) = 0,60.

3. Menghitung Likelihood

Likelihood adalah probabilitas data observasi tertentu muncul dalam kelas target tertentu :

$$P(X \mid C)=P(x^1 \mid C) \ *P(x^2 \mid C)*\ldots*P(x_n|C)$$

P(X | C): probabilitas data observasi X yang termasuk dalam kelas C.
P(x^1 | C)*P(x^2 | C)*…*P(x_n|C): probabilitas kondisional dari setiap fitur x^1 diberikan kelas C.
Sex : Female (1)
Chest Pain Type : *Typical Angina* (2)
Fasting Blood Sugar : *False* (0)
Resting Ecg : Normal (0)
Exercise Angina : (0)

- Menghitung Likelihood untuk target "Tidak Resiko"

P(X|Tidak Resiko) = P(Sex = Female | Tidak Resiko) x P (CPT = Typical Angina | Tidak Resiko) x P (FBS = False | Tidak Resiko) x P (Resting ECG = Normal | Tidak Resiko) x P (Exercise Angina = No | Tidak Resiko)

Dengan mengalikan semua probabilitas kondisional tersebut, kita mendapatkan:
P(X|Tidak Resiko) = 0,44 x 0,57 x 0,95 x 0,85 x 0,91
Perhitungan ini menghasilkan = 0,184, dengan demikian likelihood P(X|Tidak Resiko) = 0,184 menunjukkan bahwa berdasarkan fitur-fitur yang diberikan *(Sex, Chest Pain Type, Fasting Blood Sugar, Resting ECG, Exercise*

Jurnal Smart Teknologi
Vol. 6, No. 2, Januari 2025, Halaman 264 – 275
ISSN: 2774-1702, http://jurnal.unmuhjember.ac.id/index.php/JST

Angina), probabilitas bahwa data observasi tersebut termasuk dalam kelas "Tidak Resiko" adalah 18,4%.

- Menghitung Likelihood untuk target "Resiko"

P(X|Resiko) = P(Sex = Female | Resiko) x P (CPT = Typical Angina | Resiko) x P (FBS = False | Resiko) x P (Resting ECG = Normal | Resiko) x P (Exercise Angina = No | Resiko)

Dengan mengalikan semua probabilitas kondisional tersebut, kita mendapatkan :

P(X|Resiko) = 0,17 x 0,13 x 0,95 x 0,73 x 0,39

Perhitungan ini menghasilkan = 0,005, dengan demikian likelihood P(X|Tidak Resiko) = 0,005 menunjukkan bahwa berdasarkan fitur-fitur yang diberikan *(Sex, Chest Pain Type, Fasting Blood Sugar, Resting ECG, Exercise Angina)*, probabilitas bahwa data observasi tersebut termasuk dalam kelas "Tidak Resiko" adalah 0,5%.

4. Menghitung Posterior Probability

Posterior probability adalah probabilitas target kelas tertentu diberikan data observasi yang ditunjukkan :

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

P(C|X) : probabilitas target kelas C diberikan data observasi X

P(X|Ci) : likelihood probabilitas data observasi X yang termasuk dalam kelas Ci

P(Ci) : prior probabilitas, probabilitas awal dari kelas Ci

- Menghitung Posterior untuk "Tidak Resiko"

P (Tidak Resiko | X) = P (X | Tidak Resiko) x P (Tidak Resiko)

= 0,184 x 0,6714

= 0,123

- Menghitung Posterior untuk "Resiko"

P (Resiko | X) = P (X | Resiko) x P (Resiko)

= 0,005 x 0,3285

= 0,001

5. Menghitung Normalisasi

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

$$P(Ci|X) = \frac{(P(X|Ci) * P(Ci)}{(P(X)}$$

P (Tidak Resiko | X) = 0,123/(0,123+0,001) = 0,123/0,124 = 0,99

P (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".

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, cholesterol, fasting blood sugar >120, resting electrocardiographic, max heart rate, exercise angina, oldpeak, st slope dan target*. Data ini digunakan sebagai dasar untuk membangun model klasifikasi 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 klasifikasi.

B. Tahap Preprocessing

Preprocessing merupakan proses penting untuk membersihkan dan mempersiapkan data agar siap digunakan dalam model klasifikasi. Tahapan preprocessing terbagi menjadi beberapa tahapan yaitu transformasi fitur numerik dan transformasi fitur kategori.

1. Transformasi Fitur Numerik

Fitur numerik akan distandarisasi menggunakan *'Standard Scaler'*. Standarisasi ini mengubah fitur sehingga memiliki mean nol dan standar deviasi satu, yang membantu model *machine learning* berperforma lebih baik.

Jurnal Smart Teknologi
Vol. 6, No. 2, Januari 2025, Halaman 264 – 275
ISSN: 2774-1702, http://jurnal.unmuhjember.ac.id/index.php/JST

2. Transformasi Fitur Kategori

Fitur kategori akan diubah menjadi representasi *'one-hot encoding'* menggunakan *'OneHotEncoder'*. *'One-hot encoding'* mengubah setiap kategori unik menjadi kolom biner terpisah (0 atau 1), yang memungkinkan model *machine learning* untuk memahami dan memproses data kategori dengan lebih efektif.

```
# Identifikasi kolom numerik dan kategori
numeric_features = ['age', 'resting bp s', 'cholesterol', 'max heart rate', 'oldpeak']
categorical_features = ['sex', 'chest pain type', 'fasting blood sugar', 'resting ecg', 'exercise angina', 'ST slope']

# Proses preprocessing untuk data
preprocessor = ColumnTransformer(
    transformers=[
        ('num', StandardScaler(), numeric_features),
        ('cat', OneHotEncoder(handle_unknown='ignore'), categorical_features)
    ]
)
```

Gambar 9. Tahap Preprocessing
Sumber : Penulis 2024

C. Input data Training

Langkah pertama dalam melatih model adalah menginput data training, pada penelitian data training berupa file excel/xlsx. Data training diperoleh dari proses *K-Fold Cross Validation* proses ini untuk mendapatkan kinerja model yang lebih stabil dengan jumlah data 1190 akan dibagi menjadi 10 bagian yang seimbang, dengan 1071 data training yang terdiri dari 505 kelas tidak resiko, 566 kelas resiko dan 119 data testing yang terdiri dari 56 kelas tidak resiko, 63 kelas resiko.

Gambar 10. Input Data Training
Sumber : Penulis 2024

Setelah data diinput maka latih model dan unduh, untuk memperoleh model yang didapat dari data training. Model ini berbentuk dalam file pkl, model ini nantinya diujikan pada data testing.

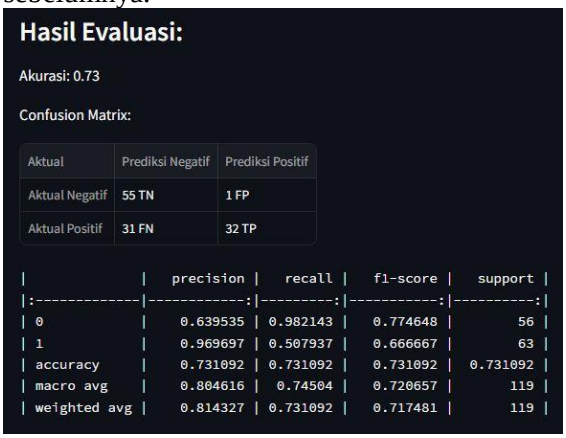
D. Input Data Testing dan Model

Input data testing adalah sekumpulan data yang digunakan untuk menguji kemampuan model klasifikasi dalam melakukan generalisasi terhadap data baru yang belum pernah dilihat sebelumnya. Data ini biasanya dipisahkan dari data pelatihan. Model dalam konteks ini adalah hasil dari proses pembelajaran yang telah dilatih menggunakan data pelatihan. Model tersebut mengandung pola dan hubungan yang telah dipelajari dari data pelatihan dan digunakan untuk membuat klasifikasi data baru.

Gambar 11. Input Data Testing dan Model
Sumber : Penulis 2024

E. Hasil Evaluasi

Hasil evaluasi model klasifikasi merupakan langkah penting dalam proses pengembangan model pembelajaran mesin, yang bertujuan untuk menilai seberapa baik model tersebut melakukan klasifikasi berdasarkan data yang belum pernah dilihat sebelumnya.



Gambar 12. Hasil Evaluasi
Sumber : Penulis 2024

Selama proses evaluasi model klasifikasi, dilakukan sepuluh uji coba untuk mengukur performa model menggunakan metrik akurasi, presisi, recall, dan F1-score.

Tabel 2. Hasil Evaluasi Seluruh Uji Coba

Langkah Uji	Akurasi	Presisi	Recall	F1-score
Uji coba 1	0,73	0,73	0,73	0,73

Jurnal Smart Teknologi
Vol. 6, No. 2, Januari 2025, Halaman 264 – 275
ISSN: 2774-1702, http://jurnal.unmuhjember.ac.id/index.php/JST

Uji coba 2	0,76	0,76	0,76	0,76
Uji coba 3	0,85	0,84	0,84	0,84
Uji coba 4	0,86	0,85	0,85	0,85
Uji coba 5	0,87	0,87	0,87	0,87
Uji coba 6	0,82	0,82	0,82	0,82
Uji coba 7	0,69	0,68	0,68	0,68
Uji coba 8	0,70	0,69	0,69	0,69
Uji coba 9	0,71	0,70	0,70	0,70
Uji coba 10	0,69	0,68	0,68	0,68

Sumber : Penulis 2024

pada uji coba ke 5 memperoleh hasil evaluasi yang cukup tinggi dengan hasil Akurasi sebesar 0,87 berarti bahwa 87% dari total klasifikasi yang dibuat oleh model adalah benar. Ini menunjukkan bahwa model secara keseluruhan memiliki kinerja yang baik dalam mengklasifikasikan data dengan benar sebagai berisiko atau tidak berisiko terkena penyakit jantung.

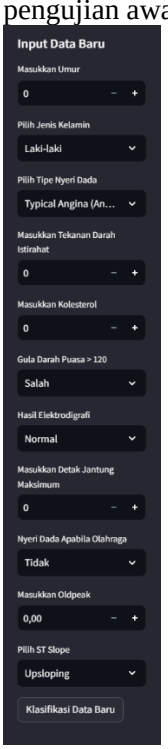
Presisi sebesar 0,87 menunjukkan bahwa dari semua klasifikasi yang dinyatakan positif oleh model, 87% di antaranya benar-benar berisiko penyakit jantung. Tingkat presisi yang tinggi mengindikasikan bahwa model memiliki tingkat kesalahan positif yang rendah, sehingga jarang mengklasifikasikan individu yang sehat sebagai berisiko.

Recall sebesar 0,87 berarti bahwa dari semua kasus yang sebenarnya berisiko penyakit jantung, model berhasil mengidentifikasi 87% di antaranya. Tingkat recall yang tinggi penting dalam konteks medis karena memastikan bahwa sebagian besar individu yang berisiko terdeteksi oleh model.

F1 Score sebesar 0,87 merupakan rata-rata harmonis dari presisi dan recall, memberikan ukuran keseimbangan antara keduanya. Nilai F1 Score yang tinggi menunjukkan bahwa model memiliki performa yang konsisten baik dalam mengidentifikasi kasus berisiko maupun meminimalkan kesalahan klasifikasi.

F. Input Data Baru

Input data baru adalah sekumpulan data yang belum pernah dilihat oleh model klasifikasi sebelumnya dan digunakan untuk menguji kemampuan model dalam membuat klasifikasi pada situasi dunia nyata. Data baru ini penting karena memungkinkan kita untuk melihat bagaimana model berperforma ketika diterapkan di luar lingkungan pelatihan dan pengujian awal.



Gambar 13. Input Data Baru
Sumber : Penulis 2024

5. KESIMPULAN DAN SARAN

A. Kesimpulan

Dengan menggunakan algoritma *Naive Bayes* pada klasifikasi penyakit gagal jantung dengan data yang tersedia hasil evaluasi menggunakan *k-fold cross-validation* dengan k=10, ada variasi yang cukup signifikan dalam hasil uji coba, dengan uji coba terbaik (ke-5) mencapai nilai metrik tertinggi memperoleh hasil evaluasi 0,87, sementara beberapa uji coba lainnya (ke-7, ke-8, dan ke-10) menunjukkan performa yang lebih rendah

B. Saran

1. Tambahkan fitur-fitur baru yang relevan seperti riwayat medis keluarga, pola makan, aktivitas fisik, dan data genetik.
2. Kolaborasi dengan ahli medis akan memberikan validasi klinis terhadap model dan memastikan relevansi medis dari hasil klasifikasi.

6. REFERENSI

A. Artikel

Alizadehsani, R., Roshanzamir, M., Abdar, M., Beykikhoshk, A., Khosravi, A., Panahiazar, M., Koohestani, A., Khozeimeh, F., Nahavandi, S., & Sarrafzadegan, N. (2019). A database for using machine learning and data mining techniques for coronary artery disease diagnosis. Scientific Data, 6(1), 1–13. https://doi.org/10.1038/s41597-019-0206-3

Hayami, R., Soni, & Gunawan, I. (2022). Klasifikasi Jamur Menggunakan Algoritma Naïve Bayes. Jurnal CoSciTech (Computer Science and Information Technology), 3(1), 28–33. https://doi.org/10.37859/coscitech.v3i1.3685

Pebdika, A., Herdiana, R., & Solihudin, D. (2023). Klasifikasi Menggunakan Metode *Naive Bayes* Untuk Menentukan Calon Penerima Pip. JATI (Jurnal Mahasiswa Teknik Informatika), 7(1), 452–458. https://doi.org/10.36040/jati.v7i1.6303

Putro, H. F., Vulandari, R. T., & Saptomo, W. L. Y. (2020). Penerapan Metode *Naive Bayes* Untuk Klasifikasi Pelanggan. Jurnal Teknologi Informasi dan Komunikasi (TIKomSiN), 8(2). https://doi.org/10.30646/tikomsin.v8i2.500

Mustofa, H., & Mahfudh, A. A. (2019). Klasifikasi Berita Hoax Dengan Menggunakan Metode *Naive Bayes*. Walisongo Journal of Information Technology, 1(1), 1. https://doi.org/10.21580/wjit.2019.1.1.3915

Setiawan, R., & Triayudi, A. (2022). Klasifikasi Status Gizi Balita Menggunakan Naïve Bayes dan K-Nearest Neighbor Berbasis Web. Jurnal Media Informatika Budidarma, 6(2), 777. https://doi.org/10.30865/mib.v6i2.3566

Veronica Agustin, A., & Voutama, A. (2023). Implementasi Data Mining Klasifikasi Penyakit Diabetes Pada Perempuan Menggunakan Naïve Bayes. JATI (Jurnal Mahasiswa Teknik Informatika), 7(2), 1002–1007. https://doi.org/10.36040/jati.v7i2.6808

Tjengharwidjaja, A., Saputra, B. D., & Michael Emmanuel, Y. M. (2024). Klasifikasi Pasien Terkena Breast Cancer Menggunakan Metode Machine Learning. Computatio : Journal of Computer Science and Information Systems, 8(1), 86–95. https://doi.org/10.24912/computatio.v8i1.15174

Ridwansyah, T. (2022). Implementasi Text Mining Terhadap Analisis Sentimen Masyarakat Dunia Di Twitter Terhadap Kota Medan Menggunakan K-Fold Cross Validation Dan Naïve Bayes Classifier. KLIK: Kajian Ilmiah Informatika dan Komputer, 2(5), 178–185. https://doi.org/10.30865/klik.v2i5.362

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. Kontribusia: Research Dissemination for Community Development, 2(1), 24-30.