

# Evaluasi Performa Algoritma Klasifikasi dalam Prediksi Gagal Jantung: Studi Kasus *Dataset Heart Failure Prediction*

*Performance Evaluation of Classification Algorithms in Heart Failure Prediction: A Case Study of the Heart Failure Prediction Dataset*

Wahyu Nugraha<sup>1</sup>, Muhamad Syarif<sup>2</sup>

<sup>1,2</sup>Fakultas Teknik dan Informatika, Universitas Bina Sarana Informatika

E-mail: <sup>1</sup> wahyu.whn@bsi.ac.id, <sup>2</sup>muhamad.mdx@bsi.ac.id

## Abstrak

Kegagalan jantung (*Heart Failure*) merupakan salah satu penyebab utama kematian dan morbiditas secara global, dengan lebih dari 26 juta orang terdampak di seluruh dunia. Deteksi dini kegagalan jantung sangat penting untuk mencegah komplikasi yang lebih serius, namun diagnosis sering kali sulit dilakukan karena gejala yang tidak spesifik. Pembelajaran mesin (*machine learning*) telah banyak digunakan untuk membantu prediksi risiko kegagalan jantung melalui analisis data medis. Berbagai algoritma klasifikasi telah diimplementasikan dalam berbagai penelitian sebelumnya, namun belum ada konsensus mengenai algoritma terbaik. Penelitian ini bertujuan untuk melakukan komparasi performa beberapa algoritma klasifikasi dalam memprediksi gagal jantung, menggunakan *dataset Heart Failure Prediction* dari Kaggle. Algoritma yang diuji termasuk *Logistic Regression*, *K-Nearest Neighbors* (KNN), *Decision Tree*, *Random Forest*, *GaussianNB*, dan *LightGBM*. Hasil menunjukkan bahwa algoritma *Random Forest* memberikan performa terbaik dengan akurasi 95 persen, sedangkan algoritma KNN menunjukkan akurasi terendah sebesar 76 persen. Penelitian ini diharapkan dapat berkontribusi dalam pengembangan sistem pendukung keputusan klinis yang lebih baik untuk prediksi kegagalan jantung.

Kata Kunci: *Heart Failure*, mesin pembelajaran algoritma klasifikasi

## Abstract

*Heart failure is one of the leading causes of mortality and morbidity globally, affecting over 26 million people worldwide. Early detection of heart failure is crucial to prevent more serious complications; however, diagnosis is often challenging due to non-specific symptoms. Machine learning has been widely used to aid in predicting heart failure risk through the analysis of medical data. Various classification algorithms have been implemented in previous research, but no consensus has been reached regarding the best algorithm. This study aims to compare the performance of several classification algorithms in predicting heart failure using the Heart Failure Prediction dataset from Kaggle. The algorithms tested include Logistic Regression, K-Nearest Neighbors (KNN), Decision Tree, Random Forest, GaussianNB, and LightGBM. The results show that the Random Forest algorithm provided the best performance with an accuracy of 95 percent, while the KNN algorithm showed the lowest accuracy at 76 percent. This research is expected to contribute to the development of better clinical decision support systems for heart failure prediction.*

Keywords: *Heart Failure*, machine learning, classification algorithms

## 1. PENDAHULUAN

Kegagalan jantung adalah salah satu masalah kesehatan global yang signifikan dan menjadi penyebab utama kematian serta morbiditas di berbagai negara. Menurut data dari Organisasi Kesehatan Dunia (WHO), lebih dari 26 juta orang di seluruh dunia mengalami kegagalan jantung, dan prevalensinya terus meningkat seiring bertambahnya usia populasi[1]. Kegagalan jantung terjadi ketika jantung tidak mampu memompa darah yang cukup untuk memenuhi kebutuhan tubuh, dan kondisi ini sering kali disebabkan oleh kombinasi beberapa faktor, termasuk hipertensi, penyakit jantung koroner, diabetes serta gaya hidup tidak sehat seperti merokok dan pola makan yang buruk [2], [3]. Deteksi dini kegagalan jantung sangat penting untuk mencegah komplikasi lebih lanjut dan memperbaiki prognosis pasien serta tantangan dalam diagnosa [4], [5]. Namun, gejala kegagalan jantung sering kali tidak spesifik, yang membuat diagnosis klinis menjadi tantangan[6]. Oleh karena itu, pendekatan berbasis pembelajaran mesin (*machine learning*) semakin banyak digunakan untuk membantu dalam prediksi risiko kegagalan jantung dengan menganalisis data medis secara otomatis[7]. Model prediksi berbasis data dapat membantu para praktisi medis dalam mengidentifikasi pasien yang berisiko tinggi mengalami kegagalan jantung lebih dini, memungkinkan intervensi medis yang lebih tepat waktu.

Dalam beberapa tahun terakhir, berbagai algoritma klasifikasi berbasis pembelajaran mesin telah diimplementasikan untuk memprediksi kegagalan jantung, seperti *Decision Tree*, *Random Forest*, *Gradient Boosting*, dan *Extra Trees*. Masing-masing algoritma memiliki karakteristik unik dalam hal penanganan data, kemampuan generalisasi, serta kecepatan komputasi[8], [9]. Karena penyakit jantung memiliki karakter yang kompleks, penyakit ini memerlukan penanganan yang hati-hati. Algoritma berbasis regresi, *knn*, *svm*, *nb*, dan *dt* digunakan untuk mengkategorikan tingkat keparahan kondisi tersebut sehingga membutuhkan *machine learning* (ML) dalam pengambilan keputusan dan prediksi dari sejumlah besar data medis[10]. Meskipun algoritma-algoritma ini telah banyak digunakan, masih belum ada konsensus yang jelas tentang algoritma mana yang memberikan performa terbaik dalam konteks prediksi kegagalan jantung.

Penelitian sebelumnya telah menunjukkan bahwa algoritma *Random Forest* dan *Gradient Boosting* cenderung unggul dalam akurasi prediksi gagal jantung dibandingkan algoritma lain. Kedua algoritma ini dikenal karena kemampuannya dalam menangani data yang kompleks dan ketidakseimbangan kelas, yang umum dalam data medis gagal jantung. Hal ini dapat meningkatkan keakuratan prediksi serta membantu intervensi dini dalam pengelolaan pasien [11]. Penelitian lain menyebutkan *Logistic Regression* sering dipilih dalam prediksi gagal jantung karena interpretabilitasnya yang tinggi, terutama di lingkungan klinis. Meskipun akurasinya mungkin sedikit lebih rendah dibandingkan algoritma seperti *Random Forest* [12]. Penerapan algoritma *machine learning* seperti *Random Forest* dan *Logistic Regression* dalam sistem pendukung keputusan klinis telah menunjukkan dampak positif terhadap diagnosis dan pengelolaan gagal jantung [13]. Efektivitas algoritma prediksi sangat bergantung pada karakteristik data yang digunakan, termasuk jumlah variabel, ketidakseimbangan kelas, dan ukuran *dataset*. Algoritma seperti *K-Nearest Neighbors (KNN)* seringkali menunjukkan performa lebih rendah dalam data besar atau data dengan ketidakseimbangan kelas, yang dapat menyebabkan kesalahan prediksi pada kategori kelas minoritas [14]. Penelitian menunjukkan bahwa data dengan jumlah sampel besar dan variabel prediktif yang relevan lebih sesuai dengan algoritma seperti *Random Forest* dan *LightGBM* dibandingkan dengan KNN [14]. Setiap algoritma memiliki keunggulan dan keterbatasan yang bervariasi bergantung pada karakteristik data, seperti distribusi variabel, ukuran *dataset*, dan tingkat ketidakseimbangan kelas[10], [15]. Oleh karena itu, tujuan dari penelitian ini adalah untuk melakukan komparasi performa berbagai algoritma klasifikasi tersebut dalam memprediksi gagal jantung.

Model evaluasi akan digunakan untuk menentukan algoritma mana yang paling efektif dalam prediksi gagal jantung. Dengan mengidentifikasi algoritma yang paling optimal, penelitian

ini diharapkan dapat berkontribusi dalam pengembangan sistem pendukung keputusan klinis yang lebih baik. Penelitian ini menggunakan dataset publik spesifik dari Kaggle, yaitu *heart failure prediction*, yang mungkin belum banyak dieksplorasi dalam penelitian sebelumnya untuk prediksi gagal jantung menggunakan metode komparatif. Penelitian ini tidak hanya menerapkan berbagai algoritma, tetapi juga melakukan analisis komparatif mendalam terhadap algoritma seperti *Logistic Regression (LR)*, *K-Nearest Neighbors (KNN)*, *Decision Tree (DT)*, *RandomForest (RF)*, *GaussianNB*, dan *Lightgbm*, untuk mengidentifikasi model terbaik. Dengan mengidentifikasi algoritma yang memberikan hasil paling optimal, penelitian ini berpotensi berkontribusi dalam pengembangan sistem pendukung keputusan klinis yang lebih efektif untuk prediksi risiko gagal jantung di masa depan. Dengan begitu, penelitian ini tidak hanya menambah wawasan mengenai kinerja berbagai algoritma, tetapi juga memberi rekomendasi spesifik untuk aplikasi klinis.

Penelitian ini menggunakan kumpulan data atau *dataset* yang digunakan untuk eksperimen pengujian model prediksi kegagalan jantung dari *Dataset Heart Failure Prediction* yang diperoleh dari *Kaggle dataset repository*. Penelitian ini akan didukung oleh analisis komparatif menggunakan beberapa algoritma klasifikasi untuk pembuatan model prediksi gagal jantung. Penelitian ini berfokus pada identifikasi algoritma yang paling optimal dalam konteks prediksi gagal jantung dengan menggunakan model evaluasi untuk mengukur keefektifan masing-masing algoritma. Dari beberapa penelitian sebelumnya menyatakan bahwa *Random Forest* dan *Logistic Regression* memiliki performa lebih baik dalam prediksi penyakit jantung jika dibandingkan dengan *K-Nearest Neighbors (KNN)* yang seringkali menunjukkan performa lebih rendah. Hasil yang ditemukan mengindikasikan kesesuaian bahwa *Random Forest* memberikan performa tertinggi (95% akurasi), sedangkan *KNN* memiliki akurasi terendah (76%).

## 2. METODE PENELITIAN

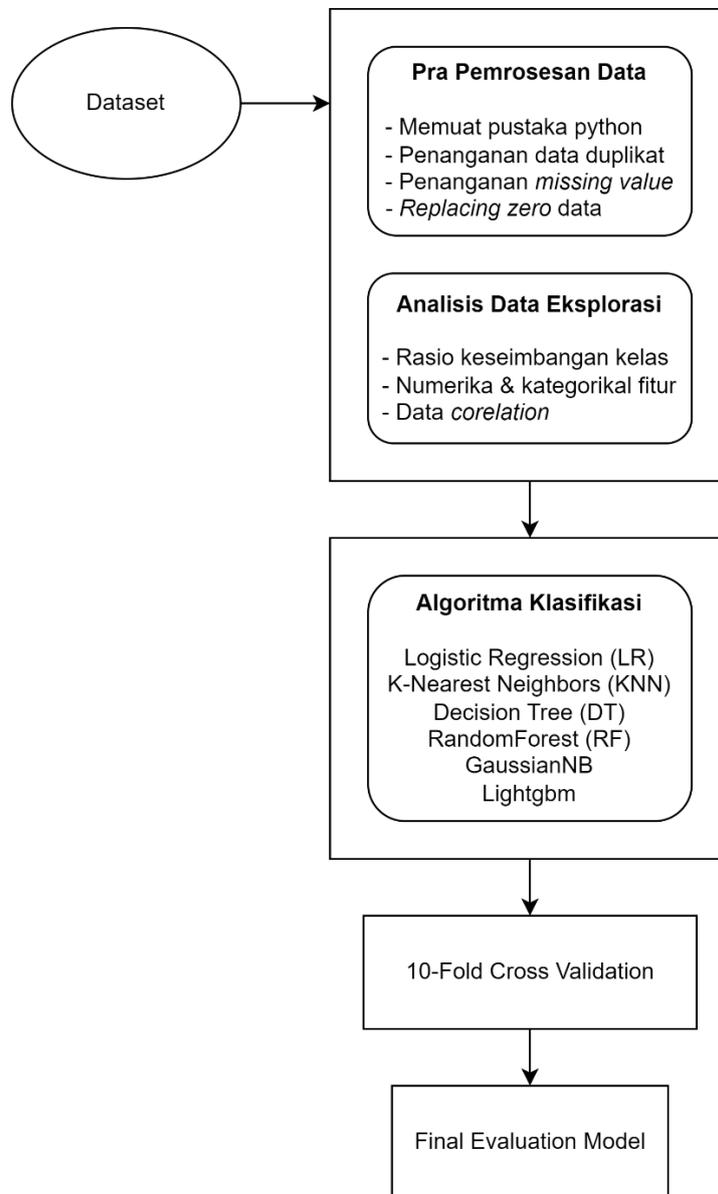
### 2.1 Pemodelan

Penelitian ini mengeksplorasi prediksi risiko gagal jantung menggunakan beberapa algoritma klasifikasi dalam model *machine learning*. Penggunaan *machine learning* untuk memprediksi risiko gagal jantung telah menjadi perhatian berbagai peneliti, yang memanfaatkan berbagai metode untuk menganalisis data medis secara efektif. Tahapan eksperimen dalam penelitian ini dijelaskan dalam Gambar 1, dimulai dari proses *import dataset* dengan bantuan pustaka Python. Setelah itu, dilakukan tahap *data preprocessing* untuk membersihkan dan mempersiapkan data yang diperlukan agar sesuai untuk analisis lanjutan.

Tahap berikutnya melibatkan *exploratory data analysis (EDA)* guna memahami pola dan karakteristik dalam dataset. Setelah EDA, model *machine learning* diuji dengan beberapa algoritma klasifikasi seperti *Logistic Regression (LR)*, *K-Nearest Neighbors (KNN)*, *Decision Tree (DT)*, *Random Forest (RF)*, *Gaussian Naive Bayes (GaussianNB)*, dan *LightGBM*. Setiap algoritma ini memiliki kelebihan dan kekurangan dalam hal kecepatan, akurasi, serta kemampuan generalisasi terhadap data.

Pada tahapan evaluasi model dilakukan dengan *cross-validation* yang merupakan teknik evaluasi dalam *machine learning* yang digunakan untuk menguji akurasi dan ketahanan model pada data baru, membantu mengurangi risiko *overfitting* (model terlalu spesifik terhadap data latih) dan *underfitting* (model kurang mampu menangkap pola data). Dalam *k-fold cross-validation*, dataset dibagi menjadi *k* subset atau *folds*. Model dilatih pada *k-1* bagian data dan diuji pada bagian yang tersisa, dengan proses ini diulang hingga setiap subset menjadi data uji satu kali. Skor dari tiap *fold* kemudian dirata-rata untuk mendapatkan skor performa akhir. Pendekatan ini memberikan estimasi yang lebih andal dibandingkan metode *train-test split* karena model diuji pada seluruh data dengan distribusi yang bervariasi [16].

Langkah terakhir dalam penelitian ini adalah melakukan komparasi performa berbagai model algoritma klasifikasi tersebut dalam memprediksi gagal jantung yang bertujuan untuk mengukur efektivitas setiap algoritma dalam memprediksi risiko gagal jantung.



Gambar 1 Alur Tahapan Penelitian

## 2.2 Pengumpulan Data

*Dataset* yang digunakan dalam penelitian ini terkait dengan penyakit gagal jantung, yang diperoleh dari *Kaggle dataset repository* dalam format CSV (*comma separated values*) yang dapat di unduh <https://www.kaggle.com/datasets/fedoriano/heart-failure-prediction>. Penyakit kardiovaskular (*Cardiovascular diseases/CVDs*) merupakan penyebab kematian nomor 1 di dunia, diperkirakan merenggut 17,9 juta nyawa setiap tahun, atau sekitar 31% dari total kematian global. Empat dari lima kematian akibat CVD disebabkan oleh serangan jantung dan *stroke*, dengan sepertiga dari kematian ini terjadi secara prematur pada orang di bawah usia 70 tahun. Gagal jantung, sebagai salah satu akibat umum dari CVD, menjadi fokus *dataset* ini, yang berisi 11 fitur untuk memprediksi kemungkinan penyakit jantung.

*Dataset* ini dibuat dengan menggabungkan berbagai *dataset* yang sebelumnya tersedia secara terpisah, namun belum pernah digabungkan sebelumnya. Dalam *dataset* ini, lima *dataset*

tentang penyakit jantung digabungkan berdasarkan 11 fitur umum, menjadikannya *dataset* penyakit jantung terbesar yang tersedia sejauh ini untuk tujuan penelitian. Lima *dataset* yang digunakan untuk penyusunannya adalah sebagai berikut: *Cleveland* dengan 303 observasi, *Hungarian* dengan 294 observasi, *Switzerland* dengan 123 observasi, *Long Beach VA* dengan 200 observasi, dan *Stalog (Heart) Dataset* dengan 270 observasi.

### 2.3 Validasi

*10-fold Cross-Validation* adalah teknik evaluasi yang umum digunakan dalam *machine learning* untuk mengukur performa model secara lebih akurat dan menghindari masalah *overfitting*. Dalam *10-fold cross-validation*, *dataset* dibagi menjadi 10 subset (*folds*). Model dilatih pada 9 *folds* dan diuji pada *fold* yang tersisa. Proses ini diulangi 10 kali, di mana setiap *fold* digunakan sekali sebagai data uji. Skor evaluasi rata-rata dari 10 iterasi tersebut memberikan estimasi performa model yang lebih stabil dan dapat diandalkan dibandingkan pembagian data uji yang dilakukan satu kali saja [17].

Metode ini dikenal efektif karena memastikan bahwa setiap data digunakan baik untuk pelatihan maupun pengujian, sehingga meminimalkan variabilitas hasil yang muncul akibat pembagian *dataset* yang berbeda. Teknik ini terutama bermanfaat pada *dataset* yang relatif kecil, karena memastikan semua data berkontribusi dalam proses pelatihan dan pengujian model [18]. Metode ini mengurangi variansi hasil dibandingkan dengan metode *train-test split* sederhana karena setiap data mendapatkan kesempatan untuk menjadi data uji, memberikan gambaran performa model yang lebih stabil dan umum. Dengan mengurangi risiko *overfitting*, *10-fold cross-validation* membantu memastikan bahwa model dapat memprediksi dengan baik pada data baru [19]. Tabel 1 menggambarkan proses *10-fold cross-validation* dimana setiap iterasi menggunakan satu *fold* sebagai *test fold*, sementara *fold* lainnya digunakan sebagai *training folds*.

Tabel 1 Proses *10-fold cross-validation*

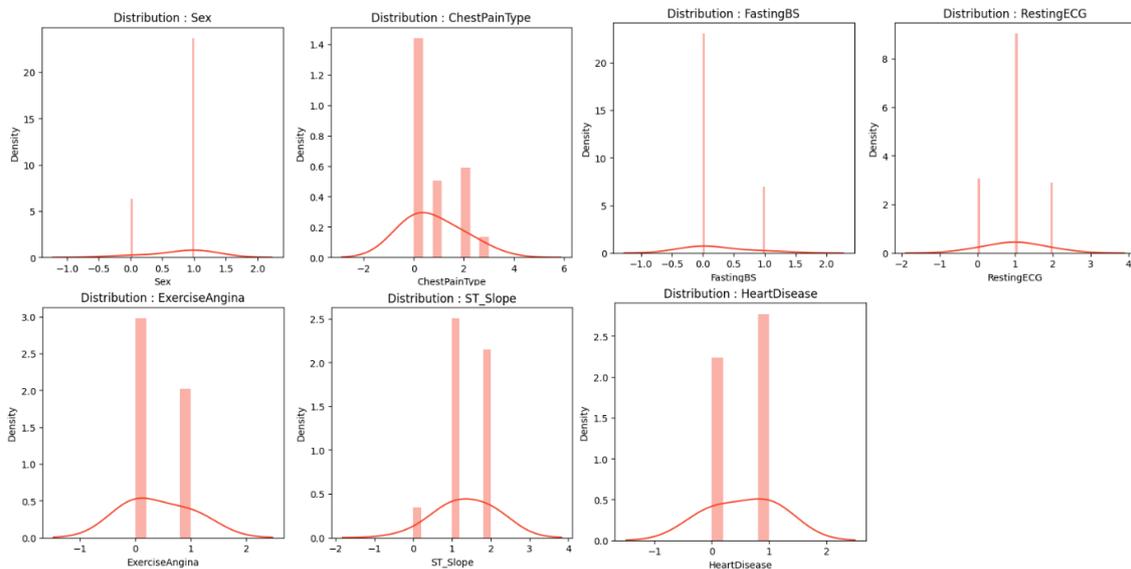
<i>Iteration</i>	<i>Training Folds</i>	<i>Test Fold</i>
<i>Fold 1</i>	<i>Fold 2-10</i>	<i>Fold 1</i>
<i>Fold 2</i>	<i>Fold 1,3-10</i>	<i>Fold 2</i>
<i>Fold 3</i>	<i>Fold 1-2,4-10</i>	<i>Fold 3</i>
<i>Fold 4</i>	<i>Fold 1-3,5-10</i>	<i>Fold 4</i>
<i>Fold 5</i>	<i>Fold 1-4,6-10</i>	<i>Fold 5</i>
<i>Fold 6</i>	<i>Fold 1-5,7-10</i>	<i>Fold 6</i>
<i>Fold 7</i>	<i>Fold 1-6,8-10</i>	<i>Fold 7</i>
<i>Fold 8</i>	<i>Fold 1-7,9-10</i>	<i>Fold 8</i>
<i>Fold 9</i>	<i>Fold 1-8,10</i>	<i>Fold 9</i>
<i>Fold 10</i>	<i>Fold 1-9</i>	<i>Fold 10</i>

## 3. HASIL DAN PEMBAHASAN

Eksperimen dari model yang dibangun dimulai dari tahap *preprocessing* untuk kemudian dilakukan observasi data untuk menemukan duplikasi data serta mengecek apakah ada data yang hilang atau missing value. Langkah selanjutnya setelah data siap dilakukan tahapan membagi fitur menjadi numerik dan kategorikal yang bertujuan untuk proses pembuatan model. Fitur kategorikal didefinisikan jika atribut memiliki kurang dari 6 elemen unik, jika tidak maka itu adalah fitur numerik. Pendekatan umum untuk pembagian fitur ini juga dapat didasarkan pada tipe data dari elemen-elemen atribut yang bersangkutan. Misalnya tipe data integer maka atribut adalah fitur numerik dan sebaliknya.

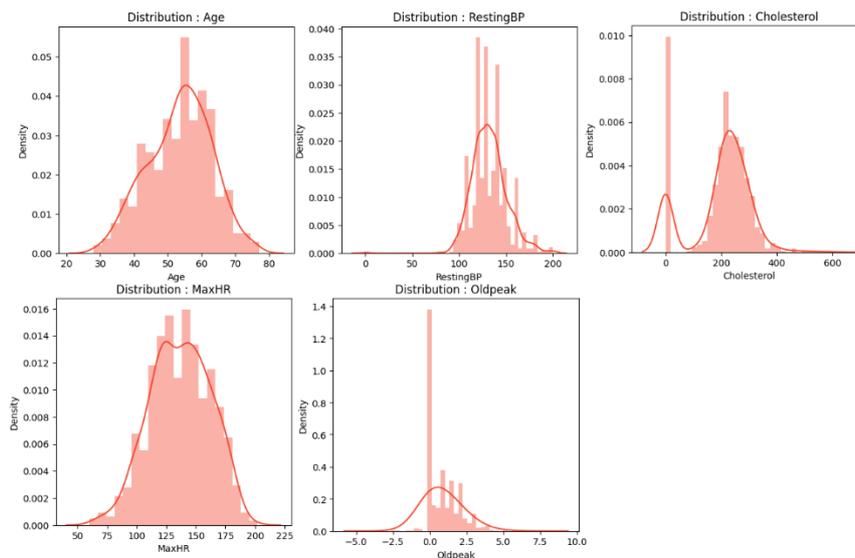
Selanjutnya menganalisa distribusi dari kategorikal fitur. Oleh karena itu, kami menggunakan *deep copy* dari *dataset* ini yang semua fiturnya telah dikonversi menjadi nilai numerik untuk keperluan visualisasi dan pemodelan. Gambar 2 menunjukkan bahwa semua fitur kategorikal hampir terdistribusi secara normal. Fitur kategorikal memiliki distribusi kategori yang relatif seragam atau seimbang. Artinya, setiap kategori dalam fitur tersebut muncul dengan

proporsi yang tidak terlalu timpang atau tidak didominasi oleh satu kategori tertentu. Kondisi ini bisa mendukung analisis karena setiap kategori terwakili secara memadai dalam data.



Gambar 2 Distribution of Categorical Features

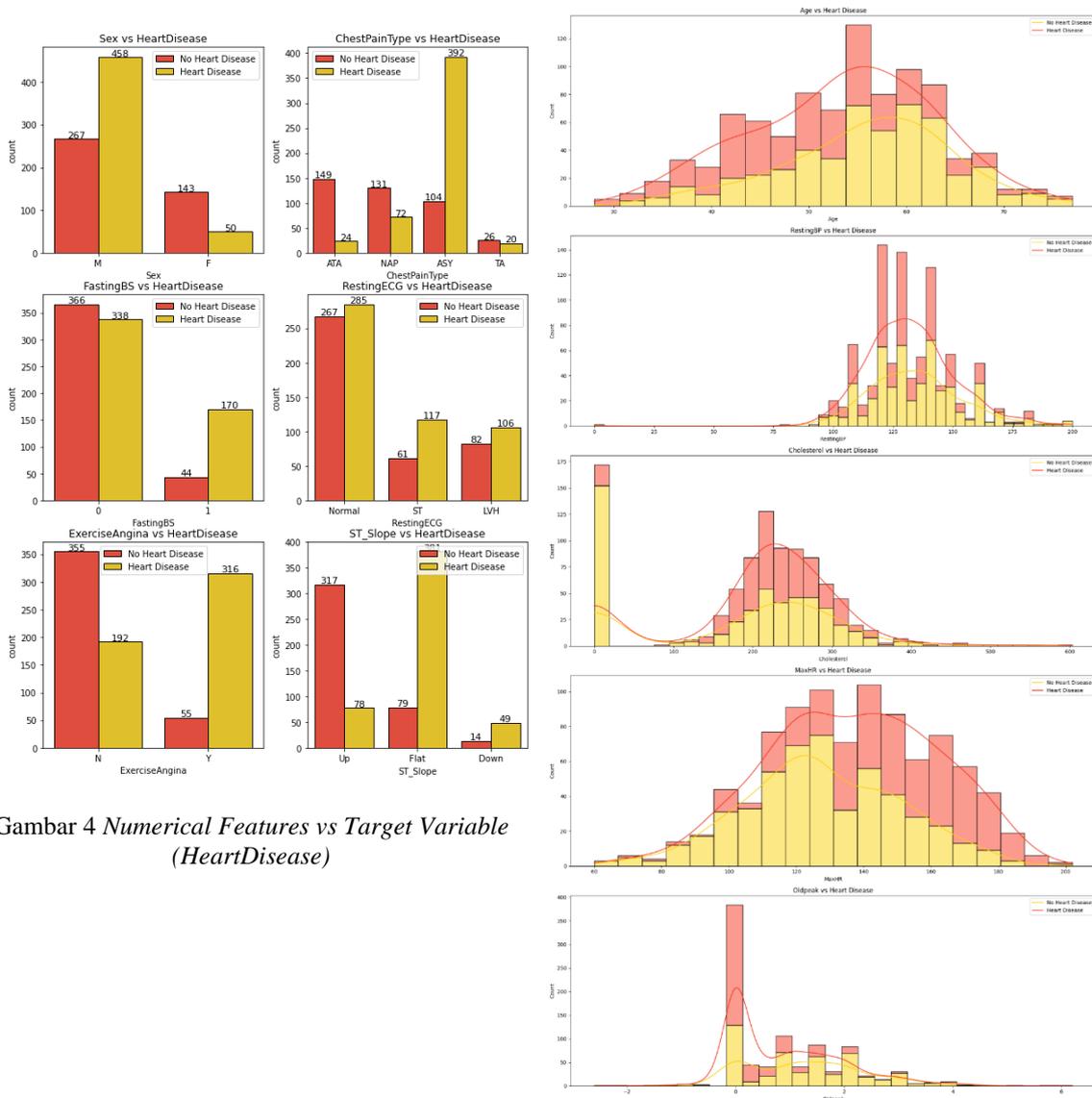
Selanjutnya menganalisa distribusi dari numerikal fitur. Gambar 3 menunjukkan visualisasi dari distribusi numerikal fitur. distribusi data untuk variabel *Oldpeak* tidak simetris dan memiliki ekor yang lebih panjang di sisi kanan (nilai yang lebih tinggi). Dalam distribusi yang *skewed* ke kanan, sebagian besar data berada di sisi kiri (nilai lebih rendah), sedangkan sedikit data memiliki nilai lebih tinggi yang membentuk ekor di kanan. Sedangkan distribusi data untuk variabel *Cholesterol* memiliki dua puncak (dua mode). Bimodal menunjukkan bahwa data tidak berkumpul di satu puncak, tetapi ada dua kelompok atau rentang nilai di mana frekuensi data relatif tinggi, yang mencerminkan adanya dua pola distribusi yang berbeda dalam data tersebut.



Gambar 3 Distribution of Numerical Features

Selanjutnya kita analisa hubungan korelasi antara kategorikal fitur terhadap variabel target. Gambar 4 dapat ditarik kesimpulan bahwa:

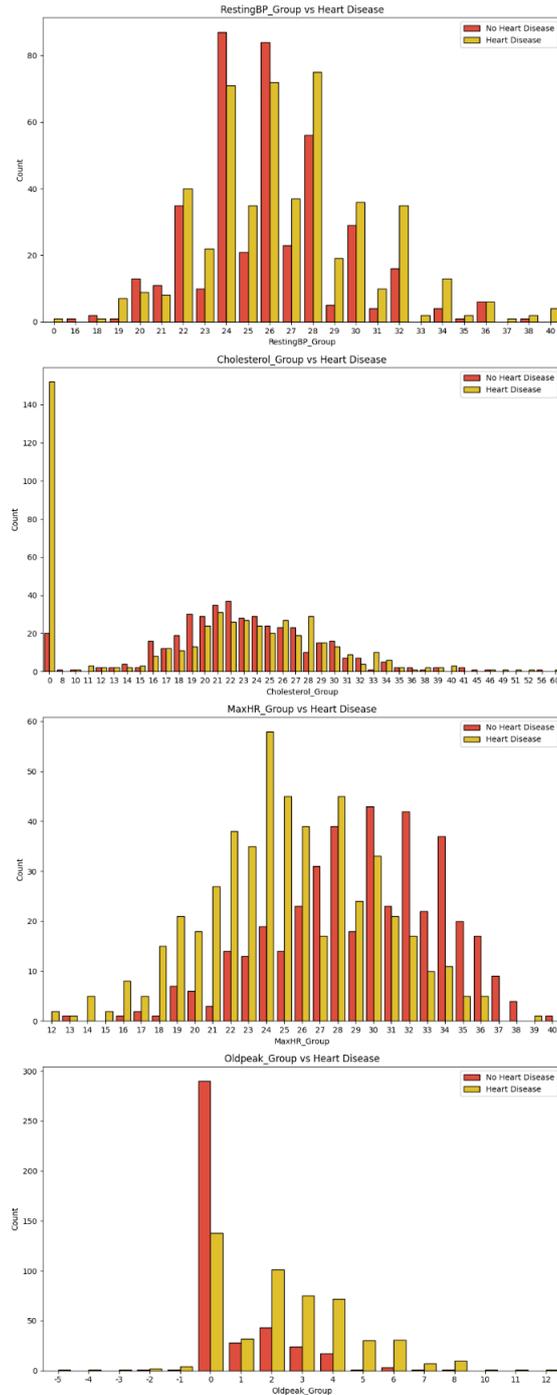
1. Laki-laki memiliki lebih banyak pasien penyakit jantung dibandingkan yang tidak memiliki penyakit jantung. Dalam kasus populasi perempuan, pasien penyakit jantung lebih sedikit dibandingkan yang tidak memiliki penyakit jantung.
2. Jenis nyeri dada ASY secara tegas menunjukkan kemungkinan besar adanya penyakit jantung
3. *Fasting Blood Sugar* cukup membingungkan. Pasien yang didiagnosis dengan *Fasting Blood Sugar* maupun yang tidak memiliki *Fasting Blood Sugar* memiliki jumlah pasien penyakit jantung yang signifikan.
4. *RestingECG* tidak menunjukkan kategori yang jelas yang menyoroti pasien penyakit jantung. Ketiga nilai tersebut memiliki jumlah pasien penyakit jantung yang tinggi.
5. *Exercise Induced Engina* Secara pasti meningkatkan kemungkinan didiagnosis dengan penyakit jantung.
6. Nilai *ST\_Slope*, *flat* menunjukkan kemungkinan yang sangat tinggi untuk didiagnosis dengan penyakit jantung. *ST\_Slope*, *down* menurun juga menunjukkan hasil yang sama, namun hanya pada beberapa titik data.



Gambar 4 Numerical Features vs Target Variable (HeartDisease)

Gambar 5 Categorical Features vs Target Variable (HeartDisease)

Analisa hubungan korelasi antara numerikal fitur terhadap variabel target. Dari gambar 5 terlihat terlalu banyak titik data unik pada fitur-fitur di atas, sulit untuk mendapatkan wawasan apa pun. Sehingga akan dilakukan konversi fitur numerik ini, kecuali Age menjadi fitur kategorikal untuk tujuan visualisasi yang lebih mudah dipahami dan mendapatkan wawasan. Untuk mengonversi fitur numerik menjadi bentuk yang lebih sederhana atau terkelompok dengan membagi nilai-nilai data (titik data) dengan angka tertentu, yaitu 5 atau 10. Hasil bagi dari pembagian ini (*quotient*) kemudian digunakan sebagai nilai representatif (konstanta) untuk data tersebut.



Gambar 6 Numerical\_Groups Features vs Target Variable (HeartDisease)

Proses ini dilakukan agar data lebih mudah dianalisis atau divisualisasikan. Pemilihan angka 5 dan 10 sebagai pembagi didasarkan pada analisis data dan intuisi, artinya angka tersebut dipilih agar dapat mengelompokkan data dengan cara yang lebih bermakna atau relevan dengan pola data yang ada. Dari Gambar 6 dapat ditarik kesimpulan bahwa:

1. Kelompok data RestingBP, terlihat dari rentang 95 (19x5) - 170 (34x5) paling rentan terdeteksi dengan penyakit jantung.
2. Tingkat kolesterol antara 160 (16x10) - 340 (34x10) sangat rentan terhadap penyakit jantung.
3. Untuk data MaxHR, penyakit jantung ditemukan di seluruh data, tetapi nilai antara 70 (14x5) - 180 (36x5) mendeteksi banyak kasus.
4. Nilai Oldpeak juga menunjukkan penyakit jantung di seluruh data. Nilai slope antara 0 (0x5/10) - 4 (8x5/10) menunjukkan kemungkinan tinggi untuk didiagnosis dengan penyakit jantung.

Secara keseluruhan Urutan atau nilai fitur untuk kasus positif pada penyakit jantung adalah sebagai berikut:

**Categorical Features (Order) :**

Sex : Male > Female

ChestPainType : ASY > NAP > ATA > TA

FastingBS : ( FBS < 120 mg/dl ) > ( FBS > 120 mg/dl)

RestingECG : Normal > ST > LVH

ExerciseAngina : Angina > No Angina

ST\_Slope : Flat > Up > Down

**Numerical Features (Range) :**

Age : 50+

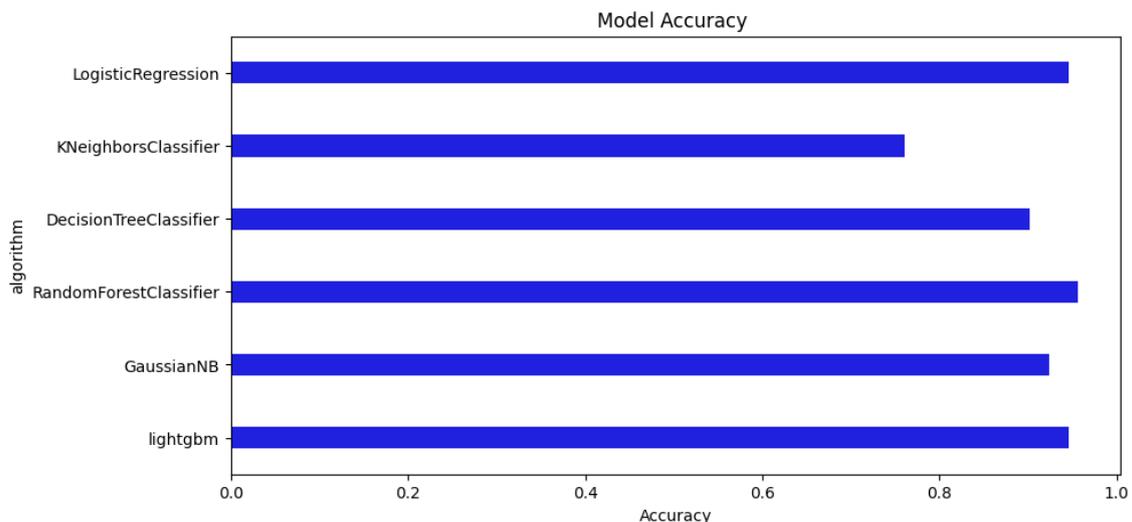
RestingBP : 95 - 170

Cholesterol : 160 - 340

MaxHR : 70 - 180

Oldpeak : 0 - 4

Setelah memahami nilai-nilai khas dari fitur-fitur tersebut, kita melanjutkan ke langkah berikutnya di mana kita memilih fitur-fitur yang sesuai untuk pemodelan. Setelah itu dilakukan proses untuk membangun model prediksi data dibagi menjadi 80% data *training* dan 20% sebagai data *testing*. Rasio antara data pelatihan (*training*) dan pengujian (*testing*) dalam pembelajaran mesin sangat penting karena memengaruhi kemampuan model dalam belajar dari data dan generalisasi terhadap data yang tidak terlihat sebelumnya [20].



Gambar 7 Barplot Hasil Pengujian Model

	algorithm object	Accuracy float64
0	LogisticRegression	0.9456521739
1	KNeighborsClassi...	0.7608695652
2	DecisionTreeClas...	0.902173913
3	RandomForestCla...	0.9565217391
4	GaussianNB	0.9239130435
5	lightgbm	0.9456521739

Gambar 8 Hasil Pengujian Model

Dari seluruh model yang dibangun akan dicari model terbaik berdasarkan hasil eksekusi. Hasil dari pengujian model dapat di lihat pada gambar 7 & 8. Studi komparatif dilakukan dengan menggunakan beberapa algoritma klasifikasi seperti *Logistic Regression* (LR), *K-Nearest Neighbors* (KNN), *Decision Tree* (DT), *RandomForest* (RF), *GaussianNB*, dan *LightGBM* untuk membangun model prediksi gagal jantung. Hasil eksperimen menunjukkan bahwa model yang dibangun menggunakan algoritma *Random Forest* memberikan kinerja terbaik dengan nilai evaluasi sebesar 95 persen. Sementara itu, model dengan akurasi terendah adalah KNN dengan nilai 76 persen, diikuti oleh *Decision Tree* dengan 90 persen, *GaussianNB* sebesar 92 persen, dan *LightGBM* sebesar 94 persen.

#### 4. KESIMPULAN DAN SARAN

Dalam studi ini, dilakukan upaya untuk memprediksi kemungkinan terjadinya penyakit jantung dengan menggunakan berbagai model klasifikasi, bertujuan untuk menentukan apakah seseorang berisiko terkena penyakit jantung berdasarkan parameter input seperti jenis kelamin, usia, dan hasil tes kesehatan lainnya. Dataset ini sangat baik untuk memahami cara menangani masalah klasifikasi biner dengan kombinasi fitur numerik dan kategorikal, di mana klasifikasi biner merujuk pada dua kemungkinan hasil, yaitu "penyakit jantung" atau "tanpa penyakit jantung," dengan fitur-fitur yang membedakan hasil tersebut mencakup data numerik (misalnya usia & tekanan darah) dan kategorikal (misalnya jenis kelamin & jenis nyeri dada).

Deteksi *outlier* (nilai ekstrem) tidak dilakukan karena peneliti tidak menemukan referensi terkait penyakit jantung. Memahami topik secara mendalam menjadi hal penting sebelum memutuskan untuk menghapus *outlier*, meskipun tes deteksi outlier menunjukkan hasil positif (yaitu, adanya data yang sangat berbeda dari nilai lainnya). Visualisasi sangat penting. Untuk pemodelan, penggunaan hiperparameter (*hyperparameter tuning*) tidak dilakukan. Padahal, penalaan *hyperparameter* bisa meningkatkan kinerja algoritma. Meskipun demikian, kinerja algoritma yang digunakan sudah cukup baik meskipun tanpa menerapkan *hyperparameter*.

#### DAFTAR PUSTAKA

- [1] World Health Organization, "Cardiovascular diseases (CVDs)," [https://www.who.int/en/news-room/fact-sheets/detail/cardiovascular-diseases-\(cvds\)](https://www.who.int/en/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)). [Accessed: Oct. 1, 2024].
- [2] F. S. Ahmad, H. Ning, J. D. Rich, C. W. Yancy, D. M. Lloyd-Jones, and J. T. Wilkins, "Hypertension, Obesity, Diabetes, and Heart Failure-Free Survival The Cardiovascular Disease Lifetime Risk Pooling Project," 2016.

- [3] H. B. Kibria and A. Matin, "The severity prediction of the binary and multi-class cardiovascular disease – A machine learning-based fusion approach," *Comput Biol Chem*, vol. 98, p. 107672, 2022, doi: <https://doi.org/10.1016/j.compbiochem.2022.107672>.
- [4] B. Bozkurt *et al.*, "Universal definition and classification of heart failure: a report of the Heart Failure Society of America, Heart Failure Association of the European Society of Cardiology, Japanese Heart Failure Society and Writing Committee of the Universal Definition of Heart Failure: Endorsed by the Canadian Heart Failure Society, Heart Failure Association of India, Cardiac Society of Australia and New Zealand, and Chinese Heart Failure Association," *Eur J Heart Fail*, vol. 23, no. 3, pp. 352–380, Mar. 2021, doi: 10.1002/ejhf.2115.
- [5] V. L. Roger, "Epidemiology of Heart Failure," *Circ Res*, vol. 128, no. 10, pp. 1421–1434, 2021, doi: 10.1161/CIRCRESAHA.121.318172.
- [6] P. A. Heidenreich *et al.*, "2022 AHA/ACC/HFSA Guideline for the Management of Heart Failure: A Report of the American College of Cardiology/American Heart Association Joint Committee on Clinical Practice Guidelines," *Circulation*, vol. 145, no. 18, pp. e895–e1032, 2022, doi: 10.1161/CIR.0000000000001063.
- [7] H. A. Taher and A. M. Abdulazeez, "Machine Learning Approaches for Heart Disease Detection: A Comprehensive Review," 2023.
- [8] L. Breiman, "Random Forests," *Mach Learn*, vol. 45, no. 1, pp. 5–32, 2001, doi: 10.1023/A:1010933404324.
- [9] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, in KDD '16. New York, NY, USA: Association for Computing Machinery, 2016, pp. 785–794. doi: 10.1145/2939672.2939785.
- [10] G. A. Ansari, S. S. Bhat, M. D. Ansari, S. Ahmad, J. Nazeer, and A. E. M. Eljialy, "Performance Evaluation of Machine Learning Techniques (MLT) for Heart Disease Prediction," *Comput Math Methods Med*, vol. 2023, 2023, doi: 10.1155/2023/8191261.
- [11] A. H. Shaker, I. A. Ibrahim, and S. K. Gharghan, "Cardiovascular diseases prediction using machine learning algorithms: A comparative study," *AIP Conf Proc*, vol. 3232, no. 1, p. 040004, Oct. 2024, doi: 10.1063/5.0236259.
- [12] R. Alizadehsani *et al.*, "A data mining approach for diagnosis of coronary artery disease," *Comput Methods Programs Biomed*, vol. 111, no. 1, pp. 52–61, 2013, doi: <https://doi.org/10.1016/j.cmpb.2013.03.004>.
- [13] A. Rajkomar *et al.*, "Scalable and accurate deep learning with electronic health records," *NPJ Digit Med*, vol. 1, no. 1, p. 18, 2018, doi: 10.1038/s41746-018-0029-1.
- [14] J. K. W *et al.*, "Artificial Intelligence in Cardiology," *J Am Coll Cardiol*, vol. 71, no. 23, pp. 2668–2679, Jun. 2018, doi: 10.1016/j.jacc.2018.03.521.
- [15] S. Okuboyejo, J. Haqbeen, and T. Ito, "Evaluating the Performance of Machine Learning Classifiers on Predicting Hypothyroidism for Public Healthcare Good," *IIAI Letters on Informatics and Interdisciplinary Research*, vol. 4, p. 1, 2023, doi: 10.52731/liir.v004.166.
- [16] M. Kuhn, "Applied predictive modeling," 2013, *Springer*.
- [17] D. Normawati and D. P. Ismi, "K-Fold Cross Validation for Selection of Cardiovascular Disease Diagnosis Features by Applying Rule-Based Datamining," *Signal and Image Processing Letters*, vol. 1, no. 2, pp. 23–35, Jul. 2019, doi: 10.31763/simple.v1i2.3.
- [18] I. Nti, O. Nyarko-Boateng, and J. Aning, "Performance of Machine Learning Algorithms with Different K Values in K-fold Cross-Validation," *International Journal of Information Technology and Computer Science*, vol. 6, pp. 61–71, Oct. 2021, doi: 10.5815/ijitcs.2021.06.05.
- [19] J. Gareth, W. Daniela, H. Trevor, and T. Robert, *An introduction to statistical learning: with applications in R*. New York: Springer, 2013.

- [20] S. Gonçalves Ivo and Silva, “Balancing Learning and Overfitting in Genetic Programming with Interleaved Sampling of Training Data,” in *Genetic Programming*, A. and H. T. and E.-U. A. Ş. and H. Bin Krawiec Krzysztof and Moraglio, Ed., Berlin, Heidelberg: Springer Berlin Heidelberg, 2013, pp. 73–84.