

Identifikasi Faktor Utama Penyebab Serangan Jantung Menggunakan Algoritma Logistic Regression

Fidinil Hikmah¹, Haerunisah², Indah Afryani³

¹Ilmu Komputer, Universitas Muhammadiyah Bima, Indonesia

²Ilmu Komputer, Universitas Muhammadiyah Bima, Indonesia

³Ilmu Komputer, Universitas Muhammadiyah Bima, Indonesia

Article Info

Article history:

Received 18-06-2025

Revised 29-06-2025

Accepted 30-06-2025

Kata Kunci:

Serangan jantung
Logistic Regression
Klasifikasi
Troponin
CK-MB

Abstrak

Penelitian ini bertujuan untuk mengidentifikasi faktor utama penyebab serangan jantung dengan menggunakan algoritma *Logistic Regression* sebagai metode klasifikasi. Data yang digunakan berasal dari dataset medis yang mencakup sejumlah variabel penting seperti usia, jenis kelamin, tekanan darah sistolik dan diastolik, kadar gula darah, serta biomarker jantung seperti CK-MB dan troponin. Model dikembangkan untuk memprediksi kemungkinan terjadinya serangan jantung dengan membagi data ke dalam dua kelas: positif (mengalami serangan jantung) dan negatif (tidak mengalami serangan jantung). Hasil evaluasi menunjukkan bahwa model memiliki akurasi sebesar 81,78% dengan nilai F1-score yang seimbang, menandakan performa yang konsisten dalam mengenali kedua kelas. Temuan ini memperkuat bahwa algoritma *Logistic Regression* tidak hanya efektif dalam klasifikasi biner, tetapi juga mampu mengidentifikasi variabel-variabel yang paling signifikan dalam memengaruhi risiko. Hasil akhir diharapkan dapat memberikan kontribusi nyata dalam pengembangan sistem pendukung keputusan medis yang lebih cepat, akurat, dan berbasis data.

This is an open access article under the [CC BY 4.0](https://creativecommons.org/licenses/by/4.0/) license.



Penulis Korespondensi:

Fidinil Hikmah
Computer Science, Faculty of Engineering and Computer Science
Universitas Muhammadiyah Bima
43400, Sarae, Bima, Indonesia
Email: fidinil@gmail.com

1. PENDAHULUAN

Penyakit jantung merupakan salah satu penyebab kematian utama di dunia, termasuk di Indonesia. Serangan jantung terjadi ketika aliran darah ke jantung terganggu, yang dapat menyebabkan kerusakan permanen pada jaringan jantung atau bahkan kematian mendadak[1]. Faktor risiko seperti hipertensi, kolesterol tinggi, obesitas, diabetes, merokok, kurang olahraga, dan riwayat keluarga berperan besar dalam meningkatkan kemungkinan terjadinya serangan jantung. Deteksi dini terhadap risiko serangan jantung menjadi sangat penting untuk mencegah kejadian fatal dan meningkatkan peluang penyelamatan pasien. Oleh karena itu, pengembangan sistem prediksi yang mampu mengidentifikasi potensi serangan jantung secara akurat sangat dibutuhkan dalam bidang medis, terutama dalam mempercepat diagnosis dan pengambilan keputusan.

Dalam dunia teknologi informasi, penerapan machine learning telah memberikan kontribusi besar dalam bidang prediksi dan klasifikasi berbagai penyakit, termasuk penyakit jantung. Machine learning memungkinkan komputer untuk belajar dari data dan mengenali pola yang sulit diidentifikasi secara manual. Salah satu pendekatan yang cukup populer adalah penggunaan algoritma *Logistic Regression*, yang mampu mengklasifikasikan data ke dalam dua kelas seperti positif (mengalami serangan jantung) dan negatif (tidak mengalami serangan jantung)[2]. *Logistic Regression* adalah algoritma statistik yang sering digunakan dalam klasifikasi biner karena kemampuannya untuk memodelkan probabilitas kejadian suatu peristiwa berdasarkan variabel input. Selain itu, algoritma ini dikenal karena kesederhanaannya, efektivitas dalam analisis data dengan jumlah besar, serta interpretasi koefisiennya yang mudah dipahami oleh tenaga medis, sehingga dapat menjembatani analisis teknis dengan keputusan klinis.

Beberapa penelitian sebelumnya telah menggunakan *Logistic Regression* dalam klasifikasi penyakit jantung dengan hasil cukup menjanjikan. Misalnya, Latifah et al melaporkan akurasi model mencapai 85,04% pada dataset Framingham dengan fitur seperti tekanan darah dan kolesterol[3]. Selain itu, Nasution et al. (2025) melaporkan akurasi 84,2% (tanpa seleksi fitur) dan 82,6% (dengan *mutual information*) pada dataset UCI, menunjukkan bahwa kualitas data dan pemilihan atribut sangat mempengaruhi performa model[4].

Penelitian ini bertujuan untuk membangun model klasifikasi penyakit serangan jantung dengan menggunakan algoritma *Logistic Regression* berdasarkan data rekam medis pasien. Data yang digunakan meliputi berbagai fitur penting seperti usia, jenis kelamin, tekanan darah sistolik dan diastolik, kadar gula darah, biomarker jantung seperti CK-MB dan Troponin, serta hasil diagnosis medis. Klasifikasi dilakukan dalam dua kelas, yaitu positif (mengalami serangan jantung) dan negatif (tidak mengalami serangan jantung). Tahapan penelitian meliputi pengumpulan data, pembersihan data, analisis eksploratif, pengkodean variabel kategorikal, pembagian data menjadi training dan testing, pembuatan model, dan evaluasi performa model. Penelitian serupa sudah banyak dilakukan, misalnya studi Anshori & Haris yang menggunakan *Logistic Regression* dengan fitur usia, tekanan darah sistolik dan diastolik, kadar gula darah, CK-MB, dan Troponin, serta variabel kategorikal seperti jenis kelamin, menghasilkan akurasi yang tinggi dalam diagnosis serangan jantung positif vs negatif[5].

Evaluasi model dilakukan menggunakan berbagai metrik seperti akurasi, presisi, recall, dan F1-score untuk mengukur performa klasifikasi secara menyeluruh. Hasil pengujian menunjukkan bahwa model *Logistic Regression* mampu mencapai akurasi tinggi, serta memiliki nilai F1-score yang seimbang untuk kedua kelas. Hal ini menunjukkan bahwa model memiliki performa yang baik, tidak hanya dalam mengenali pasien yang positif terkena serangan jantung, tetapi juga dalam menghindari kesalahan klasifikasi pada pasien yang negatif. Rokoni et al menggunakan LIME pada model *Logistic Regression* untuk menegaskan bahwa troponin dan CK-MB termasuk fitur paling berpengaruh terhadap risiko serangan jantung[6]. sehingga model ini memberikan interpretasi yang jelas tentang pengaruh

masing-masing variabel terhadap probabilitas terjadinya serangan jantung, seperti peningkatan kadar troponin dan CK-MB yang menunjukkan risiko tinggi, atau tekanan darah tinggi yang memperbesar kemungkinan diagnosis positif.

Melalui penelitian ini, diharapkan dapat memberikan kontribusi dalam pengembangan sistem pendukung keputusan medis yang efektif, khususnya dalam mendeteksi risiko serangan jantung. Model ini dapat digunakan oleh praktisi medis sebagai alat bantu dalam proses skrining awal pasien, sehingga dapat mempercepat proses diagnosis dan meningkatkan efisiensi dalam pengelolaan kasus kardiovaskular. Selain itu, hasil penelitian ini juga dapat menjadi referensi bagi pengembangan metode klasifikasi lainnya di bidang kesehatan, serta mendorong integrasi yang lebih luas antara teknologi informasi dan dunia medis dalam meningkatkan kualitas layanan kesehatan masyarakat. Dengan adanya model prediktif berbasis *machine learning*, diharapkan pengambilan keputusan medis menjadi lebih berbasis data dan lebih tepat sasaran dalam upaya menurunkan angka kematian akibat serangan jantung.

2. LANDASAN TEORI

2.1 Serangan Jantung (Heart Attack)

Serangan jantung adalah kondisi yang sangat berbahaya di mana aliran darah ke otot jantung terhambat secara tiba-tiba[7]. Hal ini disebabkan terjadinya penyumbatan karena dinding arteri koroner tertutup oleh plak lemak, kolesterol, dan zat lain. Plak ini menghambat pasokan oksigen ke jaringan jantung. Jika tidak ditangani segera, kondisi ini dapat menyebabkan kerusakan permanen pada otot jantung. Gejala umum serangan jantung adalah nyeri dada yang menjalar ke lengan atau rahang, sesak napas, mual, keringat dingin, dan lemas secara tiba-tiba.

2.2 Faktor Risiko Penyebab Serangan Jantung

Usia, jenis kelamin, kondisi medis, gaya hidup, dan genetik adalah beberapa faktor penting yang meningkatkan risiko terkenanya serangan jantung. Karena perubahan yang terjadi pada sistem kardiovaskular seseorang, risiko meningkat seiring bertambahnya usia[8]. Pria memiliki risiko yang lebih tinggi pada usia muda, sementara wanita risikonya meningkat setelah menopause. Penumpukan plak dan kerusakan arteri dapat disebabkan oleh hipertensi dan kolesterol tinggi, terutama LDL (Low-Density Lipoprotein). Risiko juga meningkat karena diabetes merusak pembuluh darah, serta Gaya hidup yang tidak sehat seperti kebiasaan merokok dapat mempersempit dan meningkatkan kemungkinan penggumpalan pembuluh darah[9]. Serta riwayat penyakit jantung dalam keluarga (genetik) juga berpengaruh memicu serangan jantung, bahkan pada orang yang menjalani gaya hidup sehat.

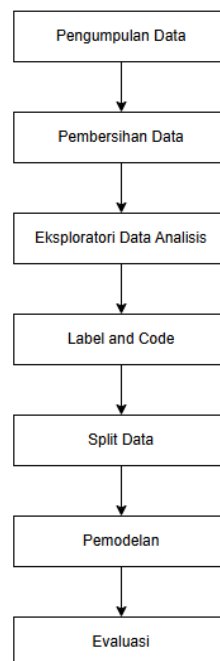
2.3 Algoritma Logistic Regression

Logistic regression merupakan algoritma dalam metode klasifikasi yang digunakan ketika variabel dependen memiliki dua kategori, seperti “benar” atau “salah”, “positif” atau “negatif”. *Logistic regression* biasa digunakan untuk memperkirakan kemungkinan suatu peristiwa terjadi berdasarkan satu atau lebih variabel bebas, yang dapat berupa data kategorikal atau numerik[10]. Algoritma ini banyak dimanfaatkan di bidang kesehatan, khususnya dalam memprediksi diagnosis serta menilai tingkat risiko terhadap suatu penyakit.

3. METHOD

Penelitian ini menggunakan metode klasifikasi. klasifikasi bekerja dengan mengelompokkan data ke dalam kategori tertentu sesuai dengan kriteria atau batasan yang telah ditetapkan sebelumnya[11]. Salah satu algoritma klasifikasi yang digunakan adalah *Logistic Regression* yang di mana algoritma ini membuat model prediksi berbasis probabilitas, dengan memetakan nilai variabel input ke dalam probabilitas keluaran kelas tertentu, dan probabilitas ini kemudian digunakan untuk menentukan kelas akhir data, seperti tingkat kemungkinan seseorang mengalami serangan jantung.

Tahapan-tahapan analisis data dalam metode klasifikasi meliputi:



Gambar 1. Alur Penelitian

1. **Pengumpulan Data:**
Data dikumpulkan dari dataset yang dapat dipercaya dan mencakup informasi seperti usia, jenis kelamin, tekanan darah, kadar gula darah, kolesterol, dan biomarker jantung seperti CK-MB dan Troponin.
2. **Pembersihan Data (Data Cleaning)**
Proses pembersihan dilakukan setelah data diperoleh untuk memastikan kualitas data yang digunakan. Penanganan nilai yang hilang (nilai yang tidak ada), penghapusan duplikasi, dan koreksi nilai yang tidak konsisten atau ekstrim termasuk dalam proses ini. Untuk menghindari kesalahan selama proses analisis dan memastikan bahwa hasil model yang diperoleh benar.
3. **Analisis Data Eksploratori (EDA)**
Analisis data dilakukan untuk memahami karakteristik dan hubungan antar variabel. Dengan menggunakan visualisasi seperti histogram, scatter plot, dan heatmap korelasi, untuk melihat pola distribusi data, menemukan anomali, dan menemukan variabel yang berpotensi meningkatkan risiko serangan jantung.
4. **Pemberian Label dan Pengkodean (Label and Code)**
Variabel kategorik seperti jenis kelamin diubah menjadi format numerik sehingga algoritma *Logistic Regression* dapat memprosesnya. Selain itu, variabel target termasuk dalam dua kategori, 0 menunjukkan kondisi yang negatif (tidak mengalami serangan jantung), dan 1 menunjukkan kondisi yang positif (mengalami serangan jantung). Kemudian melakukan pengkodean menggunakan python untuk keperluan untuk mengubah data dalam variabel yang non numerik ke numerik dan lain sebagainya.
5. **Pembagian Data (Split Data)**
Data yang telah dibersihkan dan dikodekan kemudian dibagi menjadi dua bagian yaitu Data Training dan Data Testing, dengan perbandingan umum 80:20. Tujuannya adalah untuk mengetahui seberapa baik model dapat memprediksi data baru atau data nyata di luar data pelatihan. Hal ini penting untuk mengevaluasi kemampuan model dalam membuat prediksi yang

akurat terhadap data baru, serta untuk menghindari overfitting, sehingga performa model benar-benar mencerminkan kemampuannya secara objektif.

6. Pemodelan (Modeling)

Proses pelatihan model dilakukan dengan membangun model dari Data Training dan kemudian diuji dengan Data Testing. Hasil evaluasi performa model termasuk metrik seperti akurasi, presisi, recall, dan skor F1, yang menunjukkan kemampuan model untuk memprediksi risiko serangan jantung dengan tepat.

7. Evaluasi

Proses evaluasi dilakukan untuk mengukur kinerja model pada data uji dengan menggunakan metrik seperti akurasi atau MSE. Ini memastikan bahwa model mampu bekerja dengan baik pada data baru dan tidak mengalami overfitting.

4. HASIL DAN PEMBAHASAN

4.1 Deskripsi Dataset

Data yang digunakan pada penelitian ini bersumber dari website Kaggle.com. heart attack (serangan jantung) yang terdiri dari 9 variabel dan 1319 baris data.

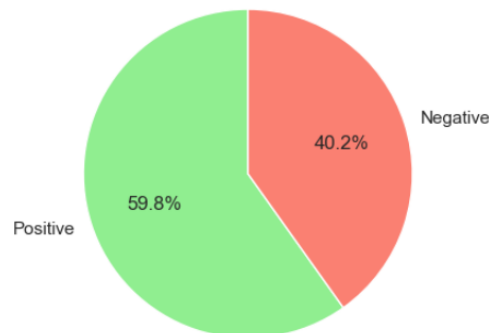
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 1319 entries, 0 to 1318
Data columns (total 9 columns):
#   Column                                Non-Null Count  Dtype
---  ---                                ---
0   Age                                  1319 non-null   int64
1   Gender                              1319 non-null   int64
2   Heart Rate                          1319 non-null   int64
3   Systolic Blood Pressure             1319 non-null   int64
4   Diastolic Blood Pressure            1319 non-null   int64
5   Blood Glucose                      1319 non-null   float64
6   CK-MB                              1319 non-null   float64
7   Troponin                            1319 non-null   float64
8   Result                             1319 non-null   object
dtypes: float64(3), int64(5), object(1)
memory usage: 92.9+ KB
```

Gambar 2. Variabel dataset

Setelah dilakukan pembersihan data yang awalnya berjumlah 1319 menjadi 1178 baris data. Variable dataset terdiri dari: Age, Gender, Heart Rate, Systolic Blood Pressure, Diastolic Blood Pressure, Blood Glucose, CK-MB, Troponin dan Result.

- Variabel Age: pada variabel ini diisi dengan usia pasien.
- Variabel Gender: pada variabel ini terdiri dari dua kategori gender yakni laki-laki (1) berjumlah 773 orang dan perempuan (0) berjumlah 405 orang.
- Variabel Heart Rate: Heart Rate (HR) atau denyut jantung adalah jumlah detakan jantung per menit (bpm - beats per minute). Variabel ini mencerminkan seberapa sering jantung memompa darah ke seluruh tubuh.
- Variabel Systolic Blood Pressure: Tekanan darah sistolik adalah tekanan maksimum dalam arteri ketika jantung berkontraksi (fase sistol). Ini adalah angka pertama atau atas dalam pembacaan tekanan darah.
- Variabel Diastolic Blood Pressure: Tekanan darah diastolik adalah tekanan dalam arteri saat jantung dalam fase relaksasi (diastol), yaitu ketika darah mengisi jantung. Ini adalah angka kedua atau bawah dalam pembacaan tekanan darah.
- Variabel Blood Glucose: Blood glucose atau kadar gula darah adalah konsentrasi glukosa dalam darah. Glukosa merupakan sumber energi utama bagi tubuh.
- Variabel CK-MB: Creatine Kinase-MB (CK-MB) adalah enzim yang dilepaskan oleh otot jantung saat terjadi kerusakan miokard, seperti pada infark miokard (serangan jantung).
- Variabel Troponin: Troponin (T atau I) adalah protein spesifik jantung yang dilepaskan ke darah saat terjadi kerusakan pada otot jantung.
- Variabel Result: Variabel Result berisi label klasifikasi atau hasil akhir diagnosis dalam suatu dataset. Pada data ini yang positive adalah 1 dan yang negative adalah 0.

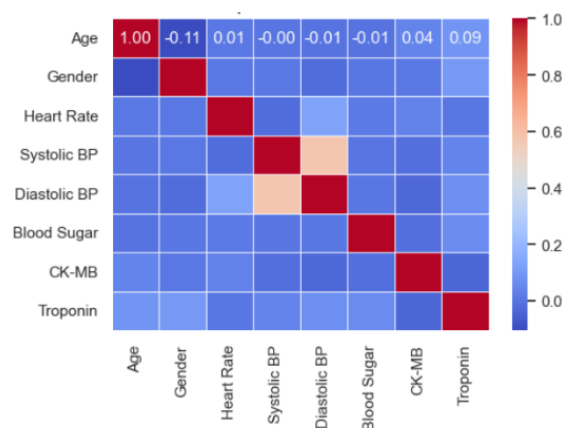
Pada dataset ini class/label yang digunakan ada 2 yakni positive dan negative yang dimana yang positive menunjukkan bahwa pasien mengalami serangan jantung, dan yang negative menunjukkan bahwa pasien tidak mengalami serangan jantung. Pada klasifikasi dataset medis class/label positive dan negative sangat umum dipake untuk mendeskripsikan terkena atau tidaknya pasien pada suatu penyakit.



Gambar 3. Distribusi Hasil Pemeriksaan Jantung

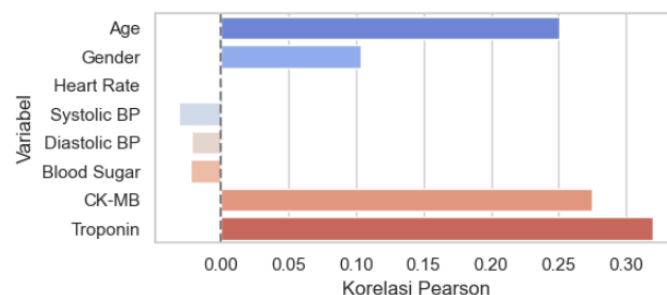
Gambar tersebut menunjukkan diagram lingkaran hasil analisis data heart attack, di mana 59,8% data yang tergolong positif terkena serangan jantung dan 40,2% negative yang tidak terkena serangan jantung.

4.2 Analisis Data Eksploratif (Exploratory Data Analysis)



Gambar 4. Korelasi Antar Variabel

Gambar diatas adalah heatmap korelasi, yang digunakan untuk menunjukkan hubungan antar variabel numerik dalam sebuah dataset. Menampilkan nama-nama variabel seperti: Age, Gender, Heart Rate, Systolic BP, Diastolic BP, Blood Sugar, CK-MB, dan Troponin. Nilai korelasi berkisar antara -1 hingga 1, yang ditampilkan dalam bentuk warna. Warna merah tua menunjukkan korelasi positif yang kuat (mendekati +1), sedangkan biru tua menunjukkan korelasi negatif atau tidak adanya korelasi (mendekati 0 atau -1). Warna yang lebih terang menandakan nilai korelasi mendekati nol, yang berarti tidak terdapat hubungan linear yang signifikan antar variabel.

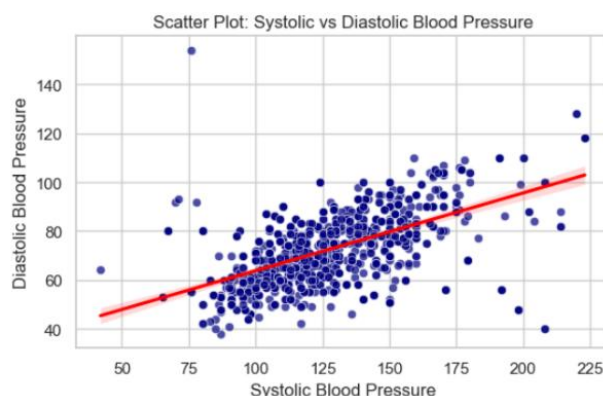


Gambar 5. Korelasi Setiap Variabel dengan Hasil

Gambar 5. adalah diagram batang horizontal yang menampilkan nilai korelasi Pearson antara beberapa variabel dengan suatu target (label penyakit serangan jantung). Sumbu X (Horizontal): Menampilkan nilai korelasi Pearson (0.00–0.30), yang menunjukkan seberapa kuat hubungan linier antara masing-masing variabel dan target. Warna biru cenderung digunakan untuk korelasi yang lebih rendah. Warna merah menunjukkan nilai korelasi yang lebih tinggi.

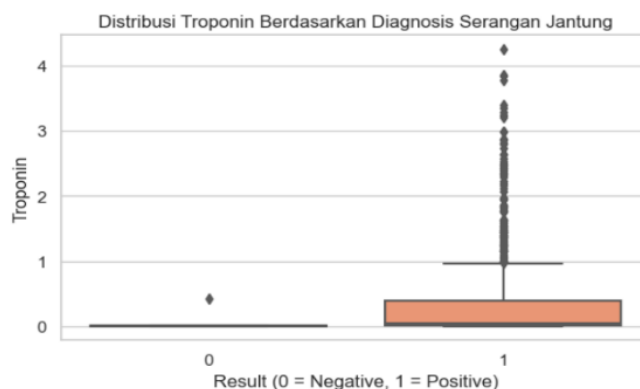
Interpretasi:

- Troponin memiliki nilai korelasi tertinggi (~0.30), diikuti oleh CK-MB (~0.25). Ini mengindikasikan bahwa kedua variabel ini memiliki hubungan yang cukup kuat terhadap target, kemungkinan besar karena keduanya merupakan biomarker penting dalam diagnosis penyakit jantung.
- Blood Sugar juga menunjukkan korelasi yang relatif lebih tinggi dibanding variabel lainnya, meskipun tidak sekuat Troponin atau CK-MB.
- Variabel seperti Age dan Gender memiliki korelasi lebih rendah, meskipun Age masih menunjukkan nilai mendekati 0.20.
- Systolic BP, Diastolic BP, dan Heart Rate memiliki nilai korelasi yang sangat kecil, menunjukkan hubungan yang lemah terhadap target.



Gambar 6. Korelasi variable Systolic dan Diastolic

Gambar 6. Menunjukkan bahwa terdapat hubungan positif antara tekanan darah sistolik dan diastolik, di mana semakin tinggi tekanan sistolik, maka tekanan diastolik cenderung ikut meningkat. Garis regresi memperkuat adanya korelasi antar keduanya. Korelasi antara tekanan darah sistolik dan diastolik yang tinggi berkaitan erat dengan risiko penyakit jantung. Tekanan darah yang tinggi (hipertensi) merupakan salah satu faktor utama yang meningkatkan beban kerja jantung, merusak pembuluh darah, dan berkontribusi pada munculnya penyakit jantung seperti gagal jantung, serangan jantung, dan stroke.



Gambar 7. Distribusi Variable Troponim

Gambar diatas adalah boxplot yang menunjukkan distribusi kadar Troponin berdasarkan diagnosis serangan jantung. Pasien dengan hasil diagnosis 0 (Negatif) mayoritas memiliki kadar troponin sangat rendah (mendekati 0), hampir tidak ada nilai yang tinggi, menunjukkan tidak ada kerusakan otot jantung. Sedangkan pasien dengan hasil diagnosis 1 (Positif) Menunjukkan kadar troponin yang jauh lebih tinggi. Boxplot memperlihatkan median yang lebih tinggi dan banyak outlier (nilai ekstrem ke atas), mencerminkan kerusakan otot jantung.

4.3 Pembagian Data (Data Splitting)

Tabel 1. Split Data

Data	Record	Variabel
Train set	942	9
Test set	236	9

Tabel di atas menunjukkan hasil proses pembagian dataset menjadi dua bagian utama: train set (data latih) dan test set (data uji). Proses ini penting dalam pengembangan model pembelajaran mesin agar model dapat belajar dari satu bagian data dan kemudian dievaluasi pada bagian lain yang tidak pernah dilihat sebelumnya. Yang dimana Train set (data latih) berisi 942 record (baris data), dan Test set (data uji) berisi 236 record.

4.4 Evaluasi Kinerja Model (Model Evaluation)

```

Confusion Matrix:
[[ 83  15]
 [ 28 110]]

Accuracy Score:
0.8177966101694916

Classification Report:
              precision    recall  f1-score   support

     0       0.75      0.85      0.79       98
     1       0.88      0.80      0.84      138

 accuracy          0.82       236
  macro avg       0.81      0.82      0.82       236
 weighted avg     0.83      0.82      0.82       236

```

Gambar 8. Akurasi data

Model klasifikasi yang diuji menunjukkan performa yang baik dengan akurasi sebesar 81.78%. Nilai F1-score yang seimbang pada kedua kelas (0 dan 1) menandakan model cukup andal dalam mengenali

baik kasus negatif maupun positif. Hasil ini menunjukkan bahwa model layak digunakan dalam skenario klasifikasi medis dengan dua kategori diagnosis.

5. KESIMPULAN

Logistic Regression dalam metode klasifikasi dapat digunakan secara efektif untuk memodelkan variabel dependen biner dan memberikan interpretasi yang jelas tentang bagaimana masing-masing faktor berpengaruh terhadap risiko serangan jantung berdasarkan data klinis pasien, yang terdiri dari data seperti usia, jenis kelamin, tekanan darah, kadar gula darah, biomarker seperti troponin, CK-MB, dan diagnosis positif atau negatif terhadap kondisi tersebut. Selain efektif dalam klasifikasi, *Logistic Regression* juga memiliki keunggulan dalam memberikan prediksi probabilitas terhadap kemungkinan terjadinya serangan jantung, serta memungkinkan untuk mengidentifikasi variabel-variabel mana yang paling signifikan memengaruhi hasil diagnosis. Koefisien dari model ini dapat dianalisis untuk memahami arah dan kekuatan pengaruh dari masing-masing faktor, misalnya apakah peningkatan kadar troponin secara signifikan meningkatkan risiko, atau apakah tekanan darah tinggi berperan besar dalam klasifikasi positif. Dengan demikian, algoritma ini sangat berguna dalam mendukung proses pengambilan keputusan klinis, terutama dalam tahap awal diagnosis dan manajemen risiko pasien. Penerapannya dalam sistem prediksi medis dapat membantu tenaga kesehatan dalam mengambil tindakan pencegahan atau penanganan yang lebih tepat dan cepat berdasarkan data yang relevan.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada dosen pembimbing atas bimbingan dan arahnya selama proses penyusunan penelitian ini. Terima kasih juga disampaikan kepada rekan-rekan mahasiswa yang telah memberikan masukan dan bantuan teknis dalam pengolahan data. Selain itu, penulis menghargai kontribusi dari pihak-pihak yang telah menyediakan dataset yang digunakan dalam penelitian ini. Seluruh dukungan yang diberikan sangat berarti dalam menyelesaikan karya ini.

INFORMASI PENDANAAN

Penulis menyatakan bahwa tidak ada pendanaan yang terlibat dalam pelaksanaan dan penulisan penelitian ini.

PERNYATAAN KONTRIBUSI PENULIS

Fidinil Hikmah berkontribusi dalam konseptualisasi penelitian, pengumpulan data, dan penulisan draf awal. Haerunisah bertanggung jawab dalam pengolahan data, analisis formal, serta visualisasi hasil. Indah Afryani melakukan peninjauan dan pengeditan naskah akhir, serta bertindak sebagai penulis korespondensi yang mengatur seluruh proses pengiriman, revisi, dan publikasi.

PERNYATAAN BENTURAN KEPENTINGAN

Penulis menyatakan tidak ada konflik kepentingan yang dapat memengaruhi hasil dan interpretasi dari penelitian ini.

REFERENSI

- [1] D. Sitanggang, N. Nicholas, V. Wilson, A. R. A. Sinaga, and A. D. Simanjuntak, "Implementasi Data Mining Untuk Memprediksi Penyakit Jantung Menggunakan Metode K-Nearest Neighbor Dan Logistic Regression," *J. Tek. Inf. dan Komput.*, vol. 5, no. 2, p. 493, 2022, doi: 10.37600/tekinkom.v5i2.698.
- [2] J. Junifer Pangaribuan, H. Tanjaya, and K. Kenichi, "Mendeteksi Penyakit Jantung Menggunakan Machine Learning Dengan Algoritma Logistic Regression," *J. Inf. Syst. Dev.*, vol. 06, no. 02, pp. 1–10, 2021.
- [3] F. A. Latifah, I. Slamet, and Sugiyanto, "Comparison of heart disease classification with logistic regression algorithm and random forest algorithm," *AIP Conf. Proc.*, vol. 2296, no. April, 2020, doi: 10.1063/5.0030579.
- [4] N. Nasution, F. B. Nasution, and M. A. Hasan, "Predicting Heart Disease Using Machine Learning : An Evaluation of Logistic Regression , Random Forest , SVM , and KNN Models on the UCI Heart Disease Dataset," vol. 9, no. 2, pp. 140–150, 2025.
- [5] M. Anshori and M. S. Haris, "Predicting Heart Disease using Logistic Regression," *Knowl. Eng. Data Sci.*, vol. 5, no. 2, p. 188, 2022, doi: 10.17977/um018v5i22022p188-196.
- [6] S. Rokoni and S. Bashar, "Heart Disease Prediction : Leveraging CNNs and," vol. 23, no. 1, pp. 52–57, 2025, doi: 10.31645/JISRC.25.23.1.6.
- [7] A. Khalisatifa, H. Dela Arum, and M. I. Jambak, "Klasifikasi Risiko Penyakit Serangan Jantung Dengan Menggunakan Algoritma C4.5," *Device*, vol. 14, no. 1, pp. 57–64, 2024, doi: 10.32699/device.v14i1.6869.
- [8] J. L. Rodgers *et al.*, "Cardiovascular risks associated with gender and aging," *J. Cardiovasc. Dev. Dis.*, vol. 6, no. 2, 2019, doi: 10.3390/jcdd6020019.
- [9] S. A. Nazli *et al.*, "Coronary risk factor profiles according to different age categories in premature coronary artery disease patients who have undergone percutaneous coronary intervention," *Sci. Rep.*, vol. 14, no. 1, pp. 1–13, 2024, doi: 10.1038/s41598-024-53539-6.
- [10] E. I. Scandea, M. Aqsha, R. Sugiarto, F. Lestari, and D. Hartanti, "Penerapan Data Mining Untuk Menganalisis Data Faktor Resiko Penyakit Jantung Menggunakan Metode Logistic Regression," pp. 683–688, 2023.
- [11] A. D. W. M. Sidik, I. Himawan Kusumah, A. Suryana, Edwinanto, M. Artiyasa, and A. Pradiftha Junfithrana, "Gambaran Umum Metode Klasifikasi Data Mining," *Fidel. J. Tek. Elektro*, vol. 2, no. 2, pp. 34–38, 2020, doi: 10.52005/fidelity.v2i2.111.