

**PERBANDINGAN KINERJA ALGORITMA LOGISTIC
REGRESSION, DECISION TREE, DAN RANDOM FOREST
DALAM KLASIFIKASI RISIKO PENYAKIT JANTUNG**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana

Program Studi S1 Informatika



disusun oleh

HABIB BAITUL HAMDİ

22.11.4571

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2026**

**PERBANDINGAN KINERJA ALGORITMA LOGISTIC
REGRESSION, DECISION TREE, DAN RANDOM FOREST
DALAM KLASIFIKASI RISIKO PENYAKIT JANTUNG**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana

Program Studi S1 Informatika



disusun oleh

HABIB BAITUL HAMDY

22.11.4571

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2026

HALAMAN PERSETUJUAN

SKRIPSI

PERBANDINGAN KINERJA ALGORITMA LOGISTIC REGRESSION, DECISION TREE, DAN RANDOM FOREST DALAM KLASIFIKASI RISIKO PENYAKIT JANTUNG

yang disusun dan diajukan oleh

HABIB BAITUL HAMDI

22.11.4571

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 27 Januari 2026

Dosen Pembimbing,



MULIA SULISTIYONO, M.Kom

NIK. 190302248

HALAMAN PENGESAHAN
SKRIPSI
PERBANDINGAN KINERJA ALGORITMA LOGISTIC REGRESSION,
DECISION TREE, DAN RANDOM FOREST DALAM KLASIFIKASI
RISIKO PENYAKIT JANTUNG

yang disusun dan diajukan oleh

HABIB BAITUL HAMDI
22.11.4571

Telah dipertahankan di depan Dewan Penguji
pada tanggal 27 Januari 2026

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

I Made Artha Agastya, S.T., M.Eng., ph.D.
NIK. 190302352



Ahlihi Masruro, S.Kom., M.Kom
NIK. 190302148



Mulia Sulistiyono, S.Kom., M.Kom
NIK. 190302248



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 27 Januari 2026

DEKAN FAKULTAS ILMU KOMPUTER



Prof. Dr. Kusriani, M.Kom.
NIK. 190302106

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : HABIB BAITUL HAMDI

NIM : 22.11.4571

Menyatakan bahwa Skripsi dengan judul berikut:

PERBANDINGAN KINERJA ALGORITMA LOGISTIC REGRESSION, DECISION TREE, DAN RANDOM FOREST DALAM KLASIFIKASI RISIKO PENYAKIT JANTUNG

Dosen Pembimbing : MULIA SULISTIYONO, M.Kom

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, 27 Januari 2026

Yang Menyatakan,


Habib Baitul Hamdi

HALAMAN PERSEMBAHAN

Sembah sujud serta syukur kepada Allah SWT atas segala nikmat, kekuatan, dan kemudahan yang telah diberikan, sehingga skripsi ini dapat terselesaikan dengan baik. Karya sederhana ini saya persembahkan untuk orang-orang istimewa dalam hidup saya:

1. Kedua Orang Tuaku Tercinta, Bapak dan Ibu. Terima kasih yang tak terhingga atas segala doa yang tiada putus, kasih sayang yang tulus, serta pengorbanan keringat dan air mata demi pendidikan saya. Skripsi ini adalah kado kecil bakti saya untuk membalas sedikit dari besarnya jasa kalian. Semoga ini menjadi awal kebahagiaan bagi kalian.
2. Teruntuk Diri Saya Sendiri, Habib Baitul Hamdi. Terima kasih karena tidak menyerah meski proses ini melelahkan. Terima kasih sudah kuat, sabar, dan terus berusaha menuntaskan apa yang sudah dimulai. *We made it.*
3. Teman-teman seperjuangan Zahran, Bagas, Evan, Anjas, Angga, Aji dan teman-teman lainnya yang mungkin tidak bisa saya sebutkan satu persatu, terima kasih atas diskusi, bantuan, kopi, dan canda tawa yang membuat masa-masa penyusunan skripsi ini menjadi lebih berwarna.

KATA PENGANTAR

Puji syukur ke hadirat Allah SWT yang telah melimpahkan rahmat, hidayah, dan karunia-Nya, sehingga penulis dapat menyelesaikan skripsi yang berjudul “Perbandingan Kinerja Algoritma Logistic Regression, Decision Tree, dan Random Forest dalam Klasifikasi Risiko Penyakit Jantung” dengan baik. Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer pada Program Studi Informatika, Fakultas Ilmu Komputer, Universitas AMIKOM Yogyakarta. Penulis menyadari bahwa dalam penyusunan skripsi ini tidak lepas dari bantuan, bimbingan, dan dukungan dari berbagai pihak. Oleh karena itu, pada kesempatan ini penulis ingin menyampaikan rasa terima kasih yang sebesar-besarnya kepada:

1. Prof. Dr. Mohammad Suryanto, M.M selaku Rektor Universitas Amikom Yogyakarta.
2. Prof. Dr. Kusriani, M.Kom selaku Dekan Fakultas Ilmu Komputer Universitas Amikom Yogyakarta.
3. Eli Pujastuti, M.kom selaku Ketua Program Studi Informatika Fakultas Ilmu Komputer Universitas Amikom Yogyakarta.
4. Bapak Mulia Sulistiyono, M.Kom., selaku Dosen Pembimbing yang telah meluangkan waktu, tenaga, dan pikiran untuk memberikan bimbingan serta arahan yang sangat berharga selama proses pengerjaan skripsi ini.
5. Tim Dosen Penguji, yang telah memberikan saran, kritik, dan masukan berharga dalam penyempurnaan skripsi ini.
6. Kedua orang tua tercinta dan keluarga besar yang senantiasa memberikan doa, kasih sayang, dukungan moral, maupun materiil yang tiada henti.

Yogyakarta, 27 Januari 2026

Habib Baitul Hamdi

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	iv
HALAMAN PERSEMBAHAN	v
KATA PENGANTAR	vi
DAFTAR ISI	vii
DAFTAR TABEL	viii
DAFTAR GAMBAR	ix
DAFTAR LAMPIRAN	x
DAFTAR LAMBANG DAN SINGKATAN	xi
DAFTAR ISTILAH	xii
INTISARI	xiii
ABSTRACT	xiv
BAB I PENDAHULUAN	1
1.1 Latar Belakang	2
1.2 Rumusan Masalah	3
1.3 Batasan Masalah	4
1.4 Tujuan Penelitian	5
1.5 Manfaat Penelitian	6

1.6	Sistematika Penulisan	7
BAB II		8
	TINJAUAN PUSTAKA	9
2.1	Studi Literatur	10
2.2	Dasar Teori	11
2.2.1	Data Mining	12
2.2.2	Machine Learning	13
2.2.3	Regresi Logistic	14
2.2.4	Decision Tree	15
2.2.5	Random Forest	16
2.2.6	Classification	17
2.2.7	Evaluasi Model	18
BAB III	METODE PENELITIAN	19
3.1	Objek Penelitian	20
3.2	Alur Penelitian	21
3.3	Alat dan Bahan	22
3.3.1	Bahan Penelitian	23
3.3.2	Alat Penelitian	24
BAB IV		25
	HASIL DAN PEMBAHASAN	26
4.1	Deskripsi Data dan Visualisasi (EDA).....	27
4.1.1	Distribusi Kelas Target	28

4.1.2	Distribusi Umur Pasien	29
4.1.3	Korelasi Antar Fitur	30
4.2	Pra-pemrosesan Data (Data Preprocessing)	31
4.2.1	Pembersihan Data Duplikat	32
4.2.2	Pengecekan Missing Value	33
4.2.3	Transformasi Data (Data Encoding)	34
4.2.4	Pembagian Data (<i>Splitting Data</i>)	35
4.2.5	Penskalaan Fitur (Feature Scaling)	36
4.3	Hasil Implementasi Algoritma	37
4.3.1	Evaluasi Confusion Matrix	38
4.3.2	Analisis Kepentingan Fitur (Feature Importance)	39
4.3.3	Perbandingan Performa Model (Model Comparison)	40
4.3.4	Perbandingan Performa Model Setelah Optimasi Hyperparameter ...	41
BAB V		42
PENUTUP		43
5.1	Kesimpulan	44
Saran		45
REFERENSI		46
LAMPIRAN		47

DAFTAR TABEL

Tabel 2.1 Confusion Matrix	42
Tabel 3.1. Deskripsi Dataset	46
Tabel 4.1 Data Duplikat	55
Tabel 4.2 Pengecekan Missing Value	56
Tabel 4.3 Hasil Transformasi Data dengan One Hot-Encoding	56
Tabel 4.4 Hasil Pembagian Data Latih dan Data Uji	57
Tabel 4.5 Sampel Data Setelah Proses Penskalaan Fitur	57
Tabel 4.6 Nilai Evaluasi metrix Ketiga Algoritma	60
Tabel 4.7 Hasil Model Setelah Hyperparameter	62

DAFTAR GAMBAR

Gambar 2. 1 ilustrasi Logistic Regression	38
Gambar 2. 2 ilustrasi Decesion Tree	39
Gambar 2. 3 ilustrasi Random Ferest	40
Gambar 3.1 Alur Penelitian	45
Gambar 4.1 Distribusi Kelas Target Data	52
Gambar 4.2 Histogram Distribusi Umur Pasien Berdasarkan Diagnosis	53
Gambar 4.3 Matriks Korelasi Antar Fitur	54
Gambar 4.4 Perbandingan <i>Confusion Matrix</i> (a) Logistic Regression, (b) Decision Tree, (c) Random Forest.....	59
Gambar 4.5 Grafik Tingkat Kepentingan Fitur pada Model <i>Random Forest</i>	60
Gambar 4.6 Grafik Perbandingan Akurasi Ketiga Algoritma.....	61

DAFTAR LAMPIRAN

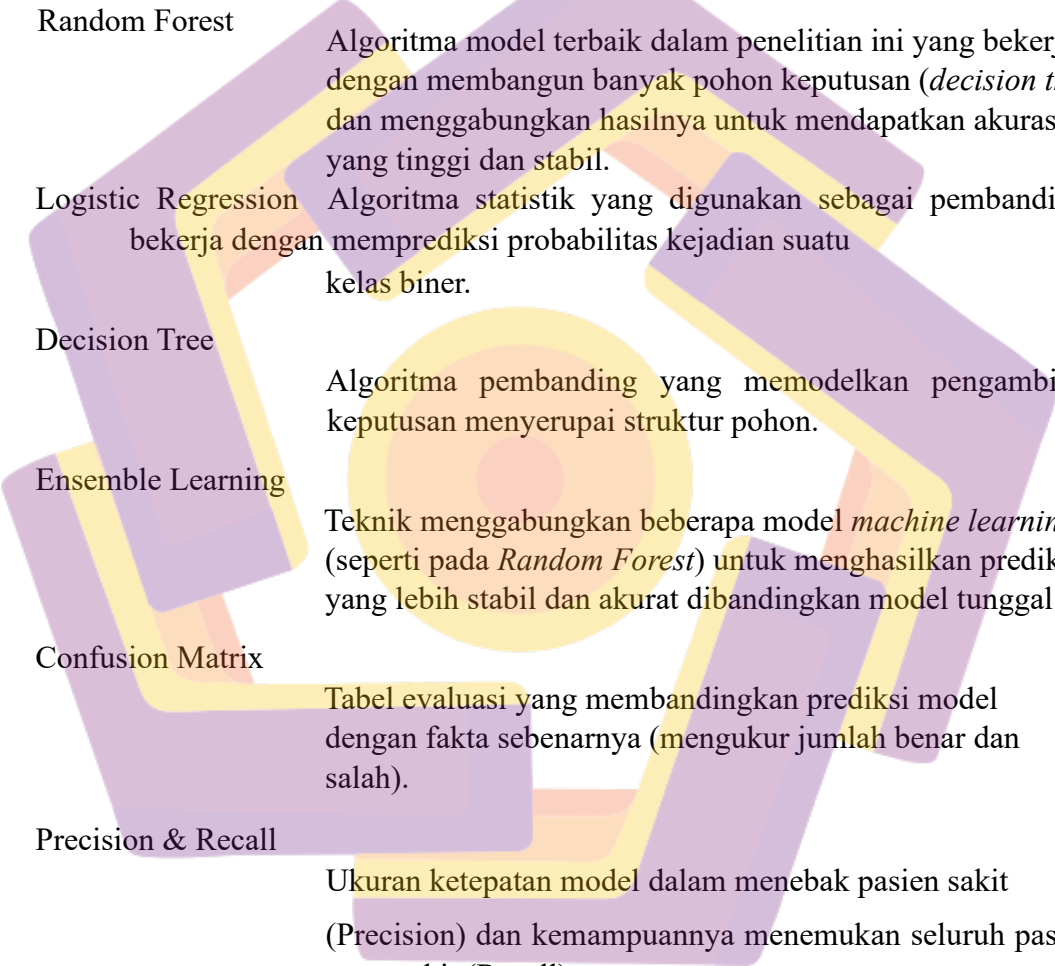
Lampiran 1. Profil obyek Penelitian	10
Lampiran 2. Dokumentasi Penelitian	11



DAFTAR LAMBANG DAN SINGKATAN

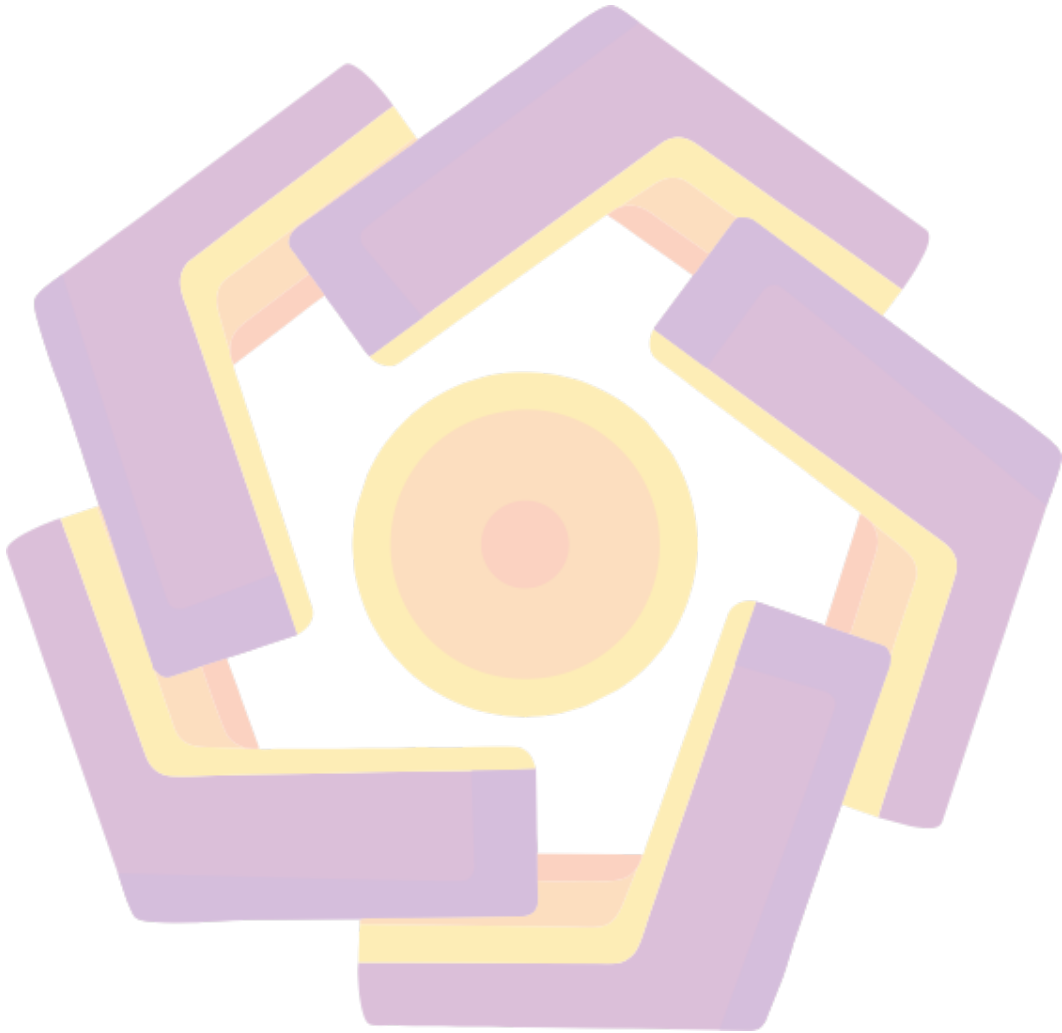
β	Koefisien regresi atau bobot variabel dalam <i>Logistic Regression</i>
β_0	Intercept atau konstanta bias pada persamaan regresi
e	Bilangan Euler (konstanta matematika ≈ 2.718)
$P(x)$	Probabilitas kejadian kelas target (0 atau 1)
S	Himpunan kasus atau kumpulan sampel data
c	Jumlah total kategori atau kelas target dalam himpunan
P_i	Rasio jumlah sampel kelas ke- i terhadap total sampel
$\log_2$	Fungsi logaritma berbasis 2 (digunakan dalam rumus Entropi)
$y(x)$	Prediksi kelas akhir pada algoritma Random Forest
x	Vektor input atau atribut data yang akan diklasifikasikan
q	Jumlah total pohon keputusan (trees) dalam Random Forest
θ_m	Variabel acak yang digunakan untuk membentuk pohon ke- m
r_m	Hasil prediksi dari pohon keputusan tunggal ke- m
TP	True Positive (Data positif yang diprediksi benar)
TN	True Negative (Data negatif yang diprediksi benar)
FP	False Positive (Data negatif yang salah diprediksi positif)
	False Negative (Data positif yang salah diprediksi negatif)
FN	

DAFTAR ISTILAH



Machine Learning	Cabang kecerdasan buatan di mana komputer belajar pola dari data untuk membuat prediksi tanpa diprogram secara eksplisit.
Classification	Teknik pembelajaran mesin untuk mengelompokkan data ke dalam kategori kelas yang berbeda (dalam penelitian ini: kelas "Sakit Jantung" atau "Normal").
Random Forest	Algoritma model terbaik dalam penelitian ini yang bekerja dengan membangun banyak pohon keputusan (<i>decision tree</i>) dan menggabungkan hasilnya untuk mendapatkan akurasi yang tinggi dan stabil.
Logistic Regression	Algoritma statistik yang digunakan sebagai pembanding, bekerja dengan memprediksi probabilitas kejadian suatu kelas biner.
Decision Tree	Algoritma pembanding yang memodelkan pengambilan keputusan menyerupai struktur pohon.
Ensemble Learning	Teknik menggabungkan beberapa model <i>machine learning</i> (seperti pada <i>Random Forest</i>) untuk menghasilkan prediksi yang lebih stabil dan akurat dibandingkan model tunggal.
Confusion Matrix	Tabel evaluasi yang membandingkan prediksi model dengan fakta sebenarnya (mengukur jumlah benar dan salah).
Precision & Recall	Ukuran ketepatan model dalam menebak pasien sakit (Precision) dan kemampuannya menemukan seluruh pasien yang sakit (Recall).
Overfitting	Kondisi di mana model terlalu "hafal" data latih sehingga performanya buruk saat diuji dengan data baru (kelemahan utama algoritma <i>Decision Tree</i>).
Preprocessing	Tahapan awal membersihkan data (seperti hapus duplikat dan <i>scaling</i>) agar siap diproses.
One-Hot Encoding	Teknik mengubah data kategori (huruf) menjadi angka

biner (0 dan 1) agar bisa dihitung oleh komputer.
Feature Importance Analisis untuk mengetahui variabel mana yang paling berpengaruh (menemukan bahwa *ST Slope* adalah fitur terpenting).



INTISARI

Penyakit jantung merupakan salah satu penyebab kematian tertinggi di dunia yang memerlukan penanganan serius. Deteksi dini sangat krusial untuk mencegah fatalitas, namun kompleksitas faktor risiko klinis sering kali menyulitkan diagnosis yang cepat dan akurat. Penelitian ini bertujuan untuk membangun dan membandingkan kinerja model *Machine Learning* dalam mengklasifikasikan risiko penyakit jantung, serta mengidentifikasi fitur medis yang paling berpengaruh terhadap prediksi tersebut. Penelitian ini menggunakan dataset sekunder yang terdiri dari data rekam medis pasien. Tiga algoritma klasifikasi diterapkan sebagai pembandingan, yaitu *Logistic Regression*, *Decision Tree*, dan *Random Forest*. Tahapan pra-pemrosesan data dilakukan secara komprehensif, meliputi penghapusan data duplikat yang menghasilkan 918 data bersih, transformasi data menggunakan *One-Hot Encoding*, pembagian data latih dan uji dengan rasio 80:20, serta penskalaan fitur menggunakan *Standard Scaler*. Evaluasi performa model diukur menggunakan metrik *Accuracy*, *Precision*, *Recall*, dan *F1-Score*. Hasil pengujian menunjukkan bahwa algoritma *Random Forest* memiliki kinerja terbaik dibandingkan dua model lainnya, dengan mencatatkan akurasi tertinggi sebesar 89,67%, *precision* 91,50%, *recall* 90,65%, dan *F1score* 91,07%. Sementara itu, *Logistic Regression* memperoleh akurasi sebesar 85,32% dan *Decision Tree* sebesar 79,34%. Analisis *Feature Importance* pada model terbaik mengungkapkan bahwa *ST Slope* (kemiringan gelombang ST) dan *Oldpeak* merupakan variabel yang paling dominan dalam memengaruhi hasil prediksi. Berdasarkan hasil tersebut, *Random Forest* direkomendasikan sebagai metode yang efektif untuk membantu tenaga medis dalam diagnosis dini penyakit jantung.

Kata kunci: Penyakit Jantung, *Machine Learning*, Klasifikasi, *Random Forest*, *Logistic Regression*, *Decision Tree*.

ABSTRACT

Heart disease is a leading cause of global mortality requiring serious attention. Early detection is crucial to preventing fatalities, yet the complexity of clinical risk factors often complicates rapid and accurate diagnosis. This study aims to build and compare the performance of Machine Learning models in classifying heart disease risk, as well as to identify the medical features that most significantly influence the prediction. This study utilizes a secondary dataset consisting of patient medical records. Three classification algorithms were applied for comparison: Logistic Regression, Decision Tree, and Random Forest. Comprehensive data preprocessing steps were conducted, including the removal of duplicate data resulting in 918 clean records, data transformation using One-Hot Encoding, splitting training and testing data with an 80:20 ratio, and feature scaling using Standard Scaler. Model performance was evaluated using Accuracy, Precision, Recall, and F1-Score metrics. The test results demonstrate that the Random Forest algorithm outperformed the other two models, recording the highest Accuracy of 89.67%, Precision of 91.50%, Recall of 90.65%, and F1-score of 91.07%. Meanwhile, Logistic Regression achieved an accuracy of 85.32%, and Decision Tree achieved 79.34%. Feature Importance analysis on the best model revealed that ST Slope and Oldpeak are the most dominant variables influencing prediction results. Based on these findings, Random Forest is recommended as an effective method to assist medical professionals in the early diagnosis of heart disease.

Keyword: Heart Disease, Machine Learning, Classification, Random Forest, Logistic Regression, Decision Tree.