

**PENERAPAN ALGORITMA DECISION TREE DAN
RANDOM FOREST UNTUK PREDIKSI RISIKO PENYAKIT
JANTUNG**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi S1 Informatika



disusun oleh

ANTONIA VANIA VIYANI

21.11.4091

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2025

**PENERAPAN ALGORITMA DECISION TREE DAN
RANDOM FOREST UNTUK PREDIKSI RISIKO PENYAKIT
JANTUNG**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

ANTONIA VANIA VIYANI

21.11.4091

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2025

HALAMAN PERSETUJUAN

SKRIPSI

PENERAPAN ALGORITMA DECISION TREE DAN RANDOM FOREST UNTUK PREDIKSI RISIKO PENYAKIT JANTUNG

yang disusun dan diajukan oleh

Antonia Vania Viyani

21.11.4091

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 12 Maret 2025

Dosen Pembimbing,

Rifda Faticha Alfa Aziza S.Kom., M.Kom
NIK. 190302392

HALAMAN PENGESAHAN

SKRIPSI

PENERAPAN ALGORITMA DECISION TREE DAN RANDOM FOREST UNTUK PREDIKSI RISIKO PENYAKIT JANTUNG

yang disusun dan diajukan oleh

Antonia Vania Viyani

21.11.4091

Telah dipertahankan di depan Dewan Penguji
pada tanggal 12 Maret 2025

Susunan Dewan Penguji

Nama Penguji

Moch Farid Fauzi, S.Kom., M.Kom
NIK. 190302284

Melany Mustika Dewi, S.Kom., M.Kom
NIK. 190302455

Rifda Faticha Alfa Aziza, S.Kom., M.Kom
NIK. 190302392

Tanda Tangan



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 12 Maret 2025

DEKAN FAKULTAS ILMU KOMPUTER



Hanif Al Fatta, S.Kom., M.Kom., Ph.D.
NIK. 190302096

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Antonia Vania Viyani
NIM : 21.11.4091

Menyatakan bahwa Skripsi dengan judul berikut:

**Penerapan Algoritma Decision Tree dan Random Forest untuk
Prediksi Risiko Penyakit Jantung**

Dosen Pembimbing : Rifda Faticha Alfa Aziza, S.Kom., 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, 12 Maret 2025

Yang Menyatakan,



Antonia Vania Viyani

HALAMAN PERSEMBAHAN

Dengan penuh rasa syukur berkat dan doa, usaha dan kerja keras akhirnya Skripsi dengan Judul “Penerapan Algoritma Decision Tree Dan Random Forest Untuk Prediksi Risiko Penyakit Jantung” ini selesai. Skripsi ini saya persembahkan untuk:

1. Tuhan Yang Maha Esa yang telah memberikan kelancaran dalam menyelesaikan penelitian ini.
2. Kedua orang tua saya yang tercinta yang senantiasa memanjatkan doa , memberikan saya semangat , dan selalu memberikan dukungan penuh kepada saya.
3. Dosen Pembimbing Ibu Rifda Faticha Alfa Aziza, S.Kom., M.Kom yang telah membimbing saya.
4. Teman-teman seperjuangan jurusan S1 Informatika yang telah memberikan banyak ilmu .

Yogyakarta, 20 Februari 2025



penulis

KATA PENGANTAR

Dengan mengucapkan puji syukur kepada Tuhan Yang Maha Esa yang telah melimpahkan segala rahmat-Nya sehingga penulis dapat menyelesaikan penyusunan skripsi sebagai salah satu persyaratan untuk memperoleh gelar sarjana strata satu atau S1 pada Program Studi Teknik Informatika Fakultas Ilmu Komputer Amikom Yogyakarta.

Adapun dalam penyusunan laporan skripsi ini, penyusun ingin mengucapkan terima kasih kepada pihak-pihak yang telah memberikan dukungan secara langsung maupun tidak langsung kepada:

1. Bapak Prof. Dr. M. Suyanto, MM selaku Rektor Universitas Amikom Yogyakarta.
2. Ibu Windha Mega Pradnya Dhuhiya, S.Kom., M.Kom selaku ketua prodi studi S1 Informatika Universitas Amikom Yogyakarta.
3. Ibu Rifda Faticha Alfa Aziza, S.Kom., M.Kom selaku dosen pembimbing yang telah membimbing saya dari awal hingga akhir penelitian.
4. Serta Keluarga dan teman teman yang tidak henti - hentinya . selalu mendukung dari belakang.

Penyusun sangat berharap adanya saran dan kritik dari semua pihak untuk kesempurnaan penyusunan skripsi ini. Hal ini ditimbulkan karena keterbatasan penyusun sebagai manusia yang tak lebih dari manusia biasa yang kerap kali tak luput dari kesalahan. Akhir kata semoga penyusunan skripsi ini dapat memberikan manfaat bagi semua pihak yang membutuhkan. Amin .

Yogyakarta, 20 Februari 2025



penulis

DAFTAR ISI

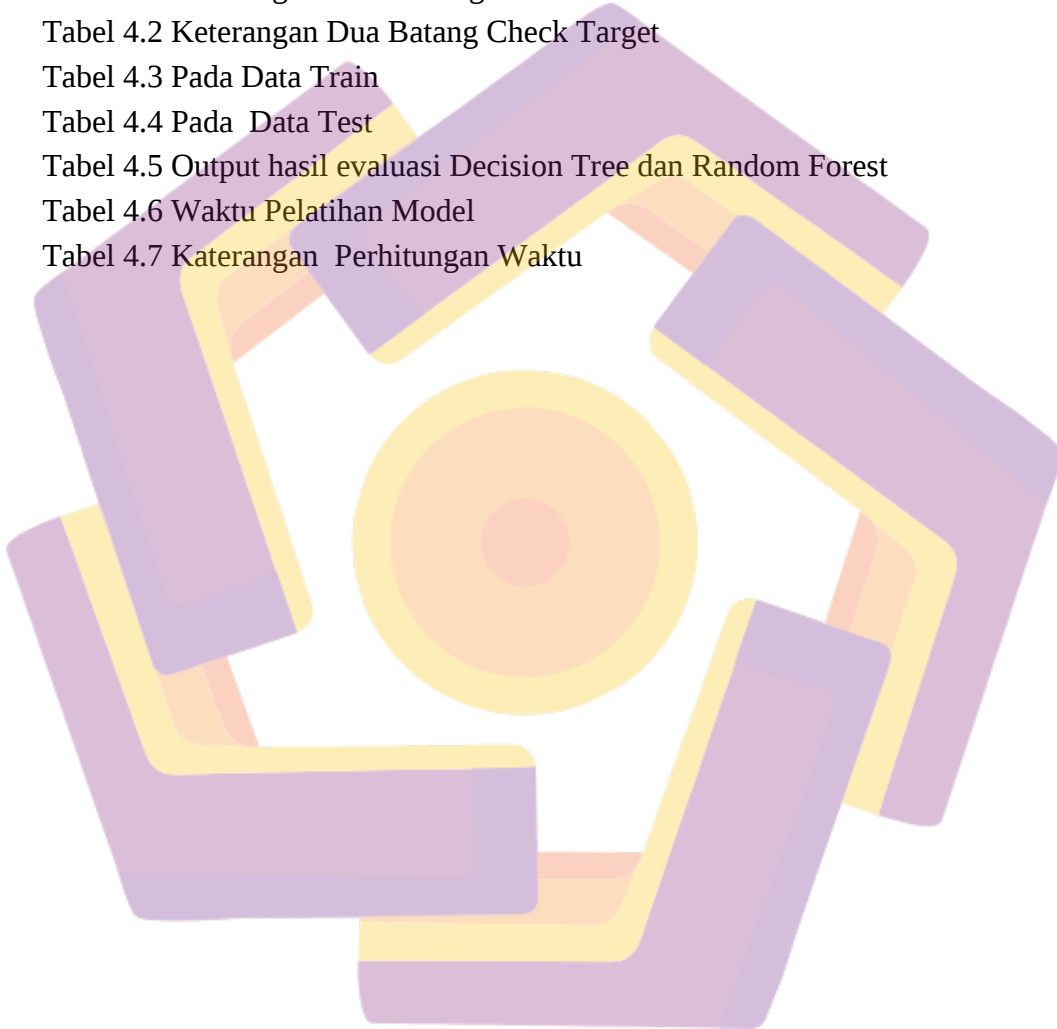
HALAMAN JUDUL	i
HALAMAN PERSETUJUAN	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	iv
HALAMAN PERSEMBAHAN	v
KATA PENGANTAR	vi
DAFTAR ISI	vii
DAFTAR TABEL	ix
DAFTAR GAMBAR	x
DAFTAR LAMBANG DAN SINGKATAN	xi
DAFTAR ISTILAH	xii
INTISARI	xiii
ABSTRACT	xiv
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah	3
1.4 Tujuan Penelitian	4
1.5 Manfaat Penelitian	4
1.6 Sistematika Penulisan	4
BAB II TINJAUAN PUSTAKA	6
2.1 Studi Literatur	6
2.2 Dasar Teori	12
2.2.1 Jantung	12
2.2.2 Machine Learning	12
2.2.3 Prediksi	12
2.2.4 Dataset	12
2.2.5 Algoritma Decision Tree	13
2.2.6 Algoritma Random Forest	14
2.2.7 Evaluasi Model	15
BAB III METODE PENELITIAN	18
3.1 Objek Penelitian	18

3.2 Alur Penelitian	19
3.3 Alat dan Bahan	23
BAB IV HASIL DAN PEMBAHASAN	24
4.1 Dataset	24
4.2 Tipe Data	24
4.3 Missing Values	25
4.4 Check Target Variabel	26
4.5. Pemilihan Fitur dan Pembagian Data Train dan Test	27
4.6 Evaluasi Model	29
4.7 Analisis Evaluasi Model	36
BAB V PENUTUP	39
5.1 Kesimpulan	39
5.2 Saran	39
REFERENSI	40



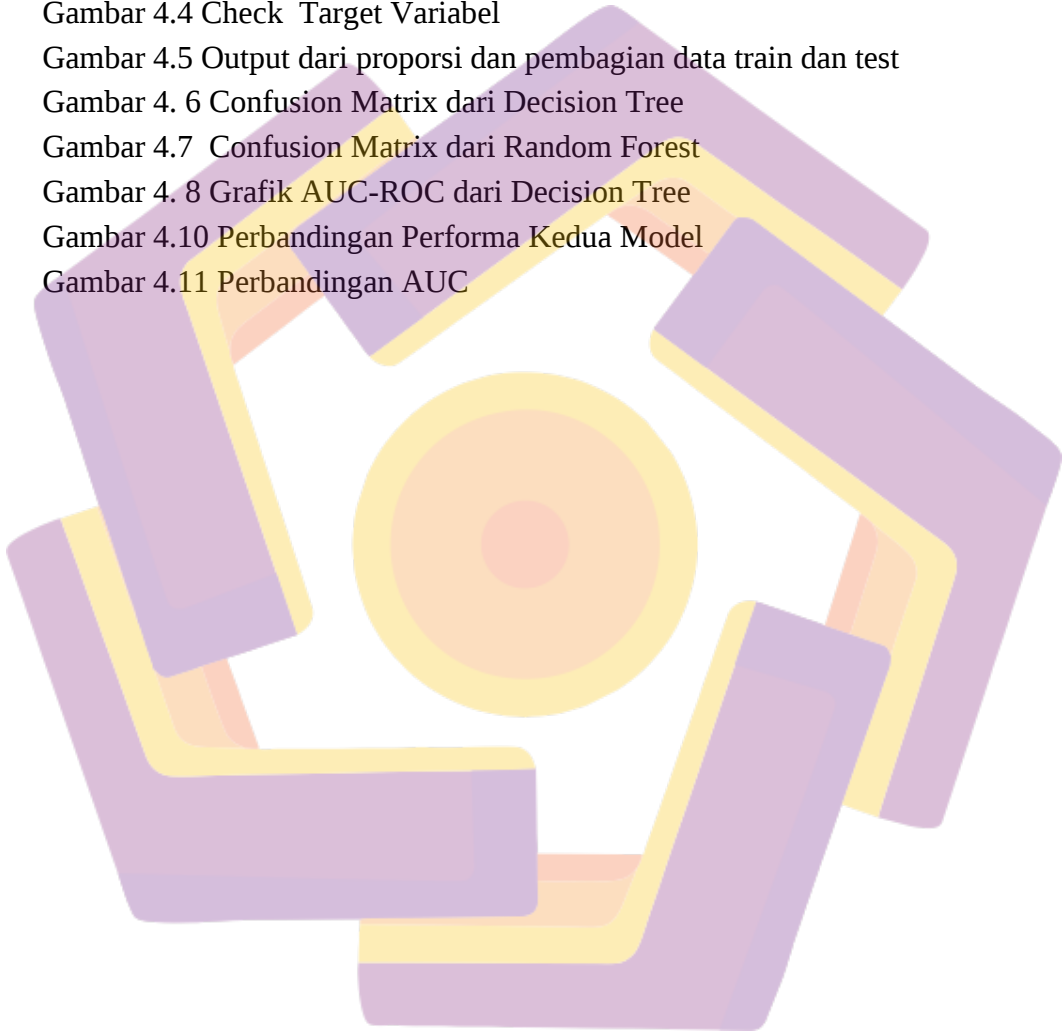
DAFTAR TABEL

Tabel 2.1 Keaslian Penelitian	8
Tabel 2.2 Keterangan Akurasi	15
Tabel 2.3 confusion matrix	16
Tabel 2.4 Keterangan Nilai AUC 0 hingga 1	17
Tabel 3.2 Atribut	20
Tabel 4.1 Keterangan Check Target	27
Tabel 4.2 Keterangan Dua Batang Check Target	27
Tabel 4.3 Pada Data Train	28
Tabel 4.4 Pada Data Test	29
Tabel 4.5 Output hasil evaluasi Decision Tree dan Random Forest	30
Tabel 4.6 Waktu Pelatihan Model	38
Tabel 4.7 Keterangan Perhitungan Waktu	38



DAFTAR GAMBAR

Gambar 3.2 Alur Penelitian	20
Gambar 4.1 Data Penyakit Jantung	24
Gambar 4.2 Info Tipe Data	25
Gambar 4.3 Missing Values.	26
Gambar 4.4 Check Target Variabel	26
Gambar 4.5 Output dari proporsi dan pembagian data train dan test	28
Gambar 4. 6 Confusion Matrix dari Decision Tree	32
Gambar 4.7 Confusion Matrix dari Random Forest	32
Gambar 4. 8 Grafik AUC-ROC dari Decision Tree	34
Gambar 4.10 Perbandingan Performa Kedua Model	36
Gambar 4.11 Perbandingan AUC	37



DAFTAR LAMBANG DAN SINGKATAN



DT	<i>Decision Tree</i>
RF	<i>Random Forest</i>
TP	<i>True Positif</i>
TN	<i>True Negatif</i>
FP	<i>False Positif</i>
FN	<i>False Negative</i>
AUC-ROC	<i>Receiver Operating Characteristic Area Under the Curve</i>
TPR	<i>True Positive Rate</i>
FPR	<i>False Positive Rate</i>

DAFTAR ISTILAH

Model algoritma *machine learning*



INTISARI

Penentuan prediksi risiko penyakit jantung merupakan proses yang sangat penting dalam dunia medis dan kesehatan. Untuk meningkatkan akurasi dan efisiensi dalam klasifikasi risiko penyakit jantung, penelitian ini berfokus pada penerapan dua model machine learning, yaitu *Decision Tree* dan *Random Forest*. *Decision Tree* karena menghasilkan model yang mudah dipahami dan menjadi pilihan yang kuat untuk memprediksi risiko penyakit jantung, memberikan kombinasi yang baik untuk akurasi. Sedangkan *Random Forest* sebagai algoritma dimana algoritma ini sebagai pencarian solusi prediksi terbaik dan memiliki keunggulan untuk pengklasifikasian dataset dengan jumlah besar. Karena fungsinya dapat digunakan untuk banyak dimensi dengan berbagai skala dan performa tinggi. Tujuan utama dari penelitian ini adalah untuk membandingkan performa kedua model tersebut dalam memprediksi risiko penyakit jantung. Perbandingan dilakukan dengan tahapan Pengumpulan Data, Pembagian Data (*Training Set* dan *Testing Set*), Penerapan Algoritma, Evaluasi Model, dan Analisis Evaluasi Model. Dataset yang akan digunakan mencakup informasi tentang jantung dapat dari kaggle. Perbandingan performa pada penelitian ini menggunakan matrix akurasi, *precision*, *recall*, *F1-Score* dan AUC-ROC untuk mengevaluasi kinerja masing-masing model. Penelitian ini menghasilkan bahwa model random forest lebih unggul dengan skor Akurasi 0.93, *precision* 0.92, *Recall* 0.92 dan *F1 Score* 0.92. Skor AUC yang digunakan untuk melihat seberapa baik model mencapai 0.97. Hasil penelitian ini diharapkan mampu memberikan rekomendasi yang bermanfaat bagi industri medis dalam memilih model yang paling tepat untuk prediksi penyakit jantung.

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

ABSTRACT

The determining heart disease risk prediction is a very important process in the medical and health world. To improve accuracy and efficiency in heart disease risk classification, this research focuses on the application of two machine learning models, namely Decision Tree and Random Forest. Decision Tree because it produces an easy-to-understand model and is a strong choice for predicting heart disease risk, providing a good combination for accuracy. Random Forest as an algorithm where this algorithm is a search for the best prediction solution and has the advantage of classifying a large number of datasets. Because of its functionality, it can be used for many dimensions with various scales and high performance. The main objective of this study is to compare the performance of the two models in predicting heart disease risk. The comparison is carried out with the stages of Data Collection, Data Sharing (Training Set and Testing Set), Algorithm Application, Model Evaluation, and Model Evaluation Analysis. The dataset to be used includes information about the heart from kaggle. Performance comparison in this study uses accuracy, precision, recall, F1-Score and AUC-ROC matrix to evaluate the performance of each model. This study found that the random forest model is superior with an Accuracy score of 0.93, Precision 0.92, Recall 0.92 and F1 Score 0.92. The AUC score used to see how good the model is reached 0.97. The results of this study are expected to provide useful recommendations for the medical industry in choosing the most appropriate model for heart disease prediction.

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