

**PENERAPAN ALGORITMA RANDOM FOREST UNTUK
DETEKSI PENYAKIT JANTUNG DAN OPTIMISASI
MENGUNAKAN FEATURE IMPORTANCE DAN
RANDOMIZED SEARCH**



SKRIPSI

**MICHAEL ALEXANDER
00000047852**

**PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG
2025**

**PENERAPAN ALGORITMA RANDOM FOREST UNTUK
DETEKSI PENYAKIT JANTUNG DAN OPTIMISASI
MENGUNAKAN FEATURE IMPORTANCE DAN
RANDOMIZED SEARCH**



Diajukan sebagai salah satu syarat untuk memperoleh
Gelar Sarjana Komputer (S.Kom.)

MICHAEL ALEXANDER
00000047852

PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG
2025

HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,

Nama : Michael Alexander

Nomor Induk Mahasiswa : 00000047852

Program Studi : Informatika

Skripsi dengan judul:

Penerapan Algoritma Random Forest untuk Deteksi Penyakit Jantung dan Optimisasi Menggunakan Feature Importance dan Randomized Search

merupakan hasil karya saya sendiri bukan plagiat dari laporan karya tulis ilmiah yang ditulis oleh orang lain, dan semua sumber, baik yang dikutip maupun dirujuk, telah saya nyatakan dengan benar serta dicantumkan di Daftar Pustaka.

Jika di kemudian hari terbukti ditemukan kecurangan/penyimpangan, baik dalam pelaksanaan maupun dalam penulisan laporan karya tulis ilmiah, saya bersedia menerima konsekuensi dinyatakan TIDAK LULUS untuk mata kuliah yang telah saya tempuh.

Tangerang, 26 Juni 2025



(Michael Alexander)

UNIVERSITAS
MULTIMEDIA
NUSANTARA

HALAMAN PENGESAHAN

Skripsi dengan judul

PENERAPAN ALGORITMA RANDOM FOREST UNTUK DETEKSI PENYAKIT JANTUNG DAN OPTIMISASI MENGGUNAKAN FEATURE IMPORTANCE DAN RANDOMIZED SEARCH

oleh

Nama : Michael Alexander
NIM : 00000047852
Program Studi : Informatika
Fakultas : Fakultas Teknik dan Informatika

Telah diujikan pada hari Selasa, 22 Juli 2025

Pukul 10.00 s/s 12.00 dan dinyatakan

LULUS

Dengan susunan penguji sebagai berikut

Ketua Sidang

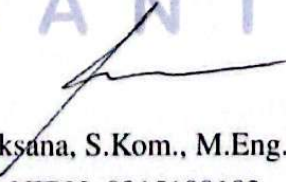
Penguji


(Moeljono Widjaja, B.Sc., M.Sc., Ph.D.) (David Agustriawan, S.Kom., M.Sc., Ph.D.)
NIDN: 0311106903 NIDN: 0525088601

Pembimbing


(Angga Aditya Permana, S.Kom., M.Kom.)
NIDN: 0407128901

Ketua Program Studi Informatika,


(Arya Wicaksana, S.Kom., M.Eng.Sc., OCA)
NIDN: 0315109103

HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

Yang bertanda tangan di bawah ini:

Nama : Michael Alexander
NIM : 00000047852
Program Studi : Informatika
Jenjang : S1
Judul Karya Ilmiah : Penerapan Algoritma Random Forest
untuk Deteksi Penyakit Jantung dan
Optimisasi Menggunakan Feature
Importance dan Randomized Search

Menyatakan dengan sesungguhnya bahwa saya bersedia:

- ☒ Saya bersedia memberikan izin sepenuhnya kepada Universitas Multimedia Nusantara untuk mempublikasikan hasil karya ilmiah saya ke dalam repositori Knowledge Center sehingga dapat diakses oleh Sivitas Akademika UMN/Publik. Saya menyatakan bahwa karya ilmiah yang saya buat tidak mengandung data yang bersifat konfidensial.
- ☐ Saya tidak bersedia mempublikasikan hasil karya ilmiah ini ke dalam repositori Knowledge Center, dikarenakan: dalam proses pengajuan publikasi ke jurnal/konferensi nasional/internasional (dibuktikan dengan *letter of acceptance*) **.
- ☐ Lainnya, pilih salah satu:
 - Hanya dapat diakses secara internal Universitas Multimedia Nusantara
 - Embargo publikasi karya ilmiah dalam kurun waktu tiga tahun.

Tangerang, 26 Juni 2025

Yang menyatakan



Michael Alexander

**Jika tidak bisa membuktikan LoA jurnal/HKI, saya bersedia mengizinkan penuh karya ilmiah saya untuk dipublikasikan ke KC UMN dan menjadi hak institusi UMN.

HALAMAN PERSEMBAHAN / MOTTO

"A good name is to be more desired than great wealth, Favor is better than silver and gold."

Proverbs 22:1 (NASB)



KATA PENGANTAR

Puji syukur atas berkat dan rahmat Tuhan Yang Maha Esa, atas terselesainya penulisan laporan skripsi ini dengan judul: Penerapan Algoritma Random Forest untuk Deteksi Penyakit Jantung dan Optimisasi Menggunakan Feature Importance dan Randomized Search. Penulisan skripsi ini dilakukan sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer pada Program Studi Informatika, Fakultas Teknik dan Informatika, Universitas Multimedia Nusantara.

Saya menyadari bahwa tanpa adanya dukungan, bimbingan, dan doa dari berbagai pihak, mulai dari masa perkuliahan hingga penyusunan skripsi ini, tentu akan sangat sulit bagi saya untuk menyelesaikannya dengan baik. Oleh karena itu, saya mengucapkan terima kasih kepada:

1. Dr. Ir. Andrey Andoko, M.Sc., selaku Rektor Universitas Multimedia Nusantara.
2. Dr. Eng. Niki Prastomo, S.T., M.Sc., selaku Dekan Fakultas Teknik dan Informatika Universitas Multimedia Nusantara.
3. Arya Wicaksana, S.Kom., M.Eng.Sc., OCA, selaku Ketua Program Studi Informatika Universitas Multimedia Nusantara.
4. Angga Aditya Permana, S.Kom., M.Kom., sebagai Pembimbing yang telah memberikan bimbingan, arahan, dan motivasi atas terselesainya tugas akhir ini.
5. Keluarga saya yang telah memberikan bantuan dukungan material dan moral, sehingga penulis dapat menyelesaikan tugas akhir ini.

Semoga karya ilmiah ini dapat bermanfaat bagi semua pihak yang membacanya.

Tangerang, 26 Juni 2025



Michael Alexander

PENERAPAN ALGORITMA RANDOM FOREST UNTUK DETEKSI PENYAKIT JANTUNG DAN OPTIMISASI MENGGUNAKAN FEATURE IMPORTANCE DAN RANDOMIZED SEARCH

Michael Alexander

ABSTRAK

Penyakit jantung merupakan salah satu penyebab utama kematian di dunia. Penelitian ini bertujuan untuk membangun model klasifikasi berbasis algoritma *Random Forest* guna mendeteksi risiko penyakit jantung menggunakan dataset dari UCI Machine Learning Repository. Model pertama dibangun menggunakan seluruh fitur, kemudian dilakukan seleksi fitur berdasarkan nilai *feature importance* untuk membentuk model kedua. Evaluasi dilakukan menggunakan metrik akurasi, ROC AUC, precision, dan recall. Hasil menunjukkan bahwa model dengan fitur terpilih memberikan efisiensi waktu pelatihan lebih baik dan akurasi lebih tinggi dibanding model awal, dengan penurunan skor AUC yang dapat ditoleransi. Penelitian ini menunjukkan bahwa penghapusan fitur yang kurang relevan mampu meningkatkan efisiensi model tanpa mengorbankan performa secara signifikan.

Kata kunci: akurasi, deteksi penyakit, feature importance, klasifikasi, random forest



**IMPLEMENTATION OF RANDOM FOREST ALGORITHM FOR HEART
DISEASE DETECTION WITH OPTIMIZATION USING FEATURE
IMPORTANCE AND RANDOMIZED SEARCH**

Michael Alexander

ABSTRACT

Heart disease is one of the leading causes of death worldwide. This study aims to develop a classification model using the Random Forest algorithm to detect the risk of heart disease, based on data from the UCI Machine Learning Repository. The first model was built using all available features, followed by a feature selection process based on feature importance to construct the second model. Evaluation was carried out using accuracy, ROC AUC, precision, and recall metrics. The results show that the reduced feature model achieved better training efficiency and higher accuracy compared to the full-feature model, with only a minor decrease in AUC score. This study demonstrates that removing less relevant features can improve model efficiency without significantly compromising classification performance.

Keywords: *accuracy, classification, disease detection, feature importance, random forest*



DAFTAR ISI

HALAMAN JUDUL	i
PERNYATAAN TIDAK MELAKUKAN PLAGIAT	ii
HALAMAN PERSETUJUAN	iii
HALAMAN PENGESAHAN	iii
HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH	iv
HALAMAN PERSEMBAHAN/MOTO	v
KATA PENGANTAR	vi
ABSTRAK	vii
ABSTRACT	viii
DAFTAR ISI	ix
DAFTAR TABEL	xi
DAFTAR GAMBAR	xii
DAFTAR KODE	xiii
DAFTAR RUMUS	xiv
DAFTAR LAMPIRAN	xv
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang Masalah	1
1.2 Rumusan Masalah	2
1.3 Batasan Permasalahan	2
1.4 Tujuan Penelitian	3
1.5 Manfaat Penelitian	3
1.6 Sistematika Penulisan	4
BAB 2 LANDASAN TEORI	5
2.1 Penyakit Jantung (Heart Disease)	5
2.2 Pembelajaran Mesin	5
2.2.1 Supervised Learning	6
2.2.2 Unsupervised Learning	6
2.2.3 Reinforcement Learning	6
2.3 Decision Tree	6
2.4 Random Forest	8
2.5 Randomized Search untuk Optimasi Hyperparameter	9
2.6 Feature Importance	10
2.7 Confusion Matrix	11
2.7.1 Confusion Matrix untuk Klasifikasi Multikelas	13
BAB 3 METODOLOGI PENELITIAN	14
3.1 Pengumpulan Data	14
3.1.1 Spesifikasi Dataset	15
3.1.2 Variabel dan Deskripsi	15
3.1.3 Target Variabel	16
3.2 Pre-processing Data	16
3.3 Pembangunan Model Pertama	18
3.4 Seleksi Fitur dengan Feature Importance	19
3.5 Pembuatan Model Kedua Berdasarkan Seleksi Fitur	20
3.5.1 Seleksi Fitur Berdasarkan Importance	21
3.5.2 Pembangunan Model Kedua	22
3.6 Perbandingan Model	22
3.6.1 Tujuan Perbandingan	23
3.6.2 Prosedur Perbandingan	23

3.6.3	Keluaran yang Diharapkan	24
BAB 4	HASIL DAN DISKUSI	25
4.1	Spesifikasi Sistem	25
4.2	Pengumpulan Data	26
4.3	Pre-Processing	26
4.4	Pembangunan Model pertama	29
4.4.1	Training	29
4.4.2	Testing	30
4.4.3	Evaluasi Akurasi Model Pertama	30
4.4.4	Evaluasi Performa Model pertama (Multiclass Classification)	32
4.5	Seleksi Fitur Berdasarkan Feature Importance	33
4.6	Pembangunan Model Kedua Berdasarkan Fitur Terpilih	34
4.6.1	Evaluasi Performa Model Kedua (Multiclass Classification)	36
4.7	Evaluasi Performa Model Sebelum dan Sesudah Seleksi Fitur	38
4.7.1	Cuplikan Kode Implementasi Model	38
4.7.2	Perbandingan Performa Model Biner	39
4.7.3	Perbandingan Performa Multiclass	39
4.7.4	Visualisasi Perbandingan	40
4.8	Hasil Eksperimen	41
4.9	Diskusi	42
BAB 5	SIMPULAN DAN SARAN	43
5.1	Simpulan	43
5.2	Saran	43
DAFTAR PUSTAKA		45



DAFTAR TABEL

Tabel 2.1	Struktur confusion matrix klasifikasi biner	12
Tabel 2.2	Confusion Matrix untuk Klasifikasi Multikelas	13
Tabel 3.1	Detail dataset UCI Heart Disease	15
Tabel 4.1	Perbandingan performa model random forest	39
Tabel 4.2	Perbandingan Performa Model Multiclass	40



DAFTAR GAMBAR

Gambar 2.1	Struktur dasar decision tree	7
Gambar 2.2	Cara kerja algoritma random forest	9
Gambar 3.1	Flowchart metodologi	14
Gambar 3.2	Flowchart proses <i>pre-processing</i>	17
Gambar 3.3	Flowchart proses pembangunan model	19
Gambar 3.4	Flowchart <i>feature importance</i>	20
Gambar 3.5	Flowchart pembuatan model kedua berdasarkan seleksi fitur	21
Gambar 3.6	Flowchart perbandingan antara kedua model	23
Gambar 4.1	Tampilan halaman dataset UCI Heart Disease pada situs Kaggle	26
Gambar 4.2	Hasil model pertama	31
Gambar 4.3	Hasil Evaluasi Model pertama Random Forest Multikelas .	32
Gambar 4.4	Visualisasi top 21 feature importances dari model random forest	34
Gambar 4.5	Confusion matrix dari model kedua (reduced features) . . .	35
Gambar 4.6	Hasil Evaluasi Model Kedua Random Forest Multikelas . .	37
Gambar 4.7	Visualisasi perbandingan performa model	41



DAFTAR KODE

Kode 4.1	Impor dataset dan konversi variabel target	27
Kode 4.2	Transformasi target	27
Kode 4.3	Pemisahan fitur dan target	27
Kode 4.4	Identifikasi tipe data	27
Kode 4.5	Pipeline transformasi	27
Kode 4.6	Pembagian data 80/20	28
Kode 4.7	Pelatihan dan pengukuran waktu	29
Kode 4.8	Prediksi dan pengukuran waktu	30
Kode 4.9	Evaluasi performa model	30
Kode 4.10	Fungsi preprocessor untuk pipeline	38
Kode 4.11	Pipeline model full dan reduced features	38
Kode 4.12	Pelatihan dan evaluasi kedua model	39



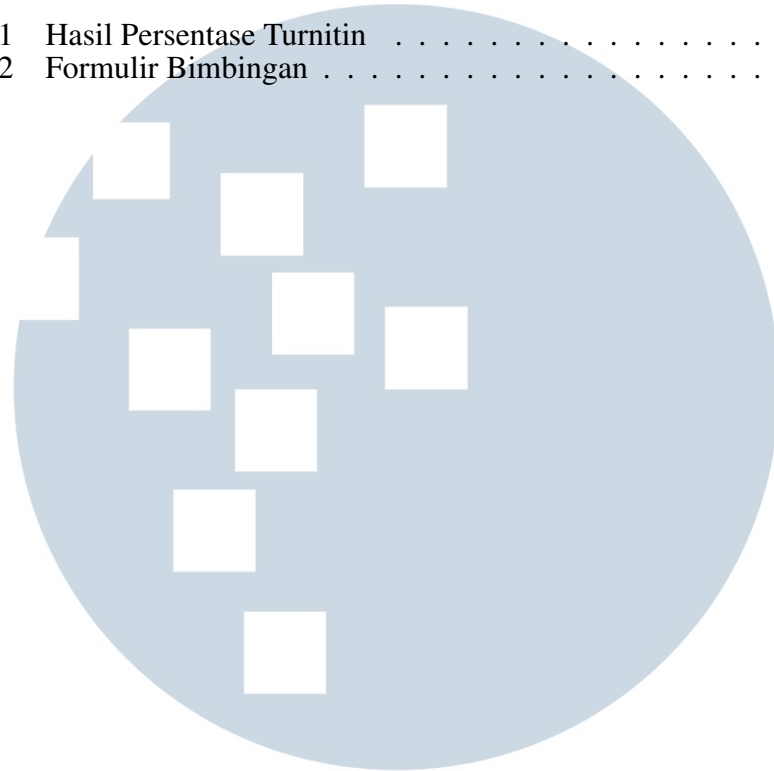
DAFTAR RUMUS

Rumus 2.1	Ilustrasi formula voting pada Random Forest	9
Rumus 2.2	<i>Accuracy</i>	12
Rumus 2.3	<i>Precision</i>	12
Rumus 2.4	<i>Recall / Sensitivity</i>	12
Rumus 2.5	<i>F1-Score</i>	12
Rumus 2.6	<i>Accuracy Multiclass</i>	13



DAFTAR LAMPIRAN

Lampiran 1	Hasil Persentase Turnitin	47
Lampiran 2	Formulir Bimbingan	54



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA