



**Komparasi Algoritma *Decision Tree* dan KNN dengan Optimasi
Feature Selection Menggunakan *Genetic Algorithm* dalam
Memprediksi Risiko Hipertensi**

LAPORAN TUGAS AKHIR

Ugniati Nuraftina (41821010055)
Rizki Laksana Putra (41821010012)
Arenga Pinnata Marzuki (41821010014)

MERCU BUANA

PROGRAM STUDI SISTEM INFORMASI

FAKULTAS ILMU KOMPUTER

UNIVERSITAS MERCU BUANA

JAKARTA

2025



**Komparasi Algoritma *Decision Tree* dan KNN dengan Optimasi
Feature Selection Menggunakan *Genetic Algorithm* dalam
Memprediksi Risiko Hipertensi**

LAPORAN TUGAS AKHIR

Ugniati Nuraftina (41821010055)
Rizki Laksana Putra (41821010012)
Arenga Pinnata Marzuki (41821010014)

Diajukan sebagai salah satu syarat untuk memperoleh gelar sarjana

PROGRAM STUDI SISTEM INFORMASI

FAKULTAS ILMU KOMPUTER

UNIVERSITAS MERCU BUANA

JAKARTA

2025

HALAMAN PERNYATAAN KARYA SENDIRI

Saya yang bertanda tangan di bawah:

Nama : Ugniati Nuraftina
NIM : 41821010055
Program Studi : Sistem Informasi
Judul Laporan Skripsi : Komparasi Algoritma Decision Tree dan KNN dengan Optimasi Feature Selection Menggunakan Genetic Algorithm dalam Memprediksi Risiko Hipertensi

Menyatakan bahwa Laporan Aplikatif/Tugas Akhir/Jurnal/Media Ilmiah ini adalah hasil karya saya sendiri dan bukan plagiat, serta semua sumber baik yang dikutip maupun dirujuk telah saya nyatakan dengan benar. Apabila ternyata ditemukan di dalam Laporan Tugas Akhir saya terdapat unsur plagiat, maka saya siap mendapatkan sanksi akademis yang berlaku di Universitas Mercu Buana.

Jakarta, 13 Januari 2025



Ugniati Nuraftina

UNIVERSIT
MERCU BUANA

LEMBAR PENGESAHAN

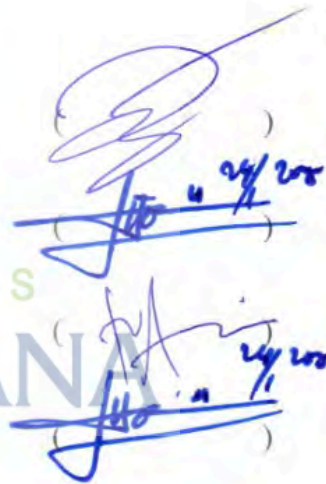
Nama Mahasiswa (1) : Ugniati Nuraftina
NIM : (41821010055)
Nama Mahasiswa (2) : Rizki Laksana Putra
NIM : (41821010012)
Nama Mahasiswa (3) : Arenga Pinnata Marzuki
NIM : (41821010014)
Judul Tugas Akhir : Komparasi Algoritma Decision Tree dan KNN dengan Optimasi Feature Selection Menggunakan Genetic Algorithm dalam Memprediksi Risiko Hipertensi

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Jakarta, 13 Januari 2025

Menyetujui

Pembimbing : Andi Nugroho, S.T, M.Kom
NIDN : 0305098303
Ketua Penguji : Lukman Hakim, ST., M.Kom
NIDN : 0327107701
Penguji 1 : Nur Ani, ST, MMSI
NIDN : 0310117801
Penguji 2 : Lukman Hakim, ST., M.Kom
NIDN : 0327107701



Handwritten signatures of the Pembimbing and Penguji members, dated 24/1/2025.

Mengetahui,



Dr. Bambang Jokonowo, S.Si., M.T.I.
Dekan Fakultas Ilmu Komputer



Dr. Ruci Meiyanti, M.Kom
Ka.Prodi Sistem Informasi

KATA PENGANTAR

Puji syukur kita panjatkan kepada Allah SWT atas berkah dan rahmatnya sehingga penulis mampu menyelesaikan Laporan Tugas Akhir yang berjudul “Komparasi Algoritma *Decision Tree* dan KNN dengan Optimasi *Feature Selection* Menggunakan *Genetic Algorithm* dalam Memprediksi Risiko Hipertensi” sebagai salah satu syarat untuk menyelesaikan program studi strata satu (S1) pada jurusan Sistem Informasi Universitas Mercu Buana.

Dalam penyusunan Laporan Tugas Akhir ini, penulis menyadari bahwa tanpa bantuan dan bimbingan dari berbagai pihak sangatlah sulit bagi penulis untuk menyelesaikan Laporan Tugas Akhir ini. Oleh karena itu, penulis mengucapkan terima kasih kepada:

1. Prof. Dr. Andi Adriansyah, M.Eng. selaku Rektor Universitas Mercu Buana.
2. Dr. Bambang Jokonowo, S.Si., M.T.I. selaku Dekan Program Studi Sistem Informasi Universitas Mercu Buana.
3. Ibu Dr. Ruci Meiyanti, S.Kom., M. Kom., selaku Ketua Program Studi Sistem Informasi Universitas Mercu Buana.
4. Bapak Andi Nugroho, S.T., M.Kom., selaku Dosen Pembimbing Tugas Akhir yang telah membimbing penulis dengan semangat, nasihat, dan ilmunya dalam penyusunan Laporan Tugas Akhir ini.
5. Lukman Hakim, ST., M.Kom dan Nur Ani, ST, MMSI selaku dosen penguji Tugas Akhir atas koreksi dan arahnya, serta masukannya.
6. Kepada pihak keluarga khususnya kedua orang tua yang tanpa henti memberikan dukungan, semangat, dan do'a.
7. Teman-teman Kommas (Wiwit, Arenga, Lana) yang selalu saling mendukung dan menyemangati hingga akhirnya Tugas Akhir ini dapat selesai tepat waktu.
8. Najma yang sudah mau menemani penulis selama mengerjakan tugas Akhir dan Uti yang sudah membantu memberikan informasi mengenai Tugas Akhir ini.
9. Teman-teman MM yang sudah memberikan dukungan dan motivasi kepada penulis.

10. Semua pihak yang telah memotivasi dan ikut memberikan bantuan yang tidak dapat penulis tuliskan satu per satu.

Akhir kata, penulis berharap semoga Allah SWT membalas segala kebaikan semua pihak yang telah membantu. Semoga Laporan Tugas Akhir ini membawa manfaat bagi pengembangan ilmu dan masyarakat.

Jakarta, 13 Januari 2025

Penulis



**HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS
AKHIR UNTUK KEPENTINGAN AKADEMIS**

Sebagai sivitas akademik Universitas Mercu Buana, saya yang bertanda tangan di bawah ini:

Nama : Ugniati Nuraftina

NIM : 41821010055

Program Studi : Sistem Informasi

Judul Laporan Skripsi : Komparasi Algoritma Decision Tree dan KNN dengan Optimasi Feature Selection Menggunakan Genetic Algorithm dalam Memprediksi Risiko Hipertensi

Demi pengembangan ilmu pengetahuan, dengan ini memberikan izin dan menyetujui untuk memberikan kepada Universitas Mercu Buana **Hak Bebas Royalti Non-Eksklusif (Non-exclusive Royalty-Free Right)** atas karya ilmiah saya yang berjudul di atas beserta perangkat yang ada (jika diperlukan).

Dengan Hak Bebas Royalti Non-Eksklusif ini Universitas Mercu Buana berhak menyimpan, mengalihmedia/format-kan, mengelola dalam bentuk pangkalan data (database), merawat, dan mempublikasikan Laporan Magang/Skripsi/Tesis/Disertasi saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Jakarta, 13 Januari 2025

Yang menyatakan,



10000
SERULUH RIBU RUPIAH
24AMX155533866
METERAI TEMPEL

Ugniati Nuraftina

ABSTRAK

Nama Mahasiswa (1) : Ugniati Nuraftina

NIM Mahasiswa (1) : 41821010055

Nama Mahasiswa (2) : Arenga Pinnata Marzuki

NIM Mahasiswa (2) : 41821010014

Nama Mahasiwa (3) : Rizki Laksana Putra

NIM Mahasiswa (3) : 41821010012

Pembimbing TA : Andi Nugroho, ST, M.Kom

Judul Tugas Akhir : Komparasi Algoritma *Decision Tree* dan KNN dengan Optimasi *Feature Selection* Menggunakan *Genetic Algorithm* dalam Memprediksi Risiko Hipertensi

Hipertensi adalah kondisi peningkatan tekanan darah di atas batas normal, dengan tekanan sistolik ≥ 140 mmHg dan diastolik ≥ 90 mmHg. Menurut data World Health Organization (WHO) tahun 2023, sekitar 1,28 miliar orang dewasa berusia 30-79 tahun di dunia mengidap hipertensi, terutama di negara berkembang, dengan 46% penderita tidak menyadari kondisinya. Penelitian ini bertujuan memprediksi risiko hipertensi menggunakan algoritma KNN dan Decision Tree dengan optimasi feature selection menggunakan Genetic Algorithm. Dataset yang digunakan terdiri dari 12 fitur dan 4240 data. Hasil penelitian menunjukkan bahwa Decision Tree memberikan performa terbaik dengan akurasi 90.6%, precision 92.3%, recall 89.2%, dan F1-Score 90.8%. Optimasi fitur berhasil meningkatkan akurasi kedua algoritma, membuktikan efektivitas Genetic Algorithm dalam seleksi fitur untuk prediksi risiko hipertensi.

Kata Kunci: *Genetic Algorithm*, *Decision Tree*, *K-Nearest Neighbors*, Hipertensi, Optimasi, Klasifikasi.

ABSTRACT

Student Name (1) : Ugniati Nuraftina

Student Number (1) : 41821010055

Student Name (2) : Arenga Pinnata Marzuki

Student Number (2) : 41821010014

Student Name (3) : Rizki Laksana Putra

Student Number (3) : 41821010012

Counsellor : Andi Nugroho, ST, M.Kom

Title : Komparasi Algoritma *Decision Tree* dan KNN dengan Optimasi *Feature Selection* Menggunakan *Genetic Algorithm* dalam Memprediksi Risiko Hipertensi

Abstract:

Hypertension is a condition characterized by elevated blood pressure levels, with systolic pressure ≥ 140 mmHg and diastolic pressure ≥ 90 mmHg. According to the World Health Organization (WHO) in 2023, approximately 1.28 billion adults aged 30-79 worldwide suffer from hypertension, primarily in developing countries, with 46% of them unaware of their condition. This study aims to predict hypertension risk using KNN and Decision Tree algorithms, optimized through feature selection with Genetic Algorithm. The dataset used consists of 12 features and 4,240 records. The results show that the Decision Tree algorithm achieved the best performance with an accuracy of 90.6%, precision of 92.3%, recall of 89.2%, and F1-Score of 90.8%. Feature optimization successfully improved the accuracy of both algorithms, demonstrating the effectiveness of Genetic Algorithm in feature selection for hypertension risk prediction.

Keywords: Genetic Algorithm, Decision Tree, K-Nearest Neighbours, Optimization, Hypertension, Classification.

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERNYATAAN KARYA SENDIRI	ii
HALAMAN PENGESAHAN	iii
KATA PENGANTAR	iv
HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI	vi
ABSTRAK	vii
ABSTRACT	viii
DAFTAR ISI	ix
DAFTAR TABEL	xii
DAFTAR GAMBAR	xiii
DAFTAR LAMPIRAN	xv
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Tujuan Penelitian	2
1.4 Batasan Masalah	3
1.5 Manfaat Penelitian	3
1.6 Sistematika Penulisan	4
BAB II TINJAUAN PUSTAKA	6
2.1 Teori/Konsep Terkait	6
2.1.1 Cross Industry Standard Process for Data Mining (CRISP-DM)	6
2.1.2 <i>Machine Learning</i>	7
2.1.3 <i>Min-Max Normalization</i>	13
2.1.4 Matlab	14
2.1.5 Python	14
2.1.6 <i>Feature Selection</i>	14
2.1.7 <i>Genetic Algorithm</i>	15
2.1.7.1 Seleksi	16
2.1.7.2 <i>Crossover</i>	17
2.1.7.3 Mutasi	18
2.1.8 <i>Confusion Matrix</i>	18

2.2 Penelitian Terdahulu	20
2.3 Analisa Literature Review	35
BAB III METODE PENELITIAN	36
3.1 Deskripsi Sumber Data.....	36
3.2 Teknik Pengumpulan Data.....	36
3.3 Diagram Alir Penelitian.....	36
3.3.1 Pengumpulan Data.....	37
3.3.2 <i>Business Understanding</i>	38
3.3.3 <i>Data Understanding</i>	38
3.3.4 <i>Data Preparation</i>	38
3.3.5 <i>Modeling</i>	39
3.3.6 <i>Evaluation</i>	40
BAB IV HASIL & PEMBAHASAN	41
4.1 Pengumpulan Data	41
4.2 <i>Business Understanding</i>	42
4.3 <i>Data Understanding</i> (EDA).....	42
4.3.1 Distribusi Risiko Hipertensi.....	43
4.3.2 Analisis Variabel Kategorikal	43
4.3.3 Analisis Variabel Numerik.....	46
4.3.4 Korelasi Antar Variabel.....	49
4.4 <i>Data Preparation</i>	53
4.4.1 <i>Data Cleaning</i>	53
4.4.2 <i>Handle Imbalanced Data</i>	55
4.4.3 <i>Data Normalization</i>	56
4.4.4 <i>Feature Selection</i> dengan <i>Genetic Algorithm</i>	64
4.4.4.1 Parameter <i>Crossover Rate</i> = 0,3.....	65
4.4.4.2 Parameter <i>Crossover Rate</i> = 0,5.....	65
4.4.4.3 Parameter <i>Crossover Rate</i> = 0,7.....	66
4.4.4.4 Parameter <i>Crossover Rate</i> = 0,9.....	66
4.5 <i>Modeling</i>	68
4.5.1 Implementasi Algoritma KNN.....	69
4.5.2 Implementasi Algoritma Decision Tree	71
4.6 <i>Evaluation</i>	73
4.6.1 Evaluasi <i>Genetic Algorithm</i>	73

4.6.2 Evaluasi Model KNN.....	73
4.6.3 Evaluasi Model <i>Decision Tree</i>	76
4.6.4 Perbandingan Hasil Evaluasi	78
BAB V KESIMPULAN DAN SARAN	80
5.1 Kesimpulan.....	80
5.2 Saran	80
DAFTAR PUSTAKA.....	82
LAMPIRAN.....	89



UNIVERSITAS
MERCU BUANA

DAFTAR TABEL

Tabel 2. 1 Confusion Matrix	18
Tabel 2. 2 Literature Review	20
Tabel 4. 1 Deskripsi Atribut Dataset	41
Tabel 4. 2 Setting Parameter Genetic Algorithm	65
Tabel 4. 3 Hasil Crossover Rate = 0,3	65
Tabel 4. 4 Hasil Crossover Rate = 0,5	65
Tabel 4. 5 Hasil Crossover Rate = 0,7	66
Tabel 4. 6 Hasil Crossover Rate = 0,9	66
Tabel 4. 7 Hasil Feature Selection	67
Tabel 4. 8 Hasil Pengujian KNN Metric Euclidean	69
Tabel 4. 9 Hasil Pengujian KNN Metric Manhattan	70
Tabel 4. 10 Hasil Pengujian KNN Metric Minkowski	70
Tabel 4. 11 Hasil Pengujian Decision Tree Preset Fine Tree	71
Tabel 4. 12 Hasil Pengujian Decision Tree Preset Medium Tree	72
Tabel 4. 13 Hasil Pengujian Decision Tree Preset Coarse Tree	72
Tabel 4. 14 Confusion Matrix Model KNN	74
Tabel 4. 15 Confusion Matrix Model Decision Tree	76
Tabel 4. 16 Perbandingan Hasil Model KNN & Decision Tree	78

UNIVERSITAS
MERCU BUANA

DAFTAR GAMBAR

Gambar 2. 1 Alir CRISP-DM.....	6
Gambar 2. 2 Percabangan Ilmu Machine Learning	8
Gambar 2. 3 Konsep Supervised Learning.....	9
Gambar 2. 4 Konsep K-Nearest Neighbors (KNN).....	10
Gambar 2. 5 Konsep Decision Tree	11
Gambar 2. 6 Flowchart Prosedur Operasi Genetic Algorithm	16
Gambar 2. 7 Roulette Wheel Selection.....	17
Gambar 2. 8 Crossover Operation.....	17
Gambar 2. 9 Mutation Operation.....	18
Gambar 3. 1 Alir CRISP-DM yang Telah Disesuaikan.....	37
Gambar 3. 2 Fitur-fitur Dalam Dataset	38
Gambar 4. 1 Import Library.....	43
Gambar 4. 2 Distribusi Pengidap Penyakit Hipertensi	43
Gambar 4. 3 Distribusi Variabel male.....	44
Gambar 4. 4 Distribusi Variabel currentSmoker	45
Gambar 4. 5 Distribusi Variabel BPMeds.....	45
Gambar 4. 6 Distribusi Variabel Diabetes	46
Gambar 4. 7 Distribusi Variabel Age.....	46
Gambar 4. 8 Distribusi Variabel cigsPerDay.....	47
Gambar 4. 9 Distribusi Variabel totchol	47
Gambar 4. 10 Distribusi Variabel sysBP	48
Gambar 4. 11 Distribusi Variabel DiaBP	48
Gambar 4. 12 Distribusi Variabel BMI	48
Gambar 4. 13 Distribusi Variabel heartRate	49
Gambar 4. 14 Distribusi Variabel glucose.....	49
Gambar 4. 15 Scatterplot Variabel sysBP & diaBP	50
Gambar 4. 16 Scatterplot Variabel BMI & totChol.....	51
Gambar 4. 17 Scatterplot Variabel Age & heartRate	51
Gambar 4. 18 Correlation Matrix	53
Gambar 4. 19 Missing Values Tiap Variabel.....	54
Gambar 4. 20 Hasil Menghapus Missing Values	54

Gambar 4. 21 Jumlah Duplicated Data.....	55
Gambar 4. 22 Rename Column.....	55
Gambar 4. 23 Implementasi SMOTE.....	56
Gambar 4. 24 Implementasi Min-Max Normalization	57
Gambar 4. 25 Hasil Normalisasi Variabel Gender.....	57
Gambar 4. 26 Hasil Normalisasi Variabel Risk.....	58
Gambar 4. 27 Hasil Normalisasi Variabel Age	58
Gambar 4. 28 Hasil Normalisasi Variabel currentSmoker	59
Gambar 4. 29 Hasil Normalisasi Variabel cigsPerDay	59
Gambar 4. 30 Hasil Normalisasi Variabel BPMeds	60
Gambar 4. 31 Hasil Normalisasi Variabel diabetes.....	61
Gambar 4. 32 Hasil Normalisasi Variabel totChol.....	61
Gambar 4. 33 Hasil Normalisasi Variabel sysBP	62
Gambar 4. 34 Hasil Normalisasi Variabel diaBP	62
Gambar 4. 35 Hasil Normalisasi Variabel BMI.....	63
Gambar 4. 36 Hasil Normalisasi Variabel heartRate	63
Gambar 4. 37 Hasil Normalisasi Variabel glucose	64
Gambar 4. 38 Pseudo Code Genetic Algorithm.....	65
Gambar 4. 39 Hasil Fitness Value Genetic Algorithm	67
Gambar 4. 40 Splitting Data Validation & Data Test	68
Gambar 4. 41 Setting Parameter KNN	69
Gambar 4. 42 Evaluasi Genetic Algorithm.....	73
Gambar 4. 43 Evaluasi Model KNN.....	74
Gambar 4. 44 ROC Curve Model KNN.....	75
Gambar 4. 45 Evaluasi Model Decision Tree.....	76
Gambar 4. 46 ROC Curve Model Decision Tree.....	77

DAFTAR LAMPIRAN

Lampiran 1 Laporan Bimbingan Asistensi Ugniati Nuraftina.....	89
Lampiran 2 Laporan Bimbingan Asistensi Rizki Laksana Putra	90
Lampiran 3 Laporan Bimbingan Asistensi Arenga Pinnata Marzuki.....	91
Lampiran 4 Curriculum Vitae Ugniati Nuraftina	92
Lampiran 5 Curriculum Vitae Rizki Laksana Putra	93
Lampiran 6 Curriculum Vitae Arenga Pinnata Marzuki.....	94
Lampiran 7 Keterangan Telah Mengikuti BNSP - Ugniati Nuraftina	95
Lampiran 8 Keterangan Telah Mengikuti BNSP - Rizki Laksana Putra	96
Lampiran 9 Keterangan Telah Mengikuti BNSP - Arenga Pinnata Marzuki	97

