

**IMPLEMENTASI ALGORITMA *RANDOM FOREST*
BERBASIS *OVERSAMPLING SMOTE*
UNTUK KLASIFIKASI PENYAKIT ALZHEIMER**



**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS MUHAMMADIYAH JEMBER
2025**

TUGAS AKHIR

**IMPLEMENTASI ALGORITMA *RANDOM FOREST*
BERBASIS *OVERSAMPLING SMOTE*
UNTUK KLASIFIKASI PENYAKIT ALZHEIMER**

Disusun untuk Melengkapi Tugas dan Memenuhi Syarat Kelulusan
Program Strata 1 Jurusan Teknik Informatika Fakultas Teknik
Universitas Muhammadiyah Jember



Oleh:
Galung Reinan Artanca
NIM. 2310651108

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS MUHAMMADIYAH JEMBER
2025**

HALAMAN PERNYATAAN ORISINALITAS

Yang bertanda tangan di bawah ini:

Nama : Galung Reinan Artanca
NIM : 2310651108
Program Studi : Teknik Informatika
Perguruan Tinggi : Universitas Muhammadiyah Jember

Dengan ini menyatakan bahwa skripsi yang berjudul:

“Implementasi Algoritma *Random Forest* Berbasis *Oversampling SMOTE* Untuk Klasifikasi Penyakit Alzheimer” adalah hasil karya tulis saya sendiri. Seluruh isi dari skripsi ini merupakan hasil pemikiran, analisis, dan penyusunan yang saya lakukan secara mandiri, tanpa menjiplak atau menyalin dari karya pihak lain yang telah dipublikasikan atau diajukan sebelumnya, baik di Universitas Muhammadiyah Jember maupun di perguruan tinggi lain, kecuali disebutkan secara jelas dalam daftar pustaka.

Apabila di kemudian hari ditemukan adanya unsur plagiarisme, penyalinan tanpa izin, atau pelanggaran hak cipta lainnya, saya bersedia menerima segala konsekuensi akademik yang berlaku di Universitas Muhammadiyah Jember.

Demikian surat pernyataan ini saya buat dengan sebenar-benarnya untuk digunakan sebagaimana mestinya.

Jember, 20 Agustus 2025

Yang membuat pernyataan,



Galung Reinan Artanca

NIM.2310651108

HALAMAN PERSETUJUAN

**IMPLEMENTASI ALGORITMA *RANDOM FOREST*
BERBASIS *OVERSAMPLING SMOTE*
UNTUK KLASIFIKASI PENYAKIT ALZHEIMER**

Oleh:

Galung Reinan Artanca

2310651108

Telah disetujui bahwa Laporan Tugas Akhir ini diajukan dalam sidang Tugas Akhir sebagai bagian dari pemenuhan syarat kelulusan guna memperoleh gelar Sarjana Komputer (S.Kom).

di

Universitas Muhammadiyah Jember

Disetujui oleh,

Pembimbing I

Pembimbing II

Agung Nilogiri S.T., M.Kom.

Wiwik Suharso, S.Kom., M.Kom.

NIDN. 0030037701

NIDN. 0006097601

HALAMAN PENGESAHAN

**IMPLEMENTASI ALGORITMA *RANDOM FOREST*
BERBASIS *OVERSAMPLING SMOTE*
UNTUK KLASIFIKASI PENYAKIT ALZHEIMER**

Oleh:

Galung Reinan Artanca

2310651108

Telah mempertanggung jawabkan Laporan Tugas Akhirnya pada Sidang Tugas Akhir pada tanggal 20 Agustus 2025 sebagai salah satu syarat kelulusan dan mendapatkan gelar Sarjana Komputer (S.Kom.)

Pembimbing I



Agung Nilogiri S.T., M.Kom.
NIDN. 0030037701

Penguji I



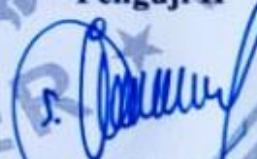
Ari Eko Wardoyo S.T., M.Kom.
NIDN. 0014027501

Pembimbing II



Wiwik Suharso, S.Kom., M.Kom.
NIDN. 0006097601

Penguji II



Ilham Saifudin, S.Pd., M.Si.
NIDN. 0731108903

**Mengesahkan,
Dekan Fakultas Teknik**



Dr. Ir. Muhtar, S.T., M.T., IPM
NIDN. 0010067301

**Mengetahui,
Ketua Program Studi Teknik
Informatika**



Rosita Yuniarti, S.Kom., M.Cs.
NIDN. 0629018601

MOTTO

*“Bahkan serigala penyendiri
melolong paling lantang dalam heningnya perjuangan,
suaranya membelah udara dingin,
menjadi lagu yang hanya dimengerti malam.*

*Serigala penyendiri berjalan sendirian,
namun tak pernah tersesat
setiap jejak kaki adalah janji,
setiap bayangan menjadi sahabat bisu.*

*Kesendirian menajamkan cakar,
perjuangan menguatkan jiwa.*

*Di tengah kesunyian,
hati menjadi baja,
tatapan pun tak tergoyahkan.*

*Dalam kegelapan,
serigala penyendiri belajar menjadi cahayanya sendiri,
menuntun langkah di tengah badai
★ tanpa bintang sebagai penunjuk. ★
Sebab kekuatan ditempa dalam kesendirian,
dan diasah oleh perjuangan.*

*Saat fajar merekah,
serigala penyendiri berdiri,
bukan sebagai mangsa malam,
melainkan sebagai penguasanya.”*

HALAMAN PERSEMBAHAN

Segala puji dan syukur penulis panjatkan ke hadirat Allah SWT, yang dengan kasih sayang dan rahmat-Nya telah memberikan kekuatan, kesabaran, serta kelapangan hati selama proses panjang penyusunan Tugas Akhir ini. Atas kehendak-Nya, setiap langkah, kegelisahan, kelelahan, dan harapan akhirnya dapat dirangkai menjadi sebuah karya ilmiah yang utuh. Tugas Akhir ini dengan penuh kerendahan hati penulis persembahkan kepada:

1. Allah SWT, Yang Maha Pendengar dan Maha Penyayang, atas segala jalan serta petunjuk yang telah diberikan kepada peneliti sehingga dapat menyelesaikan tugas akhir ini.
2. Kedua Orang Tua/Ibu dan Bapak, atas segala didikan, dukungan, pengorbanan, dan doa-doa yang di berikan hanya untuk melihat peneliti menjadi seorang sarjana.
3. Bapak Agung Nilogiri S.T., M.Kom. dan Bapak Wiwik Suharso S.Kom., M.Kom. selaku dosen pembimbing utama dan kedua, yang tak kenal lelah membimbing dan memberi arahan kepada peneliti.
4. Bapak Ari Eko Wardoyo S.T., M.Kom. dan Bapak Ilham Saifudin, S.Pd., M.Si. selaku dosen penguji utama dan kedua, yang senantiasa memberi banyak saran dan masukan, demi menyempurnakan tugas akhir ini.
5. Seluruh dosen serta karyawan Universitas Muhammadiyah Jember khususnya di program studi Teknik Informatika. Yang telah memberikan ilmu, fasilitas, dan pelayanan terbaik untuk peneliti saat menimba ilmu.
6. Teman-teman peneliti di program studi Teknik Informatika. Terima kasih atas semua bantuan, dukungan dan kenangan yang di berikan sampai detik ini.
7. Almamater tercinta Universitas Muhammadiyah Jember dan Fakultas Teknik Kebanggaan, yang telah menjadi wadah dari inovasi serta kreatifitas peneliti.

ABSTRAK

Artanca, Galung R. 2025. Implementasi Algoritma *Random Forest* Berbasis *Oversampling* SMOTE Untuk Klasifikasi Penyakit Alzheimer. Tugas Akhir. Program Sarjana. Program Studi Teknik Informatika. Universitas Muhammadiyah Jember.

Pembimbing: Agung Nilogiri S.T., M.Kom.; Wiwik Suharso, S.Kom. M.Kom.

Penyakit Alzheimer merupakan jenis demensia yang paling umum dan telah menjadi masalah kesehatan global yang serius, terutama di kalangan lansia. Deteksi dini terhadap penyakit ini sangat penting agar penderita dapat segera memperoleh perawatan yang tepat dan memperlambat perkembangan penyakit. Namun, metode diagnosis konvensional memerlukan waktu lama dan biaya tinggi. Penelitian ini bertujuan untuk mengimplementasikan algoritma *Random Forest* sebagai metode klasifikasi untuk mendeteksi penyakit Alzheimer secara lebih akurat dan efisien. Penelitian menggunakan dataset terbuka dari Kaggle yang terdiri dari 2.149 data pasien dan 35 atribut. Proses klasifikasi mengikuti metode CRISP-DM yang mencakup tahap *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modeling*, *Evaluation*, dan *Deployment*. Evaluasi dilakukan menggunakan metrik akurasi, precision, recall, dan F1-score. Penelitian ini diharapkan dapat menjadi kontribusi dalam pengembangan sistem pendukung keputusan untuk deteksi dini Alzheimer secara berbasis data.

Kata Kunci: Alzheimer, *Random Forest*, *Machine Learning*, Klasifikasi.

ABSTRACT

Artanca, Galung R. 2025. *Implementation of the SMOTE-Based Random Forest Algorithm for Alzheimer's Disease Classification*. Undergraduate Thesis. Bachelor's Program in Informatics Engineering, Universitas Muhammadiyah Jember.

Advisors: Agung Nilogiri, S.T., M.Kom.; Wiwik Suharso, S.Kom., M.Kom.

Alzheimer's disease is the most common type of dementia and has become a serious global health concern, especially among the elderly. Early detection of this disease is crucial for providing appropriate treatment and slowing its progression. However, conventional diagnostic methods are often time-consuming and costly. This study aims to implement the Random Forest algorithm as a classification method to detect Alzheimer's disease more accurately and efficiently. The research uses an open dataset from Kaggle, consisting of 2,149 patient records and 35 attributes. The classification process follows the CRISP-DM method, which includes the stages of Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment. Evaluation metrics used are accuracy, precision, recall, and F1-score. This research is expected to contribute to the development of data-driven decision support systems for early Alzheimer's detection.

Keywords: Alzheimer, Random Forest, Machine Learning, Classification.

KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Allah SWT atas segala limpahan rahmat, taufik, dan hidayah-Nya, sehingga penulis dapat menyelesaikan Tugas Akhir yang berjudul “**Implementasi Algoritma *Random Forest* Berbasis *Oversampling SMOTE* Untuk Klasifikasi Penyakit Alzheimer**”. Tugas Akhir ini disusun untuk memenuhi salah satu syarat dalam menyelesaikan pendidikan jenjang Sarjana (S1) pada Program Studi Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Jember.

Dalam proses penyusunan Tugas Akhir ini, penulis menyadari bahwa capaian ini tidak terlepas dari dukungan, doa, bimbingan, serta arahan dari berbagai pihak. Oleh karena itu, penulis mengucapkan terima kasih yang sebesar-besarnya kepada Bapak Agung Nilogiri, S.T., M.Kom., selaku Dosen Pembimbing I, Bapak Wiwik Suharso S.Kom., selaku Dosen Pembimbing II, yang telah banyak memberikan waktu, pengetahuan, dan motivasi kepada penulis selama proses bimbingan berlangsung. Ucapan terima kasih juga penulis sampaikan kepada Bapak Ari Eko Wardoyo S.T., M.Kom., selaku Dosen Penguji I, dan Bapak Ilham Saifudin, S.Pd., M.Si., selaku Dosen Penguji II, atas saran dan kritik konstruktif yang sangat membantu dalam perbaikan laporan ini.

Tidak lupa, penulis juga menyampaikan apresiasi kepada seluruh dosen dan staf di lingkungan Fakultas Teknik Universitas Muhammadiyah Jember atas segala dukungan yang diberikan selama masa studi.

Penulis menyadari bahwa laporan ini masih jauh dari kata sempurna, baik dalam penulisan maupun isi yang disampaikan. Oleh karena itu, saran dan masukan yang membangun sangat diharapkan demi penyempurnaan di masa mendatang. Semoga karya ini dapat memberikan manfaat dan menjadi referensi yang berguna bagi pembaca serta pengembangan ilmu di bidang informatika.

DAFTAR ISI

HALAMAN JUDUL	ii
HALAMAN PERNYATAAN ORISINALITAS.....	iii
HALAMAN PERSETUJUAN	iv
HALAMAN PENGESAHAN.....	v
MOTTO	vi
HALAMAN PERSEMBAHAN	vii
ABSTRAK	viii
ABSTRACT	ix
KATA PENGANTAR.....	x
DAFTAR ISI.....	xi
DAFTAR GAMBAR.....	xiv
DAFTAR TABEL	xv
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	2
1.3 Tujuan.....	3
1.4 Batasan Masalah.....	3
1.5 Manfaat	4
BAB II TINJAUAN PUSTAKA	5
2.1 Penelitian Terkait.....	5
2.2 Penyakit Demensia Alzheimer	7
2.3 <i>Machine Learning</i>	8
2.4 <i>Data Mining</i>	9
2.4.1 Klasifikasi	10

2.4.2 Estimasi	10
2.4.3 Prediksi.....	10
2.4.4 Klasterisasi (<i>Clustering</i>).....	11
2.4.5 Asosiasi	11
2.5 CRISP – DM (<i>Cross-Industry Standard Proces For Data Mining</i>).....	11
2.5.1 <i>Business Understanding</i> (Pemahaman Bisnis).....	12
2.5.2 <i>Data Understanding</i> (Pemahaman Data)	12
2.5.3 <i>Data Preparation</i> (Persiapan Data).....	12
2.5.4 <i>Modelling</i> (Pemodelan).....	12
2.5.5 <i>Evaluation</i> (Evaluasi model).....	13
2.5.6 <i>Deployment</i> (Penerapan)	13
2.6 <i>Decision Tree</i> C4.5.....	13
2.7 <i>Random Forest</i>	16
2.8 SMOTE (<i>Synthetic Minority Oversampling Technique</i>).....	18
2.9 <i>Confusion Matrix</i>	18
BAB III METODOLOGI PENELITIAN	21
3.1 Jenis Penelitian.....	22
3.2 <i>Business Understanding</i> (Pemahaman Bisnis).....	22
3.2.1 Identifikasi Masalah.....	22
3.2.2 Studi Literatur	22
3.3 <i>Data Understanding</i> (Pemahaman Data).....	23
3.3.1 Pengumpulan Data	23
3.3.2 Deskripsi dan Eksplorasi Data	23
3.4 <i>Data Preparation</i> (Persiapan Data).....	26
3.5 <i>Modeling</i> (Pemodelan).....	26
3.5.1 Sampling Data (Memilih data acak dari dataset)	26

3.5.2 Membuat Pohon Keputusan (<i>Decision Tree</i>)	28
3.5.3 Menentukan Hasil Prediksi Akhir	56
3.6 <i>Evaluation</i> (Evaluasi model).....	58
3.7 <i>Deployment</i> (Penerapan)	59
BAB IV HASIL DAN PEMBAHASAN.....	61
4.1 Lingkungan Pengujian	61
4.2 Eksplorasi <i>Dataset</i>	62
4.3 Data Preparation / Data Preprocessing.....	64
4.3.1 Seleksi Fitur	64
4.3.2 Standarisasi Fitur.....	65
4.3.3 Penanganan Ketidakseimbangan Data dengan SMOTE.....	66
4.4 Pemodelan <i>Random Forest</i>	66
4.5 Evaluasi Model.....	67
4.5.1 Evaluasi Tanpa Oversampling SMOTE	68
4.5.2 Evaluasi Menggunakan Oversampling SMOTE.....	71
4.6 Deployment.....	76
4.6.1 Penyimpanan Model.....	77
4.6.2 Pembuatan Antarmuka Aplikasi <i>Streamlit</i>	77
BAB V KESIMPULAN DAN SARAN.....	80
5.1 Kesimpulan	80
5.2 Saran.....	80
DAFTAR PUSTAKA.....	81
LAMPIRAN.....	83

DAFTAR GAMBAR

Gambar 2.1 Random Forest	17
Gambar 3.1 Tahapan Penelitian	21
Gambar 3.2 Pohon Keputusan Node Awal.....	50
Gambar 3.3 Pohon Keputusan 1.1.....	52
Gambar 3.4 Pohon Keputusan 2.1.....	53
Gambar 3.5 Struktur Pohon Keputusan	55
Gambar 3.6 <i>Data Flow Diagram</i> (Dfd) Level 0	59
Gambar 3.7 <i>Data Flow Diagram</i> (Dfd) Level 1	60
Gambar 4.1 Visualisasi Distribusi Data	62
Gambar 4.2 <i>Heatmap Korelasi</i> Fitur.....	63
Gambar 4.3 Distribusi Data Diagnosis	64
Gambar 4.4 Standarisasi Fitur.....	65
Gambar 4.5 Perbandingan Data Sebelum Dan Sesudah Diterapkan SMOTE	66
Gambar 4.6 Perbandingan Evaluasi Model.....	74
Gambar 4.7 <i>Confusion Matrix</i>	76
Gambar 4.8 Penyimpanan <i>Scaler</i> Dan Model.....	77
Gambar 4.9 <i>Import Library</i> Dan Memuat Model	78
Gambar 4.10 Tampilan Aplikasi Web	78

DAFTAR TABEL

Tabel 2.1 <i>Confusion Matrix</i>	18
Tabel 3.1 Deskripsi <i>Dataset</i>	23
Tabel 3.2 Sampling Data	27
Tabel 3.3 Tabel Atribut <i>Age</i>	28
Tabel 3.4 Tabel Atribut <i>Gender</i>	28
Tabel 3.5 Tabel Atribut <i>Ethnicity</i>	29
Tabel 3.6 Tabel Atribut <i>Educationlevel</i>	29
Tabel 3.7 Tabel Atribut <i>Bmi</i>	30
Tabel 3.8 Tabel Atribut <i>Smoking</i>	30
Tabel 3.9 Tabel Atribut <i>Alcoholconsumption</i>	30
Tabel 3.10 Tabel Atribut <i>Physicalactivity</i>	31
Tabel 3.11 Tabel Atribut <i>Dietquality</i>	31
Tabel 3.12 Tabel Atribut <i>Sleepquality</i>	31
Tabel 3.13 Tabel Atribut <i>Familyhistoryalzheimers</i>	32
Tabel 3.14 Tabel Atribut <i>Cardiovascular disease</i>	32
Tabel 3.15 Tabel Atribut <i>Diabetes</i>	32
Tabel 3.16 Tabel Atribut <i>Depression</i>	33
Tabel 3.17 Tabel Atribut <i>Headinjury</i>	33
Tabel 3.18 Tabel Atribut <i>Hypertension</i>	33
Tabel 3.19 Tabel Atribut <i>Systolicbp</i>	33
Tabel 3.20 Tabel Atribut <i>Diastolicbp</i>	34
Tabel 3.21 Tabel Atribut <i>Cholesteroltotal</i>	34
Tabel 3.22 Tabel Atribut <i>Cholesterolldl</i>	35
Tabel 3.23 Tabel Atribut <i>Cholesterolhdl</i>	35
Tabel 3.24 Tabel Atribut <i>Cholesteroltriglycerides</i>	35
Tabel 3.25 Tabel Atribut <i>Mmse</i>	36
Tabel 3.26 Tabel Atribut <i>Functionalassessment</i>	36
Tabel 3.27 Tabel Atribut <i>Memorycomplaints</i>	36
Tabel 3.28 Tabel Atribut <i>Behavioralproblems</i>	37
Tabel 3.29 Tabel Atribut <i>Adl</i>	37

Tabel 3.30 Tabel Atribut <i>Confusion</i>	38
Tabel 3.31 Tabel Atribut <i>Disorientation</i>	38
Tabel 3.32 Tabel Atribut <i>Personalitychanges</i>	38
Tabel 3.33 Tabel Atribut <i>Difficultycompletingtasks</i>	38
Tabel 3.34 Tabel Atribut <i>Forgetfulness</i>	39
Tabel 3.35 Hasil Perhitungan Node Awal	49
Tabel 3.36 Hasil Perhitungan Node 1.1	51
Tabel 3.37 Hasil Perhitungan Node 2	52
Tabel 3.38 Hasil Perhitungan Node 3	54
Tabel 3.39 Hasil Prediksi Pohon Keputusan	56
Tabel 3.40 Hasil Prediksi Pohon Keputusan 1	56
Tabel 3.41 Hasil Prediksi Pohon Keputusan 2	57
Tabel 3.42 Hasil Prediksi Pohon Keputusan 3	57
Tabel 3.43 Tabel Confusion Matrix.....	58
Tabel 4.1 Spesifikasi Perangkat Keras	61
Tabel 4.2 Spesifikasi Perangkat Lunak	61
Tabel 4.3 Tabel Hasil Pengujian Tanpa Menggunakan <i>Oversampling</i> Smote	68
Tabel 4.4 Tabel Hasil Pengujian Menggunakan <i>Oversampling</i> Smote	71