

**PERBANDINGAN ALGORITMA ARTIFICIAL NEURAL
NETWORK DAN RANDOM FOREST UNTUK ANALISIS
PENYAKIT DIABETES DENGAN VISUALISASI**

SKRIPSI

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



disusun oleh

MUHAMMAD HADI GUNAWAN

21.11.4495

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2025

**PERBANDINGAN ALGORITMA ARTIFICIAL NEURAL
NETWORK DAN RANDOM FOREST UNTUK ANALISIS
PENYAKIT DIABETES DENGAN VISUALISASI**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

MUHAMMAD HADI GUNAWAN

21.11.4495

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2025

HALAMAN PERSETUJUAN

SKRIPSI

**PERBANDINGAN ALGORITMA ARTIFICIAL NEURAL
NETWORK DAN RANDOM FOREST UNTUK ANALISIS
PENYAKIT DIABETES DENGAN VISUALISASI**

yang disusun dan diajukan oleh

Muhammad Hadi Gunawan

21.11.4495

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 8 Agustus 2025

Dosen Pembimbing,


Mulia Sulistiyono, M.KOM

NIK. 190302248

HALAMAN PENGESAHAN
SKRIPSI
PERBANDINGAN ALGORITMA ARTIFICIAL NEURAL NETWORK
DAN RANDOM FOREST UNTUK ANALISIS PENYAKIT DIABETES
DENGAN VISUALISASI

yang disusun dan diajukan oleh

Muhammad Hadi Gunawan

21.11.4495

Telah dipertahankan di depan Dewan Penguji
pada tanggal 17 September 2025

Susunan Dewan Penguji

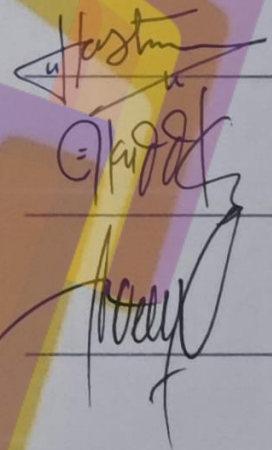
Nama Penguji

Tanda Tangan

Hastari Utama, S. Kom., M.Cs.
NIK. 190302230

Anna Baita, S.Kom., M.Kom.
NIK. 190302290

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



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 17 September 2025

DEKAN FAKULTAS ILMU KOMPUTER



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

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Muhammad Hadi Gunawan
NIM : 21.11.4495

Menyatakan bahwa Skripsi dengan judul berikut:

PERBANDINGAN ALGORITMA ARTIFICIAL NEURAL NETWORK DAN RANDOM FOREST UNTUK ANALISIS PENYAKIT DIABETES DENGAN VISUALISASI

Dosen Pembimbing : Mulia Sulistiyono., 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, 17 September 2025

Yang Menyatakan,



Muhammad Hadi Gunawan

HALAMAN PERSEMBAHAN

1. Allah SWT atas rahmat dan karunia-Nya yang senantiasa menyertai setiap langkah.
2. Ayah dan Ibu tercinta atas doa, kasih sayang, serta pengorbanan yang tiada henti.
3. Saudara-saudara dan keluarga besar yang selalu memberikan semangat dan dukungan.
4. Bapak Mulia Sulistiyono, M.Kom selaku dosen pembimbing yang dengan sabar membimbing serta memberikan arahan.
5. Para sahabat dan teman – teman yang senantiasa memberikan semangat, dukungan , dan motivasi untuk menyelesaikan tugas akhir ini.

KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Allah SWT atas limpahan rahmat, hidayah, dan Karunia+Nya, sehingga penulis dapat menyelesaikan skripsi dengan judul “Perbandingan Analisa Algoritma Artificial Neural Network dan Random Forest untuk analisis penyakit diabetes dengan visualisasi”. Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana pada Program Studi Informatika, Fakultas Ilmu Komputer, Universitas Amikom Yogyakarta.

Dengan segala kerendahan hati, penulis mengucapkan terimakasih yang sebesar-besarnya kepada semua pihak yang telah memberikan dukungan, bantuan, dan doa selama proses penyusunan skripsi ini. Ucapan terimakasih secara khusus penulis sampaikan kepada:

1. Bapak Prof. Dr. M. Suyanto, M.M., selaku Rektor Universitas Amikom Yogyakarta.
2. Ibu Prof. Dr, Kusriani, M.Kom., selaku Dekan Fakultas Ilmu Komputer.
3. Ibu Eli Pujastuti, M.Kom., selaku Ketua Program Studi Informatika.
4. Bapak Mulya Sulistiyono, M.Kom., selaku Dosen Pembimbing yang telah memberikan arahan, bimbingan, dan motivasi selama proses penyusunan skripsi ini.
5. Kedua orang tua, keluarga, serta teman – teman tercinta atas doa, semangat, dan dukungan yang selalu mengiri langkah penulis.

Penulis menyadari bahwa skripsi ini masih jauh dari kesempurnaan. Oleh karena itu, penulis sangat mengharapkan kritik dan saran yang membangun demi perbaikan di masa yang akan datang

Yogyakarta, 17 September 2025

Penulis

DAFTAR ISI

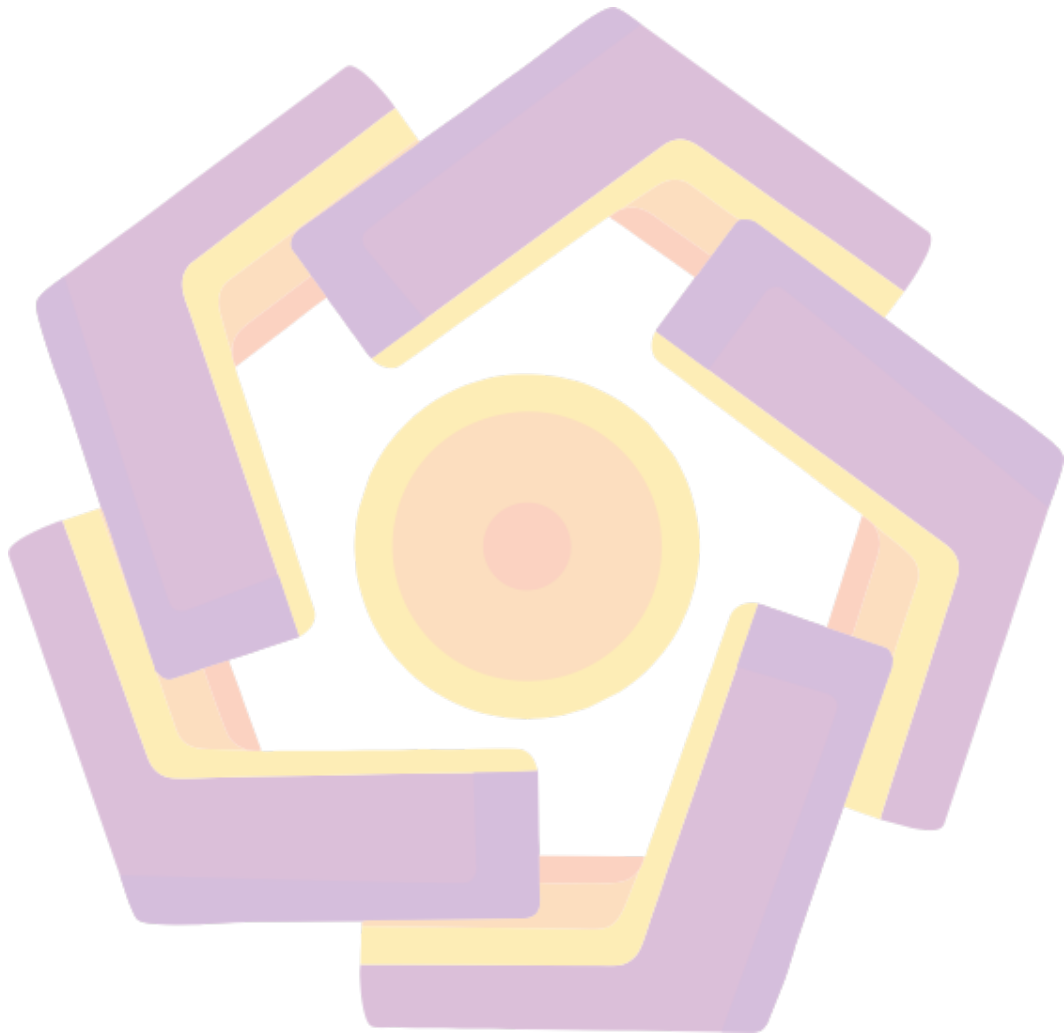
HALAMAN JUDUL	i
HALAMAN PERSETUJUAN	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	iv
HALAMAN PERSEMBAHAN	v
KATA PENGANTAR	vi
DAFTAR ISI	vii
DAFTAR TABEL	x
DAFTAR GAMBAR	xi
DAFTAR LAMPIRAN	xii
DAFTAR LAMBANG DAN SINGKATAN	xiii
DAFTAR ISTILAH	xiv
INTISARI	xv
<i>ABSTRACT</i>	xvi
BAB I PENDAHULUAN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

1.6	Sistematika Penulisan	4
BAB II TINJAUAN PUSTAKA		5
2.1	Studi Literatur	5
2.2	Dasar Teori	8
2.2.1	Penyakit Diabetes	8
2.2.2	Artificial Neural Network (ANN)	9
2.2.3	Random Forest (RF)	10
2.2.4	Preprocessing Data	11
2.2.5	SMOTE (Synthetic Minority Oversampling Technique)	11
2.2.6	Tuning Hyperparameter	12
2.2.7	Perbandingan ANN dan RF dalam Klasifikasi Diabetes	12
2.2.8	Visualisasi Data dalam Klasifikasi Diabetes	13
BAB III METODE PENELITIAN		14
3.1	Objek Penelitian	14
3.2	Alur Penelitian	14
3.2.1	Pengumpulan Data	15
3.2.2	Pre-processing Data	15
3.3	Rancangan Penelitian	18
3.4	Data dan Sumber Data	18
3.5	Tahapan Penelitian	21
3.6	Alat dan Bahan	22
BAB IV HASIL DAN PEMBAHASAN		23

4.1	Deskripsi Dataset	23
4.2	Pembagian data (Splitting Data)	24
4.3	Preprocessing Data	26
4.3.1	Data Transformation	27
4.3.2	Feature Selection	29
4.4	Rancang Model	30
4.5	Evaluasi Model	33
BAB V PENUTUP		37
5.1	Kesimpulan	37
5.2	Saran	37
REFERENSI		38
LAMPIRAN		41

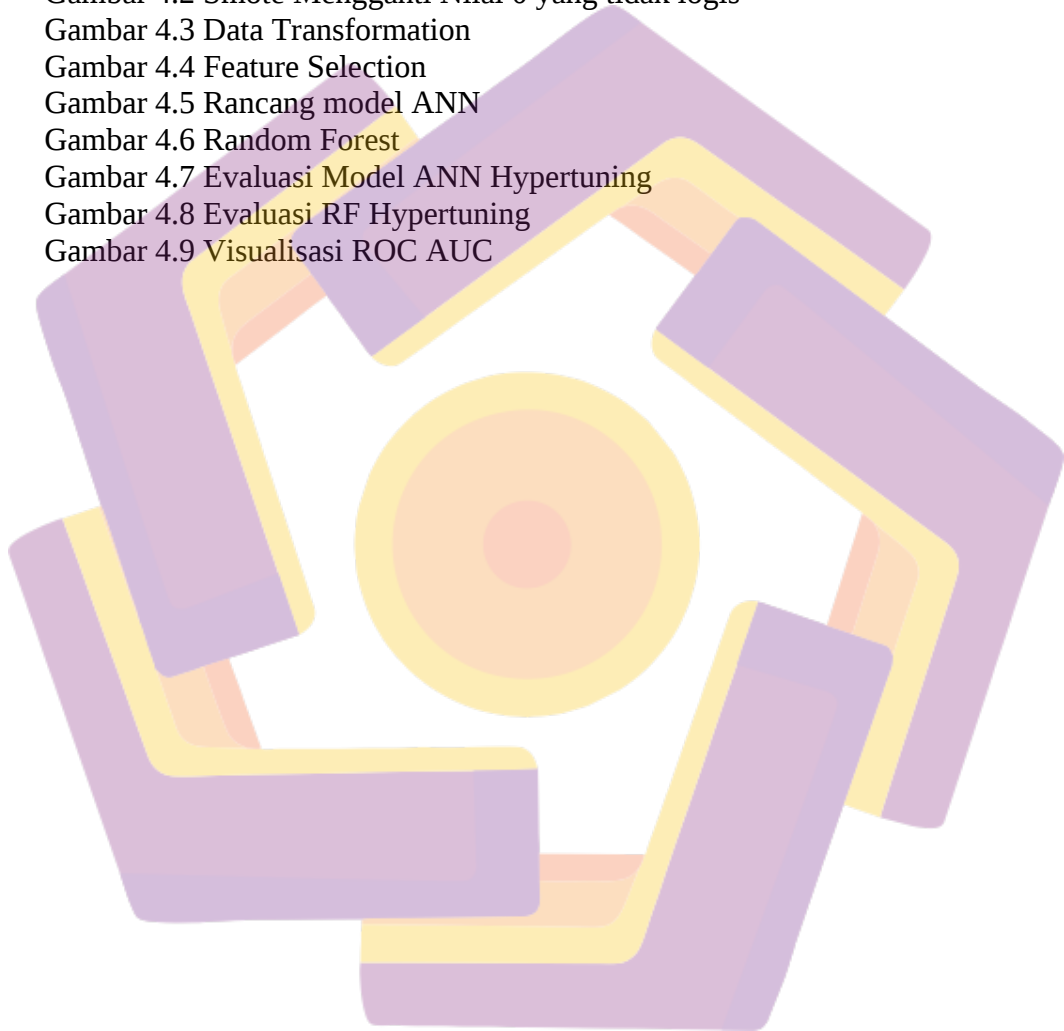
DAFTAR TABEL

Tabel 2.1 Keaslian Penelitian	6
Tabel 4.1 Atribut Utama	23
Tabel 4.2 Pembagian Data	25



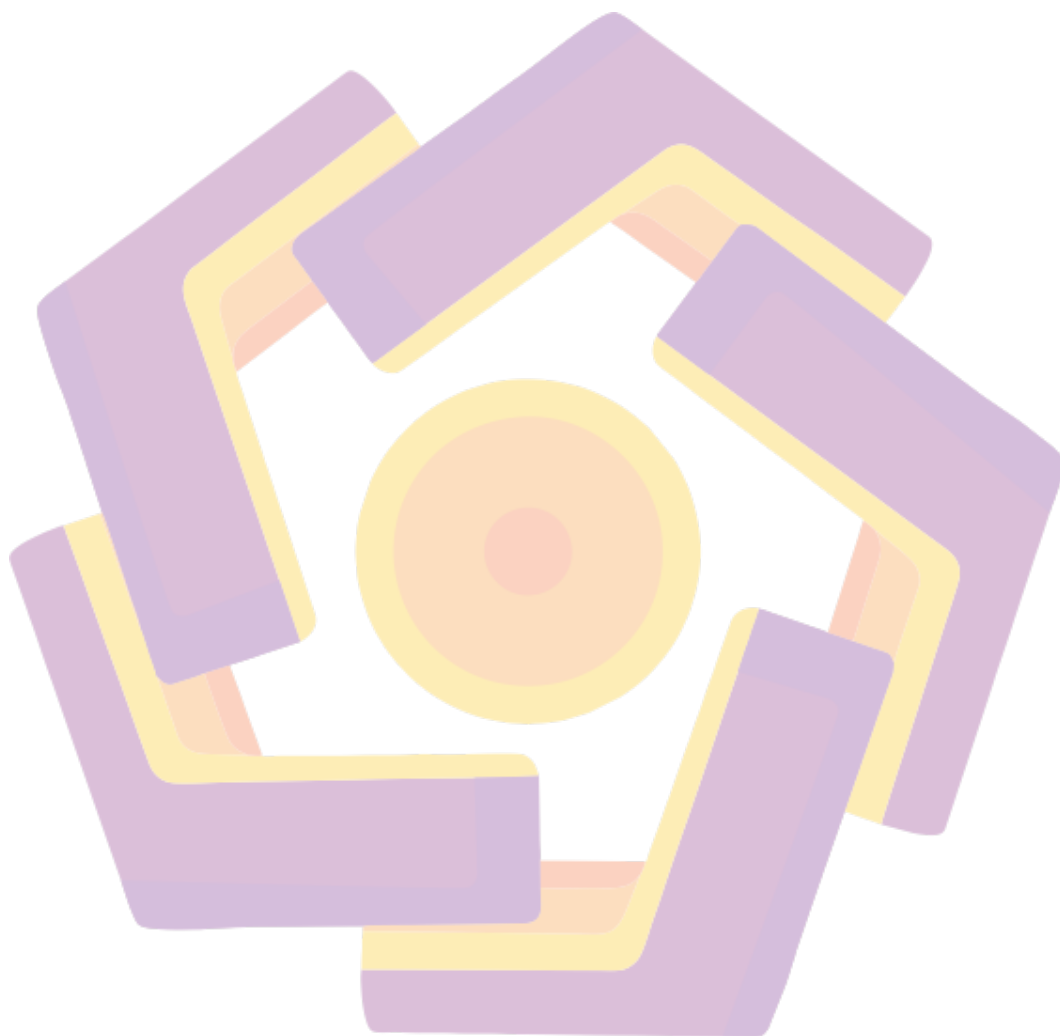
DAFTAR GAMBAR

Gambar 3.1 Alur Penelitian	14
Gambar 4.1 Splitting Data	25
Gambar 4.2 Smote Mengganti Nilai 0 yang tidak logis	27
Gambar 4.3 Data Transformation	28
Gambar 4.4 Feature Selection	29
Gambar 4.5 Rancang model ANN	30
Gambar 4.6 Random Forest	32
Gambar 4.7 Evaluasi Model ANN Hypertuning	33
Gambar 4.8 Evaluasi RF Hypertuning	34
Gambar 4.9 Visualisasi ROC AUC	35



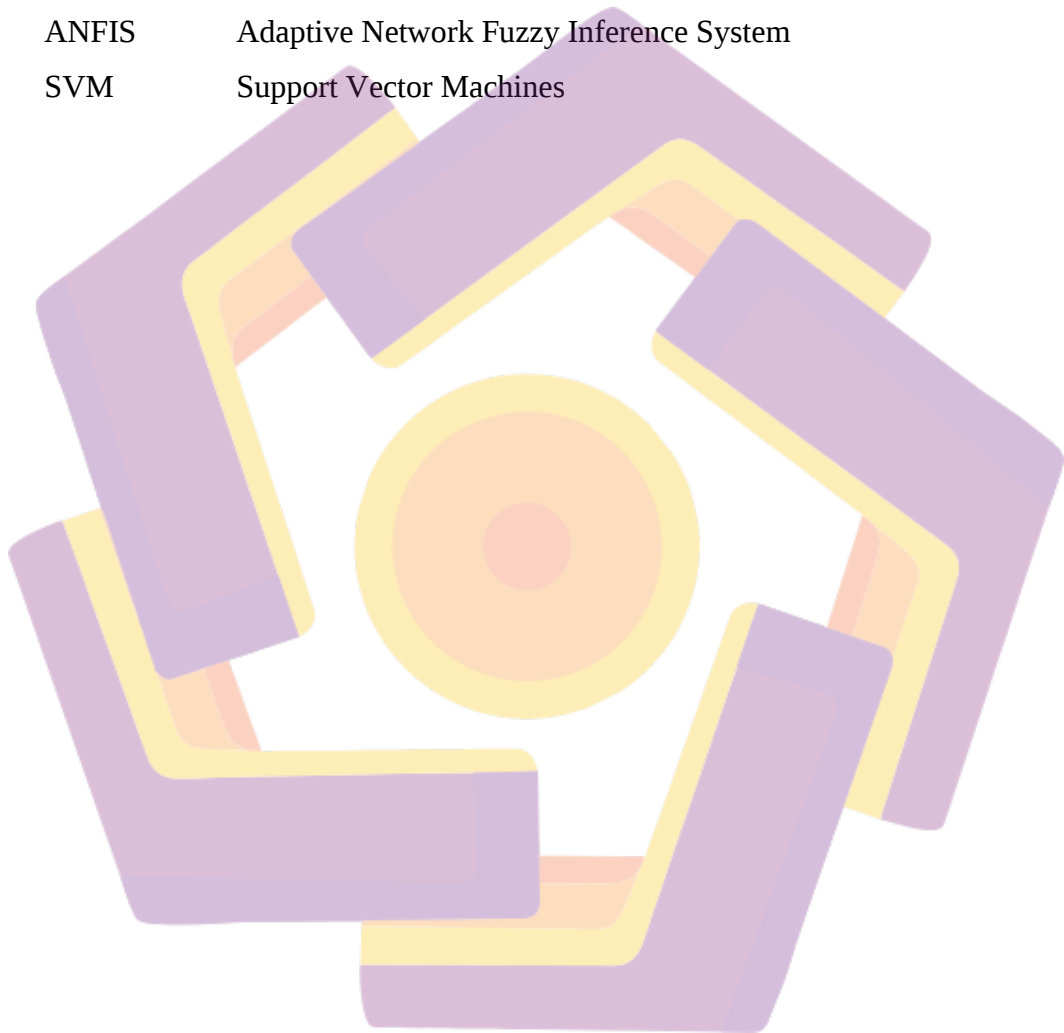
DAFTAR LAMPIRAN

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



DAFTAR LAMBANG DAN SINGKATAN

Ω	Tahanan Listrik
μ	Konstanta gesekan
ANFIS	Adaptive Network Fuzzy Inference System
SVM	Support Vector Machines



DAFTAR ISTILAH

ANN (Artificial Neural Network) : Algoritma jaringan saraf tiruan yang meniru cara kerja otak manusia.

RF (Random Forest) : Algoritma ensemble learning berbasis pohon keputusan.

SMOTE (Synthetic Minority Over-sampling Technique) : Teknik penyeimbangan data dengan membuat sampel sintetis pada kelas minoritas.

Akurasi (Accuracy) : Persentase prediksi benar dari seluruh data uji.

Presisi (Precision) : Tingkat ketepatan model dalam memprediksi kelas positif.

Recall (Sensitivity) : Kemampuan model mendeteksi semua kasus positif.

F1-Score : Rata-rata harmonis presisi dan recall.

AUC-ROC (Area Under Curve – Receiver Operating Characteristic) : Indikator kemampuan model membedakan kelas positif dan negatif.

Confusion Matrix : Tabel perbandingan hasil prediksi model dengan nilai sebenarnya. Feature Importance : Ukuran kontribusi setiap fitur dalam model Random Forest.

INTISARI

Diabetes mellitus merupakan penyakit kronis dengan prevalensi yang terus meningkat dan berpotensi menimbulkan komplikasi serius apabila tidak terdeteksi sejak dini, sehingga diperlukan metode pendukung yang lebih akurat dalam menganalisis risiko diabetes. Penelitian ini dilakukan untuk membandingkan kinerja algoritma Artificial Neural Network (ANN) dan Random Forest (RF) dalam klasifikasi diabetes dengan menggunakan dataset publik dari Kaggle. Tahapan penelitian meliputi pra-pemrosesan data melalui penanganan nilai tidak logis, normalisasi, dan penyeimbangan kelas dengan Synthetic Minority Over-sampling Technique (SMOTE), kemudian model dibangun menggunakan algoritma ANN dan RF serta dievaluasi menggunakan metrik akurasi, presisi, recall, F1-score, dan AUC-ROC dengan tambahan visualisasi berupa confusion matrix, kurva ROC, heatmap korelasi, dan feature importance. Hasil penelitian menunjukkan bahwa algoritma Random Forest memberikan performa lebih baik dengan akurasi 77,67%, presisi 75,93%, recall 81,46%, F1-score 78,59%, dan AUC 77,64%, sedangkan Artificial Neural Network memperoleh akurasi 75,67%, presisi 71,43%, recall 86,09%, F1-score 78,08%, dan AUC 75,60%. Penelitian ini diharapkan dapat memperkaya literatur akademik mengenai perbandingan algoritma machine learning di bidang kesehatan serta memberikan dasar teoritis bagi penelitian lanjutan terkait analisis penyakit diabetes.

Kata kunci: Diabetes Mellitus, ANN, Random Forest, Machine Learning, Klasifikasi

ABSTRACT

Diabetes mellitus is a chronic disease with an increasing prevalence that can lead to severe complications if not detected early, thus requiring more accurate methods to support risk analysis. This study aims to compare the performance of Artificial Neural Network (ANN) and Random Forest (RF) algorithms in diabetes classification using a public dataset from Kaggle. The research process involved data preprocessing through handling invalid values, normalization, and class balancing with the Synthetic Minority Over-sampling Technique (SMOTE), followed by model construction using ANN and RF, which were then evaluated with accuracy, precision, recall, F1-score, and AUC-ROC metrics, as well as visualizations including confusion matrix, ROC curve, correlation heatmap, and feature importance. The results showed that Random Forest achieved better performance with 77.67% accuracy, 75.93% precision, 81.46% recall, 78.59% F1-score, and 77.64% AUC, while Artificial Neural Network obtained 75.67% accuracy, 71.43% precision, 86.09% recall, 78.08% F1-score, and 75.60% AUC. This research is expected to enrich academic literature on machine learning algorithm comparison in the healthcare field and provide a theoretical foundation for further studies related to diabetes analysis..

Keyword: Diabetes Mellitus, ANN, Random Forest, Machine Learning, Classification