

**ANALISIS DETEKSI PENYAKIT DIABETES DENGAN
KELAS YANG TIDAK SEIMBANG DITANGANI
TEKNIK SMOTE-ENN PADA ALGORITMA
MACHINE LEARNING**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi S1-Teknik Komputer



disusun oleh

FRANSISKUS WILIS ALIT NARANTOKO

21.83.0647

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2026

**ANALISIS DETEKSI PENYAKIT DIABETES DENGAN
KELAS YANG TIDAK SEIMBANG DITANGANI
TEKNIK SMOTE-ENN PADA ALGORITMA
MACHINE LEARNING**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi S1-Teknik Komputer



disusun oleh

FRANSISKUS WILIS ALIT NARANTOKO

21.83.0647

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2026

HALAMAN PERSETUJUAN

SKRIPSI

**ANALISIS DETEKSI PENYAKIT DIABETES DENGAN
KELAS YANG TIDAK SEIMBANG DITANGANI
TEKNIK SMOTE-ENN PADA ALGORITMA
MACHINE LEARNING**

yang disusun dan diajukan oleh

Fransiskus Wilis Alit Narantoko

21.83.0647

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 03 Februari 2026

Dosen Pembimbing,

Dr. Dony Arivus, S.S., M.Kom.

NIK. 190302128

HALAMAN PENGESAHAN

SKRIPSI

**ANALISIS DETEKSI PENYAKIT DIABETES DENGAN
KELAS YANG TIDAK SEIMBANG DITANGANI
TEKNIK SMOTE-ENN PADA ALGORITMA
MACHINE LEARNING**

yang disusun dan diajukan oleh

Fransiskus Wilis Alit Narantoko

21.83.0647

Telah dipertahankan di depan Dewan Penguji
pada tanggal 24 Februari 2026

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Senie Destya, S.T., M.Kom.
NIK. 190302312

Wahid Miftahul Ashari, S.Kom., M.T.
NIK. 190302452

Dr. Dony Ariyus, S.S., M.Kom.
NIK. 190302128



Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 03 Maret 2026

DEKAN FAKULTAS ILMU KOMPUTER



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

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Fransiskus Wilis Alit Narantoko
NIM : 21.83.0647

Menyatakan bahwa Skripsi dengan judul berikut:

ANALISIS DETEKSI PENYAKIT DIABETES DENGAN KELAS YANG TIDAK SEIMBANG DITANGANI TEKNIK SMOTE-ENN PADA ALGORITMA MACHINE LEARNING

Dosen Pembimbing : Dr. Dony Ariyus S.S., 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, 03 Februari 2026

Yang Menyatakan,



Fransiskus Wilis Alit Narantoko

HALAMAN PERSEMBAHAN

Segala puji dan rasa syukur saya haturkan kepada Tuhan Yang Maha Esa atas limpahan rahmat, kasih, dan bimbingan-Nya, sehingga saya dapat menyelesaikan skripsi ini dengan baik. Dengan penuh rasa hormat dan kerendahan hati, karya sederhana ini saya persembahkan kepada:

1. Orang tua tercinta, atas doa, kasih, dan dukungan yang tiada henti dalam setiap langkah hidup saya.
2. Prof. Dr. M. Suyanto., M.M selaku Rektor Universitas Amikom Yogyakarta.
3. Bapak Dr. Dony Ariyus S.S., M.Kom. selaku Dosen Pembimbing dari Teknik Komputer yang telah memberikan bimbingan dan bantuan kepada penulis.
4. Keluarga besar, yang senantiasa memberi semangat, perhatian, dan dorongan moral dalam perjalanan ini, serta tempat berbagi kebahagiaan atas pencapaian yang
5. Sahabat dan teman seperjuangan, atas kebersamaan, motivasi, dan dukungan yang membuat perjalanan ini lebih bermakna.

KATA PENGANTAR

Puji syukur saya panjatkan ke hadirat Allah SWT atas limpahan rahmat, taufik, serta karunia-Nya, sehingga saya dapat menuntaskan penyusunan skripsi yang berjudul “Analisis Deteksi Penyakit Diabetes Dengan Ketidakseimbangan Kelas Yang Ditangani Teknik SMOTE ENN Pada Algoritma Machine Learning”. Karya ilmiah ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer pada Program Studi Teknik Komputer, Universitas Amikom Yogyakarta.

Dalam proses penyusunan skripsi ini, saya mendapatkan banyak dukungan, bimbingan, dan bantuan dari berbagai pihak. Oleh sebab itu, dengan penuh rasa hormat dan kerendahan hati, saya menyampaikan ucapan terima kasih yang sebesar-besarnya kepada:

1. Orang tua tercinta, atas doa, kasih, dan dukungan yang tiada henti dalam setiap langkah hidup saya.
2. Prof. Dr. M. Suyanto., M.M selaku Rektor Universitas Amikom Yogyakarta.
3. Bapak Dr. Dony Ariyus S.S., M.Kom. selaku Dosen Pembimbing dari Teknik Komputer yang telah memberikan bimbingan dan bantuan kepada penulis.

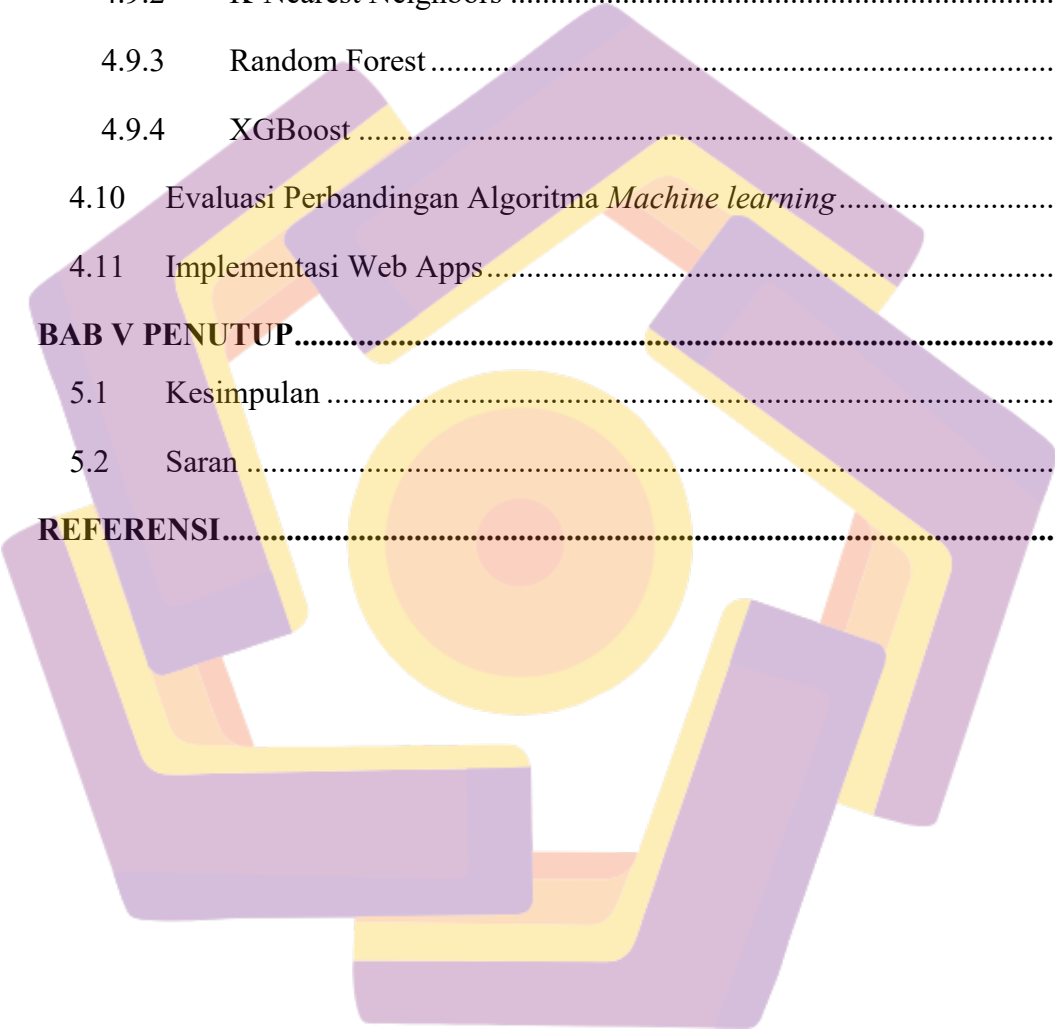
Yogyakarta, 03 Februari 2026

Fransiskus Wilis Alit Narantoko

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 GAMBAR.....	x
DAFTAR TABEL	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	5
BAB II TINJAUAN PUSTAKA	7
2.1 Studi Literatur	7
2.2 Dasar Teori.....	15
2.2.1 Supervised Learning	15
2.2.2 Machine learning	15
2.2.3 Penyakit Diabetes.....	16
2.2.4 <i>Label Encoder</i>	17

2.2.5	<i>StandardScaler</i>	17
2.2.6	<i>SMOTE ENN</i>	18
2.2.7	<i>Logistic Regession</i>	19
2.2.8	<i>K-Nearest Neighbors</i>	20
2.2.9	Random Forest	20
2.2.10	XGBoost	21
2.2.11	Evaluation Model.....	22
BAB III METODE PENELITIAN		24
3.1	Objek Penelitian.....	24
3.2	Alur Penelitian	25
3.2.1	Data Collection	26
3.2.2	Praprocessing Data.....	27
3.2.3	Algoritma <i>Machine learning</i>	31
3.2.4	Evaluation Model.....	34
3.2.5	Implementation of Web Apps.....	35
3.3	Alat dan Bahan.....	35
3.3.1	Bahan	35
3.3.2	Alat/Instrumen	36
BAB IV HASIL DAN PEMBAHASAN		38
4.1	Data Collection	38
4.2	Exploratory Data Analysis (EDA).....	42
4.3	Data Cleaning.....	46
4.4	Label Encoder	50
4.5	StandardScaler	52
4.6	Data Split	54



4.7	<i>SMOTE ENN</i>	55
4.8	Distribusi Dataset.....	56
4.9	Klasifikasi Algoritma <i>Machine learning</i>	58
4.9.1	Logistic Regression.....	58
4.9.2	K-Nearest Neighbors	60
4.9.3	Random Forest	62
4.9.4	XGBoost	64
4.10	Evaluasi Perbandingan Algoritma <i>Machine learning</i>	66
4.11	Implementasi Web Apps.....	74
BAB V PENUTUP		80
5.1	Kesimpulan	80
5.2	Saran	81
REFERENSI		82

DAFTAR GAMBAR

Gambar 3. 1 Alur Penelitian	30
Gambar 4. 1 Kode Import Dataset Untuk Menampilkan Tabel	43
Gambar 4. 2 Semua Fitur Dalam Dataset	44
Gambar 4. 3 Kode Memeriksa Struktur Data	45
Gambar 4. 4 Struktur Data	46
Gambar 4. 5 Kode Melihat Kelas Diabetes Dalam Dataset	47
Gambar 4. 6 Kode Melihat Statistik Fitur Numerik	47
Gambar 4. 7 Hasil Statistik Fitur Numerik	48
Gambar 4. 8 Kode Menghitung Korelasi Dengan Tampilan Heatmap	49
Gambar 4. 9 Hasil Heatmap Korelasi Antar Fitur	50
Gambar 4. 10 Kode Melihat Data Missing Value	51
Gambar 4. 11 Jumlah Missing Value Sebelum Data Cleanig	52
Gambar 4. 12 Kode Menangani Missing Value	53
Gambar 4. 13 Hasil Memeriksa Missing Value	54
Gambar 4. 14 Kode Menangani Data Duplicate	55
Gambar 4. 15 Kode Label Encoding	55
Gambar 4. 16 Hasil Penggunaan Label Encoder	56
Gambar 4. 17 Kode Standarisasi Menggunakan StandardScaler	57
Gambar 4. 18 Hasil Standarisasi Menggunakan StandardScaler	58
Gambar 4. 19 Kode Data Split	59
Gambar 4. 20 Kode Penerapan Augmentasi SMOTE-ENN	60
Gambar 4. 21 Distribusi Kelas Sebelum Augmentasi SMOTE-ENN	61
Gambar 4. 22 Distribusi Kelas Sesudah Augmentasi SMOTE-ENN	62
Gambar 4. 23 Kode Pelatihan Model Algoritma Logistic Regression	63
Gambar 4. 24 Classification Report Model Algoritma Logistic Regression	64
Gambar 4. 25 Kode Pelatihan Model Algoritma K-Nearest Neighbors	65
Gambar 4. 26 Classification Report Model Algoritma K-Nearest Neighbors	66
Gambar 4. 27 Kode Pelatihan Model Algoritma Random Forest	67
Gambar 4. 28 Classification Report Model Algoritma Random Forest	68

Gambar 4. 29 Kode Pelatihan Model Algoritma XGBoost	69
Gambar 4. 30 Classification Report Model Algoritma XGBoost	70
Gambar 4. 31 Perbandingan Performa Pada Setiap Model Algoritma	71
Gambar 4. 32 Confusion Matrix Pada Setiap Model	73
Gambar 4. 33 Receiver Operating Characteristic Pada Setiap Model Algoritma	74
Gambar 4. 34 K-Fold Validation Pada Setiap Model Algoritma	75
Gambar 4. 35 Learning Curve Setiap Algoritma	77
Gambar 4. 36 Kode Import Library Pada App.py	78
Gambar 4. 37 Kode Fungsi Prediksi Pada app.py	79
Gambar 4. 38 Route Menuju Templates Pada app.py	80
Gambar 4. 39 Tampilan awal web	81
Gambar 4. 40 Tampilan Form Cek Diabetes	82
Gambar 4. 41 Tampilan Hasil Cek Diabetes	83



DAFTAR TABEL

Table 2. 1. Keaslian Penelitian	11
Table 3. 1. Confusion Matrix	39
Table 4. 1. Peforma Pada Setiap Model Algoritma	70



INTISARI

Diabetes melitus merupakan salah satu penyakit kronis dengan prevalensi tinggi yang dapat menimbulkan berbagai komplikasi serius apabila tidak terdeteksi sejak dini. Penerapan machine learning dalam deteksi diabetes sering menghadapi permasalahan ketidakseimbangan kelas (class imbalance), di mana jumlah data penderita diabetes jauh lebih sedikit dibandingkan data non-diabetes, sehingga menurunkan kemampuan model dalam mengenali kelas minoritas. Penelitian ini bertujuan untuk menganalisis efektivitas teknik SMOTE-ENN (Synthetic Minority Over-sampling Technique with Edited Nearest Neighbors) dalam menangani ketidakseimbangan kelas serta membandingkan kinerja beberapa algoritma machine learning dalam mendeteksi penyakit diabetes. Dataset yang digunakan adalah Behavioral Risk Factor Surveillance System (BRFSS) tahun 2015 yang terdiri dari 253.680 data dengan 22 variabel indikator kesehatan. Tahapan penelitian meliputi preprocessing data, penerapan SMOTE-ENN, pelatihan model menggunakan algoritma Logistic Regression, K-Nearest Neighbors, Random Forest, dan XGBoost, serta evaluasi model menggunakan metrik accuracy, precision, recall, F1-score, dan AUC-ROC. Hasil penelitian menunjukkan bahwa penerapan SMOTE-ENN mampu meningkatkan performa model dalam mendeteksi kelas minoritas diabetes. Berdasarkan hasil evaluasi, algoritma Random Forest memberikan performa terbaik dengan nilai akurasi sebesar 95,85%, sehingga dinilai paling efektif dalam mendeteksi penyakit diabetes. Model terbaik selanjutnya diimplementasikan dalam bentuk aplikasi web berbasis Flask untuk membantu prediksi risiko diabetes secara interaktif.

Kata kunci: Diabetes Melitus, Machine Learning, SMOTE-ENN, Random Forest, Klasifikasi.

ABSTRACT

Diabetes mellitus is a chronic disease with high prevalence that can cause various serious complications if not detected early. The application of machine learning in diabetes detection often overcomes the problem of class imbalance, where the number of data of diabetics is much less than the data of non-diabetics, thus reducing the model's ability to recognize minority classes. This study aims to analyze the effectiveness of the SMOTE-ENN (Synthetic Minority Over-sampling Technique with Edited Nearest Neighbors) technique in handling class performance and compare the performance of several machine learning algorithms in detecting diabetes. The dataset used is the Behavioral Risk Factor Surveillance System (BRFSS) in 2015 which consists of 253,680 data with 22 health indicator variables. The research stages include data preprocessing, application of SMOTE-ENN, model training using the Logistic Regression, K-Nearest Neighbors, Random Forest, and XGBoost algorithms, as well as model evaluation using accuracy, precision, recall, F1-score, and AUC-ROC metrics. The results showed that the application of SMOTE-ENN improved the model's performance in detecting minority diabetes classes. Based on the evaluation, the Random Forest algorithm performed best with an accuracy of 95.85%, making it the most effective in detecting diabetes. The best model was then implemented in a Flask-based web application to assist with interactive diabetes risk prediction.

Keyword: Diabetes Mellitus, Machine Learning, SMOTE-ENN, Random Forest, Classification.