

**OPTIMASI DIAGNOSIS DIABETES MENGGUNAKAN
MACHINE LEARNING DENGAN SMOTE DAN
FEATURE SELECTION**

SKRIPSI NON REGULER - SCIENTIST

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



disusun oleh

REFDIAN SHELLA SELINA

21.11.4462

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2025

**OPTIMASI DIAGNOSIS DIABETES MENGGUNAKAN
MACHINE LEARNING DENGAN SMOTE DAN
FEATURE SELECTION**

SKRIPSI NON REGULER - SCIENTIST

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



disusun oleh

REFDIAN SHELLA SELINA

21.11.446

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2025

HALAMAN PERSETUJUAN

SKRIPSI NON REGULER - SCIENTIST

**OPTIMASI DIAGNOSIS DIABETES MENGGUNAKAN
MACHINE LEARNING DENGAN SMOTE DAN
FEATURE SELECTION**

yang disusun dan diajukan oleh

Refdian Shella Selina

21.11.4462

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 13 Februari 2025

Dosen Pembimbing,



Majid Rahardj, S.Kom., M.Eng
NIK.190302393

HALAMAN PENGESAHAN

SKRIPSI

OPTIMASI DIAGNOSIS DIABETES MENGGUNAKAN
MACHINE LEARNING DENGAN SMOTE DAN
FEATURE SELECTION

yang disusun dan diajukan oleh

Refdian Shella Selina

21.11.4462

Telah dipertahankan di depan Dewan Penguji
pada tanggal 13 Februari 2025

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Harvoko, S.Kom., M.Cs
NIK. 190302286

Dr. Emigawaty, M.Kom
NIK. 190302226

Mauid Rahardi, S.Kom., M.Eng
NIK. 190302393

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 13 Februari 2025

DEKAN FAKULTAS ILMU KOMPUTER



Hanif Al Fatta, S.Kom., M.Kom., Ph.D.
NIK. 190302096

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Refdian Shella Selina
NIM : 21.11.4462

Menyatakan bahwa Skripsi dengan judul berikut:

Optimasi Diagnosis Diabetes menggunakan Machine Learning dengan SMOTE dan Feature Selection

Dosen Pembimbing : Majid Rahardi, S.Kom., M.Eng

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, 13 Februari 2025

Yang Menyatakan,



Refdian Shella Selina

HALAMAN PERSEMBAHAN

Segala puji bagi Allah SWT atas segala rahmat, karunia, dan hidayah-Nya yang telah diberikan sehingga penulis dapat menyelesaikan skripsi ini dengan baik. Skripsi ini dipersembahkan kepada :

1. Bapak Prof. Dr. M. Suyanto, MM, selaku Rektor Universitas AMIKOM Yogyakarta.
2. Bapak Hanif Al Fatta, S.Kom., M.Kom., Ph.D., selaku Dekan Fakultas Ilmu Komputer.
3. Ibu Windha Mega Pradnya Duhita, M.Kom, selaku Ketua Program Studi Informatika.
4. Bapak Majid Rahardi, S.Kom., M.Eng, selaku Dosen Pembimbing, atas arahan, dukungan, dan kesabaran selama proses penelitian dan penulisan skripsi ini.
5. Bapak Haryoko, S.Kom., M.Cs dan Ibu Dr. Emigawaty, M.Kom, selaku Dosen Penguji, atas masukan dan bimbingan yang sangat berharga selama proses penyusunan skripsi ini.
6. Orang tua tercinta, yang telah memberikan doa, dukungan moral, dan motivasi yang tak ternilai selama masa studi.
7. Teman-teman dan rekan seperjuangan, atas kebersamaan dan semangat yang terus diberikan dalam setiap langkah perjalanan akademik ini.

Penulis menyadari bahwa dalam penyusunan laporan akhir ini masih terdapat kekurangan, sehingga saran dan kritik yang membangun sangat penulis harapkan. Harapannya, laporan akhir ini dapat memberikan manfaat dan menjadi sumber inspirasi bagi pembaca.

KATA PENGANTAR

Puji syukur ke hadirat Allah SWT atas segala rahmat, karunia, dan hidayah-Nya sehingga penulis dapat menyelesaikan skripsi non reguler – scientist yang berjudul "Optimasi Diagnosis Diabetes menggunakan Machine Learning dengan SMOTE dan Feature Selection" ini dengan baik. Dalam kesempatan ini, penulis ingin menyampaikan rasa terima kasih yang sebesar-besarnya kepada berbagai pihak yang telah memberikan dukungan dan bimbingan selama proses penyusunan skripsi ini :

1. Bapak Majid Rahardi, S.Kom., M.Eng, selaku Dosen Pembimbing, yang dengan penuh kesabaran telah memberikan arahan, masukan, dan dukungan dalam setiap tahapan penelitian ini.
2. Bapak Haryoko, S.Kom., M.Cs dan Ibu Dr. Emigawaty, M.Kom, selaku Dosen Penguji, yang telah memberikan kritik dan saran yang sangat berharga untuk penyempurnaan skripsi ini.
3. Orang tua tercinta yang senantiasa memberikan doa, dukungan moral, dan material selama masa studi.
4. Semua pihak yang telah membantu, baik secara langsung maupun tidak langsung, yang tidak dapat disebutkan satu per satu.

Penulis menyadari bahwa skripsi ini masih jauh dari sempurna. Oleh karena itu, kritik dan saran yang membangun sangat penulis harapkan untuk pengembangan di masa yang akan datang. Semoga skripsi ini dapat memberikan manfaat bagi pembaca dan menjadi kontribusi kecil bagi perkembangan ilmu pengetahuan.

Yogyakarta, 13 Februari 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
INTISARI	viii
ABSTRACT	ix
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
BAB II TINJAUAN PUSTAKA	3
2.1 Studi Literatur	3
BAB III METODE PENELITIAN	6
3.1 <i>Dataset</i>	6
3.2 <i>Exploratory Data Analysis (EDA)</i>	8
3.3 <i>Data Preprocessing</i>	9
3.4 <i>Classification Modelling</i>	13
3.5 <i>Performance Evaluation</i>	14
BAB IV HASIL DAN PEMBAHASAN	15
BAB V PENUTUP	20
REFERENSI	21
LAMPIRAN	23

INTISARI

Penelitian ini mengkaji penggunaan *machine learning* dalam pengklasifikasian diabetes dengan tujuan meningkatkan akurasi diagnosis. Tiga algoritma klasifikasi—*Random Forest*, *Logistic Regression*, dan *Decision Tree*—telah dievaluasi, di mana *Random Forest* mencapai tingkat akurasi tertinggi sebesar 97,53% dan menunjukkan kinerja yang lebih baik dibandingkan algoritma lainnya. Untuk mengatasi ketidakseimbangan kelas, penelitian ini menerapkan *Synthetic Minority Over-sampling Technique (SMOTE)*, yang secara signifikan meningkatkan kemampuan model dalam mendeteksi *instance* pada kelas minoritas. *Feature selection* juga dilakukan untuk mengoptimalkan performa model. Metodologi penelitian ini mengikuti pendekatan terstruktur, mencakup *Exploratory Data Analysis (EDA)*, *Data Preprocessing*, *Model Classification*, dan *Performance Evaluation*. Hasil penelitian ini menggarisbawahi pentingnya pemilihan teknik yang sesuai untuk meningkatkan presisi dan keandalan diagnosis diabetes. Dengan mengintegrasikan *SMOTE* dan *feature selection*, penelitian ini menunjukkan peningkatan signifikan dalam akurasi dan efisiensi sistem diagnosis diabetes, yang berpotensi memberikan dampak positif terhadap hasil kesehatan. Temuan ini menekankan pentingnya penerapan metode *machine learning* yang canggih dalam sektor kesehatan untuk mendukung deteksi dan pengelolaan penyakit secara lebih efektif.

Kata kunci: *Machine Learning*, Klasifikasi Diabetes, *SMOTE*, *Feature Selection*, Evaluasi Model.

ABSTRACT

This study explores the use of machine learning for diabetes categorization, aiming to enhance diagnostic accuracy. Three classification algorithms—Random Forest, Logistic Regression, and Decision Tree—are evaluated, with Random Forest achieving the highest accuracy of 97.53%, outperforming the others. To address class imbalance, the research employs the Synthetic Minority Over-sampling Technique (SMOTE), improving the model's ability to detect minority class instances. Feature selection is also implemented to optimize model performance. The methodology follows a structured approach, including Exploratory Data Analysis (EDA), Data Preprocessing, Model Classification, and Performance Evaluation. The findings highlight the importance of selecting appropriate techniques to improve the precision and reliability of diabetes diagnosis. By incorporating SMOTE and feature selection, the study demonstrates significant improvements in the accuracy and efficiency of diabetes diagnostic systems, potentially leading to better health outcomes. These results emphasize the value of integrating advanced machine learning methods in healthcare for more effective disease detection and management.

Keyword: Machine Learning, Diabetes Classification, SMOTE, Feature Selection, Model Evaluation.