KLASIFIKASI PENYAKIT DIABETES *MELLITUS* MENGGUNAKAN METODE *SYNTHETIC MINORITY OVER-SAMPLING TECHNIQUE* (SMOTE) *RANDOM FOREST*

SKRIPSI

Diajukan untuk memenuhi salah satu syarat kelulusan
di Program Studi Sains Data

Disusun Oleh:

NOFA AULIYATUL MAULIDIYYAH

20083010029

UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN" JAWA TIMUR

FAKULTAS ILMU KOMPUTER

PROGRAM STUDI SAINS DATA

SURABAYA

2024

LEMBAR PENGESAHAN

KLASIFIKASI PENYAKIT *DIABETES MELLITUS* MENGGUNAKAN *METODE SYNTHETIC MINORITY OVER-SAMPLING TECHNIQUE* (SMOTE) *RANDOM FOREST*

SKRIPSI

Diajukan untuk memenuhi salah satu syarat memperoleh gelar Sarjana Sains Data
pada : Senin, 15 juli 2024

Program Studi S-1 Sains Data
Fakultas Ilmu Komputer
Universitas Pembangunan Nasional Veteran Jawa Timur
Surabaya

Oleh :

NOFA AULIYATUL MAULIDIYYAH
NPM. 20083010029

Disetujui oleh Tim Penguji Skripsi :

Penguji 1	Penguji 2
Tresna Maulana Fahrudin, S.ST., M.T.	**Amri Muhaimin, S.Stat., M.Stat., M.S.**
NIP. 199305012022030100 7	NIP. 21119950723270
Pembimbing 1	Pembimbing 2
Trimono, S.Si., M.Si	**Aviolla Terza Damaliana, S.Si., M.Stat**
NIP 199509082022031003	NIP 199408022022032015

Mengetahui,

Fakultas Ilmu Komputer	Program Studi Sains Data
Dekan,	Fakultas Ilmu Komputer
	Koordinator,
Prof. Dr. Ir. Novirina Hendrasarie, MT	**Dr.Eng.Ir.Dwi Arman Prasetya, ST.,MT.,IPU.**
NIP 196811261994032001	NIP. 198012052005011002

Surabaya, Juli, 2024

SURAT PERNYATAAN

Saya yang bertanda tangan di bawah ini:

Nama	: Nofa Auliyatul Maulidiyyah
NPM	: 20083010029
Program Studi	: Sains Data

Menyatakan bahwa judul Skripsi / Tugas Akhir sebagai berikut:

Klasifikasi Penyakit *Diabetes Mellitus* Menggunakan Metode *Synthetic Minority Over-Sampling Technique* (SMOTE) Random Forest

Bukan merupakan plagiat dari Skripsi/ Tugas Akhir/ Penelitian orang lain dan juga bukan merupakan produk/ *software*/ hasil karya yang saya beli dari orang lain

Saya juga menyatakan bahwa Skripsi/ Tugas Akhir ini adalah pekerjaan saya sendiri, kecuali yang dinyatakan dalam Daftar Pustaka, dan tidak pernah diajukan untuk syarat memperoleh gelar di Universitas Pembangunan Nasional "Veteran" Jawa Timur maupun di institusi pendidikan lain.
Jika ternyata dikemudian hari pernyataan ini terbukti tidak benar, maka Saya bertanggung jawab penuh dan siap menerima segala konsekuensi, termasuk pembatalan ijazah dikemudian hari

Surabaya, 30 April 2024

Hormat Saya

Nofa Auliyatul Maulidiyyah
NPM.

ABSTRAK

KLASIFIKASI PENYAKIT *DIABETES MELLITUS* MENGGUNAKAN METODE *SYNTHETIC MINORITY OVER-SAMPLING TECHNIQUE* (SMOTE) *RANDOM FOREST*

Nama Mahasiswa / NPM	**:**	**Nofa Auliyatul Maulidiyyah / 20083010029**
Program Studi	**:**	**Sains Data, FASILKOM,UPN Veteran Jatim**
Dosen Pembimbing 1	**:**	**Trimono, S.Si., M.Si**
Dosen Pembimbing 2	**:**	**Aviolla Terza Damaliana, S.Si., M.Stat**

Abstrak

Indonesia menempati peringkat kelima di dunia, dengan 19,47 juta orang diperkirakan menderita diabetes pada tahun 2021, dan akan meningkat menjadi 23,32 pada tahun 2030. Penyakit diabetes dapat dicegah dengan melakukan deteksi dini dan melakukan pendekatan machine learning untuk memprediksi penyakit tersebut. Namun sering kali prediksi menjadi kurang akurat karena data yang tersedia tidak seimbang dalam distribusi kelas. Penelitian ini bertujuan untuk menangani masalah ketidakseimbangan data serta membandingkan performa model dalam memprediksi penyakit diabetes. Metode yang digunakan yaitu metode *Synthetic Minority Oversampling Technique* (SMOTE) untuk mengatasi ketidakseimbangan data, *Random Forest* untuk mengklasifikasikan penyakit diabetes dan *Confusion Matrix* untuk menghitung evaluasi kinerja model. Hasil penelitian menunjukkan bahwa model yang dilatih dengan data yang melalui proses SMOTE menunjukkan akurasi lebih tinggi yaitu 98% dibandingkan model tanpa SMOTE yang menghasilkan akurasi 93%. Penelitian menunjukkan bahwa penggunaan SMOTE dalam proses pelatihan model *Random Forest* secara signifikan mengurangi kesalahan prediksi dan meningkatkan akurasi klasifikasi.

Kata kunci: *Diabetes Mellitus, Machine Learning, SMOTE, Random Forest, Confusin Matrix*

ABSTRACT

CLASSIFICATION OF DIABETES MELLITUS USING THE SYNTHETIC MINORITY OVER-SAMPLING TECHNIQUE (SMOTE) RANDOM FOREST METHOD

Student Name / NPM	**:**	**Nofa Auliyatul Maulidiyyah / 20083010029**
Study Program	**:**	**Sains Data, FASILKOM,UPN Veteran Jatim**
Advisor 1	**:**	**Trimono, S.Si., M.Si**
Advisor 2	**:**	**Aviolla Terza Damaliana, S.Si., M.Stat**

Abstract

Indonesia ranks fifth globally, with an estimated 19.47 million people suffering from diabetes in 2021, projected to rise to 23.32 million by 2030. Diabetes can be prevented by carrying out early detection and using a machine learning approach to predict the disease. However, predictions often become less accurate because the available data is not balanced in the class distribution. This research aims to address the problem of data imbalance and compare the performance of models in predicting diabetes. The methods used are the Synthetic Minority Oversampling Technique (SMOTE) method to overcome data imbalances, Random Forest to classify diabetes, and Confusion Matrix to calculate model performance evaluations. The research results show that the model trained with data that went through the SMOTE process showed higher accuracy, namely 98%, compared to the model without SMOTE, which produced an accuracy of 93%. Research shows that the use of SMOTE in the Random Forest model training process significantly reduces prediction errors and increases classification accuracy.

Keywords : *Diabetes Mellitus, Machine Learning, SMOTE, Random Forest, Confusin Matrix*

KATA PENGANTAR

Segala puja dan puji syukur atas kehadirat Allah SWT, yang telah melimpahkan rahmat dan hidayahnya kepada kita semua, sehingga penulis dapat menyelesaikan Tugas Akhir yang merupakan persyaratan dalam menyelesaikan mata kuliah Skripsi pada Program Studi S1 Sains Data di Universitas Pembangunan Nasional "Veteran" Jawa Timur.

Shalawat dan salam semoga tetap tercurahkan kepada Nabi besar kita Nabi Muhammad SAW yang telah membimbing kita dari jalan kegelapan menuju jalan terang benderang yakni addinul islam ahlussunna waljamaah.

Dalam penyusunan Skripsi ini, penulis tidak ada kata yang dapat penulis ucapkan, kecuali berterimakasih yang sebesar-besarnya kepada semua pihak yang telah membantu dalam penulisan ini baik secara langsung maupun tidak langsung. Maka dalam kesempatan ini, penulis mengucapkan terimakasih kepada :

1. Dua orang paling berjasa dalam hidup penulis, ayah dan ibu. Terimakasih atas kepercayaan yang telah diberikan atas izin merantau dari kalian, serta cinta, do'a, motivasi, semangat, dan nasihat-nasihat nya. Semoga Allah SWT selalu menjaga kalian dalam kebaikan dan kemudahan.
2. Prof. Dr. Ir. Akhmad Fauzi, M.MT., IPU selaku Rektor Universitas Pembangunan Nasional "Veteran" Jawa Timur.
3. Ibu Prof. Dr. Ir. Novirina Hendrasarie, M.T.selaku Dekan Fakultas Ilmu Komputer Universitas Pembangunan Nasional "Veteran" Jawa Timur
4. Bapak Dr. Eng. Ir. Dwi Arman Prasetya, ST., MT., IPU selaku Koordinator Program Studi Sains Data Universitas Pembangunan Nasional "Veteran" Jawa Timur.
5. Bapak Trimono, S.Si., M.Si dan Ibu Aviolla Terza Damaliana, S.Si., M.Stat selaku Dosen Pembimbing 1 dan 2.
6. Bapak dan Ibu dosen Program Studi Sains Data UPN "Veteran" Jawa Timur yang sudah berkenan untuk memberikan waktu untuk berkontribusi dalam penelitian ini.
7. Inisial AR yang memaksa untuk namanya dimasukkan ke dalam buku skripsi, terimakasih atas bantuan terhadap jalannya pembuatan skripsi ini.

8. Terakhir, kepada diri saya sendiri, terimakasih sudah berjuang dan bertahan sejauh ini. Apresiasi sebesar-besarnya karena bertanggung jawab untuk menyelesaikan apa yang telah dimulai. Terimakasih untuk tidak menyerah dalam hal sesulit apapun dalam proses penyusunan skripsi ini. Tetap bersyukur dan rendah hati.

Penulis menyadari bahwa masih terdapat banyak kekurangan dalam penyusunan Skripsi ini, namun penulis berharap semoga Skripsi ini dapat memberikan kontribusi terhadap perkembangan ilmu pengetahuan, khususnya dalam bidang ilmu sains data.

Surabaya, 05 Juli 2024

Nofa Auliyatul Maulidiyyah

DAFTAR ISI

DAFTAR GAMBAR

DAFTAR TABEL

DAFTAR LAMPIRAN