



UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAKARTA

**KLASIFIKASI TIPE PENYAKIT DIABETES
MENGUNAKAN ALGORITMA *SUPPORT VECTOR
MACHINE* (SVM)**

SKRIPSI

**NUR AFIIFAH AZ-ZAHRA
2010511085**

**PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS PEMBANGUNAN NASIONAL “VETERAN” JAKARTA
2023**



UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAKARTA

**KLASIFIKASI TIPE PENYAKIT DIABETES
MENGUNAKAN ALGORITMA *SUPPORT VECTOR
MACHINE* (SVM)**

SKRIPSI

**Diajukan Sebagai Salah Satu Syarat Untuk Memperoleh
Gelar Sarjana Komputer**

**NUR AFIIFAH AZ-ZAHRA
2010511085**

**PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN" JAKARTA**

PERNYATAAN ORISINALITAS

PERNYATAAN ORISINALITAS

Tugas akhir ini adalah hasil karya sendiri, dan semua sumber yang dikutip maupun yang dirujuk telah saya nyatakan dengan benar.

Nama : Nur Afifah Az-Zahra

NIM : 2010511085

Tanggal : 20 Juli 2024

Bilamana dikemudian hari ditemukan ketidaksesuaian dengan pernyataan saya ini, maka saya bersedia dituntut dan diproses sesuai dengan ketentuan yang berlaku.

Depok, 20 Juli 2024

Yang menyatakan,



(Nur Afifah Az-Zahra)

PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS

PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS

Sebagai civitas akademik Universitas Pembangunan Nasional “Veteran” Jakarta, saya yang bertanda tangan di bawah ini :

Nama : Nur Afifah Az-Zahra

NIM : 2010511085

Fakultas : Ilmu Komputer

Program Studi : Informatika

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Pembangunan Nasional “Veteran” Jakarta Hak Bebas Royalti Non eksklusif (*Non-exclusive Royalty Free Right*) atas karya ilmiah saya yang berjudul :

KLASIFIKASI TIPE PENYAKIT DIABETES MENGGUNAKAN ALGORITMA *SUPPORT VECTOR MACHINE* (SVM).

Dengan Hak Bebas Royalti ini Universitas Pembangunan Nasional “Veteran” Jakarta berhak menyimpan, mengalih media/formatkan, mengelola dalam bentuk pangkalan data (database), merawat, dan mempublikasikan Tugas Akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta. Demikian Pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Depok

Pada tanggal : 20 Juli 2024

Yang menyatakan,



(Nur Afifah Az-Zahra)

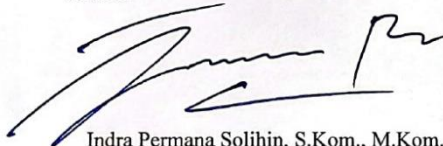
LEMBAR PENGESAHAN

LEMBAR PENGESAHAN

Skripsi diajukan oleh :

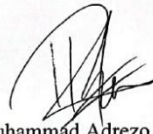
Nama : Nur Afifah Az-Zahra
NIM : 2010511085
Program Studi : S-1 Informatika
Judul Skripsi : **KLASIFIKASI TIPE PENYAKIT DIABETES MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE (SVM).**

Telah berhasil dipertahankan di hadapan Tim Penguji dan diterima sebagai bagian dari persyaratan yang diperlukan untuk memperoleh gelar Sarjana Komputer pada Program Studi S-1 Informatika, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional "Veteran" Jakarta.



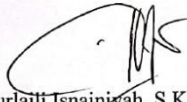
Indra Permana Solihin, S.Kom., M.Kom.

Penguji I



Muhammad Adrezo, S.Kom., M.Sc

Penguji II



Ika Nurlaili Isnainiyah, S.Kom., M.Sc.

Pembimbing I

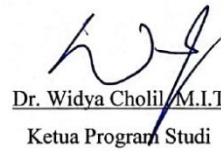


Nurul Afifah Arifuddin, S.Pd., M.T.

Pembimbing II



Prof. Dr. Ir. Supriyanto, ST., M.Sc., IPM
Dekan



Dr. Widya Cholil, M.I.T
Ketua Program Studi

Ditetapkan di : Jakarta

Tanggal Ujian : 15 Juli 2024

KATA PENGANTAR

Puji syukur penulis panjatkan kehadirat Allah SWT atas limpahan rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan penelitian ini yang berjudul "Klasifikasi Tipe Penyakit Diabetes Menggunakan Algoritma *Support Vector Machine* (SVM)". Penelitian ini tidak akan dapat terselesaikan tanpa dukungan dan bantuan dari berbagai pihak. Oleh karena itu, penulis ingin menyampaikan penghargaan dan terima kasih kepada semua pihak yang telah memberikan dukungan, baik secara moral maupun material. Ucapan terima kasih penulis sampaikan kepada:

1. Bapak Prof.Dr.Ir Supriyanto, S.T., M.Sc selaku dekan Fakultas Ilmu Komputer UPN Veteran Jakarta
2. Ibu Ika Nurlaili Isnainiyah, S.Kom.,M.Sc selaku dosen pembimbing I tugas akhir yang telah memberikan bimbingan, arahan, dan saran serta motivasinya yang sangat berharga selama proses penelitian ini.
3. Ibu Nurul Afifah Arifuddin, S.Pd., M.T, selaku dosen pembimbing II tugas akhir yang telah memberikan bimbingan, arahan, dan saran serta motivasinya yang sangat berharga selama proses penelitian ini.
4. Rekan-rekan dan sahabat, yang telah memberikan dukungan dan motivasi.
5. Keluarga tercinta, yang selalu memberikan doa, dukungan, dan semangat yang tiada henti.

Penulis menyadari bahwa penelitian ini masih memiliki keterbatasan dan kekurangan. Oleh karena itu, kritik dan saran yang membangun sangat penulis harapkan demi kesempurnaan penelitian ini di masa mendatang. Semoga penelitian ini dapat memberikan manfaat dan kontribusi nyata bagi pengembangan ilmu pengetahuan, khususnya dalam bidang ilmu komputer. Akhir kata, semoga penelitian ini bermanfaat bagi semua pihak yang berkepentingan dan dapat memberikan wawasan baru dalam dunia bisnis dan teknologi.

Jakarta, 10 Juli 2024

Penulis

KLASIFIKASI TIPE PENYAKIT DIABETES MENGUNAKAN ALGORITMA *SUPPORT VECTOR MACHINE* (SVM)

Nur Afifah Az-Zahra

ABSTRAK

Diabetes melitus adalah penyakit kronis yang menjadi salah satu penyebab utama kematian di Indonesia. Pada tahun 2019, diabetes merupakan penyebab kematian tertinggi ketiga di Indonesia dengan sekitar 57,42 kematian per 100.000 penduduk. Prevalensi kasus diabetes di Indonesia meningkat signifikan, termasuk peningkatan prevalensi pada anak-anak sebanyak 70 kali lipat sejak tahun 2010. Mengingat dampak serius terhadap kesehatan masyarakat, klasifikasi yang akurat antara diabetes tipe 1 dan tipe 2 sangat diperlukan untuk mencegah dampak yang merugikan. Penggunaan algoritma Support Vector Machine (SVM) dalam machine learning menawarkan solusi untuk klasifikasi cepat dan akurat berdasarkan data klinis. Penelitian ini bertujuan mengetahui kinerja algoritma SVM dalam klasifikasi diabetes berdasarkan data pasien dari Puskesmas Pekayon Jaya, Bekasi, Jawa Barat. Hasil menunjukkan bahwa penggunaan SMOTE dan *hyperparameter tuning* C, Gamma, dan Kernel secara signifikan meningkatkan akurasi model. Percobaan terbaik menunjukkan akurasi data latih 99%, validasi 89%, dan uji 91% dengan parameter optimal $C = 1000$, $\text{Gamma} = \text{scale}$, dan kernel RBF. Hasil validasi yang lebih mendekati performa test memberikan indikasi bahwa model tidak overfit pada data validasi dan kemungkinan besar akan memberikan performa yang serupa pada data yang benar-benar baru. Kesimpulannya, metode SMOTE dan *hyperparameter tuning* efektif dalam menangani ketidakseimbangan data dan meningkatkan akurasi model. Pembagian data yang tepat juga berpengaruh positif terhadap akurasi.

Kata Kunci: Support Vector Machine, Diabetes, Klasifikasi, SMOTE, Imbalance Dataset

Classification of Diabetes Types Using Support Vector Machine (SVM) Algorithm

Nur Afiifah Az-Zahra

ABSTRACT

Diabetes mellitus is a chronic disease that has become one of the leading causes of death in Indonesia. In 2019, diabetes was the third highest cause of death in Indonesia, with approximately 57.42 deaths per 100,000 population. The prevalence of diabetes cases in Indonesia has increased significantly, including a seventy-fold increase in prevalence among children since 2010. Given the serious impact on public health, accurate classification between type 1 and type 2 diabetes is essential to prevent adverse effects. The use of Support Vector Machine (SVM) algorithms in machine learning offers a solution for fast and accurate classification based on clinical data. This study aims to determine the performance of the SVM algorithm in classifying diabetes based on patient data from Puskesmas Pekayon Jaya, Bekasi, West Java. The results show that using SMOTE and hyperparameter tuning significantly improves model accuracy. The best experiment showed training data accuracy of 99%, validation accuracy of 89%, and test accuracy of 91%, with optimal parameters $C = 1000$, $\text{Gamma} = \text{scale}$, and kernel RBF. In conclusion, the SMOTE method and hyperparameter tuning are effective in handling data imbalance and improving model accuracy. Proper data partitioning also positively impacts accuracy.

Keywords: *Support Vector Machine, Diabetes, Classification, SMOTE, Imbalance Dataset*

Daftar Isi

PERNYATAAN ORISINALITAS	i
PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS	ii
LEMBAR PENGESAHAN	iii
KATA PENGANTAR	iv
ABSTRAK	v
Daftar Isi	vii
Daftar Tabel	x
Daftar Gambar	xi
Daftar Lampiran	xii
BAB I	1
PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Ruang Lingkup	3
1.4 Tujuan Penelitian.....	3
1.5 Manfaat Penelitian.....	3
1.6 Luaran yang diharapkan	4
1.7 Sistematika Penulisan.....	4
BAB II	6
TINJAUAN PUSTAKA	6
2.1 Diabetes	6
2.1.1 Definisi	6
2.1.2 Tipe Diabetes	7
2.2 Praproses Data	7
2.3 Normalisasi <i>Z-Score</i>	8
2.4 <i>Hyperparameter Tuning</i>	9
2.5 <i>Machine Learning</i>	10
2.6 Klasifikasi.....	11
2.7 Algoritma Support Vector Machine (SVM).....	11
2.8 SMOTE (Synthetic Minority Over-sampling Technique).....	14
2.9 Python.....	14
2.10 Evaluasi	15

2.10.1 Akurasi	16
2.10.2 Precision	16
2.10.3 <i>Recall</i> (Sensitifitas)	16
2.10.4 Specificity	16
2.10.5 Skor F1	16
2.10 Penelitian Terkait	17
BAB III.....	20
METODOLOGI PENELITIAN	20
3.1 Kerangka Berpikir	20
3.1.1 Studi Literatur	21
3.1.2 Perumusan Masalah	21
3.1.3 Pengumpulan Data	21
3.1.4 Praproses Data.....	21
3.1.5 Seleksi Fitur	22
3.1.6 Penerapan SMOTE (<i>Synthetic Minority Over-sampling Technique</i>)22	
3.1.7 Pembagian Data	22
3.1.8 Pembuatan Model.....	22
3.1.9 Evaluasi	23
3.1.10 Kesimpulan	24
3.2 Alat Pendukung Penelitian	24
3.2.1 Perangkat Keras	24
3.2.2 Perangkat Lunak.....	24
3.3 Tempat dan Waktu Penelitian	24
3.4 Jadwal Penelitian.....	25
BAB IV	26
HASIL DAN PEMBAHASAN	26
4.1 Pengumpulan Data	26
4.2 Pra-proses Data.....	28
4.2.1 Pembersihan Data.....	28
4.2.2 Memeriksa Duplikasi Data.....	29
4.2.3 Transformasi Data.....	30
4.2.4 Normalisasi Data	32
4.3 Seleksi Fitur.....	33

4.4 Keseimbangan Data.....	33
4.5 Pembagian Data.....	34
4.6 Pembuatan Model Klasifikasi.....	35
4.7 Evaluasi Model.....	39
4.7.1 Hasil Evaluasi Model Percobaan 1	40
4.7.2 Evaluasi Model Percobaan 2.....	42
4.7.3 Evaluasi Model Percobaan 3	44
4.7.4 Evaluasi Model Percobaan 4.....	46
4.7.5 Evaluasi Model Percobaan 5	48
4.7.6 Evaluasi Model Percobaan 6.....	50
4.8 Analisis Hasil	52
BAB V.....	54
KESIMPULAN DAN SARAN	54
5.1 Kesimpulan.....	54
5.2 Saran	54
Daftar Pustaka.....	56
LAMPIRAN.....	59
Lampiran 1. Surat Izin Riset dari FIK	59
Lampiran 2. Surat Izin Riset Dinas Kesehatan.....	60
Lampiran 3. Transkrip Wawancara Dokter	61
Lampiran 4. Data Diri Dokter	62
Lampiran 5. Dokumentasi Wawancara	63
Lampiran 6. Tautan Pengerjaan Penelitian	64

Daftar Tabel

Tabel 2.1 Confusion Matrix	15
Tabel 2.2 Matriks Review Jurnal	17
Tabel 3.1 Jadwal Penelitian.....	25
Tabel 4.1 Sampel Dataset.....	26
Tabel 4.2 Deskripsi Atribut.....	27
Tabel 4.3 Mean Imputation	29
Tabel 4.4 Data Sebelum Transformasi	30
Tabel 4.5 Data Sesudah Transformasi	31
Tabel 4.6 Data Hasil OneHotEncoding.....	31
Tabel 4.7 Tipe Data.....	32
Tabel 4.8 Jumlah Data.....	34
Tabel 4.9 Pembagian Data Train dan Data Test	34
Tabel 4.10 Pembagian Data K-Fold CV	35
Tabel 4.11 Hasil Pemilihan Parameter.....	36
Tabel 4.12 Skema Percobaan Model.....	39
Tabel 4.13 Hasil K-Fold CV Percobaan 1	40
Tabel 4.14 Confusion Matrix Data Uji Percobaan 1	41
Tabel 4.15 Hasil Metrik Kinerja Data Uji Percobaan 1	42
Tabel 4.16 Hasil K-Fold CV Percobaan 2	42
Tabel 4.17 Confusion Matrix Data Uji Percobaan 2.....	43
Tabel 4.18 Hasil Metrik Kinerja Data Uji Percobaan 2	43
Tabel 4.19 Hasil K-Fold CV Percobaan 3	44
Tabel 4.20 Confusion Matrix Data Uji Percobaan 3.....	45
Tabel 4.21 Hasil Metrik Kinerja Data Uji Percobaan 3	45
Tabel 4.22 Hasil K-Fold CV Percobaan 4	46
Tabel 4.23 Confusion Matrix Data Uji 4	47
Tabel 4.24 Hasil Metrik Kinerja Data Uji Percobaan 4.....	47
Tabel 4.25 Hasil K-Fold CV Percobaan 5	48
Tabel 4.26 Confusion Matrix Data Uji 5	49
Tabel 4.27 Hasil Metrik Kinerja Data Uji Percobaan 5	49
Tabel 4.28 Hasil K-Fold CV Percobaan 6	50
Tabel 4.29 Confusion Matrix Data Uji 6	51
Tabel 4.30 Hasil Metrik Kinerja Data Uji Percobaan 6.....	51
Tabel 4.31 Analisis Hasil Akurasi	52

Daftar Gambar

Gambar 2.1 Garis Pemisah Kedua Kelas	12
Gambar 2.2 Hyperplane pada SVM	13
Gambar 3.1 Tahapan Penelitian	20
Gambar 3.2 Propose Model.....	23
Gambar 4.1 Jumlah Missing Value	28
Gambar 4.2 Proses Normalisasi Data	32
Gambar 4.3 Hasil Normalisasi Data.....	33
Gambar 4.4 Proses Tuning Model	36
Gambar 4.5 Cetak Hasil Parameter	36
Gambar 4.6 Penggunaan Model I	38
Gambar 4.7 Penggunaan Model II	38

Daftar Lampiran

Lampiran 1. Surat Izin Riset dari FIK.....	59
Lampiran 2. Surat Izin Riset Dinas Kesehatan	60
Lampiran 3. Transkrip Wawancara Dokter.....	61
Lampiran 4. Data Diri Dokter	62
Lampiran 5. Dokumentasi Wawancara	63
Lampiran 6. Tautan Pengerjaan Penelitian	64