



**KLASIFIKASI PENYAKIT DIABETES MENGGUNAKAN METODE  
SUPPORT VEKTOR MACHINE (SVM)**

**TUGAS AKHIR**

**MUHAMMAD MUMTAZ RAMADHAN**

**1910511102**

**UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAKARTA**

**FAKULTAS ILMU KOMPUTER**

**PROGRAM STUDI S-1 INFORMATIKA**

**2023**



**UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAKARTA**

**KLASIFIKASI PENYAKIT DIABETES MENGGUNAKAN METODE  
SUPPORT VEKTOR MACHINE (SVM)**

**TUGAS AKHIR**

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

**MUHAMMAD MUMTAZ RAMADHAN**

**1910511102**

**UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAKARTA**

**FAKULTAS ILMU KOMPUTER**

**PROGRAM STUDI S-1 INFORMATIKA**

**2023**

## LEMBAR PERSETUJUAN

Yang bertanda tangan di bawah ini:

Nama : Muhammad Mumtaz Ramadhan

NIM : 1910511102

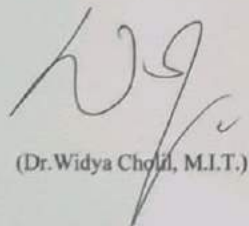
Program Studi : S1- Informatika

Judul : Klasifikasi Penyakit Diabetes Menggunakan Metode Support  
Vektor Machine (SVM)

Dinyatakan telah memenuhi syarat dan menyetujui untuk ujian sidang skripsi.

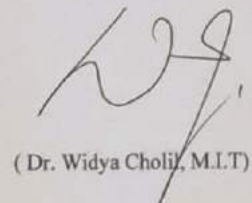
Jakarta, 18 Oktober 2023

Mengetahui,  
Ketua Program Studi



(Dr. Widya Cholil, M.I.T.)

Menyetujui,  
Dosen Pembimbing



( Dr. Widya Cholil, M.I.T.)

## PERNYATAAN ORISINALITAS

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

Nama : Muhammad Mumtaz Ramadhan

NIM : 1910511102

Tanggal : 19 Desember 2023

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

Jakarta, 19 Desember 2023



(Muhammad Mumtaz Ramadhan)

Muhammad Mumtaz Ramadhan, 2023  
*KLASIFIKASI PENYAKIT DIABETES MENGGUNAKAN METODE SUPPORT VEKTOR  
MACHINE (SVM)*  
Universitas Pembangunan Nasional Veteran Jakarta, Fakultas Ilmu Komputer, S1 Informatika  
[www.upnvj.ac.id – www.library.upnvj.ac.id – www.repository.upnvj.ac.id]

ii

**PERNYATAAN PERSETUJUAN PUBLIKASI  
SKRIPSI UNTUK KEPENTINGAN AKADEMIS**

Sebagai civitas akademik Universitas Pembangunan Nasional Veteran  
Jakarta, saya yang bertanda tangan dibawah ini:

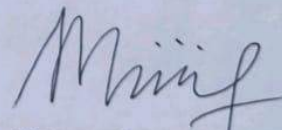
Nama : Muhammad Mumtaz Ramadhan  
NIM : 1910511102  
Fakultas : Ilmu Komputer  
Program studi : S1 Informatika

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

**Klasifikasi Penyakit Diabetes Menggunakan Metode Support Vektor Machine (SVM)**

Dengan Hak Bebas royalty ini Universitas Pembangunan Nasional Veteran Jakarta berhak menyimpan, mengai media/formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat dan mempublikasikan skripsi saya selama mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta. Demikian pernyataan ini saya buat dengan sebenarnya.

Yang menyatakan,



(Muhammad Mumtaz Ramadhan)

Dibuat di : Jakarta

Pada Tanggal : 19 Desember 2023

Muhammad Mumtaz Ramadhan, 2023  
**KLASIFIKASI PENYAKIT DIABETES MENGGUNAKAN METODE SUPPORT VEKTOR  
MACHINE (SVM)**  
Universitas Pembangunan Nasional Veteran Jakarta, Fakultas Ilmu Komputer, S1 Informatika  
[www.upnvj.ac.id – www.library.upnvj.ac.id – www.repository.upnvj.ac.id]

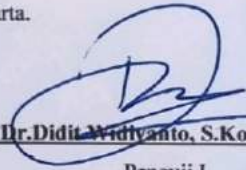
iii


## LEMBAR PENGESAHAN

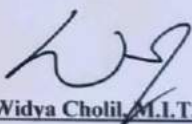
Tugas Akhir ini diajukan oleh:

Nama : Muhammad Mumtaz Ramadhan  
NIM : 1910511102  
Program Studi : S1 Informatika  
Judul Tugas Akhir : Klasifikasi Penyakit Diabetes Menggunakan Metode Support Vektor Machine (SVM)

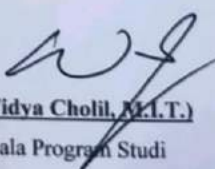
Telah berhasil dipertahankan dihadapan Tim Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana pada Program Studi Informatika, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional Veteran Jakarta.

  
(Dr. Didit Widlyanto, S.Kom, M.Si.)  
Penguji I

  
(Rio Wirawan, S.Kom., MMSL.)  
Penguji II

  
(Dr. Widya Cholil, M.I.T.)  
Pembimbing

  
(Herly Krisnanik, S.Kom., MM)  
Plt Dekan

  
(Dr. Widya Cholil, M.I.T.)  
Kepala Program Studi

Ditetapkan di : Jakarta  
Tanggal Ujian : 16 Desember 2023



Muhammad Mumtaz Ramadhan, 2023  
**KLASIFIKASI PENYAKIT DIABETES MENGGUNAKAN METODE SUPPORT VEKTOR MACHINE (SVM)**  
Universitas Pembangunan Nasional Veteran Jakarta, Fakultas Ilmu Komputer, S1 Informatika  
[www.upnvj.ac.id – www.library.upnvj.ac.id – www.repository.upnvj.ac.id]

iv

# **Klasifikasi Penyakit Diabetes Menggunakan Metode Support Vektor Machine (SVM)**

**Muhammad Mumtaz Ramadhan**

## **ABSTRAK**

Diabetes merupakan penyakit yang cukup berbahaya. Ada beberapa indikator atau faktor yang mempengaruhi penyakit diabetes. Seperti misalnya usia, jenis kelamin, berat badan, dan sebagainya. Tenaga kesehatan atau Dokter cukup kesulitan dalam menentukan faktor apakah yang lebih memungkinkan seseorang terkena penyakit diabetes dan pasien manakah yang bisa dikatakan positive dan negative terkena diabetes. Tujuan dari penelitian ini adalah untuk memudahkan dokter atau tenaga kesehatan agar lebih mudah dalam melakukan prediksi, keputusan, dan atau melakukan klasifikasi terhadap pasien yang positive diabetes dan negative. Penelitian ini menggunakan metode SVM dan KNN. Kemudian dibandingkan, metode manakah yang lebih baik dalam melakukan prediksi atau klasifikasi penyakit diabetes. Berdasarkan Hasil penelitian yang telah dilakukan nilai akurasi yang didapat dengan model SVM cukup baik, karena hasil metrik evaluasi memperoleh hasil lebih dari 80% dengan nilai akurasi yang didapat adalah 88% dengan cross val accuracy sebesar 89%, precision sebesar 83%, recall sebesar 100%, F1 Score sebesar 90%, dan ROC yang diperoleh sebesar 83%. Sedangkan Akurasi terbaik yang diperoleh dengan model KNN adalah 99% dengan K1. Cross Val Accuracy yang diperoleh sebesar 96%, precision sebesar 100%, recall yang diperoleh sebesar 98%, F1 sebesar 99%, dan ROC yang diperoleh sebesar 83%. Dengan demikian dalam penelitian ini metode KNN lebih baik dibandingkan dengan SVM dalam melakukan prediksi atau klasifikasi penyakit diabetes.

**Kata kunci: Diabetes, Klasifikasi, Prediksi, SVM, KNN**

# **Klasifikasi Penyakit Diabetes Menggunakan Metode Support Vektor Machine (SVM)**

**Muhammad Mumtaz Ramadhan**

## ***ABSTRACT***

*Diabetes is a quite dangerous disease. There are several indicators or factors that influence diabetes. For example, age, gender, weight, and so on. Health workers or doctors have difficulty determining what factors are more likely for someone to develop diabetes and which patients can be said to be positive or negative for diabetes. The aim of this research is to make it easier for doctors or health experts to make predictions, decisions, and/or classify patients who are positive for diabetes and negative. This research uses SVM and KNN methods. Then compared, which methods are better for predicting or classifying diabetes. Based on the results of the research that has been done, the accuracy value obtained with the SVM model is quite good, because the results of the evaluation metrics obtained results of more than 80% with the accuracy value obtained is 88% with cross val accuracy of 89%, precision of 83%, recall of 100%, F1 Score of 90%, and ROC obtained of 83%. While the best accuracy obtained with the KNN model is 99% with KI. Cross Val Accuracy obtained was 96%, precision was 100%, recall obtained was 98%, F1 was 99%, and ROC obtained was 83%. Thus, in this research, the KNN method is better than SVM in predicting or classifying diabetes..*

***Keywords: Diabetes, Classification, Prediction, SVM, KNN***



## **KATA PENGANTAR**

Puji syukur segala puji bagi Allah SWT yang telah memberikan nikmat dan hidayah-Nya sehingga penulis dapat menyelesaikan tugas akhir ini dengan judul “*Klasifikasi Penyakit Diabetes Menggunakan Metode Support Vektor Machine (SVM)*” tepat pada waktunya. Penyelesaian tugas akhir ini pula tidak lepas dari bimbingan dan dorongan dari berbagai pihak. Dalam kesempatan ini, penulis ingin menyampaikan ucapan terima kasih kepada:

- a. Ibu Erly Krisnanik, S.Kom.,MM., selaku Plt dekan Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jakarta.
- b. Ibu Dr. Widya Cholil, M.I.T.selaku Kepala Program Studi S1 Informatika Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jakarta sekaligus Dosen Pembimbing Skripsi saya.
- c. Semua pihak yang telah memberikan bantuan dan dukungan serta kerja sama yang baik sehingga laporan tugas akhir ini dapat diselesaikan dengan baik

## DAFTAR ISI

LEMBAR PERSETUJUAN.....	i
PERNYATAAN ORISINALITAS.....	ii
PERNYATAAN PERSETUJUAN PUBLIKASI.....	iii
LEMBAR PENGESAHAN.....	iv
ABSTRAK .....	v
KATA PENGANTAR .....	vii
DAFTAR ISI .....	viii
DAFTAR TABEL .....	xi
DAFTAR GAMBAR.....	xii
DAFTAR RUMUS .....	xiv
DAFTAR LAMPIRAN.....	xv
BAB I.....	1
PENDAHULUAN .....	1
1.1. Latar Belakang .....	1
1.2. Rumusan Masalah .....	2
1.3. Batasan Penelitian .....	2
1.4. Tujuan Penelitian .....	2
1.5. Manfaat Penelitian .....	3
1.6. Luaran yang diharapkan .....	3
1.7. Sistematika Penulisan.....	3
BAB II.....	5
TINJAUAN PUSTAKA.....	5
2.1. Penelitian Terdahulu .....	5
2.2. Landasan Teori.....	9
2.2.1 Diabetes.....	9
2.2.2 Machine Learning.....	9
2.2.3 CRISP-DM.....	11
2.2.4 EDA (Exploratory Data Analysis).....	13
2.2.5 Sweetviz .....	14
2.2.6 Autoviz.....	17
2.2.7 Pandas Profilling .....	18
2.2.8 Support Vector Machine .....	19

2.2.9	KNN ( K-Nearest Neighbors) .....	21
2.2.10	ANOVA .....	22
2.2.11	Data Preprocessing .....	23
2.2.12	Evaluasi Model.....	24
BAB III.....		26
METODOLOGI PENELITIAN .....		26
3.1.	<i>Data Acquisition</i> .....	26
3.2.	<i>Data Understanding</i> .....	31
3.3.	<i>Data Preprocessing</i> .....	31
3.4.	Splitting Data Training dan Testing .....	32
3.5.	Data Normalization .....	33
3.6.	Model Evaluation .....	33
3.7.	Alat Bantu Penelitian.....	34
3.8.	Jadwal Rencana Penelitian .....	34
BAB IV .....		36
PEMBAHASAN.....		36
4.1.	Data Acquisition .....	36
4.2.	Data Understanding.....	41
4.3.	Data Preprocessing.....	53
4.3.1	Data Transformation.....	53
4.3.2	Correlation of Features .....	57
4.3.3	Features Selection.....	62
4.4.	Splitting Dataset.....	65
4.5.	Data Normalization .....	66
4.6.	Model Evaluation .....	68
4.6.1	Model Evaluation (SVM).....	68
4.6.2	Model Evaluation (KNN).....	72
4.7.	Perbandingan Model evaluation SVM dan KNN .....	79
BAB V.....		80
KESIMPULAN DAN SARAN .....		80
5.1.	Kesimpulan .....	80
5.2.	Saran.....	81
DAFTAR PUSTAKA .....		82
RIWAYAT HIDUP .....		86

LAMPIRAN .....	87
Lampiran I : Dataset .....	87
Lampiran II : Hasil Turnitin .....	88

## DAFTAR TABEL

Tabel 1 Penelitian yang relevan.....	5
Tabel 2 <i>Confusion Matrix</i> .....	24
Tabel 3 Atribut Dataset.....	27
Tabel 4 Preview Dataset.....	30
Tabel 5 Jadwal Rencana Penelitian .....	35
Tabel 6 Library yang digunakan dalam Penelitian .....	37
Tabel 7 Model Evaluation SVM dalam bentuk Tabel .....	71
Tabel 8 Model Evaluation SVM dan KNN .....	79
Tabel 9 Perbandingan metode SVM dan KNN berdasarkan metrik evaluasi .....	79

## DAFTAR GAMBAR

Gambar 1	Contoh Penggunaan EDA .....	13
Gambar 2	Contoh Visualisasi dengan Sweetviz .....	16
Gambar 3	Contoh Visualisasi dengan Autoviz .....	17
Gambar 4	Contoh Visualisasi dengan Pandas Profiling .....	19
Gambar 5	Ilustrasi Konvolusi.....	20
Gambar 6	Metode Penelitian .....	26
Gambar 7	Perintah Untuk Mengambil dan Menampilkan Dataset.....	36
Gambar 8	Tampilan Dataset dengan Library Pandas .....	37
Gambar 9	Perintah untuk melihat missing value pada data .....	41
Gambar 10	Hasil Pencarian Missing Value pada data.....	42
Gambar 11	Informasi dataset dengan ' <i>df.info()</i> ' .....	42
Gambar 12	Perintah untuk membuat plot distribusi atribut data.....	43
Gambar 13	Hasil Plotting Distribusi atribut data yang berbeda.....	45
Gambar 14	Hasil Plotting Distribusi atribut data yang berbeda.....	45
Gambar 15	Visualisasi menggunakan Tabel kontigensi antara variable gender dan class.....	46
Gambar 16	Perintah yang digunakan untuk menginstal sweetviz .....	47
Gambar 17	Perintah yang digunakan untuk import library Sweetviz.....	47
Gambar 18	Perintah untuk buat objek Sweetviz .....	48
Gambar 19	Output ketika Objek Sweetviz berhasil dibuat .....	48
Gambar 20	Perintah untuk menampilkan laporan di browser dengan ekstensi .html .....	48
Gambar 21	Output Jika report Sweetviz berhasil digenerate .....	48
Gambar 21	Output report Sweetviz berupa file html.....	48
Gambar 23	Tampilan Visualisasi data dengan Sweetviz .....	49
Gambar 24	Perintah untuk menginstal Autoviz .....	50
Gambar 25	Perintah untuk import library Autoviz.....	50
Gambar 26	Perintah untuk but objek Autoviz.....	50
Gambar 27	Perintah untuk menampilkan hasil autoviz .....	50
Gambar 28	Tampilan Visualisasi data dengan Autoviz.....	51
Gambar 29	Tampilan pembuatan EDA dengan Pandas Profiling .....	52
Gambar 30	Hasil EDA dengan Pandas Profiling pada file html .....	52
Gambar 31	Fitur filtering variable dengan Pandas Profiling pada file html .....	53
Gambar 32	Perintah untuk transformasi data mengubah kelas 'positive' menjadi 1 dan kelas 'negative' menjadi 0 .....	54
Gambar 33	Perintah untuk memisahkan variabel target dengan variabel lainnya .....	55
Gambar 34	Perintah untuk melakukan storing features .....	55
Gambar 35	Perintah untuk melakukan emcoding.....	56
Gambar 36	Perbandingan sebelum dan sesudah Transformasi data.....	56
Gambar 37	Perbandingan sebelum dan sesudah Transformasi data (berdasarkan tipe data).....	57
Gambar 38	Korelasi antar fitur .....	58

Gambar 39 Perintah untuk melihat correlation features dengan menggunakan grafik batang .....	58
Gambar 40 Korelasi antar fitur dengan diagram batang .....	59
Gambar 41 Hasil Matriks Korelasi antar fitur dalam Tabel .....	60
Gambar 42 Tabel Korelasi Dalam bentuk C-Map .....	60
Gambar 43 fungsi untuk pembuatan Tabel Korelasi Dalam bentuk C-Map .....	61
Gambar 44 Source Code untuk featur selection .....	63
Gambar 45 Hasil Featur Selection .....	64
Gambar 46 Fitur Keseluruhan .....	64
Gambar 47 Fitur Yang akan digunakan ke dalam model .....	65
Gambar 48 Fitur Yang tidak akan digunakan ke dalam model .....	65
Gambar 49 Perintah untuk membagi data menjadi set pelatihan (training & testing) .....	65
Gambar 50 Perintah untuk melakukan normalisasi .....	67
Gambar 51 Perbandingan sebelum dan sesudah dilakukan normalisasi pada fitur 'Age' .....	67
Gambar 52 Perintah untuk pembuatan model SVM .....	68
Gambar 53 Perintah untuk Cross-Validation .....	69
Gambar 54 Hasil K-Fold untuk Model SVM .....	70
Gambar 55 Perintah untuk model evaluation SVM .....	70
Gambar 56 Model Evaluation SVM dalam bentuk Classification Report .....	72
Gambar 57 Pembuatan model KNN .....	73
Gambar 58 Daftar Akurasi Prediksi Untuk Setiap Jumlah Tetangga (N_Neighbors) 1-9 .....	74
Gambar 59 Perintah untuk mengurutkan nilai K terbesar hingga terkecil .....	75
Gambar 60 Daftar Akurasi Prediksi dengan KNN Yang Diurutkan Dari Tertinggi Ke Terendah .....	75
Gambar 61 Perintah untuk membuat metric evaluasi akurasi model KNN .....	76
Gambar 62 Perintah untuk menghitung metrik evaluasi lainnya KNN .....	76
Gambar 63 Perintah membuat Cross Validation untuk KNN .....	77
Gambar 64 Perintah untuk menampilkan hasil Fold Model KNN .....	77
Gambar 65 Perintah untuk menampilkan hasil evaluasi yang disimpan dalam DataFrame .....	78
Gambar 66 Metrik Evaluasi Report Model KNN .....	78

## DAFTAR RUMUS

Rumus 1 Rumus fungsi untuk memodelkan SVM .....	21
Rumus 2 Akurasi Score.....	25
Rumus 3 Recall Score.....	25
Rumus 4 Precision Score .....	25
Rumus 5 F1 Score.....	25



## DAFTAR LAMPIRAN

Lampiran I	: Dataset.....	87
Lampiran II	: Hasil Turnitin.....	88