



**PENERAPAN ALGORITMA K-NEAREST NEIGHBOR DAN
SUPPORT VECTOR MACHINE UNTUK MEMPREDIKSI
PENYAKIT DIABETES**

LAPORAN TUGAS AKHIR

Renaldi Khairul Fadhil

41821010097

**UNIVERSITAS
MERCU BUANA**

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS ILMU KOMPUTER
UNIVERSITAS MERCU BUANA
JAKARTA
2025**



**PENERAPAN ALGORITMA K-NEAREST NEIGHBOR DAN
SUPPORT VECTOR MACHINE UNTUK MEMPREDIKSI
PENYAKIT DIABETES**

LAPORAN TUGAS AKHIR

Renaldi Khairul Fadhil

41821010097

Diajukan sebagai salah satu syarat untuk memperoleh gelar sarjana

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS ILMU KOMPUTER
UNIVERSITAS MERCU BUANA
JAKARTA**

2025

HALAMAN PENYATAAN KARYA SENDIRI

Saya yang bertanda tangan di bawah ini:

Nama : Renaldi Khairul Fadhil
NIM : 41821010097
Program Studi : Sistem Informasi
Judul Laporan Skripsi : Penerapan Algoritma K-Nearest Neighbor Dan Support Vector Machine Untuk Memprediksi Penyakit Diabetes

Menyatakan bahwa Laporan Aplikatif/Tugas Akhir/Jurnal/Media Ilmiah ini adalah hasil karya saya sendiri dan bukan plagiat, serta semua sumber baik yang dikutip maupun dirujuk telah saya nyatakan dengan benar. Apabila ternyata ditemukan di dalam Laporan Tugas Akhir saya terdapat unsur plagiat, maka saya siap mendapatkan sanksi akademis yang berlaku di Universitas Mercu Buana.

Jakarta, 28 Juni 2025



Renaldi Khairul Fadhil

UNIVERSITAS
MERCU BUANA

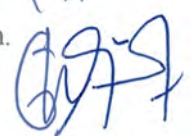
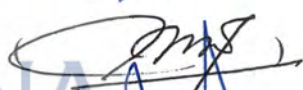
LEMBAR PENGESAHAN

Nama Mahasiswa (1) : Renaldi Khairul Fadhil
NIM : (41821010097)
Judul Tugas Akhir : PENERAPAN ALGORITMA K-NEAREST
NEIGHBOR DAN SUPPORT VECTOR
MACHINE UNTUK MEMPREDIKSI PENYAKIT
DIABETES

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Mercu Buana.

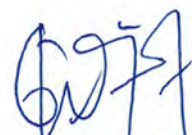
Jakarta, 28 Juli 2025

Menyetujui,

Pembimbing : Abdi Wahab, S.Kom, MT
NIDN : 0305068502 ()
Ketua Penguji : Wawan Gunawan, S.Kom, MT., M.Kom.
NIDN : 0424108104 ()
Penguji 1 : Misni, S.Kom, M.Kom
NIDN : 0413046802 ()
Penguji 2 : Suhendra, S.Kom, M.Kom
NIDN : 0308019002 ()

Mengetahui,


Dr. Bambang Jokonowo, S.Si., M.T.I.
Dekan Fakultas Ilmu Komputer


Wawan Gunawan, S.Kom, MT., M.Kom.
Ka.Prodi Sistem Informasi

KATA PENGANTAR

Puji syukur kita panjatkan kehadirat Allah Subhanahu Wa Ta'ala, yang telah melimpahkan berkat dan karunia-Nya sehingga penulis dapat menyelesaikan Laporan Metodologi Penelitian Teknologi Informasi dengan baik. Dalam kesempatan ini, saya ingin menyampaikan laporan yang telah saya lakukan selama beberapa waktu dengan judul “Penerapan Algoritma K-Nearest Neighbor Dan Support Vector Machine Untuk Memprediksi Penyakit Diabetes“. Laporan ini disusun dengan tujuan sebagai syarat kelulusan dalam fakultas ilmu komputer universitas Mercu Buana.

Saya ingin menyampaikan terimakasih yang sebesar besarnya kepada :

1. Kepada Allah SWT, Tuhan Yang Maha Esa, atas limpahan rahmat, taufik, dan hidayah-Nya, sehingga penulis dapat menyelesaikan Tugas Akhir ini dengan baik dan lancar.
2. Bapak Lukman Hakim, S.T., M.Kom selaku Ketua Program Studi Sistem Informasi.
3. Bapak Abdi Wahab, S.Kom, MT selaku Dosen Pembimbing Tugas Akhir, yang telah memberikan arahan, bimbingan, dan masukan yang sangat berarti selama proses penyusunan tugas akhir ini.
4. Ibu Yunita Sartika Sari, S.Kom, M.Kom selaku Dosen Pembimbing Akademik, yang memberikan dukungan selama proses penyusunan tugas akhir ini.
5. Seluruh dosen dan staf di Fakultas Ilmu Komputer Universitas Mercu Buana yang telah memberikan ilmu dan fasilitas selama masa perkuliahan.
6. Kedua orang tua tercinta, yang selalu memberikan doa, semangat, dan dukungan moril maupun materiil yang tidak ternilai harganya.
7. Kepada sahabat-sahabat dan rekan seperjuangan yang telah memberikan dukungan, semangat, serta warna dalam perjalanan akademik penulis, khususnya Muhammad Fadhil Darussalam, Fairuz Fadhil, Ferenc Janos Suparmadi, Daffa Abinaya Putra, serta sahabat saya yang luar biasa, Ridwan Ahmad Zulkarnaen. Juga kepada seluruh teman-teman Program Studi

Sistem Informasi, terkhusus keluarga besar HIMSISFO yang tidak dapat disebutkan satu per satu—terima kasih atas kebersamaan dan dukungan kalian yang selalu penulis banggakan dan cintai.

Penulis menyadari sepenuhnya bahwa Tugas Akhir ini masih jauh dari kata sempurna. Meskipun telah disusun dengan penuh dedikasi, ketekunan, dan upaya maksimal, penulis tidak menutup kemungkinan masih terdapat kekurangan baik dari segi isi, penyajian data, maupun tata penulisan. Oleh karena itu, penulis dengan rendah hati sangat mengharapkan adanya kritik, masukan, dan saran yang bersifat membangun dari para pembaca, dosen, maupun pihak-pihak yang memiliki kompetensi di bidang ini, demi perbaikan dan penyempurnaan di masa mendatang. Akhir kata, penulis berharap semoga Tugas Akhir ini tidak hanya menjadi syarat akademik semata, tetapi juga dapat memberikan manfaat yang nyata, memperluas wawasan, serta menjadi referensi yang berguna bagi para pembaca, peneliti, maupun pihak-pihak lain yang memiliki kepentingan dan minat terhadap topik yang dibahas dalam penelitian ini.

Jakarta, 19 Juni 2025



Renaldi Khairul Fadhil



UNIVERSITAS
MERCU BUANA

HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS

Sebagai sivitas akademik Universitas Mercu Buana, saya yang bertanda tangan di bawah ini:

Nama : Renaldi Khairul Fadhil
NIM : 41821010097
Program Studi : Sistem Informasi
Judul Laporan Skripsi : Penerapan Algoritma K-Nearest Neighbor Dan Support Vector Machine Untuk Memprediksi Penyakit Diabetes

Demi pengembangan ilmu pengetahuan, dengan ini memberikan izin dan menyetujui untuk memberikan kepada Universitas Mercu Buana **Hak Bebas Royalti Non-Eksklusif (*Non-exclusive Royalty-Free Right*)** atas karya ilmiah saya yang berjudul di atas beserta perangkat yang ada (jika diperlukan).

Dengan Hak Bebas Royalti Non-Eksklusif ini Universitas Mercu Buana berhak menyimpan, mengalihmedia/format-kan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan mempublikasikan Laporan Magang/Skripsi/Tesis/Disertasi saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

UNIVERSITAS
MERCU BUANA

Jakarta, 28 Juni 2025

Yang menyatakan,



Renaldi Khairul Fadhil

ABSTRAK

Nama : Renaldi Khairul Fadhil
NIM : 41821010097
Pembimbing TA : Abdi Wahab, S.Kom, MT
Judul Tugas Akhir : Penerapan Algoritma K-Nearest Neighbor Dan Support Vector Machine Untuk Memprediksi Penyakit Diabetes

Prevalensi diabetes mellitus di seluruh dunia terus meningkat secara signifikan, menjadikannya sebagai salah satu masalah kesehatan utama yang membutuhkan deteksi dini dan penanganan yang efektif. Penelitian ini bertujuan untuk membandingkan kinerja algoritma Support Vector Machine (SVM) dan K-Nearest Neighbor (KNN) dalam memprediksi risiko diabetes menggunakan data sekunder dari Kaggle yang terdiri dari 768 data pasien dengan sembilan atribut medis utama. Metode yang digunakan meliputi tahap pembersihan data, normalisasi, pemilihan fitur, serta penerapan teknik oversampling SMOTE (Synthetic Minority Over-sampling Technique) untuk menyeimbangkan data yang tidak seimbang, sehingga meningkatkan akurasi prediksi. Selanjutnya, kedua algoritma dievaluasi melalui metrik akurasi, presisi, recall, dan F1-score. Hasil menunjukkan bahwa SVM unggul dalam hal akurasi dataset yang kompleks, sedangkan KNN lebih efektif pada data dengan distribusi seragam dan jumlah kecil. Penerapan SMOTE terbukti mampu meningkatkan performa model dengan meningkatkan kemampuan klasifikasi terhadap data minoritas. Kesimpulan dari penelitian ini adalah bahwa SVM lebih optimal untuk prediksi diabetes pada data yang kompleks, sementara KNN dapat digunakan sebagai alternatif yang fleksibel. Penelitian ini memberikan kontribusi dalam pengembangan sistem pendukung keputusan berbasis data untuk diagnosis dan manajemen risiko diabetes yang lebih akurat.

Kata Kunci: Prediksi Diabetes, Support Vector Machine, K-Nearest Neighbor, SMOTE, Oversampling.

ABSTRACT

Nama : Renaldi Khairul Fadhil
NIM : 41821010097
Pembimbing TA : Abdi Wahab, S.Kom, MT
Judul Tugas Akhir : Penerapan Algoritma K-Nearest Neighbor Dan Support Vector Machine Untuk Memprediksi Penyakit Diabetes

The prevalence of diabetes mellitus worldwide continues to increase significantly, making it one of the major health problems that require early detection and effective treatment. This study aims to compare the performance of Support Vector Machine (SVM) and K-Nearest Neighbor (KNN) algorithms in predicting diabetes risk using secondary data from Kaggle consisting of 768 patient data with nine main medical attributes. The methods used include data cleaning, normalization, feature selection, and the application of the SMOTE (Synthetic Minority Over-sampling Technique) oversampling technique to balance unbalanced data, thereby improving prediction accuracy. Furthermore, both algorithms are evaluated through accuracy, precision, recall, and F1-score metrics. Results show that SVM excels in terms of accuracy on complex datasets, while KNN is more effective on data with uniform distribution and small numbers. The application of SMOTE is proven to improve the performance of the model by increasing the classification ability of minority data. The conclusion of this study is that SVM is more optimal for diabetes prediction on complex data, while KNN can be used as a flexible alternative. This research contributes to the development of a data-based decision support system for more accurate diagnosis and risk management of diabetes.

Keywords: Diabetes Prediction, Support Vector Machine, K-Nearest Neighbor, SMOTE, Oversampling.

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERNYATAAN KARYA SENDIRI	ii
HALAMAN PENGESAHAN	iii
KATA PENGANTAR.....	iv
HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH	vi
ABSTRAK	vii
ABSTRACT	viii
DAFTAR ISI.....	ix
DAFTAR TABEL	xi
DAFTAR GAMBAR.....	xii
DAFTAR LAMPIRAN	xiv
BAB I PENDAHULUAN.....	1
1.1 Latar belakang	1
1.2 Rumusan Masalah	2
1.3 Tujuan Penelitian.....	2
1.4 Manfaat Penelitian.....	3
1.5 Batasan Masalah.....	3
1.6 Sistematika Penulisan.....	4
BAB II TINJAUAN PUSTAKA.....	5
2.1 Teori/Konsep Terkait	5
2.1.1 Diabetes Melitus	5
2.1.2 K-Nearest Neighbor.....	6
2.1.3 Support Vector Machine (SVM).....	7
2.1.4 SMOTE.....	7
2.1.5 Confusion Matrix	8
2.2 Penelitian Terdahulu.....	10
2.3 Analisis Literatur Riview	20
BAB III METODE PENELITIAN	21
3.1 Deskripsi Sumber Data.....	21

3.2 Teknik Pengumpulan Data	21
3.3 Diagram Alir Penelitian.....	22
3.4 Jadwal Penelitian	25
BAB IV HASIL PENELITIAN.....	26
4.1 Establish Ground Truth	26
4.2 Mengumpulkan Data dan Data Pre-Processing.....	26
4.2.1 Pengumpulan Data	26
4.2.2 Pre Processing.....	27
4.2.3 Feature Selection.....	28
4.2.4 Splitting the Data Frame	29
4.3 Pengembangan Model	30
4.3.1 Train Test Split	30
4.3.2 K-Nearest Neighbor	31
4.3.3 Support Vector Machine.....	32
4.4 Penerapan SMOTE.....	34
4.4.1 K-Nearest Neighbor Oversampling Dengan Teknik SMOTE	34
4.4.2 Support Vector Machine Oversampling Dengan Teknik SMOTE	35
4.5 Evaluasi Model Sebelum dan Sesudah SMOTE	37
4.5.1 Evaluasi Sebelum SMOTE	37
4.5.2 Evaluasi Setelah SMOTE	42
4.6 Pembahasan	47
BAB V KESIMPULAN DAN SARAN	58
5.1 Kesimpulan.....	58
5.2 Saran	58
DAFTAR PUSTAKA	60
LAMPIRAN.....	62

DAFTAR TABEL

Tabel 2. 1 Confusion Matrix	8
Tabel 2. 2 Penelitian Terdahulu	10
Tabel 3. 2 Jadwal Pengerjaan Penelitian	25
Tabel 4. 1 Splitting Data Latih dan Data Uji	50
Tabel 4. 2 Tabel Evaluasi Kinerja Model K-Nearest Neighbor	51
Tabel 4. 3 Hasil Evaluasi Kinerja Model Support Vector Machine	51
Tabel 4. 4 Evaluasi Kinerja Model K-Nearest Neighbor SMOTE Oversampling	52
Tabel 4. 5 Evaluasi Kinerja Model Support Vector Machine SMOTE Oversampling	53
Tabel 4. 6 Perbandingan 80% Training dan 20% Testing	54
Tabel 4. 7 Perbandingan 70% Training dan 30% Testing	55
Tabel 4. 8 Perbandingan 80% Training dan 20% Testing SMOTE Oversampling	56
Tabel 4. 9 Perbandingan 70% Training dan 30% Testing SMOTE Oversampling	56



DAFTAR GAMBAR

Gambar 2. 1 Euclidean distance	6
Gambar 2. 2 Minkowski.....	6
Gambar 2. 3 Koefisien Korelasi Pearson p	7
Gambar 2. 4 Kesamaan cosine	7
Gambar 2. 5 Rumus Akurasi	9
Gambar 2. 6 Rumus Presisi.....	9
Gambar 2. 7 Rumus Recall	9
Gambar 2. 8 Rumus F1-Score.....	10
Gambar 3. 1 Diagram Alir Penelitian	22
Gambar 4. 1 Informasi Dataset	27
Gambar 4. 2 Load Data	27
Gambar 4. 3 Cleansing Data	28
Gambar 4. 4 Handling Missing Data.....	28
Gambar 4. 5 Feature Selection.....	29
Gambar 4. 6 Splitting the Data Frame	29
Gambar 4. 7 Train Test Split 80%train & 20%test.....	30
Gambar 4. 8 Train Test Split 70%train & 30%test.....	30
Gambar 4. 9 Proses Pemodelan K-Nearest Neighbor	31
Gambar 4. 10 Proses Pembuatan Model Support Vector Machine	32
Gambar 4. 11 KNN Oversampling Dengan Teknik SMOTE.....	34
Gambar 4. 12 SVM Oversampling Dengan Teknik SMOTE.....	35
Gambar 4. 13 Hasil Klasifikasi K-Nearest Neighbor Data Uji 20%.....	37
Gambar 4. 14 Hasil Klasifikasi K-Nearest Neighbor Data Uji 30.....	38
Gambar 4. 15 Hasil Klasifikasi Support Vector Machine Data Uji 20%	39
Gambar 4. 16 Hasil Klasifikasi Support Vector Machine Data Uji 30%	40
Gambar 4. 17 Hasil Klasifikasi K-Nearest Neighbor Data Uji 20% SMOTE Oversampling	42
Gambar 4. 18 Hasil Klasifikasi K-Nearest Neighbor Data Uji 30% SMOTE Oversampling	44

Gambar 4. 19 Hasil Klasifikasi Support Vector Machine Data Uji 20% SMOTE	
Oversampling	45
Gambar 4. 20 Hasil Klasifikasi Support Vector Machine Data Uji 30% SMOTE	
Oversampling	46



DAFTAR LAMPIRAN

Lampiran 1 Kartu Asistensi Bimbingan.....	62
Lampiran 2 Curriculum Vitae	63
Lampiran 3 BNSP	64

