

**IMPLEMENTASI K-NEAREST NEIGHBOR DALAM
PEMODELAN KLASIFIKASI POTENSI PENYAKIT DIABETES
MILETUS**

SKRIPSI SARJANA SISTEM INFORMASI

Oleh:

Muhammad Ichsan

217006516100



**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNOLOGI KOMUNIKASI DAN INFORMATIKA
UNIVERSITAS NASIONAL**

2025

IMPLEMENTASI K-NEAREST NEIGHBOR DALAM PEMODELAN KLASIFIKASI POTENSI PENYAKIT DIABETES MILETUS

Karya ilmiah sebagai salah satu syarat untuk memperoleh gelar
Sarjana Sistem Informasi dari Fakultas Teknologi Komunikasi dan Informatika



**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS TEKNOLOGI KOMUNIKASI DAN INFORMATIKA
UNIVERSITAS NASIONAL**

2025

HALAMAN PENGESAHAN

TUGAS AKHIR

IMPLEMENTASI K-NEAREST NEIGHBOR DALAM PEMODELAN KLASIFIKASI POTENSI PENYAKIT DIABETES MILETUS



Dosen Pembimbing 1

(Frenda Farahdinna, S.Kom., M.Kom.)

HALAMAN KEASLIAN TUGAS AKHIR

Saya menyatakan dengan sesungguhnya bahwa Tugas Akhir dengan judul :

IMPLEMENTASI K-NEAREST NEIGHBOR DALAM PEMODELAN KLASIFIKASI

POTENSI PENYAKIT DIABETES MILETUS

Yang dibuat untuk melengkapi salah satu persyaratan menjadi Sarjana Komputer pada Program Studi Sistem Informasi Fakultas Teknologi Komunikasi dan Informatika Universitas Nasional, sebagaimana yang saya ketahui adalah bukan merupakan tiruan atau publikasi dari Tugas Akhir yang pernah diajukan atau dipakai untuk mendapatkan gelar di lingkungan Universitas Nasional maupun perguruan tinggi atau instansi lainnya, kecuali pada bagian – bagian tertentu yang menjadi sumber informasi atau acuan yang dicantumkan sebagaimana mestinya.

Jakarta, 28 Februari 2025


Muhammad Ichsan

217006516100

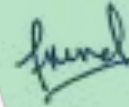
LEMBAR PERSETUJUAN TUGAS AKHIR

Tugas Akhir dengan judul :

IMPLEMENTASI K-NEAREST NEIGHBOR DALAM PEMODELAN KLASIFIKASI POTENSI PENYAKIT DIABETES MELITUS

Dibuat untuk melengkapi salah satu persyaratan menjadi Sarjana Komputer pada Program Studi Sistem Informasi, Fakultas Teknologi Komunikasi dan Informatika Universitas Nasional. Tugas Akhir ini diajukan pada Sidang Review Akhir Semester Ganjil 2024-2025 pada tanggal 26 Februari Tahun 2025

Dosen Pembimbing 1



Frenda Farahdinna, S.Kom., M.Kom

NIDN: 0305119402

Ketua Program Studi



Dr. Andriantingih, S.Kom., M.M.S.I

NIDN: 0303097902

LEMBAR PERSETUJUAN JUDUL YANG TIDAK ATAU YANG DIREVISI

Nama : Muhammad Ichsan
 NPM : 217006516100
 Fakultas/Akademi : Fakultas Teknologi Komunikasi dan Informatika
 Program Studi : Sistem Informasi
 Tanggal Sidang : 26 Februari 2025

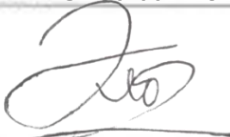
JUDUL DALAM BAHASA INDONESIA :

IMPLEMENTASI K-NEAREST NEIGHBOR DALAM PEMODELAN
 KLASIFIKASI POTENSI PENYAKIT DIABETES MILETUS

JUDUL DALAM BAHASA INGGRIS :

IMPLEMENTATION OF K-NEAREST NEIGHBOR IN MODELING THE
 CLASSIFICATION OF POTENTIAL DIABETES MILETUS DISEASE

TANDA TANGAN DAN TANGGAL

Pembimbing I	Ka. Prodi	Mahasiswa
TGL : 28 Februari 2025	TGL : 28 Februari 2025	TGL : 28 Februari 2025
		

KATA PENGANTAR

Puji syukur kehadiran Allah SWT yang senantiasa melimpahkan rahmat, hidayah, dan karunia-Nya sehingga penulis dapat menyelesaikan skripsi ini dengan judul " Implementasi K-Nearst Neighbor Dalam Pemodelan Klasifikasi Potensi Penyakit Diabetes Miletus INFORMASI KESEHATAN" sebagai salah satu syarat untuk memperoleh gelar Sarjana Sistem Informasi.

Skripsi ini bertujuan untuk mengembangkan analisis berbasis data dalam sistem informasi kesehatan guna membantu klasterisasi pasien berdasarkan tingkat risiko penyakit jantung. Dengan memanfaatkan algoritma K-Nearst Neighbor penelitian ini diharapkan dapat memberikan kontribusi nyata terhadap pengelolaan data kesehatan yang lebih efisien dan mendukung pengambilan keputusan medis yang lebih tepat.

Penyelesaian skripsi ini tentunya tidak terlepas dari bimbingan, dukungan, dan doa dari berbagai pihak. Oleh karena itu, dengan penuh rasa hormat, penulis ingin menyampaikan ucapan terima kasih kepada:

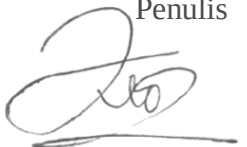
1. Orang tua tercinta yang selalu memberikan doa, dukungan, dan semangat tanpa henti.
2. Bapak/Ibu pembimbing skripsi yang telah dengan sabar memberikan arahan, kritik, dan masukan yang membangun selama proses penyusunan skripsi ini
3. Dosen dan staf akademik Program Studi Sistem Informasi yang telah membekali penulis dengan ilmu pengetahuan dan pengalaman selama masa studi.
4. Rekan-rekan mahasiswa dan teman seperjuangan, yang telah memberikan semangat dan bantuan baik langsung maupun tidak langsung dalam menyelesaikan skripsi ini.

Penulis menyadari bahwa skripsi ini masih jauh dari sempurna. Oleh karena itu, kritik dan saran yang membangun sangat penulis harapkan demi perbaikan di masa

mendatang. Semoga skripsi ini dapat bermanfaat bagi semua pihak, khususnya dalam pengembangan sistem informasi kesehatan di Indonesia.

Jakarta, 22 November 2024

Penulis



Muhammad Ichsan



ABSTRAK

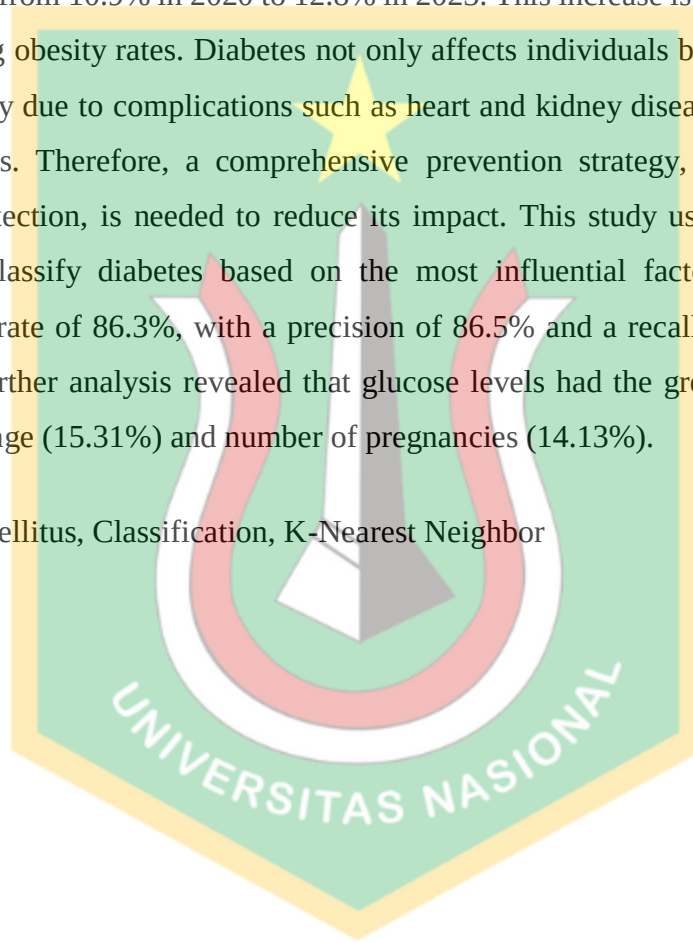
Diabetes Mellitus (DM) merupakan salah satu penyakit tidak menular dengan prevalensi yang terus meningkat secara global, termasuk di Indonesia. Penyakit ini ditandai dengan kadar gula darah tinggi akibat gangguan produksi atau fungsi insulin. Menurut laporan International Diabetes Federation (IDF), pada tahun 2020 terdapat sekitar 463 juta penderita diabetes di seluruh dunia, dan angka ini terus meningkat setiap tahunnya. Indonesia menjadi salah satu negara dengan jumlah penderita diabetes tertinggi di Asia Tenggara, dengan prevalensi yang meningkat dari 10,9% pada 2020 menjadi 12,8% pada 2023. Peningkatan ini disebabkan oleh perubahan gaya hidup tidak sehat dan meningkatnya angka obesitas. Diabetes tidak hanya berdampak pada individu tetapi juga membebani sistem kesehatan nasional, terutama akibat komplikasi seperti penyakit jantung dan ginjal yang meningkatkan angka kematian dan biaya pengobatan. Oleh karena itu, strategi pencegahan yang komprehensif, termasuk edukasi gaya hidup sehat dan deteksi dini, sangat diperlukan untuk mengurangi dampaknya. Penelitian ini menggunakan algoritma K-Nearest Neighbor (KNN) untuk mengklasifikasikan penyakit diabetes berdasarkan faktor yang paling berpengaruh. Hasil penelitian menunjukkan tingkat akurasi klasifikasi sebesar 86,3%, dengan precision 86,5% dan recall 77,1%, yang menunjukkan performa model yang baik. Analisis lebih lanjut mengungkap bahwa kadar glukosa memiliki pengaruh terbesar terhadap diabetes (18,94%), diikuti oleh usia (15,31%) dan jumlah kehamilan (14,13%).

Kata kunci: Diabetes Mellitus, Klasifikasi, K-Nearest Neighbor

ABSTRACT

Diabetes Mellitus (DM) is a non-communicable disease with a prevalence that continues to increase globally, including in Indonesia. This disease is characterized by high blood sugar levels due to impaired insulin production or function. According to the International Diabetes Federation (IDF) report, in 2020 there were around 463 million people with diabetes worldwide, and this number continues to increase every year. Indonesia is one of the countries with the highest number of diabetes sufferers in Southeast Asia, with a prevalence increasing from 10.9% in 2020 to 12.8% in 2023. This increase is due to changes in unhealthy lifestyles and increasing obesity rates. Diabetes not only affects individuals but also burdens the national health system, especially due to complications such as heart and kidney disease which increase mortality rates and medical costs. Therefore, a comprehensive prevention strategy, including healthy lifestyle education and early detection, is needed to reduce its impact. This study uses the K-Nearest Neighbor (KNN) algorithm to classify diabetes based on the most influential factors. The results showed a classification accuracy rate of 86.3%, with a precision of 86.5% and a recall of 77.1%, indicating good model performance. Further analysis revealed that glucose levels had the greatest influence on diabetes (18.94%), followed by age (15.31%) and number of pregnancies (14.13%).

Keywords: Diabetes Mellitus, Classification, K-Nearest Neighbor



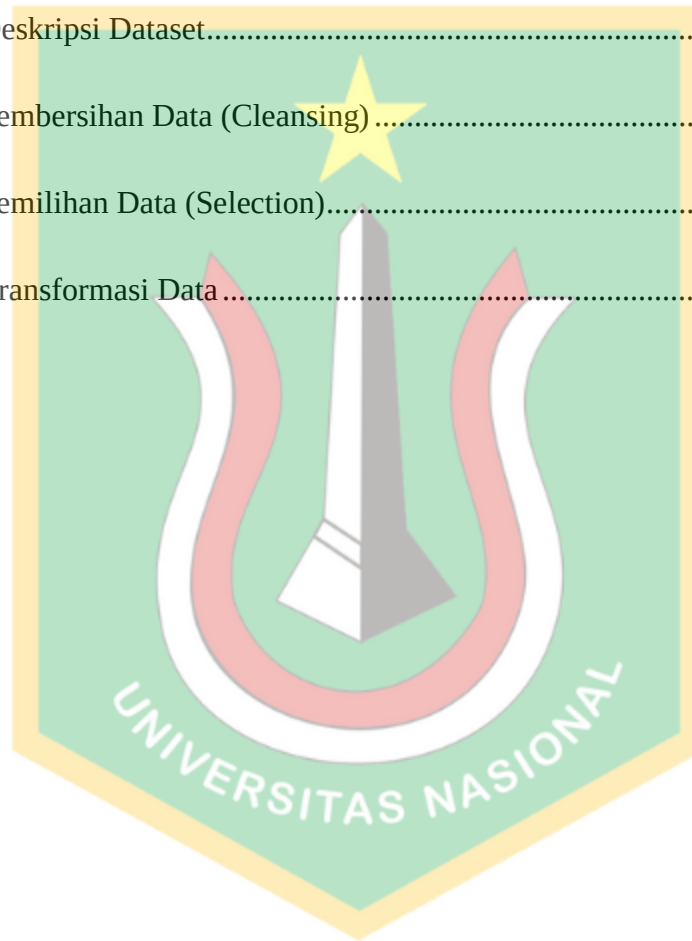
DAFTAR ISI

HALAMAN PENGESAHAN	i
HALAMAN KEASLIAN TUGAS AKHIR.....	ii
LEMBAR PERSETUJUAN TUGAS AKHIR.....	iii
LEMBAR PERSETUJUAN JUDUL YANG TIDAK ATAU YANG DIREVISI	iv
KATA PENGANTAR	v
ABSTRACT	viii
DAFTAR ISI	ix
DAFTAR TABEL	xi
DAFTAR GAMBAR	xii
BAB I PENDAHULUAN	1
1.1. Latar Belakang.....	1
1.2 Identifikasi Masalah	2
1.3 Tujuan Penelitian.....	2
1.4 Batasan Masalah.....	2
1.5 Manfaat Penelitian.....	3
BAB II TINJAUAN PUSTAKA.....	4
2.1 Teori Dasar	4
2.1.1 Data Mining	4
2.1.2 CRISP-DM.....	4
2.1.3 Diabetes Mellitus	6
2.1.4 K-Nearest Neighbor	6
2.1.5 Confusion Matrix	7
2.2 Penelitian Terdahulu.....	9
BAB III METODE PENELITIAN.....	18
3.1 Jenis Penelitian	18

3.2	Waktu Penelitian	18
3.3	Penentuan Subjek Penelitian	18
3.4	Fokus Penelitian	19
3.5	Sumber Data	19
3.6	Teknik Pengumpulan Data	19
3.7	Desain Penelitian	19
3.7.1	Identifikasi Masalah	21
3.7.2	Studi Literatur dan Pengumpulan Data	21
3.7.3	CRISP-DM	21
3.7.4	Kesimpulan	22
BAB IV	HASIL DAN PEMBAHASAN	23
4.1	<i>Data Preparation</i>	23
4.2	<i>Modeling</i>	30
4.3	<i>Evaluation</i>	30
4.4	<i>Deployment</i>	31
BAB V	PENUTUP	33
5.1	Kesimpulan	33
5.2	Saran	33
DAFTAR PUSTAKA		34
LAMPIRAN		35

DAFTAR TABEL

Tabel 2.1 Confusion Matriks.....	8
Tabel 2.2 Studi Literatur Penelitian Terdahulu	9
Tabel 3.1 Waktu Penelitian.....	18
Tabel 4.1 Dataset Diabetes	23
Tabel 4.2 Deskripsi Dataset.....	24
Tabel 4.3 Pembersihan Data (Cleansing).....	24
Tabel 4.4 Pemilihan Data (Selection).....	24
Tabel 4.5 Transformasi Data.....	25



DAFTAR GAMBAR

Gambar 2. 1 CRISP-DM	4
Gambar 3.1 Desain Penelitian	20
Gambar 4.1 Pembagian Data Latih dan Data Uji.....	26
Gambar 4.2 Pemodelan.....	30
Gambar 4.3 Confusion Matrix.....	31
Gambar 4.4 Implementasi Model Klasifikasi Menggunakan Streamlit	32

