

SKRIPSI

**Penerapan Algoritma *K-Nearest Neighbor (KNN)* dan *Naive Bayes*
untuk Klasifikasi Diabetes Mellitus**



Oleh:

Difa Ageng Sukmana

NIM:19083000130

**PROGRAM S1 SISTEM INFORMASI
FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS MERDEKA
MALANG
2023**

**PENERAPAN ALGORITMA K-NEAREST NEIGHBOR (KNN)
DAN NAÏVE BAYES UNTUK KLASIFIKASI DIABETES
MELITUS**



SKRIPSI

Dibuat untuk Memenuhi Salah Satu Persyaratan
Memperoleh Gelar Sarjana Sistem Informasi (S.Kom)

Oleh :

Difa Ageng Sukmana
NIM: 19083000130

**PROGRAM S1 SISTEM INFORMASI
FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS MERDEKA
MALANG
2023**

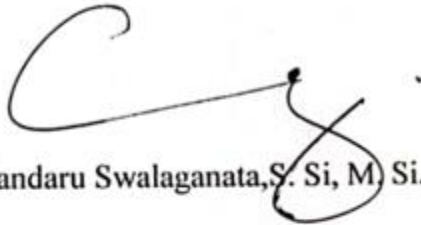
LEMBAR PERSETUJUAN SKRIPSI

Nama : Difa Ageng Sukmana
NIM : 19083000130
Universitas : Merdeka Malang
Fakultas : Teknologi Informasi
Program Studi : Sistem Informasi
Judul Skripsi : Penerapan Algoritma *K-Nearest Neighbor*
(*KNN*) dan *Naive Bayes* untuk Klasifikasi
Diabetes Mellitus

Malang , 21 Agustus 2023

DISETUJUI DAN DITERIMA :

Ketua Program Studi Sistem Informasi



Galandaru Swalaganata, S. Si, M. Si.

Dosen Pembimbing



Arif Saivul Affandi, S.Kom., M.M.

Dekan



Dr. Mardiana Andarwati, M.Si.

LEMBAR PENGESAHAN

Dipersiapkan dan disusun oleh

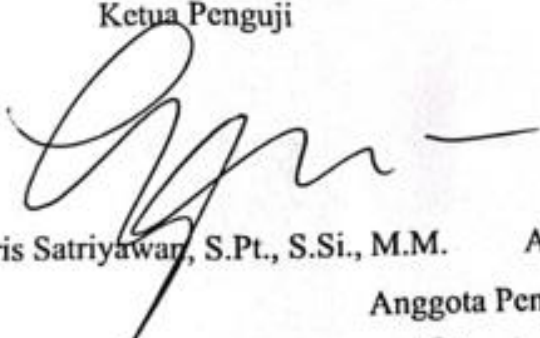
Nama : Difa Ageng Sukmana
NIM : 19083000130
Judul : Penerapan Algoritma *K-Nearest Neighbor (KNN)*
dan *Naive Bayes* untuk Klasifikasi Diabetes
Mellitus

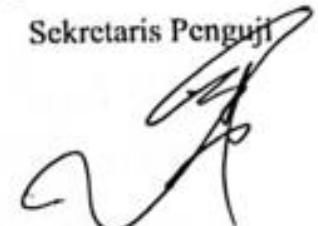
Telah dipertahankan di depan Dosen Penguji pada,

Hari : Senin
Tanggal : 21 Agustus 2023
Tempat : Fakultas Teknologi Informasi Universitas
Merdeka Malang
Susunan Dewan Penguji


Ketua Penguji

Sekretaris Penguji


Haris Satriyawan, S.Pt., S.Si., M.M.


Arif Saivul Affandi, S.Kom., M.M.

Anggota Penguji


Viry Puspaning Ramadhan, A.Md.Li., S.Kom., M.Kom.

Skripsi ini Telah Diterima Sebagai Salah Satu Persyaratan Untuk Memperoleh
Gelar Sarjana Sistem Informasi



Malang, 21 Agustus 2023
Dekan Fakultas Teknologi Informasi
Universitas Merdeka Malang

Dr. Mardiana Andarwati, M.Si.

PERNYATAAN ORISINALITAS

Yang bertanda tangan dibawah ini:

Nama : Difa Ageng Sukmana
NIM : 19083000130
Program Studi : S1 Sistem Informasi
Bidang Kajian : Machine Learning
Judul Tugas Akhir : Penerapan Algoritma *K-Nearest Neighbor (KNN)* dan *Naive Bayes* untuk Klasifikasi Diabetes Mellitus

Alamat Rumah Asal : Jl. Kacapiring No.20, Karangates, Kec.
Sumberpucung, Kab. Malang

No. Telp/Email : 081336115002
: sukmanadifa7@gmail.com

Dengan ini menyatakan bahwa saya benar-benar melakukan penelitian dan penulisan skripsi tersebut di atas benar-benar karya saya dan tidak melakukan plagiasi. Jika saya melakukan plagiasi maka saya bersedia untuk dicabut gelar akademik saya.

Demikian surat pernyataan ini di buat untuk dipergunakan sebagaimana mestinya.

Malang, 31 Agustus 2023

Peneliti,



Difa Ageng Sukmana

MOTTO

“SEMANGAT DALAM MENAPAKI JEJAK PENGETAHUAN.”

LEMBAR PERSEMBAHAN

Dengan tulus dan Ikhlas, penulis mengucapkan rasa Syukur dan terima kasih kepada Allah SWT, Tuhan Yang Maha Esa, atas segala Rahmat, hidayah, karunia-Nya sehingga penulis dapat menyelesaikan penelitian ini. Penulis juga ingin mengucapkan terima kasih yang sebesar-besarnya kepada:

1. Kepada keluarga, terima kasih atas doa, dukungan, dan kasih sayang yang tak terhingga yang menjadi motivasi penulis untuk menyelesaikan penelitian ini.
2. Saya ingin mengucapkan terima kasih kepada Bapak Arif Saivul Affandi, S.Kom., M.M. selaku dosen pembimbing yang telah memberikan arahan, bimbingan, dan masukan berharga dalam penelitian ini.
3. Serta kepada semua dosen dan rekan mahasiswa, yang telah berbagi pengetahuan dan pengalaman dalam perjalanan kuliah dan proses penelitian ini.
4. Penghargaan juga saya sampaikan kepada semua sahabat dan keluarga yang turut serta mendukung penelitian ini, dalam berbagai bentuk kontribusi.

Akhir kata, penulis ingin menyampaikan terima kasih kepada semua yang telah ikut berperan dalam proses penelitian ini. Semoga penelitian ini bisa menambah literatur dan pengetahuan di bidang Kesehatan dan teknologi informasi.

Dengan rendah hati, saya mengucapkan rasa hormat dan terima kasih.

Malang, 21 Agustus 2023



Difa Ageng Sukmana

KATA PENGANTAR

Puji Syukur penulis panjatkan ke hadirat Allah SWT atas rahmat dan hidayah-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul “Penerapan Algoritma *K-Nearest Neighbor (KNN)* dan *Naive Bayes* untuk Klasifikasi Diabetes Mellitus”. Dalam rangka menyelesaikan studi di Program Studi Sistem Informasi, Fakultas Teknologi Informasi, Universitas Merdeka Malang. Penulis ingin mengucapkan rasa terima kasih yang sebesar besarnya kepada:

1. Ibu Dr. Mardiana Andarwati, M.Si selaku Dekan Fakultas Teknologi Informasi Universitas Merdeka Malang.
2. Bapak Galandaru Swalaganata, S.Si., M.Si selaku Ketua Program S1 Sistem Informasi Fakultas Teknologi Informasi Universitas Merdeka Malang.
3. Bapak Arif Saivul Affandi, S.Kom., M.M. selaku dosen pembimbing yang telah memberi arahan, bimbingan, dan masukan dalam menyelesaikan skripsi ini.
4. Bapak dan Ibu Dosen Program S1 Sistem Informasi Fakultas Teknologi Informasi Universitas Merdeka Malang.
5. Bapak dan Ibu Karyawan S1 Sistem Informasi Fakultas Teknologi Informasi Universitas Merdeka Malang.
6. Ibu dan Bapak serta seluruh keluarga yang senantiasa memberikan dukungan secara penuh yang menjadi motivasi penulis untuk menyelesaikan skripsi ini.
7. Teman-teman mahasiswa Program S1 Sistem Informasi Fakultas Teknologi Informasi Universitas Merdeka Malang.

Semua bantuan dan dukungan yang diberikan oleh pihak-pihak di atas sangat berarti bagi penulis dalam menyelesaikan skripsi ini. Penulis menyadari bahwa daftar ini mungkin tidak lengkap, namun penulis mengucapkan terima kasih secara tulus. Akhir kata, penulis berharap skripsi ini dapat memberikan kontribusi positif bagi perkembangan ilmu pengetahuan.

Malang, 21 Agustus 2023



Difa Ageng Sukmana

DAFTAR ISI

LEMBAR PERSETUJUAN SKRIPSI.....	ii
LEMBAR PENGESAHAN	iii
PERNYATAAN ORISINALITAS.....	iv
MOTTO.....	v
LEMBAR PERSEMBAHAN	vi
KATA PENGANTAR.....	vii
DAFTAR ISI	viii
DAFTAR GAMBAR	x
DAFTAR TABEL.....	xi
DAFTAR LAMPIRAN	xii
ABSTRAK	xiii
ABSTRACT.....	xiv
BAB I PENDAHULUAN	1
Latar Belakang	1
1.1 Rumusan Masalah	2
1.2 Tujuan Penelitian.....	2
1.3 Manfaat Penelitian.....	2
1.4 Batasan Masalah.....	3
BAB II KAJIAN PUSTAKA	4
2.1 Landasan Teori	4
2.2 Penelitian Terdahulu.....	10
BAB III METODE PENELITIAN.....	13
3.1 Desain Penelitian.....	13
3.2 Pengumpulan data	13
3.3 Pre-Processing data	14
3.4 Processing data	14
3.5 Evaluasi	15
3.6 Analisis Hasil	15
BAB IV PEMBAHASAN DAN HASIL PENELITIAN	16
4.2 Processing data	16

4.1.1 Drop Duplicate	16
4.1.2 Handling Missing Value	17
4.1.3 Detecting Outlier	18
4.2 Processing data	22
4.2.1 Split Data.....	22
4.2.2 Modeling	23
4.3 Evaluasi Performa	23
4.4 Rancangan Pembaruan Dataset	29
BAB V PENUTUP	32
5.1 Kesimpulan.....	32
5.2 Saran.....	32
DAFTAR PUSTAKA	33
HASIL CEK PLAGIASI	36
LAMPIRAN.....	37

DAFTAR GAMBAR

Gambar 2.1 Logo Python	4
Gambar 2.2 Confusion Matrix	8
Gambar 3.1 Desain Penelitian.....	13
Gambar 3.2 Dataset Penelitian.....	14
Gambar 4.1 Jumlah Isi Kolom	16
Gambar 4.2 Total Duplicate Data	17
Gambar 4.3 Penghapusan Data Duplikat	17
Gambar 4.4 Check Data Kosong.....	18
Gambar 4.5 Drop Data Kosong.....	18
Gambar 4.6 Grafik Outcome Sebelum Penghapusan Outlier	19
Gambar 4.7 Boxplot Data Sebelum Penghapusan Outlier	19
Gambar 4.8 Penghapusan Outlier Dengan IQR	20
Gambar 4.9 Grafik Outcome Sesudah Penghapusan Outlier	20
Gambar 4.10 Boxplot Data Sesudah Penghapusan Outlier.....	21
Gambar 4.11 Split Data.....	22
Gambar 4.12 Import library model	23
Gambar 4.13 Modeling	23
Gambar 4.14 Confusion Matrix Pada KNN.....	24
Gambar 4.15 Confusion Matrix Naïve Bayes	25
Gambar 4.16 Perbandingan Akurasi	26
Gambar 4.17 Perbandingan Presisi	27
Gambar 4.18 Perbandingan Recall.....	27
Gambar 4.19 Perbandingan F1 Score	28
Gambar 4.20 Flowchart Rancangan Pembaruan Dataset.....	30

DAFTAR TABEL

Tabel 2. 1 Penelitian Terdahulu	10
Tabel 4.1 Tabel Hasil Preprocessing Data	21
Tabel 4.2 Performa Model	29

DAFTAR LAMPIRAN

Lampiran 1 Halaman Riwayat Hidup	37
Lampiran 2 Halaman Bimbingan Skripsi.....	38
Lampiran 3 Halaman Persetujuan Skripsi.....	39
Lampiran 4 Halaman Revisi Ujian Skripsi 1	40
Lampiran 5 Halaman Revisi Ujian Skripsi 2	41
Lampiran 6 Nota Dinas Terkait Penelitian.....	42
Lampiran 7 Surat Izin Penelitian.....	43
Lampiran 8 Pemrograman Python	44

ABSTRAK

Penelitian ini fokus pada penerapan algoritma K-Nearest Neighbor (KNN) dan Naive Bayes untuk klasifikasi diabetes mellitus. Diabetes mellitus menjadi permasalahan kesehatan global yang serius, memerlukan diagnosis dini yang akurat untuk mencegah komplikasi lebih lanjut. KNN dan Naive Bayes adalah metode klasifikasi yang memiliki kinerja baik dalam berbagai kasus.

Tujuan dari penelitian ini adalah membandingkan performa KNN dan Naive Bayes dalam klasifikasi diabetes serta mengimplementasikan perbandingan model kedua algoritma ini. Penelitian ini menggunakan analisis data pada dataset dari Dinas Kesehatan Kabupaten Malang dengan total 3811 sampel.

Hasil penelitian menunjukkan bahwa model KNN memiliki kinerja sedikit lebih baik dibandingkan Naive Bayes dalam mengklasifikasikan data diabetes. KNN memiliki akurasi 91,57% sementara Naive Bayes mencapai 90,78%. Meskipun perbedaan ini tidak signifikan, KNN memiliki nilai presisi dan F1-Score yang lebih baik. Di sisi lain, Naive Bayes memiliki recall yang sedikit lebih tinggi daripada KNN.

ABSTRACT

This research focuses on the application of K-Nearest Neighbor (KNN) and Naive Bayes algorithms for diabetes mellitus classification. Diabetes mellitus is a serious global health problem, requiring accurate early diagnosis to prevent further complications. KNN and Naive Bayes are classification methods that have good performance in various cases.

The purpose of this research is to compare the performance of KNN and Naive Bayes in diabetes classification and implement a model comparison of these two algorithms. This research uses data analysis on a dataset from the Malang District Health Office with a total of 3811 samples.

The results showed that the KNN model performed slightly better than Naive Bayes in classifying diabetes data. KNN has an accuracy of 91.57% while Naive Bayes reaches 90.78%. Although this difference is not significant, KNN has better precision and F1-Score values. On the other hand, Naive Bayes has a slightly higher recall than KNN.