

**IMPLEMENTASI ALGORITMA K-NEAREST
NEIGHBOR (KNN) DAN ALGORITMA RANDOM
FOREST DALAM MENGIDENTIFIKASI PENYAKIT
DIABETES**

SKRIPSI SARJANA

Oleh :

Virly Diranisha

207064516105



PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNOLOGI KOMUNIKASI DAN
INFORMATIKA

UNIVERSITAS NASIONAL

2023

**IMPLEMENTASI ALGORITMA K-NEAREST
NEIGHBOR (KNN) DAN ALGORITMA RANDOM
FOREST DALAM MENGIDENTIFIKASI PENYAKIT
DIABETES**

SKRIPSI SARJANA

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

Disusun Oleh :

Virly Diranisha

207064516105



PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNOLOGI KOMUNIKASI DAN
INFORMATIKA
UNIVERSITAS NASIONAL

2023

HALAMAN PERNYATAAN ORISINALITAS

Skripsi ini adalah hasil karya saya sendiri, dan semua sumber yang dikutip maupun dirujuk telah saya nyatakan dengan benar. Bilamana di kemudian hari ditemukan bahwa karya tulis ini menyalahi peraturan yang ada berkaitan etika dan kaidah penulisan karya ilmiah yang berlaku, maka saya bersedia dituntut dan diproses sesuai dengan ketentuan yang berlaku.

Yang menyatakan,

Nama : Virly Diranisha


NIM : 207064516105


Tanda Tangan :



Tanggal : 23 Februari 2024

Mengetahui

Pembimbing I : Dr. Agung Triayudi, S.Kom., M.Kom. ()

Pembimbing II : Ratih Titi Komala Sari, S.T., M.M., MMSI. ()

HALAMAN PENGESAHAN

**IMPLEMENTASI ALGORITMA K-NEAREST
NEIGHBOR (KNN) DAN ALGORITMA RANDOM
FOREST DALAM MENGIDENTIFIKASI PENYAKIT
DIABETES**

Oleh

Virly Diranisha

207064516105

Program Studi Informatika

Fakultas Teknologi Komunikasi dan Informatika

Disetujui pada Tanggal 23 Februari 2024

Pembimbing I



(Dr. Agung Triayudi, S.Kom., M.Kom)
NIDN. 0107019009

Pembimbing II



(Ratih Titi Komala Sari, S.T., M.M., MMSI)
NIDN. 0301038302

**HALAMAN PERNYATAAN PERSETUJUAN
PUBLIKASI TUGAS AKHIR UNTUK
KEPENTINGAN AKADEMIS**

Sebagai sivitas akademik Program Studi Informatika, Fakultas Teknologi Komunikasi dan Informatika, saya yang bertanda tangan di bawah ini:

Nama : Virly Diranisha

NIM : 207064516105

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Fakultas Teknologi Komunikasi dan Informatika, Hak Bebas Royalti Noneksklusif (*Non-exclusive Royalti Free Right*) atas karya ilmiah saya yang berjudul:

**IMPLEMENTASI ALGORITMA K-NEAREST NEIGHBOR (KNN) DAN
ALGORITMA RANDOM FOREST DALAM MENGIDENTIFIKASI
PENYAKIT DIABETES**

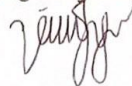
Beserta perangkat yang ada (jika diperlukan). Dengan Hak ini Fakultas Teknologi Komunikasi dan Informatika berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (database), merawat, dan mempublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Jakarta

Pada tanggal : 23 Februari 2024

Yang menyatakan



(Virly Diranisha)

KATA PENGANTAR

Segala rasa syukur saya panjatkan kepada Allah subhanahu wa ta'ala atas limpahan rahmat, karunia, dan petunjuk-Nya yang telah diberikan kepada penulis, memungkinkan penyelesaian skripsi ini dengan sukses. Mengenai judul penelitian ini, yaitu “IMPLEMENTASI ALGORITMA K-NEAREST NEIGHBOR (KNN) DAN ALORITMA RANDOM FOREST DALAM MENGIDENTIFIKASI PENYAKIT DIABETES”. Salah satu tujuan penelitian ini sebagai syarat memperoleh gelar Sarjana Komputer (S.Kom) dari Program Studi Informatika. di Universitas Nasional. Skripsi ini diambil berdasarkan hasil penelitian yang dilakukan dan penulis juga mencari dan menganalisa berbagai sumber bacaan dan referensi. Penulis menyadari jika tanpa bimbingan dan dukungan dari semua pihak dalam proses penelitian ini, tidak dapat selesai tepat pada waktunya dan dengan hasil yang baik. Izinkanlah penulis menyampaikan ucapan terima kasih yang tak terhingga kepada :

1. Bapak Dr. Agung Triayudi, S.Kom., M.Kom. Selaku pembimbing 1 dan Ibu Ratih Titi Komalasari S.T., M.M., MMSI. Selaku dosen pembimbing 2 yang telah menyediakan waktu dan pikirannya dalam membantu penulis menyelesaikan skripsi ini.
2. Ayah, mama, nenek, kaka, bang iyo, bang iman, ade malikha, dan yunus yang telah memberikan dukungan moril dan materil, serta doa kepada penulis sehingga skripsi ini dapat selesai dengan baik.
3. Teman-teman saya wulan, nadia, acha dan teman himti lainnya yang saling menyemangati saling membantu dan berjuang bersama dalam mengerjakan skripsi.
4. Seluruh dosen FTKI Universitas Nasional yang telah memberikan pelajaran selama menempuh pendidikan.
5. Seluruh staff dan karyawan FTKI Universitas Nasional yang telah melayani penulis selama kuliah. Serta semua pihak turut berkontribusi pada terwujudnya penulisan skripsi ini. Penulis mengakui bahwa penulisan ini masih memiliki kekurangan. Sebagai penutup, diharapkan

skripsi ini dapat memberikan manfaat bagi penulis dan semua pihak yang terlibat dan pembacanya.

Jakarta, 10 November
2023

Penulis



HALAMAN PENGESAHAN

TUGAS AKHIR

IMPLEMENTASI ALGORITMA K NEAREST NEIGHBOR DAN
ALGORITMA RANDOM FOREST DALAM MENGIDENTIFIKASI
PENYAKIT DIABETES



Virly Diranisha

207064516105

Dosen Pembimbing 1



Dr. Agung Triayudi, S.Kom., M.Kom.

Dosen Pembimbing 2



Ratih Titi Komala Sari, S.T., M.M., MMSI.

PERNYATAAN KEASLIAN TUGAS AKHIR

Saya menyatakan dengan sesungguhnya bahwa Tugas Akhir dengan judul :

IMPLEMENTASI ALGORITMA K NEAREST NEIGHBOR DAN ALGORITMA RANDOM FOREST DALAM MENGIDENTIFIKASI PENYAKIT DIABETES

Yang dibuat untuk melengkapi salah satu persyaratan menjadi Sarjana Komputer pada Program Studi Informatika 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, 21 Februari 2024



Virly Diranisha

207064516105


LEMBAR PERSETUJUAN TUGAS AKHIR

Tugas Akhir dengan judul :

IMPLEMENTASI ALGORITMA K NEAREST NEIGHBOR DAN ALGORITMA RANDOM FOREST DALAM MENGIDENTIFIKASI PENYAKIT DIABETES

Dibuat untuk melengkapi salah satu persyaratan menjadi Sarjana Komputer pada Program Studi Informatika, Fakultas Teknologi Komunikasi dan Informatika Universitas Nasional. Tugas Akhir ini diujikan pada Sidang Akhir Semester Ganjil 2023-2024 pada tanggal 21 Februari Tahun 2024

Dosen Pembimbing 1



Dr. Agung Triayudi, S.Kom., M.Kom.

NIDN. 0107019009

Ketua Program Studi



Ratih Titi Komala Sari, S.T., M.M., MMSI.

NIDN. 0301038302



LEMBAR PERSETUJUAN JUDUL YANG TIDAK ATAU YANG DIREVISI

Nama : Virly Diranisha
NPM : 207064516105
Fakultas/Akademi : Fakultas Teknologi Komunikasi dan Informatika
Program Studi : Informatika
Tanggal Sidang : 21 Februari 2024



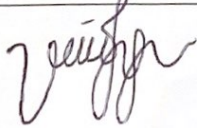
JUDUL DALAM BAHASA INDONESIA :

IMPLEMENTASI ALGORITMA K NEAREST NEIGHBOR DAN ALGORITMA
RANDOM FOREST DALAM MENGIDENTIFIKASI PENYAKIT DIABETES

JUDUL DALAM BAHASA INGGRIS :

IMPLEMENTATION OF K NEAREST NEIGHBOR ALGORITHM AND RANDOM
FOREST ALGORITHM IN IDENTIFYING DIABETES

TANDA TANGAN DAN TANGGAL

Pembimbing 1	Ka. Prodi	Mahasiswa
TGL : 21 Februari 2024	TGL : 21 Februari 2024	TGL : 21 Februari 2024
		

LEMBAR PERSETUJUAN JUDUL YANG TIDAK ATAU YANG DIREVISI

Nama : Virly Diranisha
NPM : 207064516105
Fakultas/Akademi : Fakultas Teknologi Komunikasi dan Informatika
Program Studi : Informatika
Tanggal Sidang : 21 Februari 2024

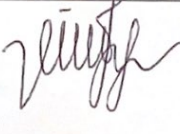
JUDUL DALAM BAHASA INDONESIA :

IMPLEMENTASI ALGORITMA K NEAREST NEIGHBOR DAN ALGORITMA
RANDOM FOREST DALAM MENGIDENTIFIKASI PENYAKIT DIABETES

JUDUL DALAM BAHASA INGGRIS :

IMPLEMENTATION OF K NEAREST NEIGHBOR ALGORITHM AND RANDOM
FOREST ALGORITHM IN IDENTIFYING DIABETES

TANDA TANGAN DAN TANGGAL

Pembimbing 2	Ka. Prodi	Mahasiswa
TGL : 21 Februari 2024	TGL : 21 Februari 2024	TGL : 21 Februari 2024
		

ABSTRAK

Diabetes salah satu jenis penyakit yang tidak dapat ditularkan antar orang (PTM) saat ini menjadi ancaman kesehatan utama di seluruh dunia. Selama ini gejala diabetes hanya di diagnosis oleh orang sesuai dengan karakteristik fisik yang diketahui tanpa dukungan bukti faktual atau pertimbangan medis lainnya. Dengan kemajuan teknologi, memungkinkan penggunaan algoritma dapat menyelesaikan berbagai macam masalah. Salah satu kecerdasan buatan (AI) yaitu machine learning berkonsentrasi pada pembuatan sistem yang dapat belajar dari data. Penelitian ini menggunakan algoritma K-Nearest Neighbor (KNN) dan Random Forest yang dapat dimanfaatkan sebagai algoritma pengujian untuk mengidentifikasi penyakit diabetes. Membandingkan kedua algoritma antara algoritma KNN dan Algoritma Random Forest. Dengan membagi data testing dan data training menggunakan perbandingan 20% : 80% data dari total data 768 menjadi 614 data *training* dan data *test* 20% dari total data 768 menjadi 154 data. Hasil dari evaluasi akurasi yang didapatkan dari Confusion Matrix menunjukkan bahwa Algoritma Random Forest memiliki nilai akurasi terbaik 80%, Presisi 80%, Recall 58% dan F1-Score 67% dengan estimator 100 pohon, kriteria entropy, min sample leaf 5, max features 0.2, max depth 5. Sedangkan Algoritma KNN akurasi yang didapatkan sebesar 75%, Presisi 67%, Recall 62% dan F1-Score 64% dari nilai $K=7$. Berdasarkan hasil Perbandingan dari kedua algoritma tersebut menunjukkan bahwa nilai akurasi yang didapatkan lebih besar algoritma Random Forest walaupun nilai yang didapat tidak jauh berbeda. Data Hasil ditampilkan dengan framework streamlit menampilkan visualisasi data dan percobaan prediksi hasil.

Kata Kunci : Klasifikasi, Diabetes, Perbandingan, Random Forest, K-Nearest Neighbors, Streamlit.

ABSTRACT

Diabetes, a type of disease that cannot be transmitted between people (NCD), is currently a major health threat throughout the world. So far, diabetes symptoms have only been diagnosed by people according to known physical characteristics without the support of factual evidence or other medical considerations. With advances in technology, it is possible to use algorithms to solve various kinds of problems. One type of artificial intelligence (AI), namely machine learning, concentrates on creating systems that can learn from data. This research uses the K-Nearest Neighbor (KNN) and Random Forest algorithms which can be used as testing algorithms to identify diabetes. Comparing the two algorithms between the KNN algorithm and the Random Forest algorithm. By dividing the testing data and training data using a ratio of 20%: 80% of the data from a total of 768 data becomes 614 training data and 20% test data from a total of 768 data becomes 154 data. The results of the accuracy evaluation obtained from the Confusion Matrix show that the Random Forest Algorithm has the best accuracy value of 80%, Precision 80%, Recall 58% and F1-Score 67% with an estimator of 100 trees, entropy criteria, min sample leaf 5, max features 0.2 , max depth 5. Meanwhile, the accuracy obtained by the KNN Algorithm is 75%, Precision 67%, Recall 62% and F1-Score 64% from the value $K=7$. Based on the comparison results of the two algorithms, it shows that the accuracy value obtained is greater than the Random Forest algorithm, although the values obtained are not much different. Results data is displayed using a streamlit framework displaying data visualization and experimental prediction results.

Keywords : Classification, Diabetes, Comparison, Random Forest, K-Nearest Neighbours, Streamlit.

DAFTAR ISI

HALAMAN PERNYATAAN ORISINALITAS	Error! Bookmark not defined.
KATA PENGANTAR.....	iv
ABSTRAK.....	vi
ABSTRACT.....	vii
DAFTAR ISI.....	viii
DAFTAR GAMBAR.....	x
DAFTAR TABEL.....	xii
BAB I.....	1
PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Identifikasi Masalah.....	4
1.3 Tujuan.....	4
1.4 Batasan Masalah.....	5
1.5 Kontribusi.....	5
BAB II.....	6
TINJAUAN PUSTAKA.....	6
2.1 Diabetes Melitus.....	6
2.2 Machine Learning.....	7
2.3 Algoritma K-Nearest Neighbor.....	9
2.4 Algoritma Random Forest.....	10
2.5 Python.....	12
2.6. Streamlit.....	13
2.7 Studi Literatur.....	14
2.7.1 Penelitian Terdahulu.....	14
BAB III.....	17
METODE PENELITIAN.....	17
3.1 Waktu Penelitian.....	17
3.2 Alat dan Bahan.....	17
3.3 Fokus Penelitian.....	18
3.4 Sumber Data.....	18
3.5 Tahap Penelitian.....	18

3.5.1	Identifikasi Masalah	19
3.5.2	Studi Literatur	19
3.5.3	Pengambilan Sumber Data	20
3.5.4	Perancangan Program	21
3.6	Metode Pengujian.....	23
BAB IV	26
HASIL DAN PEMBAHASAN	26
4.1	Implementasi Algoritma K-Nearest Neighbor	26
4.1.1	Perhitungan Manual Algoritma K-Nearest Neighbor	26
4.1.2	Tahap Pemodelan Machine Learning Algoritma K-Nearest Neighbor	28
4.2	Implementasi Algoritma Random Forest	31
4.2.1	Perhitungan Manual Algoritma Random Forest	31
4.2.2	Tahap Pemodelan Machine Learning Algoritma Random Forest	35
4.3	Evaluasi Pengujian Performa Menggunakan Metode Confusion Matrix	38
4.4	Tampilan Interface Pada Sistem Streamlit	40
4.1.1	Menu Algoritma	40
4.1.2	Visualisasi Data.....	41
4.1.3	Hasil Perbandingan.....	44
4.5	Prediksi	45
BAB V	46
KESIMPULAN DAN SARAN	46
DAFTAR PUSTAKA	47
LAMPIRAN	50

DAFTAR GAMBAR

Gambar 1.1 Diagram Penderita Diabetes	2
Gambar 2.1 Ilustrasi KNN	10
Gambar 2.2 Ilustrasi Random Forest	11
Gambar 3.1 Flowchart Tahap Penelitian	19
Gambar 3.2 Dataset Diabetes	20
Gambar 3.3 Pemanggilan Dataset	21
Gambar 3.4 Flowchart Program	22
Gambar 4.1 Nilai Akurasi K=7	28
Gambar 4.2 Akurasi K=2 – K=19 KNN	29
Gambar 4.3 Grafik Value Akurasi KNN	29
Gambar 4.4 Confusion Matrix KNN	30
Gambar 4.5 Hasil Uji KNN	30
Gambar 4.6 Plot Bar Prediksi KNN	31
Gambar 4.7. Pohon Keputusan Root Node	33
Gambar 4.8 Pohon Keputusan Root Node 1	34
Gambar 4.9 Ilustrasi Pohon Keputusan	35
Gambar 4.10 Model Pohon Random Forest	37

Gambar 4.11 Conf Matrix Random Forest	37
Gambar 4.12 Hasil Uji Random Forest	37
Gambar 4.13 Plot Bar Prediksi KNN	38
Gambar 4.14 Menu Algoritma	40
Gambar 4.15 Visualisasi Dataset	41
Gambar 4.16 Visualisasi Algoritma KNN	42
Gambar 4.17 Visualisasi Random Forest	43
Gambar 4.18 Diagram Uji Kedua Algoritma	44
Gambar 4.19 Tampilan Menu Prediksi	45



DAFTAR TABEL

Tabel. 3.1 Waktu Penelitian	17
Tabel 3.2 Variabel Independen	20
Tabel 3.3 Variabel Dependen	21
Tabel 3.4 Confusion Matrix	24
Tabel 4.1 Sampel Data Training	26
Tabel 4.2 Sampel Data Testing	27
Tabel 4.3 Data Training dan Testing	27
Tabel 4.4 Hasil Perhitungan	27
Tabel 4.5 Sample Training Random Forest	31
Tabel 4.6. Sampel Bootstrapped Dataset	31
Tabel 4.7. Perhitungan entropy dan gain	32
Tabel 4.8. Perhitungan entropy dan gain	33
Tabel 4.9 Confusion Matrix Hasil Penguji	38