

**PENDETEKSI PENYAKIT DIABETES MENGGUNAKAN MACHINE LEARNING
DENGAN ALGORITMA K-NEAREST NEIGHBOR (K-NN)**

SKRIPSI



disusun oleh

Arvi Pramudyantoro

18.11.1987

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2022**

**PENDETEKSI PENYAKIT DIABETES MENGGUNAKAN MACHINE LEARNING
DENGAN ALGORITMA K-NEAREST NEIGHBOR (K-NN)**

SKRIPSI

untuk memenuhi sebagian persyaratan
mencapai gelar Sarjana
pada Program Studi Informatika



disusun oleh

Arvi Pramudyantoro

18.11.1987

**PROGRAM SARJANA
PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA
2022**

PERSETUJUAN

SKRIPSI

PENDETEKSI PENYAKIT DIABETES MENGGUNAKAN MACHINE LEARNING DENGAN ALGORITMA K-NEAREST NEIGHBOR (K-NN)

yang dipersiapkan dan disusun oleh

Arvi Pramudyantoro

18.11.1987

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 28 Januari 2022

Dosen Pembimbing,

Rumini, M. Kom
NIK. 190302246

PENGESAHAN

SKRIPSI

PENDETEKSI PENYAKIT DIABETES MENGGUNAKAN MACHINE LEARNING DENGAN ALGORITMA K-NEAREST NEIGHBOR (K-NN)

yang dipersiapkan dan disusun oleh

Arvi Pramudyantoro

18.11.1987

telah dipertahankan di depan Dewan Penguji
pada tanggal 17 Februari 2022

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Sri Ngudi Wahyuni., S.T, M.Kom
NIK. 190302060

Arifiyanto Hadinegoro., S.Kom, M.T
NIK. 190302289

Rumini, M.Kom
NIK. 190302246

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal 17 Februari 2022

DEKAN
FAKULTAS ILMU KOMPUTER



Hanif Al Fatta, M.Kom
NIK. 190302096

PERNYATAAN

Saya yang bertandatangan dibawah ini menyatakan bahwa, skripsi ini merupakan karya saya sendiri (ASLI), dan isi dalam skripsi ini tidak terdapat karya yang pernah diajukan oleh orang lain untuk memperoleh gelar akademis di suatu institusi pendidikan tinggi manapun, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis dan/atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Segala sesuatu yang terkait dengan naskah dan karya yang telah dibuat adalah menjadi tanggungjawab saya pribadi.

Yogyakarta, 18 Maret 2022

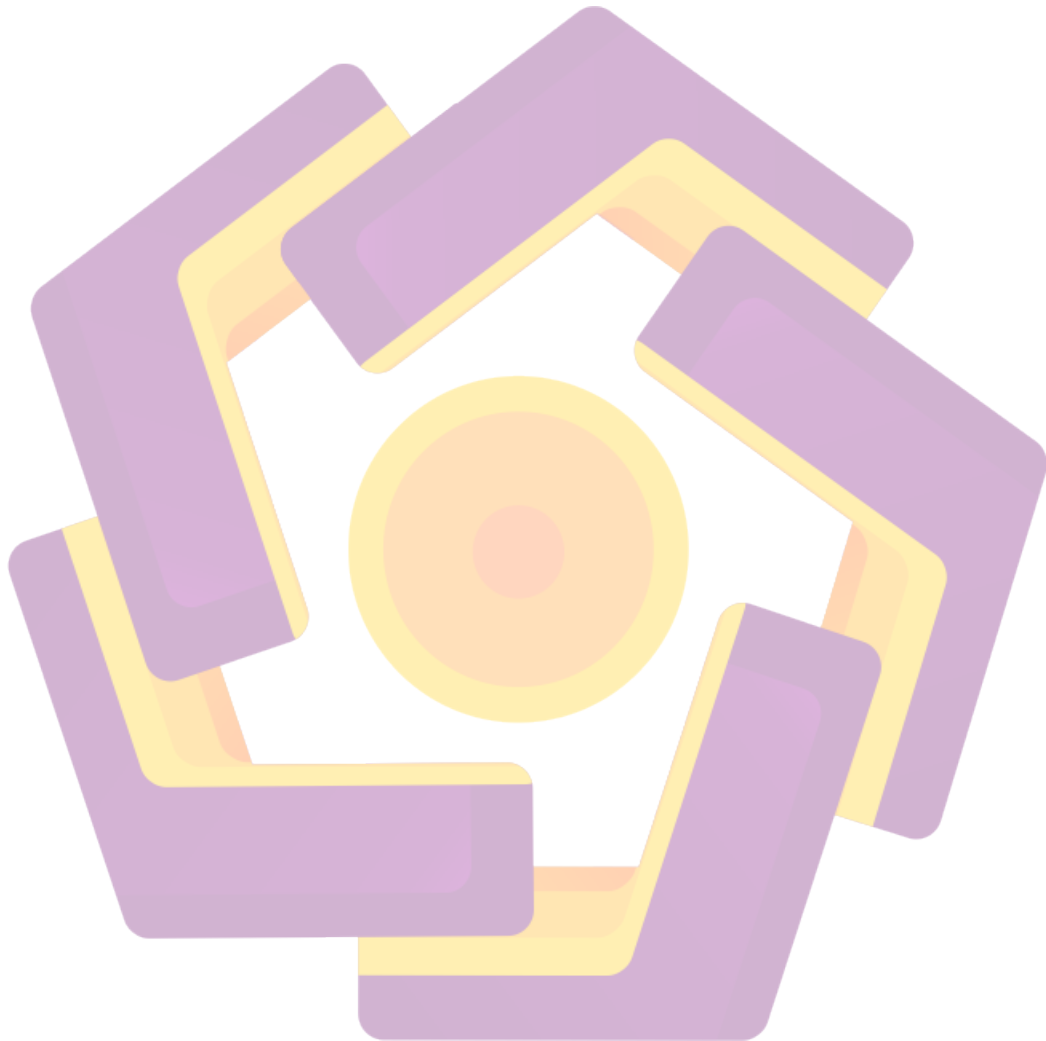


Arvi Pramudyantoro
NIM. 18.11.1987

MOTTO

The only limit to our realization of tomorrow will be our doubts of today. Let us move forward with strong and active faith. – Franklin Roosevelt

(Satu-satunya batasan untuk meraih mimpi kita adalah keragu-raguan kita akan hari ini. Marilah kita maju dengan keyakinan yang aktif dan kuat.) - Franklin Roosevelt



PERSEMBAHAN

Puji syukur saya ucapkan kehadiran Allah SWT yang telah melimpahkan rahmat dan karunia-Nya sehingga saya dapat menyelesaikan penyusunan skripsi yang berjudul “Pendeteksi Penyakit Diabetes Menggunakan Machine Learning Dengan Algoritma K-Nearest Neighbor (K-NN)”, dan diajukan sebagai salah satu syarat untuk mendapat gelar Sarjana pada Prodi Informatika, Fakultas Ilmu Komputer, Universitas Amikom Yogyakarta.

Saya mengucapkan terimakasih kepada pihak yang terlibat langsung maupun tidak langsung atas kelarnya skripsi ini:

1. Kedua orang tua saya yaitu Bapak Sugiya dan Ibu Suyati tercinta, dan juga para kakak-kakak saya yang selalu mendoakan, memberi nasehat dan semangat sehingga dapat menyelesaikan skripsi dengan baik.
2. Keluarga besar yang sudah membantu dalam pengerjaan serta selalu mendoakan.
3. Dosen pembimbing, Ibu Rumini, M.Kom yang telah membimbing dan membantu penulis dari awal sampai akhir pembuatan skripsi.
4. Seluruh dosen dan staff Universitas Amikom Yogyakarta yang telah membantu semasa perkuliahan dengan ilmu dan fasilitasnya.
5. Terimakasih untuk Mardhotillah, you are the best partner.
6. Teman – teman yang selalu membantu penulis dalam menyelesaikan permasalahan yaitu, Rafly Inggar Dean Syach, dan Abimanyu Kurnia Aidul.
7. Teman – teman kelas IF 03, yang menemani penulis pada masa pengerjaan skripsi ini.
8. Teman – teman semasa SMK dan SMP yang selalu mendoakan, berjuang bersama, dan menghibur penulis.
9. Teman - teman Himpunan Mahasiswa Informatika (HMIF) yang senang membagikan ilmunya kepada penulis.
10. Dan untuk semuanya yang tidak bisa penulis sebutkan satu persatu.

KATA PENGANTAR

Puji syukur penulis panjatkan kehadirat Allah SWT yang telah melimpahkan rahmat serta hidayah-Nya, tidak lupa shalawat serta salam selalu penulis panjatkan kepada junjungan kita Nabi Muhammad SAW, yang telah menuntun umatnya sehingga penulis dapat menyelesaikan skripsi dengan baik.

Skripsi yang berjudul “Pendeteksi Penyakit Diabetes Menggunakan Machine Learning Dengan Algoritma K-Nearest Neighbor (K-NN)” disusun sebagai salah satu syarat utama untuk menyelesaikan program sarjana pada Fakultas Ilmu Komputer Universitas AMIKOM Yogyakarta.

Pengerjaan skripsi ini tidak lepas dari bantuan berbagai pihak. Oleh karena itu, penulis ingin menyampaikan rasa hormat dan terima kasih kepada:

1. Prof. Dr. M. Suyanto, MM. selaku Rektor Universitas AMIKOM Yogyakarta.
2. Ibu Windha Mega Pradnya D, M.Kom selaku Ketua Program Studi Informatika Universitas AMIKOM Yogyakarta dan selaku Pembina HMIF.
3. Ibu Rumini, M. Kom selaku dosen wali serta dosen pembimbing yang telah membantu penulis dengan saran dan waktunya.

Penulis menyadari sepenuhnya penelitian ini masih terdapat kekurangan, maka dari itu kritik dan saran serta masukan dari berbagai pihak akan penulis terima dengan lapang dada sebagai perbaikan karya – karya selanjutnya. Semoga skripsi yang sederhana ini dapat bermanfaat bagi pihak yang membutuhkan

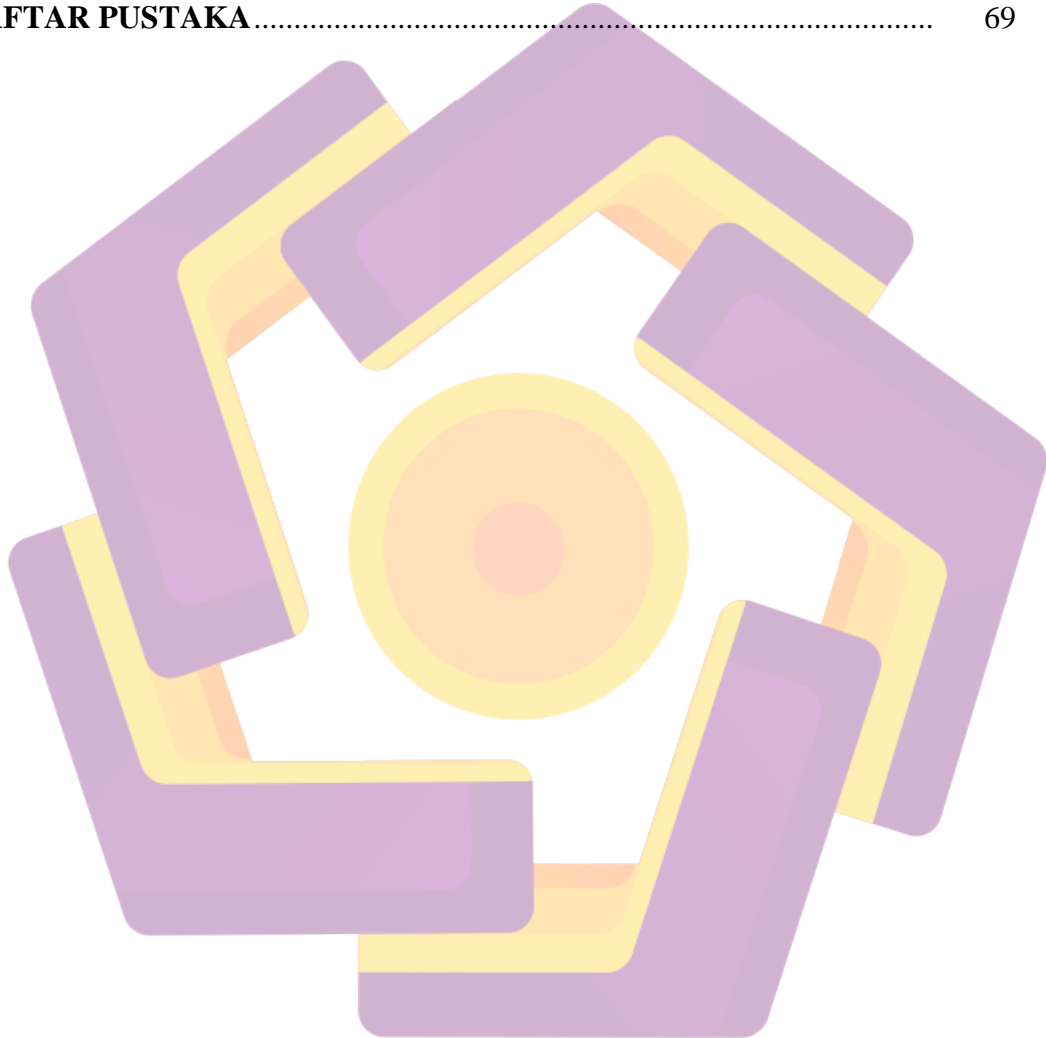
Yogyakarta,

Arvi Pramudyantoro

DAFTAR ISI

JUDUL	ii
PERSETUJUAN	iii
PENGESAHAN	iv
PERNYATAAN	v
MOTTO	vi
PERSEMBAHAN	vii
KATA PENGANTAR	viii
DAFTAR ISI	ix
DAFTAR TABEL	xi
DAFTAR GAMBAR	xii
INTISARI	xiv
ABSTRACT	xv
BAB I PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	3
1.3. Batasan Masalah.....	3
1.4. Urgensi Penelitian	3
1.5. Maksud dan Tujuan Penelitian.....	4
1.6. Manfaat Penulisan	4
1.7. Metode Penelitian.....	5
1.8. Sistematika Penulisan.....	6
BAB II LANDASAN TEORI	8
2.1 Kajian Pustaka.....	8
2.2 Dasar Teori	11
BAB III METODOLOGI PENELITIAN	25
3.1. Alur Penelitian.....	25
3.2. Analisis Kebutuhan	31
3.3. Analisis Prediksi dengan Algoritma K-Nearest Neighbor	32
3.4. Evaluasi	45
3.5. Analisis Kebutuhan Sistem	48

BAB IV IMPLEMENTASI DAN PEMBAHASAN	52
4.1. Deskripsi Implementasi.....	52
4.2. Implementasi Pembuatan Model.....	52
4.3. Implementasi Pembuatan Sistem	62
BAB V PENUTUP	68
5.1. Kesimpulan.....	68
5.2. Saran.....	68
DAFTAR PUSTAKA.....	69



DAFTAR TABEL

Tabel 2. 1 Perbandingan penelitian	10
Tabel 2. 2 Perbandingan penelitian (lanjutan)	11
Tabel 2. 3 Confusion Matrix	18
Tabel 3. 1 Perangkat Untuk Membangun Sistem.....	31
Tabel 3. 2 Perangkat Untuk Menjalankan Sistem.....	32
Tabel 3. 3 Model data training	33
Tabel 3. 4 Model data testing	39
Tabel 3. 5 Nilai Minimal dan Maksimal Data Latih.....	39
Tabel 3. 6 Nilai Minimal Dan Maksimal Data Uji.....	39
Tabel 3. 7 Model data Latih setelah Normalisasi Data.....	40
Tabel 3. 8 Model data Uji setelah Normalisasi Data.....	41
Tabel 3. 9 Model data Latih setelah Normalisasi Data.....	41
Tabel 3. 10 Hasil dari Perhitungan Euclidean Distance.....	42
Tabel 3. 11 Hasil uji Nilai $k=5$	43
Tabel 3. 12 Jumlah Kelas dan Hasil Prediksi.....	44
Tabel 3. 13 Contoh hasil perbandingan prediksi dengan aktual	45
Tabel 3. 14 Model confusion matrix	46
Tabel 3. 15 menunjukkan Confusion Matrix dengan perhitungan pada Excel	46
Tabel 4. 1 <i>Confusion Matrix</i> pengujian pertama.....	59
Tabel 4. 2 <i>Confusion Matrix</i> pengujian kedua	60
Tabel 4. 3 <i>Confusion Matrix</i> pengujian ketiga.....	61

DAFTAR GAMBAR

Gambar 3. 1 Alur Penelitian.....	25
Gambar 3. 2 <i>Flowchart</i> Proses	27
Gambar 3. 3 <i>Data Flow Diagram</i> Level 0	29
Gambar 3. 4 <i>Data Flow Diagram</i> Level 1	30
Gambar 3. 5 Diagram Pasien positif dan negatif Diabetes.....	35
Gambar 3. 6 Diagram Perbandingan kadar glukosa.....	35
Gambar 3. 7 Diagram Pasien positif dan negatif Diabetes.....	36
Gambar 3. 8 Diagram Perbandingan lemak tubuh	36
Gambar 3. 9 Diagram Perbandingan Insulin	37
Gambar 3. 10 Diagram Perbandingan BMI pasien	37
Gambar 3. 11 Diagram umur pada pasien.....	38
Gambar 3. 12 Rancangan antarmuka Halaman Beranda.....	48
Gambar 3. 13 Rancangan antarmuka Halaman Informasi	49
Gambar 3. 14 Rancangan antarmuka Halaman Proses Diagnosis.....	50
Gambar 3. 15 Rancangan antarmuka Hasil Prediksi	50
Gambar 3. 16 Rancangan antarmuka Halaman Tentang	51
Gambar 4. 1 Penginstallan Numpy.....	53
Gambar 4. 2 <i>Import Library</i> Phyton yang di butuhkan.....	53
Gambar 4. 3 Memasukkan dataset pada model Python.....	54
Gambar 4. 4 Menampilkan informasi dari dataset	54
Gambar 4. 5 Mencari nilai yang kosong pada data	55
Gambar 4. 6 Mengubah data 0 menjadi nilai rata-rata	55
Gambar 4. 7 Memisahkan data menjadi x dan y	56
Gambar 4. 8 Normalisasi Data	56
Gambar 4. 9 Pembagian data uji dan data latih.....	57
Gambar 4. 10 Pembuatan KNN.....	57
Gambar 4. 11 Menampilkan Tabel <i>Confusion Matrix</i>	58
Gambar 4. 12 Label <i>Confusion Matrix</i>	58
Gambar 4. 13 Menampilkan Tabel Hasil klasifikasi.....	59
Gambar 4. 14 Penyimpanan Model.....	63
Gambar 4. 15 Penginstallan Virtualenv	63
Gambar 4. 16 Penginstalan Flask.....	64

Gambar 4. 17 Python *App.py*.....64
Gambar 4. 18 Halaman Beranda65
Gambar 4. 19 Halaman Informasi66
Gambar 4. 20 Halaman Prediksi.....66
Gambar 4. 21 Halaman Tentang.....67
Gambar 4. 22 Halaman Hasil Prediksi67



INTISARI

Kesehatan merupakan hal terpenting dalam kehidupan ini. Jika kesehatan manusia sedang terganggu, maka aktivitas sehari - harinya juga akan terganggu. Salah satu penyakit yang mengganggu aktivitas manusia adalah *Diabetes Mellitus*. Dari data *International Diabetes Federation*, Indonesia menduduki peringkat ke - 7 dunia dengan jumlah penderita *Diabetes Mellitus* lebih dari 10 juta jiwa. *World Health Organization* memprediksi jumlah penderita *Diabetes Mellitus* di Indonesia akan mencapai 21 juta jiwa di tahun 2030. Angka penderita *Diabetes Mellitus* yang semakin meningkat menandakan tingkat kesadaran penduduk Indonesia yang masih rendah. Untuk itulah, diperlukan pemahaman dan pengetahuan tentang gejala-gejala awal *Diabetes Mellitus* guna mencegah secara dini penyakit tersebut. Karena semakin dini penyakit terdeteksi akan semakin besar juga kesempatan untuk sembuh. Untuk melakukan itu, peneliti akan memanfaatkan informasi dari situs *Kaggle* untuk mendapatkan dataset yang nantinya dapat diproses dengan algoritma K-Nearest Neighbor dalam memprediksi penyakit *Diabetes Mellitus*.

Algoritma K-Nearest Neighbor merupakan salah satu algoritma yang dapat digunakan untuk mengklasifikasi berdasarkan mayoritas dari nilai k -tetangga terdekat. Tujuan dari algoritma ini adalah untuk mengklasifikasi objek baru berdasarkan *atribut* dan sampel – sampel dari data latih atau *training data*. Dengan algoritma K-Nearest Neighbor, akan digunakan untuk memprediksi penyakit *Diabetes Mellitus* berdasarkan data – data yang sesuai pada pengguna. Hasil prediksi akan berupa kelas positif dan negatif.

Pengujian sistem ini dilakukan sebanyak 3 kali. Total jumlah data yang digunakan yaitu sebanyak 768 data. Dari hasil pengujian didapatkan akurasi tertinggi pada pengujian pertama yaitu dengan akurasi 84 %. Pada pengujian pertama ini menggunakan perbandingan data training dan data testing sebesar 90% dan 10% (691 data *training* dan 77 data *testing*).

Kata Kunci: Prediksi, Algoritma K-Nearest Neighbor, Diabetes Mellitus

ABSTRACT

Health is the most important thing in this life. If human health is being disturbed, then daily activities will also be disrupted. One of the diseases that interfere with human activities is Diabetes Mellitus. Based on data from the International Diabetes Federation, Indonesia is ranked 7th in the world with more than 10 million people with Diabetes Mellitus. The World Health Organization predicts that the number of people with Diabetes Mellitus in Indonesia will reach 21 million in 2030. The increasing number of people with Diabetes Mellitus indicates the level of awareness of the Indonesian population is still low. For that, it is necessary to understand and know about the early symptoms of Diabetes Mellitus in order to prevent the disease early. Because the earlier the disease is detected, the greater the chance for recovery. To do that, researchers will use information from the Kaggle website to get a dataset that can later be accessed by yahoo K-Nearest Neighbor in predicting Diabetes Mellitus.

The K-Nearest Neighbor algorithm is one of the algorithms that can be used to classify based on the majority of k-nearest neighbor values. The purpose of this algorithm is to classify new objects based on attributes and samples from training data or training data. The K-Nearest Neighbor algorithm will be used to predict Diabetes Mellitus based on data that is suitable for the user. Prediction results will be positive and negative classes.

Testing of this system was carried out 3 times. The total amount of data used is 768 data. From the test results obtained the highest accuracy in the first test with an accuracy of 84%. In this first test using a comparison of training data and testing data of 90% and 10% (691 training data and 77 testing data).

Keywords: *Prediction, K-Nearest Neighbor, Diabetes Mellitus*