

**ANALISIS KLASIFIKASI PENYAKIT DIABETES
MENGUNAKAN ALGORITMA *K-NEAREST NEIGHBOR***

SKRIPSI



Disusun oleh:

**Ferdian Estu Anantama
18.11.2559**

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

**ANALISIS KLASIFIKASI PENYAKIT DIABETES
MENGUNAKAN ALGORITMA *K-NEAREST NEIGHBOR***

SKRIPSI

Diajukan kepada Fakultas Ilmu Komputer Universitas AMIKOM Yogyakarta
untuk memenuhi salah satu syarat memperoleh gelar Sarjana Komputer
Pada Jenjang Program Sarjana – Program Studi Informatika



Disusun oleh:

Ferdian Estu Anantama
18.11.2559

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

PERSETUJUAN

SKRIPSI

**IMPLEMENTASI *K-NEAREST NEIGHBOR* (KNN) UNTUK
PENGENALAN WAJAH ORANG BERMASKER**

yang dipersiapkan dan disusun oleh

Ferdian Estu Anantama

18.11.2559

Telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal 31 Agustus 2022

Dosen Pembimbing

Arif Dwi Laksito, M.Kom

NIK. 190302150

PENGESAHAN

SKRIPSI

**ANALISIS KLASIFIKASI PENYAKIT DIABETES
MENGUNAKAN ALGORITMA *K-NEAREST NEIGHBOR***

yang dipersiapkan dan disusun oleh

Ferdian Estu Anantama

18.11.2559

Telah dipertahankan di depan Dewan Penguji
pada tanggal 31 Agustus 2022

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Wiwi Widayani, M.Kom
NIK. 190302272

Krisnawati, S.Si., M.T.
NIK. 190302038

Arif Dwi Laksito, M.Kom
NIK. 190302150

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer

DEKAN FAKULTAS ILMU KOMPUTER

Hanif Al Fatta, S.Kom., 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 akademik di suatu institusi pendidikan tinggi manapun. Dan sepengetahuan saya, tidak terdapat karya atau pendapat yang pernah ditulis dan/atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah skripsi ini dan disebutkan dalam daftar pustaka.

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

Yogyakarta, 7 September 2022



Ferdian Estu Anantama

NIM 18.11.2559

MOTTO

“Orang yang suka berkata jujur mendapatkan tiga hal: kepercayaan, cinta, dan rasa hormat.”

(Ali bin Abi Thalib)

“Lihatlah orang-orang yang di bawahmu dalam urusan harta dunia, dan jangan sekali-kali melihat yang berada di atasmu, supaya kamu tidak meremehkan karunia Allah yang diberikan kepadamu.”

(Nabi Muhammad SAW)

“Jadilah orang yang bermartabat, jujur dan selalu menyampaikan kebenaran.”

(Umar bin Khatab)

“Untuk menang besar, terkadang Anda harus mengambil risiko yang besar pula”

(Bill Gates)

“Resiko Terbesar adalah ketika tidak mengambil resiko apapun. Didunia yang berubah sangat cepat ini, Satu-satu nya Strategi yang dijamin pasti gagal adalah tidak mengambil resiko sama sekali”

(Mark Zuckerberg)

“Pada saatnya semua pasti mengalami kegagalan, tetapi orang yang mampu belajar dari kegagalan adalah orang yang berani untuk sukses”

(Ferdian Estu Anantama)

PERSEMBAHAN

Puji syukur saya ucapkan kehadiran Allah SWT yang telah melimpahkan rahmat dan karunia-Nya sehingga saya dapat menyelesaikan penyusunan skripsi ini dan skripsi ini kupersembahkan kepada :

1. Ibu Ismardani, Seorang wanita hebat tangguh dan penyangga kepada semua anaknya. Wanita yang hebat dalam membentuk karakter anak – anaknya. Dengan tanpa gelar akademik sekalipun tetap menjadi suksesor sarjana bagi anak – anaknya. Izinkan aku membuatmu bangga ketika aku sukses nanti. Doaku sepanjang waktu kepadamu.
2. Bapak Sugiyono, sang tulang punggung keluarga. Walaupun dengan usiamu yang tidak muda lagi tetap gigih memperjuangkan pendidikan setinggi-tingginya kepada anak-anaknya. Semoga engkau diberi kesehatan wal afiat oleh Allah swt, tunggu sebentar lagi putera sulungmu bakal membahagiakanmu. Terima kasih untuk semuanya, Bapak!
3. Kakakku Fatma Nur Rahmawati dan kedua adikku Galang Tri Mahardika dan Garin Bintang Aditama yang selalu memberikan support dan kenyamanan saat berada dirumah. Semoga kita sama-sama sukses dikemudian hari.
4. Bapak Arif Dwi Laksito ,M.Kom sebagai dosen pembimbing yang telah banyak membantu saya berupa bimbingan, motivasi, kritik dan saran atas terselesaikannya skripsi ini.
5. Seluruh dosen Universitas Amikom Yogyakarta yang telah mengajar dan mendidik serta memberikan banyak hal pengalaman kepada saya.
6. Partner saya AS yang ada di solo, terima kasih telah membuat hari hariku jadi berwarna dan yang tidak pernah bosan memberikan doa dan motivasi untuk kebaikanku. Semoga niat baikku dan niat baikmu akan dipercepat dikemudian hari. Sukses selalu buat kamu.
7. Teman kaum burjois dan teman marno boys, terima kasih sahabat perjuangan yang telah banyak membantu dari materi sampe motivasi. Sukses semuanya dan semoga kalian cepat menyusul untuk lulus!

KATA PENGANTAR

Puji syukur atas kehadiran Allah SWT yang telah melimpahkan rahmat dan hidayahnya sehingga penulis dapat menyelesaikan skripsi yang berjudul “Analisis Klasifikasi Penyakit Diabetes Menggunakan Algoritma K-Nearest Neighbor”. Pembuatan Skripsi ini menjadi syarat untuk memenuhi kelulusan S1 Informatika di Universitas Amikom Yogyakarta. Skripsi ini selesai berkat dukungan dari berbagai pihak yang telah memberikan bantuan berupa materi maupun ilmu pengetahuan. Oleh karena itu, penulis mengucapkan terima kasih kepada :

1. Bapak Prof. Dr. M. Suyanto MM selaku rektor Universitas Amikom Yogyakarta.
2. Ibu Windha Mega Pradnya D, M.Kom selaku ketua program studi S1 Informatika.
3. Bapak Arif Dwi Laksito, M.kom selaku dosen pembimbing Skripsi yang telah membimbing dan mengarahkan dalam pembuatan Skripsi.
4. Seluruh dosen, staff dan karyawan Universitas Amikom Yogyakarta.
5. Orangtua dan keluarga yang sudah mendoakan dan mendukung sampai akhir.
6. Teman-teman di universitas amikom yogyakarta yang sudah memberikan dukungannya.

Penulis berharap, semoga skripsi ini dapat memberikan manfaat bagi pembaca. Penulis menyadari jika dalam pembuatan dan penyusunan skripsi ini masih ada kekurangan. Jika ada kritik dan saran berharap semoga dapat menyempurnakan skripsi ini.

Yogyakarta, 7 September 2022

Penulis

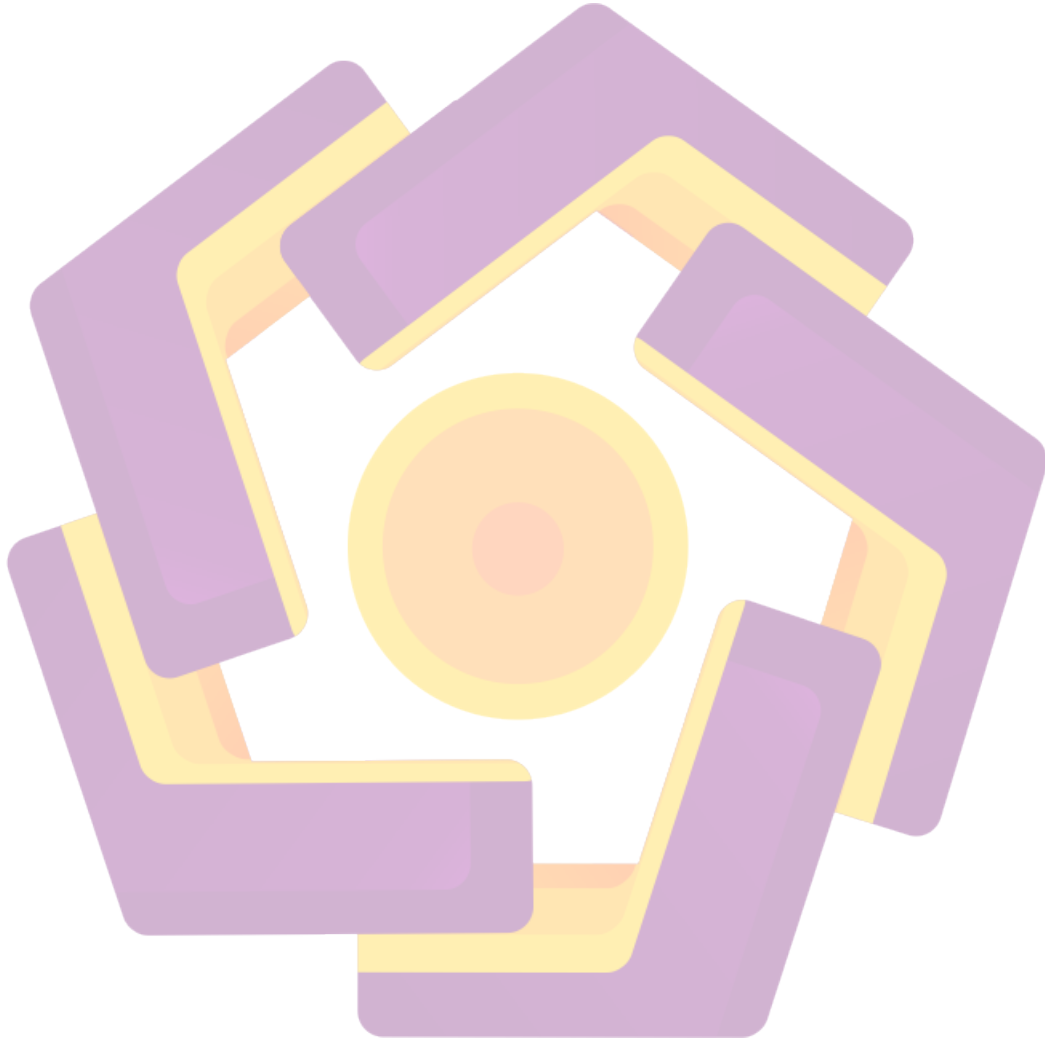
DAFTAR ISI

HALAMAN 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 Masalah	1
1.2 Rumusan Masalah.....	2
1.3 Batasan Masalah	2
1.4 Tujuan Penelitian	2
1.5 Manfaat Penelitian	2
1.6 Sistematika Penulisan	3
BAB II LANDASAN TEORI.....	4
2.1 Tinjauan Pustaka.....	4
2.2 Dasar Teori.....	7
2.2.1 <i>Machine Learning</i>	7
2.2.2 Klasifikasi.....	8
2.2.3 Python	9
2.2.4 K-Nearest Neighbor (KNN)	10
2.2.5 Penyakit Diabetes	11
2.2.6 <i>Balancing Data</i>	12

2.2.7	<i>Random Oversampling (ROS)</i>	12
2.2.8	<i>Confusion Matrix</i>	13
BAB III	METODOLOGI PENELITIAN	15
3.1	Alur Penelitian	15
3.2	Alat Penelitian.....	16
3.3	Pengumpulan Data	16
3.4	<i>Pre-Proccesing Data</i>	18
3.5	<i>Balancing Data</i>	18
3.6	Split Data	19
3.7	Modeling Data	19
3.8	Evaluasi Data	20
BAB IV	pembahasan	21
4.1	Pengumpulan Data	21
4.2	Pre-Processing Data	22
4.3	Set Variabel (x,y).....	23
4.4	Balancing Data	24
4.5	Split Data	26
4.6	Evaluasi Data	28
BAB V	PENUTUP	35
5.1	Kesimpulan	35
5.2	Saran	36
DaftAR	PUSTAKA	37

DAFTAR TABEL

Tabel 2. 1 Tabel Perbandingan.....	5
Tabel 3. 1 Kebutuhan Perangkat Keras dan Perangkat Lunak.....	16
Tabel 4. 1 Perbandingan Hasil Uji Performa KNN	33



DAFTAR GAMBAR

Gambar 2. 1 Penggambaran KNN	10
Gambar 2. 2 Flowchart Algoritma <i>Random Oversampling</i>	13
Gambar 2. 3 Prosedur confusion matrix	14
Gambar 3. 1 Diagram Alur Penelitian.....	15
Gambar 3. 2 Contoh Dataset Diabetes	17
Gambar 3. 3 Contoh Dataset Diabetes	17
Gambar 3. 4 fungsi <i>isna</i> dan <i>sum</i>	18
Gambar 3. 5 <i>Librry balancing data</i>	18
Gambar 3. 6 fungsi <i>corr</i>	19
Gambar 4. 1 Dataset Diabetes	22
Gambar 4. 2 Load dan Explore Data.....	23
Gambar 4. 3 Hasil kode fungsi <i>isna()</i> dan <i>sum()</i>	23
Gambar 4. 4 Hasil Hitung Korelasi.....	24
Gambar 4. 5 Set Variabel Independen dan Dependen	24
Gambar 4. 6 Kode Fungsi <i>Ros</i>	25
Gambar 4. 7 Hasil <i>Balancing Data</i>	25
Gambar 4. 8 Kode Fungsi Split Data	26
Gambar 4. 9 Hasil Split Data	26
Gambar 4. 10 Kode Fungsi KNN (N=3).....	27
Gambar 4. 11 Kode Fungsi KNN (N=5).....	27
Gambar 4. 12 Kode Fungsi KNN (N=7).....	27
Gambar 4. 13 Kode Fungsi KNN (N=9).....	28
Gambar 4. 14 Hasil evaluasi data imbalanced (n=3)	29
Gambar 4. 15 Hasil evaluasi data imbalanced (n=5)	29
Gambar 4. 16 Hasil evaluasi data imbalanced (n=7)	30
Gambar 4. 17 Hasil evaluasi data imbalanced (n=9)	30

Gambar 4.18 Hasil evaluasi data balance (n=3) 31
Gambar 4. 19 Hasil evaluasi data balance (n=5) 31
Gambar 4. 20 Hasil evaluasi data balance (n=7) 32
Gambar 4. 21 Hasil evaluasi data balance (n=9) 32



INTISARI

Diabetes atau biasa disebut penyakit gula dapat diidap oleh siapapun tidak hanya dari faktor keturunan, selain itu terlalu banyak mengkonsumsi gula berkepanjangan merupakan salah satu faktor penyebab penyakit diabetes. Gula itu sendiri memiliki peran yang penting yaitu sumber penyimpanan energi di dalam tubuh, tetapi jika mengonsumsi secara tidak wajar menyebabkan kekebalan pada hormon insulin dan gula darah tubuh menjadi terlalu tinggi. Gula darah tubuh yang tinggi dan tidak terkontrol menyebabkan penyakit diabetes, jika tidak diperhatikan dengan baik penyakit ini dapat menyebabkan terjadinya komplikasi penyakit lainnya.

Metode yang digunakan peneliti pada penelitian ini adalah K-Nearest Neighbors. K-Nearest Neighbors merupakan salah satu metode klasifikasi pada data mining dan juga menjadi algoritma *supervised learning* pada *machine learning* dengan metode pengelompokan suatu hal berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut. K-Nearest Neighbors digunakan untuk mendiagnosa penyakit diabetes. Selain tidak ada asumsi terhadap data, metode ini juga mudah dipahami dan diimplementasikan. K-Nearest Neighbors dapat mengklasifikasi penyakit berdasarkan data pembelajaran dan data pengujian yang dilakukan.

Pada penelitian ini berfokus terhadap pengukuran performa metode *K-Nearest Neighbor* (KNN) dengan berbagai macam nilai K pada objek 768 data. Dataset yang digunakan pada penelitian ini bersumber dari dataset *National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK)* yang dibagikan oleh Mehmet Akturk pada situs *Kaggle* dimana dataset ini terdiri dari 8 atribut umum (*Pregnancies, Glucose, Blood Pressure, Skin Thickness, Insulin, BMI, Diabetes Pedigree Function, Age*) dan 1 kelas (*Outcome*).

Kata kunci: Diabetes, *K-Nearest Neighbors*, *Machine Learning*

ABSTRACT

Diabetes or commonly called sugar disease can be suffered by anyone not only from hereditary factors, besides that consuming too much sugar for a long time is one of the factors causing diabetes. Sugar itself has an important role, namely a source of energy storage in the body, but if you consume it unnaturally, it causes immunity to the hormone insulin and the body's blood sugar becomes too high. High and uncontrolled blood sugar of the body leads to diabetic diseases, if not properly observed this disease can lead to the occurrence of complications of other diseases.

The method used by researchers in this study was K-Nearest Neighbors. K-Nearest Neighbors is one of the classification methods in data mining and also becomes a supervised learning algorithm in machine learning with a method of grouping things based on learning data that is closest to the object. K-Nearest Neighbors is used to diagnose diabetes. In addition to no assumptions about the data, this method is also easy to understand and implement. K-Nearest Neighbors can classify diseases based on learning data and testing data performed.

This study focuses on measuring the performance of the K-Nearest Neighbor (KNN) method with various K values on objects of 768 data. The dataset used in this study is sourced from the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) dataset shared by Mehmet Akturk on the Kaggle site where this dataset consists of 8 common attributes (Pregnancies, Glucose, Blood Pressure, Skin Thickness, Insulin, BMI, Diabetes Pedigree Function, Age) and 1 class (Outcome).

Keyword: *Diabetes, K-Nearest Neighbors, Machine Learning*