

TUGAS AKHIR

**PERBANDINGAN ALGORITMA K-NEAREST NEIGHBOR
DAN GAUSSIAN NAÏVE BAYES PADA KLASIFIKASI
PENYAKIT DIABETES MELITUS**



**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS MUHAMMADIYAH JEMBER
2021**

TUGAS AKHIR

**PERBANDINGAN ALGORITMA K-NEAREST NEIGHBOR
DAN GAUSSIAN NAÏVE BAYES PADA KLASIFIKASI
彭YAKIT DIABETES MELITUS**

Diajukan Kepada Universitas Muhammadiyah Jember Sebagai Persyaratan Untuk
Menyelesaikan Program Studi Teknik Informatika dan Mendapatkan Gelar Strata

Satu (S-1)



Oleh:

PUPUT TRI RAHAYU

1710651023

PROGRAM STUDI TEKNIK INFORMATIKA

FAKULTAS TEKNIK

UNIVERSITAS MUHAMMADIYAH JEMBER

2021

HALAMAN PERSETUJUAN

PERBANDINGAN ALGORITMA K-NEAREST NEIGHBOR DAN GAUSSIAN NAÏVE BAYES PADA KLASIFIKASI PENYAKIT DIABETES MELITUS

Diujukan Oleh :

PUPUT TRI RAHAYU

NIM. 1710651023

Disetujui oleh :

Pembimbing I

Pembimbing II



Daryanto, S.Kom., M.Kom

NIDN: 0707077203



Qurrota A'yun, M.Pd

NIDN. : 0703069002

HALAMAN PENGESAHAN

PERBANDINGAN ALGORITMA K-NEAREST NEIGHBOR DAN GAUSSIAN NAÏVE BAYES PADA KLASIFIKASI PENYAKIT DIABETES MELITUS

Diajukan Oleh :

PUPUT TRI RAHAYU

NIM. 1710651023

Telah disetujui dan disahkan di Jember pada tanggal ... bulan tahun 2021

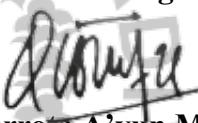
oleh:

Pembimbing I


Daryanto, S.Kom., M.Kom

NIDN: 0707077203

Pembimbing II


Ourrota A'yun M.Pd

NIDN. : 0703069002

Pengaji I


Hardian Oktavianto, S.Si., M.Kom

NIDN: 0722108105

Pengaji II


Ari Eko Wardoyo, S.T., M.Kom

NIDN: 0014027501

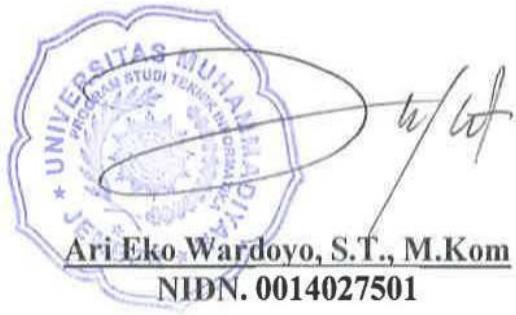
Mengetahui,

Mengesahkan,
Dekan Fakultas Teknik

Mengetahui,
Ketua Program Studi
Teknik Informatika



* Dr. Nanang Saiful Rizal, S.T., MT
NIDN. 0705047806



* Ari Eko Wardoyo, S.T., M.Kom
NIDN. 0014027501

HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini :

Nama : PUPUT TRI RAHAYU
Nomor Induk Mahasiswa : 1710651023
Institusi : Program Studi Teknik Informatika, Fakultas
Teknik Universitas Muhammadiyah Jember

Menyatakan dengan sesungguhnya karya ilmiah berupa tugas akhir yang berjudul **“Perbandingan Algoritma K-Nearest Neighbor Dan Gaussian Naïve Bayes Pada Klasifikasi Penyakit Diabetes Melitus”** adalah murni hasil tanpa adanya tekanan dan paksaan dari pihak manapun hasil buatan atau karya penulis sendiri, bukan hasil plagiasi ataupun mengambil karya orang lain kecuali dalam bentuk kutipan yang telah dicantumkan sumbernya. Sekian pernyataan ini penulis buat tanpa adanya tekanan ataupun paksaan dari pihak manapun.

Jember, ★ September 2021



PUPUT TRI RAHAYU
NIM. 1710651023

HALAMAN PERSEMBAHAN

Alhamdulillah bersyukur atas segala rahmat, taufiq, dan hidayahnya, atas ridho dan karunianya penulis dapat menyelesaikan skripsi ini dengan baik, sebagai ucapan terima kasih penulis persembahkan kepada:

1. Bapak Ari Eko Wardoyo, S.T., M.Kom, selaku Ketua Program Studi Teknik Informatika Fakultas Teknik Universitas Muhammadiyah Jember yang turut memberikan semangat sehingga saya dapat dengan segera menyelesaikan tugas akhir ini.
2. Bapak Ibu dosen pembimbing 1 dan 2 saya yang telah meluangkan waktu dan membantu saya dalam menyusun dan menyelesaikan tugas akhir ini. Yaitu bapak Daryanto, S.Kom., M.Kom. selaku dosen pembimbing 1 dan Ibu Qurrota A'yun M.Pd. selaku dosen pembimbing 2.
3. Bapak dosen yang telah menguji saya dalam tahap seminar proposal dan juga seminar tugas akhir. Hardian Oktavianto, S.Si., M.Kom. selaku dosen penguji 1 dan bapak Ari Eko Wardoyo, S.T., M.Kom. Selaku dosen penguji 2.
4. Kepada ayah dan mama tersayang yang selama ini telah membantu dalam bentuk perhatian dan pengertian, kasih sayang, semangat, serta doa yang tidak pernah berhenti mengalir demi kelancaran dan kesuksesan peneliti dalam menyelesaikan penelitian ini. Kemudian terimakasih banyak untuk kakak-kakak tercinta mami Afran dan tante Dya yang selalu memberi saran serta dukungan kepada peneliti.
5. Kang Minhee dan rivalnya mas gembul tercinta yang telah sangat mempengaruhi penyusunan penelitian ini.
6. Sahabat-sahabat, Chicken University, Bestai, Teen, Ribka, anak pak Dar, anak mas Farid, anak kos uye, JungMo, Serim, Allen, Wonjin, Youngtae, SeongMin, Hyungjun, Woobin, Namjoon, mas Suga, Jungpal, Han Seojoon, Joonki appa dan Avenged Sevenfold, yang telah menjadi penyemangat bagi peneliti dukungan, semangat, motivasi, serta doa hingga peneliti dapat menyelesaikan tugas akhir ini dengan baik.
7. Serta rekan-rekan perjuanganku seluruh Angkatan Teknik Informatika 17 .

MOTTO

“Find Your Goal, By Any Way It Will Be Reached!”

(Cupu)

Arti:

“Temukan tujuanmu, dengan cara apa pun itu akan tercapai”

“Running Towards My Dream With No Regrets”

(Koo Jung Mo)

Arti:

“Berlari Menuju MimpiKu Tanpa Penyesalan”



KATA PENGANTAR

Alhamdulillah atas segala rahmat dan taufik, dan hidayahnya, atas ridho Allah S.W.T., sehingga penulis dapat menyelesaikan skripsi ini dengan judul **“Perbandingan Algoritma K-Nearest Neighbor Dan Gaussian Naïve Bayes Pada Klasifikasi Penyakit Diabetes Melitus”** Skripsi ini diajukan sebagai salah satu persyaratan menyelesaikan Program Studi Teknik Informatika Fakultas Teknik Universitas Muhammadiyah Jember pada kesempatan kali ini penulis menyampaikan terima kasih kepada :

1. Bapak Ari Eko Wardoyo, S.T., M.Kom, selaku Ketua Program Studi Teknik Informatika Fakultas Teknik Universitas Muhammadiyah Jember yang turut memberikan semangat sehingga saya dapat dengan segera menyelesaikan skripsi ini.
2. Bapak Ibu dosen pembimbing 1 dan 2 saya yang telah meluangkan waktu dan membantu saya dalam menyusun dan menyelesaikan skripsi ini. Yaitu bapak Daryanto, S.Kom., M.Kom. selaku dosen pembimbing 1 dan Ibu Qurrota A'yun M.Pd. selaku dosen pembimbing 2.
3. Bapak Ibu dosen yang telah menguji saya dalam tahap seminar proposal dan juga seminar tugas akhir. Hardian Oktavianto, S.Si., M.Kom. selaku dosen penguji 1 dan bapak Ari Eko Wardoyo, S.T., M.Kom. Selaku dosen penguji 2.
4. Seluruh teman-teman mahasiswa program studi Teknik informatika tahun angkatan 2017 yang telah membantu dan bersama-sama selama masa perkuliahan di Universitas Muhammadiyah Jember.

Penulis menyadari masih ada banyak dan kesalahan dan kekurangan dalam penulisan tugas akhir ini, oleh karena itu kritik dan saran akan sangat membantu penulis dalam perbaikan dan penyelesaian tugas akhir ini.

Jember, September 2021

Penulis

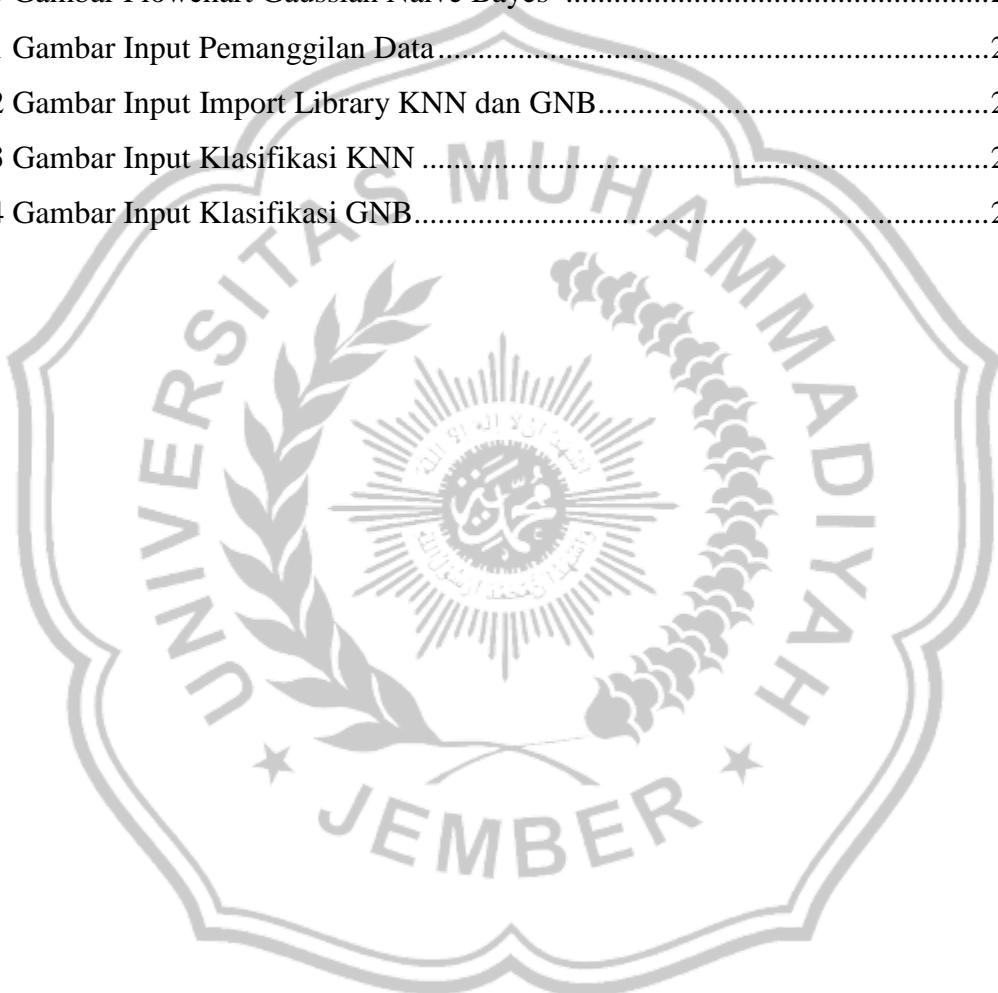
DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN	ii
HALAMAN PENGESAHAN.....	iii
HALAMAN PERNYATAAN.....	iv
HALAMAN PERSEMBERANH	v
MOTTO	vi
ABSTRAK	vii
ABSTRACT	viii
KATA PENGANTAR.....	ix
DAFTAR ISI.....	x
DAFTAR GAMBAR.....	xiii
DAFTAR TABEL	xiv
DAFTAR SIMBOL.....	xv
DAFTAR RUMUS	xvi
BAB 1 PENDAHULUAN	
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Tujuan Penelitian	3
1.4 Batasan Masalah	4
1.5 Manfaat Penelitian	4
BAB 2 TINJAUAN PUSTAKA	
2.1 Penelitoan Terdahulu	5
2.2 Diabetes Melitus	5
2.3 Data Mining	7
2.4 Naïve Bayes	9
2.4.1 Gaussian Naïve Bayes	9
2.5 K- Nearest Neighbor	10
2.6 Confusion Matrix	10
2.6.1 Accuracy	12

2.6.2 Sensitivity	12
2.6.3 Specificity	12
2.6.4 Precision	13
2.6.5 Error rate	13
2.7 Cross Validation.....	13
2.8 Phyton	13
2.9 Jupiter Notebook	14
BAB 3 METODOLOGI PENELITIAN	
3.1 Kerangka Penlitian	15
3.2 Studi Literatur	15
3.3 Pengumpulan Data	15
3.4 Klasifikasi	16
3.5 Validasi dan Evaluasi	16
3.6 Perhitungan Algoritma K-Nearest Neighbor	17
3.7 Perhitungan Algoritma Gaussian Naïve Bayes	22
BAB IV HASILDAN PEMBAHASAN	
4.1. Pengumpulan Data.....	27
4.2. Hasil Klasifikasi.....	27
4.2.1 Hasil Klasifikasi K-Nearest Neighbor.....	30
4.2.2 Hasil Klasifikasi Gaussian Naïve Bayes	32
4.3. Hasil Dan Evaluasi.....	34
BAB V KESIMPULAN DAN SARAN	
5.1 Kesimpulan	35
5.2 Saran	35
DAFTAR PUSTAKA	36
LAMPIRAN	38
BIOADATA PENULIS	51

DAFTAR GAMBAR

3.1 Gambar Kerangka Kerja Penelitian	15
3.2 Gambar Flowchart K-Nearest Neighbor	17
3.3 Gambar Flowchart Gaussian Naïve Bayes	22
4.1 Gambar Input Pemanggilan Data.....	27
4.2 Gambar Input Import Library KNN dan GNB.....	27
4.3 Gambar Input Klasifikasi KNN	28
4.4 Gambar Input Klasifikasi GNB.....	28



DAFTAR TABEL

2.1 Tabel Penelitian Terdahulu	5
2.2 Tabel Ukuran Tekanan Darah	8
2.3 Tabel Confusion Matrix	12
3.1 Tabel Atribut Dataset Beserta Deskripsinya.....	16
3.2 Tabel Data Latih K-Nearest Neighbor	17
3.3 Tabel Data Uji K-Nearest Neighbor	18
3.4 Tabel Hasil Jarak Euclidean Antar Data Latih dan Data Uji	19
3.5 Tabel Mayoritas Data	20
3.6 Tabel Hasil Klasifikasi KNN	21
3.7 Tabel Hasil Confusion Matrix.....	21
3.8 Tabel Data Latih Gaussian Naïve Bayes	22
3.9 Tabel Data Uji Gaussian Naïve Bayes	23
3.10 Tabel Probabilitas Kategori	23
3.11 Tabel Hasil Pencarian Nilai Mean	24
3.12 Tabel Total Jarak Mean Kehamilan Usia Kategori Positif (1)	25
3.13 Tabel Total Jarak Mean Kategori Negatif (0)	25
3.14 Tabel Hasil Klasifikasi Gaussian Naïve Bayes.....	25
3.15 Tabel Hasil Confusion Matrix.....	25
4.1 Tabel Data Diabetes Melitus.....	27
4.2 Tabel Pembagian Data Trining dan Data Testing	27
4.3 Tabel Confusion Matrix pada Algoritma KNN dengan CV 2	29
4.4 Tabel Hasil Klasifikasi pada KNN pada K2 dengan $kfold$ 2-10	31
4.5 Tabel Confusion Matrix pada Algoritma GNB dengan CV 2	32
4.6 Tabel Hasil Klasifikasi pada Algortma GNB dengan $kfold$ 2-10	32
4.7 Hasil Rata-rata KNN dengan K2 dan Uji Cross Validation	33
4.8 Hasil Rata-rata KNN dengan K3 dan Uji Cross Validation	33
4.9 Hasil Rata-rata KNN dengan K4 dan Uji Cross Validation	33
4.10 Hasil Rata-rata KNN dengan K5 dan Uji Cross Validation.....	33
4.11 Hasil Rata-rata pada GNB dengan Uji Cross Validation.....	34
4.12 Tabel Hasil Accuracy, Sensitivity, Specificity, Precision dan Error Rate.....	34

DAFTAR SIMBOL

X	: Data dengan kelas yang belum diketahui
H	: Hipotesis data X merupakan suatu kelas spesifik
$P(H X)$: Probabilitas hipotesis H berdasarkan kondisi X
$P(X)$: Probabilitas hipotesis H
$P(X H)$: Probabilitas X berdasarkan kondisi pada hipotesis H
$P(X)$: Probabilitas X
P	: Peluang
X_i	: Atribut ke i
Y	: Kelas yang dicari
y_j	: Sub kelas Y yang dicari
μ	: Mean, menyatakan rata rata dari seluruh atribut
σ	: Deviasi standar, menyatakan varian dari seluruh atribut
n	: Jumlah sampel
$d(x_i, x_j)$: Jarak Euclidean
x_{1i}	: Data Training
x_{2i}	: Data testing
TP	: Banyaknya data dengan nilai benar positif dan nilai prediksi positif.
FP	: Banyaknya data yang memiliki nilai benar negatif dan nilai prediksi positif.
FN	: Banyaknya data yang memiliki nilai benar positif dan nilai prediksi negatif.
TN	: Banyaknya data dengan nilai benar negatif dan nilai prediksi negatif.
Nc	: Jumlah pada seluruh data c
N	: keseluruhan data

DAFTAR RUMUS

Perhitungan BMI.....	9
Nilai Probabilitas Tertinggi.....	11
Rumus <i>Densitas Gauss</i>	11
Menghitung <i>Mean</i>	12
Menghitung Standar Deviasi.....	12
Rumus Euclidean Distance	13
Rumus <i>Accuracy</i>	14
Rumus <i>Sensitivity</i>	15
Rumus <i>Specificity</i>	15
Rumus <i>Precision</i>	15
Rumus <i>Error Rate</i>	15