

TUGAS AKHIR
PERBANDINGAN METODE K-NEAREST NEIGHBOR DAN
GAUSSIAN NAIVE BAYES UNTUK KLASIFIKASI
PENYAKIT STROKE



Oleh :
Dinda Ulfatul Maula Rachmad
1710651024

TEKNIK INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS MUHAMMADIYAH JEMBER
2021

HALAMAN JUDUL

**PERBANDINGAN METODE K-NEAREST NEIGHBOR DAN
GAUSSIAN NAIVE BAYES UNTUK KLASIFIKASI
PENYAKIT STROKE**

Disusun untuk Melengkapi dan Memenuhi Syarat Kelulusan Guna Meraih Gelar
Sarjana Komputer Teknik Informatika Universitas Muhammadiyah Jember



Oleh :

Dinda Ulfatul Maula Rachmad

1710651024

**TEKNIK INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS MUHAMMADIYAH JEMBER**

2021

HALAMAN PERSETUJUAN

PERBANDINGAN METODE K-NEAREST NEIGHBOR DAN GAUSSIAN NAIVE BAYES UNTUK KLASIFIKASI PENYAKIT STROKE

Oleh :

Dinda Ulfatul Maula Rachmad

(1710651024)

Telah disetujui bahwa Laporan Tugas Akhir ini untuk diajukan pada Sidang
Tugas Akhir sebagai salah satu syarat kelulusan dan mendapatkan gelar

Sarjana Komputer (S.Kom)

di

Universitas Muhammadiyah Jember

Disetujui oleh :

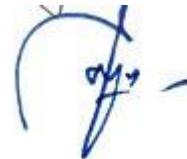
Pembimbing I



Hardian Oktavianto, S.Si., M.Kom.

NIDN. 0722108105

Pembimbing II



Miftahur Rahman, S.Kom., M.Kom

NIDN. 0724039201

HALAMAN PENGESAHAN

**PERBANDINGAN METODE K-NEAREST NEIGHBOR DAN GAUSSIAN
NAIVE BAYES UNTUK KLASIFIKASI PENYAKIT STROKE**

Oleh :

**Dinda Ulfatul Maula Rachmad
(1710651024)**

Telah mempertanggung jawabkan Laporan Tugas Akhirnya pada Sidang Tugas Akhir tanggal 6 November 2021 sebagai salah satu syarat kelulusan dan mendapatkan gelar Sarjana Komputer (S.Kom)

di

Universitas Muhammadiyah Jember

Disetujui oleh :

Penguji I



Wiwik Suharso, S.Kom.,M.Kom.

NIDN. 0006097601

Pembimbing I



Hardian Oktavianto, S.Si., M.Kom.

NIDN. 0722108105

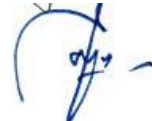
Penguji II



Reni Umilasari, S.Pd.,M.Si.

NIDN. 0728079101

Pembimbing II



Miftahur Rahman, S.Kom.,M.Kom

NIDN. 0724039201

Mengesahkan,
Dekan Fakultas Teknik


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

Mengetahui,
Ketua Program Studi
Teknik Informatika


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

LEMBAR PERNYATAAN KEASLIAN

Yang bertanda tangan dibawah ini:

NAMA : DINDA ULFATUL MAULA RACHMAD
NIM : 1710651024
INSTITUSI : Strata-1 Program Studi Teknik Informatika Fakultas
Teknik, Universitas Muhammadiyah Jember

Menyatakan bahwa Tugas Akhir yang berjudul **“Perbandingan Metode K-Nearest Neighbor Dan Gaussian Naive Bayes Untuk Klasifikasi Penyakit Stroke”** bukan merupakan Tugas Akhir orang lain baik sebagian maupun keseluruhan kecuali dalam bentuk kutipan yang telah disebutkan sumbernya.

Demikian surat pernyataan ini dibuat dengan sebenar-benarnya dan apabila pernyataan ini tidak benar penulis bersedia mendapatkan sanksi akademik.

Jember, 6 November 2021

Yang Menyatakan,



Dinda Ulfatul Maula Rachmad

NIM. 1710651024

MOTTO

“Pembelajaran tidak didapat dengan kebetulan. Ia harus dicari dengan semangat dan disimak dengan tekun “

(Abigail Adams)

“Jangan malu dengan kegagalanmu, belajarlh darinya dan mulai lagi”

(Richard Branson)

“Berbuat baiklah tanpa perlu alasan “



UNGKAPAN TERIMA KASIH

Puji syukur kepada Allah S.W.T atas rahmat dan hidayah-Nya sehingga penulis dapat menyelesaikan laporan tugas akhir ini. Atas segala upaya, bimbingan, dan arahan dari semua pihak, penulis mengucapkan terimakasih sebesar-besarnya kepada:

1. Allah SWT berkat segala ridho, rahmat dan hidayah-Nya penulis dapat menyelesaikan segala urusan dalam menyusun laporan Tugas Akhir dan diberikan kesempatan mendapatkan gelar Sarjana Komputer.
2. Bapak Dr. Nanang Saiful Rizal, S.T., M.T selaku Dekan Fakultas Teknik Universitas Muhammadiyah Jember.
3. Bapak Ari Eko Wardoyo, S.T., M.Kom selaku Kepala Program Studi Teknik Informatika di Universitas Muhammadiyah Jember.
4. Bapak Hardian Oktivianto, S.Si., M.Kom selaku dosen pembimbing 1 tugas akhir yang telah memberikan bimbingan dari awal sampai akhir dengan penuh kesabaran sehingga tugas akhir ini dapat terselesaikan.
5. Bapak Miftahur Rahman, S.Kom., M.Kom selaku dosen pembimbing 2 tugas akhir yang telah memberikan bimbingan dari awal sampai akhir dengan penuh kesabaran sehingga tugas akhir ini dapat terselesaikan.
6. Bapak Wiwik Suharso, S.Kom., M.Kom selaku dosen penguji 1 yang telah memberikan kritik, saran dan masukan yang membangun dalam penelitian ini.
7. Ibu Reni Umilasari, S.Pd., M.Si selaku dosen penguji 2 yang telah memberikan kritik, saran dan masukan yang membangun dalam penelitian ini.
8. Kedua orang tua saya Bapak Dede Rachmad dan Ibu Nurita, terimakasih atas support dan doanya dalam penyusunan laporan tugas akhir ini.
9. Teman-teman seperjuangan yang senasib dan setanah air,” terimakasih atas doa dan dukungannya selama ini, dan terima kasih atas kebersamaannya.

PERBANDINGAN METODE K-NEAREST NEIGHBOR DAN GAUSSIAN NAIVE BAYES UNTUK KLASIFIKASI PENYAKIT STROKE

Dinda Ulfatul Maula Rachmad, Hardian Oktavianto, Miftahur Rahman

Prodi Teknik Informatika Fakultas Teknik Universitas Muhammadiyah Jember

Abstrak

Stroke adalah kondisi bahaya yang perlu ditangani secepatnya, karena sel otak dapat mati dalam hitungan menit. Kematian secara mendadak mungkin terjadi apabila seorang pasien mengalami yang sangat parah. Penanganan yang cepat dapat mengurangi tingkat kerusakan pada otak dan kemungkinan timbulnya komplikasi. Maka perlu dilakukan prediksi orang tersebut terkena penyakit stroke atau tidak. Salah satu cara untuk memprediksi penyakit stroke yaitu menggunakan klasifikasi. Penyakit stroke perlu diklasifikasi agar dapat memprediksi penyakit dengan akurat. Hasil prediksi yang akurat membantu praktisi kesehatan dalam mengambil keputusan dengan tepat. Metode klasifikasi yang digunakan pada penelitian ini yaitu membandingkan algoritma *K-Nearest Neighbor* dan *Gaussian Naive Bayes*. Hasil yang diperoleh dari penelitian ini pada algoritma *K-Nearest Neighbor* didapatkan dengan hasil akurasi sebesar 68,30%, presisi sebesar 67,20% dan recall sebesar 73,34%, sedangkan pada algoritma *Gaussian Naive Bayes* mendapatkan hasil akurasi sebesar 74,45%, presisi sebesar 74,01% dan *recall* sebesar 75,71%. Dari perbandingan akurasi, presisi dan *recall* dapat dilihat bahwa terdapat peningkatan akurasi sebesar 6,15%, presisi sebesar 6,81% dan *recall* sebesar 2,37%, sehingga membuktikan bahwa kinerja algoritma *Gaussian Naive Bayes* lebih baik.

Kata Kunci : Klasifikasi penyakit, Stroke, *Gaussian Naive Bayes*, *K-Nearest Neighbor*.

**COMPARISON OF K-NEAREST NEIGHBOR AND GAUSSIAN NAIVE
BAYES METHODS FOR STROKE DISEASE CLASSIFICATION**

*Dinda Ulfatul Maula Rachmad, Hardian Oktavianto, Miftahur Rahman
Informatics Engineering Study Program, Faculty of Engineering, University of
Muhammadiyah Jember*

Abstract

Stroke is a dangerous condition that needs to be treated as soon as possible, because brain cells can die in a matter of minutes. Sudden death is possible if a patient has a very severe condition. Prompt treatment can reduce the level of damage to the brain and the possibility of complications. So it is necessary to predict whether the person has a stroke or not. One way to predict stroke is to use classification. Stroke needs to be classified in order to predict the disease accurately. Accurate prediction results help health practitioners in making the right decisions. The classification method used in this study is to compare the K-Nearest Neighbor and Gaussian Naive Bayes algorithms. The results obtained from this study on the K-Nearest Neighbor algorithm were obtained with an accuracy of 68.30%, a precision of 67.20% and a recall of 73.34%, while the Gaussian Naive Bayes algorithm obtained an accuracy of 74.45%, precision is 74.01% and recall is 75.71%. From the comparison of accuracy, precision and recall, it can be seen that there is an increase in accuracy of 6.15%, precision of 6.81% and recall of 2.37%, thus proving that the performance of the Gaussian Naive Bayes algorithm is better.

Keywords: Disease classification, Stroke, Gaussian Naive Bayes, K-Nearest Neighbor.

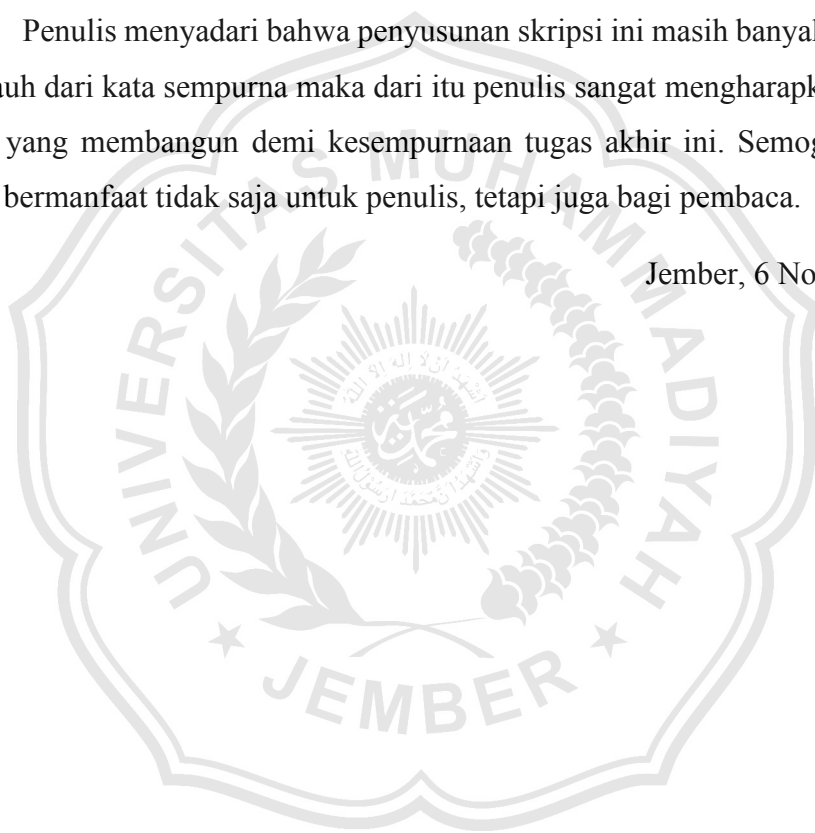
KATA PENGANTAR

Puji syukur kehadirat Allah SWT., yang telah melimpahkan rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan Tugas Akhir Skripsi yang berjudul “Perbandingan Metode *K-Nearest Neighbor* Dan *Gaussian Naive Bayes* Untuk Klasifikasi Penyakit Stroke”. Skripsi ini disusun untuk memenuhi salah satu syarat untuk menyelesaikan pendidikan Strata Satu (S1) pada Jurusan Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Jember.

Penulis menyadari bahwa penyusunan skripsi ini masih banyak kekurangan dan jauh dari kata sempurna maka dari itu penulis sangat mengharapkan kritik dan saran yang membangun demi kesempurnaan tugas akhir ini. Semoga tulisan ini dapat bermanfaat tidak saja untuk penulis, tetapi juga bagi pembaca.

Jember, 6 November 2021

Penulis



DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN	ii
LEMBAR PERNYATAAN KEASLIAN	iv
MOTTO	v
UNGKAPAN TERIMA KASIH	vi
Abstrak.....	vii
Abstract.....	viii
KATA PENGANTAR.....	ix
DAFTAR ISI.....	xi
DAFTAR TABEL	xiv
DAFTAR GAMBAR.....	xvi
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Tujuan Penelitian.....	3
1.4 Batasan Masalah.....	3
1.5 Manfaat Penelitian.....	3
BAB II TINJAUAN PUSTAKA.....	4
2.1 Penelitian Terdahulu.....	4
2.2 Stroke.....	6
2.3 Data Mining.....	6
2.3.1 Tahap-tahap Data Mining	7

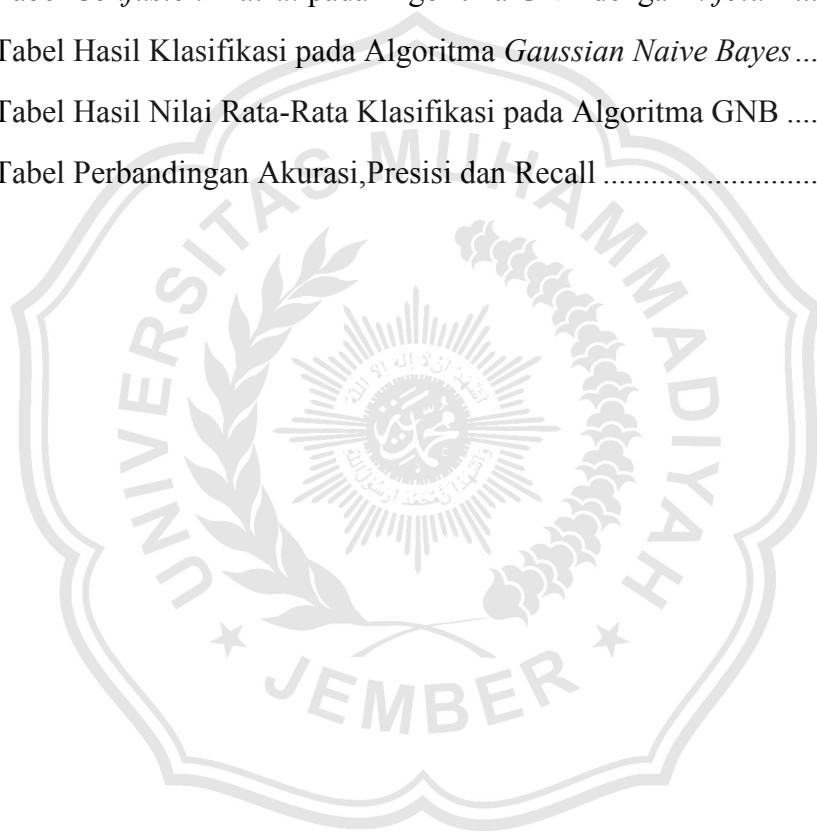
2.3.2 Tipe-tipe Data	8
2.4 Klasifikasi.....	9
2.5 Algoritma <i>K-Nearest Neighbor (K-NN)</i>	10
2.6 Naive Bayes.....	10
2.7 Gaussian Naive Bayes	11
2.7.1 Mean	12
2.7.2 Standar Deviasi	12
2.8 <i>Confusion Matrix</i>	12
2.9 <i>Cross Validation</i>	13
2.10 <i>Python</i>	13
2.11 <i>Jupyter Notebook</i>	14
BAB III METODE PENELITIAN	15
3.1 Tahapan Penelitian	15
3.2 Studi Literatur.....	15
3.3 Pengumpulan Data	15
3.1 <i>Preprocessing Data</i>	16
3.4 Klasifikasi.....	17
3.5 Validasi dan Evaluasi	17
3.6 Perhitungan Algoritma <i>K-Nearest Neighbor</i>	18
3.7 Implementasi Gaussian Naive Bayes	23
3.7.1 Menghitung Nilai Probabilitas.....	24
3.7.2 Mencari Nilai Mean	24
3.7.3 Mencari Standar Deviasi.....	25
3.7.4 Mencari Nilai Probabilistik.....	26
3.7.5 Mencari Hasil Klasifikasi	26

BAB IV HASIL DAN PEMBAHASAN	27
4.1 Pengumpulan Data	27
4.2 Preprocessing Data	28
4.3 Hasil Klasifikasi	28
4.3.1 Hasil Klasifikasi <i>K-Nearest Neighbor</i>	31
4.3.2 Hasil Klasifikasi Gaussian Naive Bayes.....	39
4.4 Hasil dan Evaluasi	42
BAB V KESIMPULAN DAN SARAN	43
5.1 Kesimpulan.....	43
5.2 Saran.....	43
DAFTAR PUSTAKA	45
LAMPIRAN.....	47
Lampiran 1. Data sesudah <i>preprocessing</i>	47
Lampiran 2. Hasil Nilai Akurasi Pengujian <i>k-fold</i>	63
Lampiran 3. Input Klasifikasi <i>K-Nearest Neighbor</i>	71
Lampiran 4. Input Klasifikasi Gaussian Naive Bayes.....	75
Lampiran 5. Lembar Revisi Sidang Tugas Akhir Penguji 1	77
Lampiran 6. Lembar Revisi Sidang Tugas Akhir Penguji 2	78

DAFTAR TABEL

2.1 Tabel Penelitian Terdahulu	5
2.2 Tabel <i>Confusion Matrix</i>	13
3.1 Tabel Atribut Dataset	16
3.2 Tabel Tranformasi Data	17
3.3 Tabel Data Latih K-Nearest Neighbor	19
3.4 Tabel Data Uji	20
3.5 Tabel Hasil Jarak Euclidean antar Data Latih dan Data Uji	21
3.6 Tabel Mayoritas Data	22
3.7 Tabel Probabilitas Kategori.....	24
3.8 Tabel Hasil Pencarian Nilai Mean	24
3.9 Tabel Total Jarak Mean Age Positive(1)	25
3.10 Tabel Total Jarak Mean Age Negative(0).....	25
4.1 Tabel Data Penyakit Stroke.....	27
4.2 Tabel Hasil Preprocessing Data	28
4.3 Tabel K-Fold 2 Skenario Pertama.....	31
4.4 Tabel Confusion Matrix Algoritma K-Nearest Neighbor 2-fold.....	31
4.5 Tabel K-Fold 3 Skenario Pertama.....	32
4.6 Tabel Confusion Matrix Algoritma K-Nearest Neighbor 3-fold.	32
4.7 Tabel K-Fold 4 Skenario Pertama.....	32
4.8 Tabel Confusion Matrix Algoritma K-Nearest Neighbor 4-fold.	33
4.9 Tabel K-Fold 5 Skenario Pertama.....	33
4.10 Tabel Confusion Matrix Algoritma K-Nearest Neighbor 5-fold.	33
4.11 Tabel K-Fold 6 Skenario Pertama.....	34
4.12 Tabel Confusion Matrix Algoritma K-Nearest Neighbor 6-fold	34
4.13 Tabel K-Fold 7 Skenario Pertama.....	35
4.14 Tabel Confusion Matrix Algoritma K-Nearest Neighbor 7-fold	35
4.15 Tabel K-Fold 8 Skenario Pertama.....	36

4.16 Tabel Confusion Matrix Algoritma K-Nearest Neighbor 8-fold	36
4.17 Tabel K-Fold 9 Skenario Pertama.....	37
4.18 Tabel Confusion Matrix Algoritma K-Nearest Neighbor 9-fold	37
4.19 Tabel K-Fold 10 Skenario Pertama.....	38
4.20 Tabel Confusion Matrix Algoritma K-Nearest Neighbor 10-fold	38
4.21 Tabel Nilai Akurasi, Presisi, Recall pada Percobaan K	39
4.22 Tabel <i>Confusion Matrix</i> pada Algoritma GNB dengan <i>k-fold</i> 2	40
4.23 Tabel Hasil Klasifikasi pada Algoritma <i>Gaussian Naive Bayes</i>	41
4.24 Tabel Hasil Nilai Rata-Rata Klasifikasi pada Algoritma GNB	41
4.25 Tabel Perbandingan Akurasi, Presisi dan Recall	42



DAFTAR GAMBAR

3.1 Gambar Tahap Penelitian	15
3.2 Gambar Flowchart <i>K-Nearest Neighbor</i>	18
3.3 Gambar Flowchart Gaussin Naive Bayes	23
4.2 Gambar Input Fold	28
4.3 Gambar Input Membagi Data.....	29
4.4 Gambar Input Nilai K pada <i>K-Nearest Neighbor</i>	29
4.5 Gambar Input Data Uji dan Data Prediksi	29
4.6 Gambar Input Confusion Matrix.....	29
4.7 Gambar Input Akurasi, Presisi, Recall.....	29
4.8 Gambar Input Fold	30
4.9 Gambar Input Gaussian Naive Bayes	30
4.10 Gambar Input Data Uji dan Data Prediksi	30
4.11 Gambar Input Akurasi, Presisi, Recall.....	30