

PERBANDINGAN ALGORITMA KLASIFIKASI DATA MINING UNTUK PREDIKSI PENYAKIT STROKE

Laporan Tugas Akhir

Diajukan Untuk Memenuhi
Persyaratan Guna Meraih Gelar Sarjana Strata 1
Teknik Informatika Universitas Muhammadiyah Malang



Aidia Khoiriyah Firdausy
201910370311260

Bidang Minat
Data Science

**PROGRAM STUDI INFORMATIKA FAKULTAS
TEKNIK**

UNIVERSITAS MUHAMMADIYAH MALANG

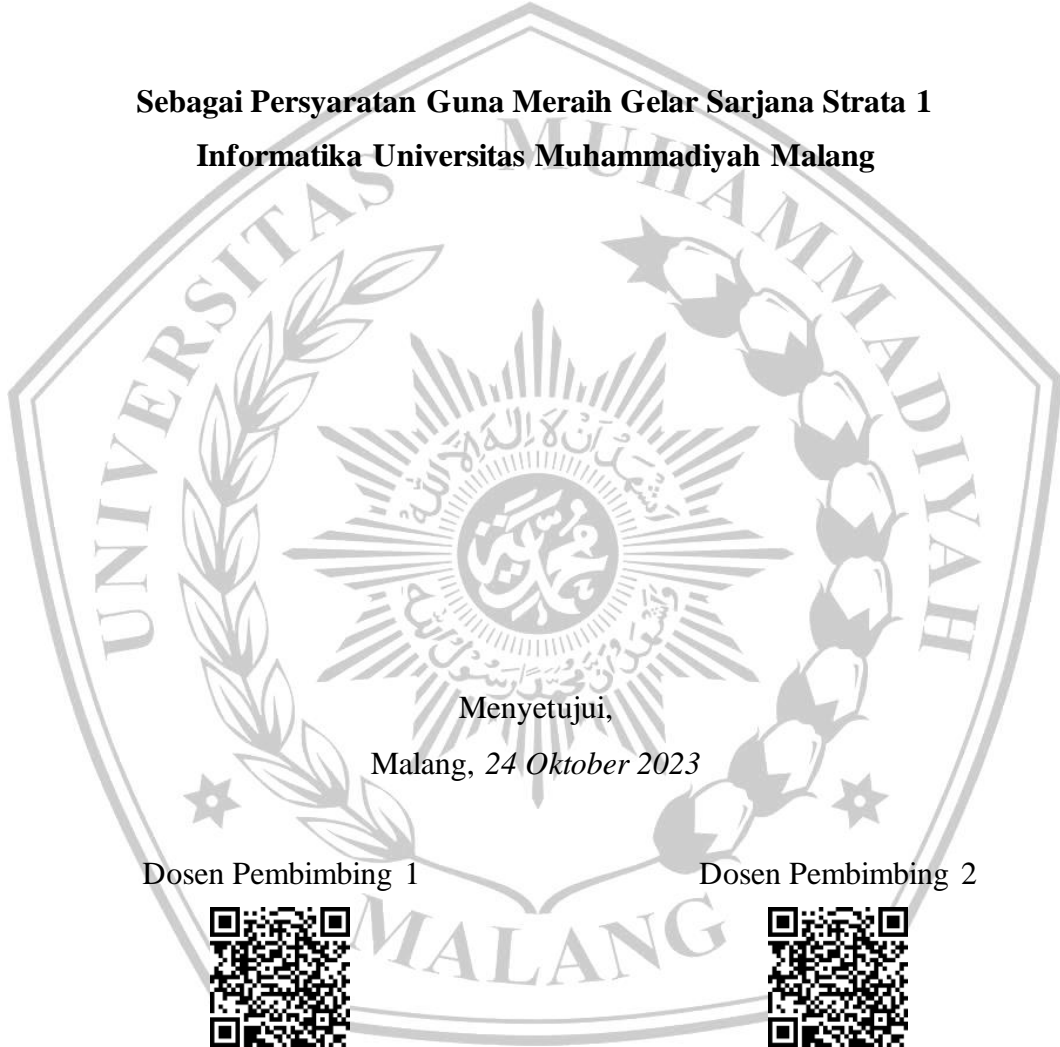
2023

LEMBAR PERSETUJUAN

PERBANDINGAN ALGORITMA KLASIFIKASI DATA MINING UNTUK PREDIKSI PENYAKIT STROKE

TUGAS AKHIR

Sebagai Persyaratan Guna Meraih Gelar Sarjana Strata 1
Informatika Universitas Muhammadiyah Malang



Menyetujui,

Malang, 24 Oktober 2023

Dosen Pembimbing 1



Ir. Yufis Azhar S.Kom., M.Kom.

NIP. 10814100544PNS.

Dosen Pembimbing 2



Vinna Rahmayanti S S.Si., M.Si

NIP. 180306071990PNS.

LEMBAR PENGESAHAN
PERBANDINGAN ALGORITMA KLASIFIKASI DATA
MINING UNTUK PREDIKSI PENYAKIT STROKE
TUGAS AKHIR

Sebagai Persyaratan Guna Meraih Gelar Sarjana Strata 1
Informatika Universitas Muhammadiyah Malang

Disusun Oleh :

Aidia Khoiriyah Firdausy

201910370311260

Tugas Akhir ini telah diuji dan dinyatakan lulus melalui sidang majelis penguji
pada tanggal 24 Oktober 2023

Menyetujui,

Dosen Penguji 1



Ir. Gita Indah Marthasari ST., M.Kom

NIP. 10806110442PNS.

Dosen Penguji 2



Ir. Galih Wasis Wicaksono S.kom.

M.Cs.

NIP. 10814100541PNS.

Mengetahui,
Ketua Jurusan Informatika



Ir. Galih Wasis Wicaksono S.kom. M.Cs.

NIP. 10814100541PNS.



LEMBAR PERNYATAAN

Yang bertanda tangan dibawah ini :

NAMA : Aidia Khoiriyah Firdausy

NIM : 201910370311260

FAK./JUR. : Informatika

Dengan ini saya menyatakan bahwa Tugas Akhir dengan judul **“PERBANDINGAN ALGORITMA KLASIFIKASIDATA MINING UNTUK PREDIKSI PENYAKIT STROKE”** beserta seluruh isinya adalah karya saya sendiri dan bukan merupakan karya tulis orang lain, baik sebagian maupun seluruhnya, kecuali dalam bentuk kutipan yang telah disebutkan sumbernya.

Demikian surat pernyataan ini saya buat dengan sebenar-benarnya. Apabila kemudian ditemukan adanya pelanggaran terhadap etika keilmuan dalam karya saya ini, atau ada klaim dari pihak lain terhadap keaslian karya saya ini maka saya siap menanggung segala bentuk resiko/sanksi yang berlaku.

Mengetahui,
Dosen Pembimbing

Malang, 24 Oktober 2023
Yang Membuat Pernyataan



Ir. Yufis Azhar S.Kom., M.Kom.

Aidia Khoiriyah Firdausy

KATA PENGANTAR

Dengan memanjatkan puji syukur kehadiran Allah SWT. Atas limpahan rahmat serta hidayah-Nya sehingga peneliti dapat menyelesaikan Tugas Akhir yang berjudul: “ **Perbandingan Algoritma Klasifikasi Data Mining Untuk Prediksi Penyakit Stroke** “. Di dalam tulisan ini disajikan pokok-pokok bahasan yang meliputi latar belakang, metode penelitian, serta hasil dan pembahasan yang telah didapat dari proses masih banyak kekurangan dan keterbatasan. Oleh karena itu peneliti mengharapkan saran yang membangun agar tulisan ini bermanfaat bagi perkembangan ilmu pengetahuan.

Malang, 18 Agustus 2023



Aidia Khoiriyah Firdausy



DAFTAR ISI

LEMBAR PERSETUJUAN	ii
LEMBAR PENGESAHAN	iii
LEMBAR PERNYATAAN	iv
LEMBAR PERSEMBAHAN	v
ABSTRAK	vi
ABSTRACT	vii
KATA PENGANTAR	viii
DAFTAR ISI	ix
DAFTAR GAMBAR	x
DAFTAR TABEL	xi
Pendahuluan	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Tujuan Penelitian	3
1.4 Batasan Masalah	3
2.1 Penelitian Terdahulu	4
2.2 Stroke	5
2.3 Metode Data Mining	6
2.4 Metode Decision Tree C4.5	6
2.5 Logistic Regression (Linier Regression)	7
2.6 Metode Random Forest / Random Forest Method	8
2.7 SVM (Support Vector Machine)	9
2.8 K-Nearest Neighbours (KNN)	9
2.9 Naïve Bayes	10
BAB III	12
METODOLOGI PENELITIAN	12
3.1 Dataset	12
3.2 Preprocessing	12

3.2.1	ColumnTransformer	12
3.2.2	Splitting Data	13
3.2.3	Visualisasi Data	13
3.3	Modeling dan Evaluasi	14
3.3.1	K-Nearest Neighbors (KNN)	14
3.3.2	Naïve Bayes	15
3.3.3	Logistic Regression	16
3.3.4	Decision Tree	17
3.3.5	Random Forest	17
3.3.6	Support Vector Machine (SVM)	18
BAB IV	20
HASIL DAN PEMBAHASAN	20
4.1	Preprocessing	20
4.1.1	Column Transformer	21
4.1.2	Split Data	22
4.1.3	Visualisasi Data	22
4.2	Modelling	23
4.2.1	K-Nearest Neighbors	23
4.2.2	Naïve Bayes	23
4.2.3	Logistic Regression	24
4.2.4	Decision Tree	24
4.2.5	Random Forest	25
4.2.6	Support Vector Machine	25
4.1	Evaluasi Model	25
BAB V	28
KESIMPULAN	28
5.1	Kesimpulan	28
5.2	Saran	28
DAFTAR PUSTAKA	29

DAFTAR GAMBAR

Gambar 1 Visualisasi Data	13
Gambar 2 Flowchart KNN	14
Gambar 3 Flowchart Naive bayes	15
Gambar 4 Flowchart Logistic Regression	16
Gambar 5 Flowchart Decision Tree	17
Gambar 6 Flowchart Random Forest	17
Gambar 7 Flowchart SVM	18
Gambar 8 Ilustrasi Support Vector Machine	19
Gambar 9 Source Code Library	20
Gambar 10 Import dataset	21
Gambar 11 Load Dataset	21
Gambar 12 coloumn transformer	21
Gambar 13 Split data	22
Gambar 14 Hasil Visualisasi Data	22
Gambar 15 Source Code KNN	23
Gambar 16 Source Code Naive Bayes	23
Gambar 17 Source code logistic regression	24
Gambar 18 Source code decision tree	24
Gambar 19 Source code random forest	25
Gambar 20 Source code SVM	25
Gambar 21 Source Code Confusion Matrix	25
Gambar 22 Output Confusion Matrix	26
Gambar 23 Source Code Visualisasi Hitung Prediksi	27
Gambar 24 Visualisasi Hasil Prediksi	27

DAFTAR TABEL

Tabel 1 Penelitian Terdahulu.....	5
-----------------------------------	---

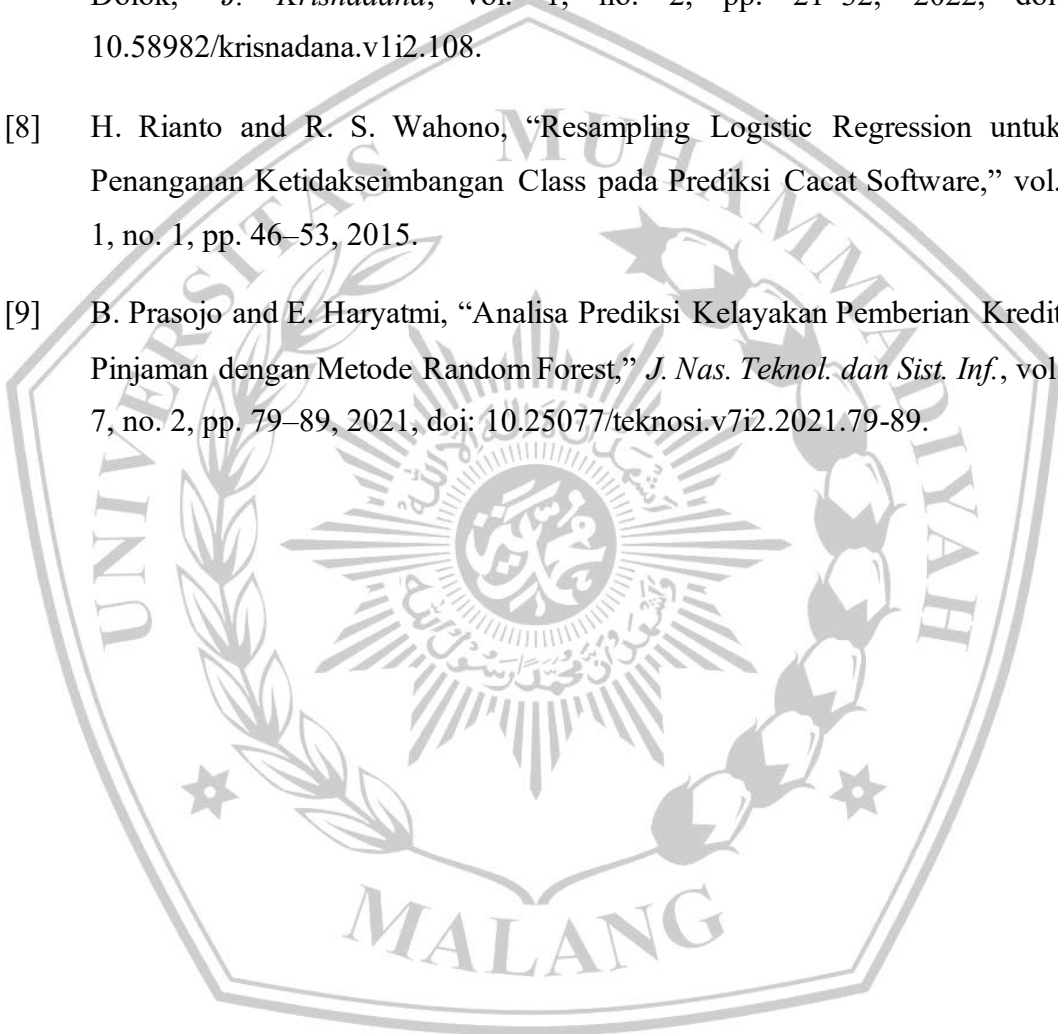


DAFTAR PUSTAKA

- [1] H. Hikmayanti Handayani, K. Ahmad Baihaqi, and U. Buana Perjuangan Karawang, "Implementasi Algoritma Logistic Regression Untuk Klasifikasi Penyakit Stroke," *Syntax J. Inform.*, vol. 12, no. 01, pp. 15–23, 2023.
- [2] A. F. Riany and G. Testiana, "Penerapan Data Mining untuk Klasifikasi Penyakit Stroke Menggunakan Algoritma Naïve Bayes," *J. SAINTEKOM*, vol. 13, no. 1, pp. 42–54, 2023, doi: 10.33020/saintekom.v13i1.352.
- [3] A. Byna and M. Basit, "Penerapan Metode Adaboost Untuk Mengoptimasi Prediksi Penyakit Stroke Dengan Algoritma Naïve Bayes," *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 9, no. 3, pp. 407–411, 2020, doi: 10.32736/sisfokom.v9i3.1023.
- [4] L. Ghani, L. K. Mihardja, and D. Delima, "Faktor Risiko Dominan Penderita Stroke di Indonesia," *Bul. Penelit. Kesehat.*, vol. 44, no. 1, pp. 49–58, 2016, doi: 10.22435/bpk.v44i1.4949.49-58.
- [5] D. Ulfatul, M. Rachmad, H. Oktavianto, and M. Rahman, "Jurnal Smart Teknologi Perbandingan Metode K-Nearest Neighbor Dan Gaussian Naive Bayes Untuk Klasifikasi Penyakit Stroke Comparison Of K-Nearest Neighbor And Gaussian Naive Bayes Methods For Stroke Disease Classification Jurnal Smart Teknologi bidang pengenalan pola , kategori teks , pengolahan objek dan lain lain , karena sederhana dalam pengolahannya dan dapat melakukan training data dalam jumlah yang besar . Penerapan algoritma K-Nearst Neighbor dan Gaussian Naive Bayes sebagai proses klasifikasi yang dirancang untuk mendiagnosa penyakit dengan mengklasifikasi gejala yang dialami pasien dengan memperoleh nilai akurasi tertinggi . Maka dari itu , penulis memutuskan menggunakan algoritma K-Nearst Neighbor dan Gaussian Naive Bayes pada proses klasifikasi dalam melakukan perbandingan dan menentukan algoritma yang memiliki akurasi lebih tinggi dalam proses klasifikasi," vol. 3, no. 4, pp. 405–412, 2022.
- [6] F. Dwi Meliani Achmad, Budanis, Slamet, "Klasifikasi Data Karyawan

Untuk Menentukan Jadwal Kerja Menggunakan Metode Decision Tree,” *J. IPTEK*, vol. 16, no. 1, pp. 18–23, 2012, [Online]. Available: <http://jurnal.itats.ac.id/wp-content/uploads/2013/06/3.-BUDANIS-FINAL-hal-17-23.pdf>.

- [7] S. F. Damanik, A. Wanto, and I. Gunawan, “Penerapan Algoritma Decision Tree C4.5 untuk Klasifikasi Tingkat Kesejahteraan Keluarga pada Desa Tiga Dolok,” *J. Krisnadana*, vol. 1, no. 2, pp. 21–32, 2022, doi: 10.58982/krisnadana.v1i2.108.
- [8] H. Rianto and R. S. Wahono, “Resampling Logistic Regression untuk Penanganan Ketidakseimbangan Class pada Prediksi Cacat Software,” vol. 1, no. 1, pp. 46–53, 2015.
- [9] B. Prasojo and E. Haryatmi, “Analisa Prediksi Kelayakan Pemberian Kredit Pinjaman dengan Metode Random Forest,” *J. Nas. Teknol. dan Sist. Inf.*, vol. 7, no. 2, pp. 79–89, 2021, doi: 10.25077/teknosi.v7i2.2021.79-89.



FORM CEK PLAGIARISME LAPORAN TUGAS AKHIR

Nama Mahasiswa : Aidia Khoiriyah Firdausy
NIM : 201910370311260
Judul TA : Perbandingan Algoritma Klasifikasi Data Mining Untuk
Prediksi Penyakit Stroke

Hasil Cek Plagiarisme dengan Turnitin

No.	Komponen Pengecekan	Nilai Maksimal Plagiarisme (%)	Hasil Cek Plagiarisme (%) *
1.	Bab 1 – Pendahuluan	10 %	8 %
2.	Bab 2 – Tinjauan Pustaka	25 %	16 %
3.	Bab 3 – Analisis dan Perancangan	25 %	15 %
4.	Bab 4 – Implementasi dan Pengujian	15 %	13 %
5.	Bab 5 – Kesimpulan dan Saran	5 %	5 %
6.	Makalah Tugas Akhir	20%	15 %

*) Hasil cek plagiarism diisi oleh pemeriksa (staf TU)

*) Maksimal 5 kali (4 Kali sebelum ujian, 1 kali sesudah ujian)

Mengetahui,

Pemeriksa (Staff TU)


dary

()