

**DATA MINING UNTUK PREDIKSI PENYAKIT STROKE PADA SEMUA
USIA MENGGUNAKAN METODE K-NEAREST NEIGHBOUR**



**UNIVERSITAS
MERCU BUANA**

Muhammad Bima Aditya 41819010040

Muhammad Yan Silmi Kaaffah 41819010104

**UNIVERSITAS
MERCU BUANA**

PROGRAM STUDI SISTEM INFORMASI

FAKULTAS ILMU KOMPUTER

UNIVERSITAS MERCU BUANA

JAKARTA

2022

**DATA MINING UNTUK PREDIKSI PENYAKIT STROKE PADA SEMUA
USIA MENGGUNAKAN METODE K-NEAREST NEIGHBOUR**



**UNIVERSITAS
MERCU BUANA**

Proposal Tugas Akhir

Diajukan Untuk Capaian Pembelajaran Matakuliah

TUGAS AKHIR

Oleh:

Muhammad Bima Aditya 41819010040

Muhammad Yan Silmi Kaaffah 41819010104

PROGRAM STUDI SISTEM INFORMASI

FAKULTAS ILMU KOMPUTER

UNIVERSITAS MERCU BUANA

JAKARTA

2022

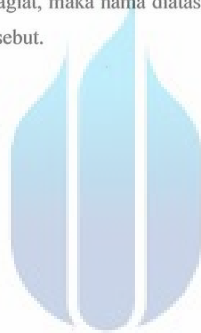
LEMBAR PERNYATAAN ORISINALITAS

SURAT PERNYATAAN ORISINALITAS

Yang bertanda tangan dibawah ini:

Nama Mahasiswa (1) : Muhammad Bima Aditya
NIM (41819010040)
Nama Mahasiswa (2) : Muhammad Yan Silmi Kaffah
NIM (41819010104)
Judul Tugas Akhir : Data Mining Untuk Prediksi Penyakit Stroke Pada
Semua Usia Menggunakan Metode K-Nearest
Neighbor

Menyatakan bahwa laporan jurnal ini adalah hasil karya nama yang tercantum diatas dan bukan plagiat (tidak *copy paste* sumber lain). Apabila ternyata ditemukan di dalam Tugas Akhir ini terdapat unsur plagiat, maka nama diatas siap untuk mendapatkan sanksi akademik yang terkait dengan hal tersebut.



Jakarta, 02 Februari 2023



Muhammad Bima Aditya

UNIVERSITAS
MERCU BUANA

SURAT PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR

SURAT PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR

Sebagai mahasiswa Universitas Mercu Buana, saya yang bertanda tangan di bawah ini :

Nama Mahasiswa (1) : Muhammad Bima Aditya
NIM (41819010040)
Nama Mahasiswa (2) : Muhammad Yan Silmi Kaffah
NIM (41819010104)
Judul Tugas Akhir : Data Mining Untuk Prediksi Penyakit Stroke Pada Semua Usia Menggunakan Metode K-Nearest Neighbor

Dengan ini memberikan izin dan menyetujui untuk memberikan kepada Universitas Mercu Buana **Hak Bebas Royalti Noneksklusif** (*None-exclusive Royalty Free Right*) atas karya ilmiah saya yang berjudul diatas beserta perangkat yang ada (jika diperlukan).

Dengan Hak Bebas Royalti/Noneksklusif ini Universitas Mercu Buana berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat dan mempublikasikan tugas akhir saya.

Selain itu, demi pengembangan ilmu pengetahuan di lingkungan Universitas Mercu Buana, saya memberikan izin kepada Peneliti di Lab Riset Fakultas Ilmu Komputer, Universitas Mercu Buana untuk menggunakan dan mengembangkan hasil riset yang ada dalam tugas akhir untuk kepentingan riset dan publikasi selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Jakarta, 02 Februari 2023

UNIVERSITAS
MERCU BUANA



Muhammad Bima Aditya

LEMBAR PERSETUJUAN

Nama Mahasiswa : Muhammad Bima Aditya
NIM : 41819010040
Nama Mahasiswa : Muhammad Yan Silmi Kaaffah
NIM : 41819010104
Judul Tugas Akhir : DATA MINING UNTUK PREDIKSI PENYAKIT
STROKE PADA SEMUA USIA
MENGUNAKAN METODE K-NEAREST
NEIGHBOUR

Tugas Akhir ini telah diperiksa dan disetujui

Jakarta, 23 - Desember - 2022

Menyetujui,



(Sarwati Rahayu, ST., MMSI)

Dosen Pembimbing

LEMBAR PENGESAHAN

Nama Mahasiswa (1) : Muhammad Bima Aditya
NIM (41819010040)
Nama Mahasiswa (2) : Muhammad Yan Silmi Kaaffah
NIM (41819010104)
Judul Tugas Akhir : DATA MINING UNTUK PREDIKSI PENYAKIT
STROKE PADA SEMUA USIA
MENGUNAKAN METODE K-NEAREST
NEIGHBOUR

Tugas Akhir ini telah diperiksa dan disidangkan sebagai salah satu persyaratan untuk memperoleh gelar Sarjana pada Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Mercu Buana.

Jakarta, 7 Januari 2023

Menyetujui,

Pembimbing Sarwati Rahayu, ST., MMSI

Penguji 1 Andi Nugroho, S.T., M. Kom T A S

Penguji 2 Bagus Priambodo, S.Kom., MTI

Penguji 3 Sulis Sandiwarno, S.Kom, M.Kom

Mengetahui,


(Yunita Sartika Sari, S.Kom., M.Kom)
Sek. Prodi Sistem Informasi


((Dr. Ruci Meiyanti, M.Kom)
Ka. Prodi Sistem Informasi

KATA PENGANTAR

Puji syukur kita panjatkan kehadiran Allah Subhanahu wa ta'ala, yang telah melimpahkan rahmat dan karunianya sehingga saya sebagai penulis dapat menyelesaikan Metodologi Penelitian Teknologi Informasi yang berjudul “DATA MINING UNTUK PREDIKSI PENYAKIT STROKE PADA SEMUA USIA MENGGUNAKAN METODE *K-NEAREST NEIGHBOUR*”

Penulis menyadari bahwa Laporan Tugas Akhir ini tidak akan selesai tanpa bantuan dan bimbingan dari berbagai pihak. Oleh karena itu, penulis mengucapkan terima kasih kepada:

1. Ibu Sarwati Rahayu, ST., MMSI selaku dosen pembimbing yang telah memberikan bimbingan, bantuan, serta saran sehingga penulis dapat menyelesaikan laporan ini.
2. Ibu Ratna Mutu Manikam, S.Kom., M.T selaku Kepala Program Studi Sistem Informasi
3. Bapak Indra Ranggadara, S.Kom, MT selaku dosen pengampu
4. Ibu Puji Rahayu, Dr, M. Kom selaku dosen pengampu
5. Semua Bapak dan Ibu dosen Program Studi Informasi
6. Kepada kedua orangtua yang selalu memberikan doa, motivasi dan semangat untuk menyelesaikan laporan ini.
7. Sahabat dan semua pihak yang telah membantu namun tidak bisa disebutkan satu persatu

Akhir kata, penulis berharap segala bentuk saran serta masukan bahkan kritik yang membangun dari berbagai pihak. Semoga laporan tugas akhir ini dapat bermanfaat bagi semua pihak dan khususnya dalam bidang teknologi informasi.

Jakarta, 23 - Desember -2022



Muhammad Bima Aditya

DAFTAR ISI

HALAMAN SAMPUL	i
HALAMAN JUDUL.....	i
LEMBAR PERNYATAAN ORISINALITAS	ii
SURAT PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR	iii
LEMBAR PERSETUJUAN	iv
LEMBAR PENGESAHAN.....	v
ABSTRAK.....	vi
ABSTRACT	vii
KATA PENGANTAR	viii
DAFTAR ISI	ix
DAFTAR TABEL	xi
DAFTAR GAMBAR	xii
LAMPIRAN	xiii
BAB I PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	2
1.3 Tujuan Penelitian.....	2
1.4 Batasan Masalah.....	3
a. Data.....	3
b. Metode.....	3
c. Hasil.....	3
1.5 Manfaat Penelitian.....	4
a. Teoritis/akademis	4
b. Praktis.....	4
1.6 Sistematika Penulisan.....	4
BAB II TINJAUAN PUSTAKA	5
2.1 Teori/Konsep Terkait.....	5
2.1.1 DATA MINING	5

2.1.2	CONFUSION MATRIX	5
2.1.3	DATASET	6
2.1.4	METODE K-NEAREST NEIGHBOR.....	6
2.1.5	Jupyter Notebook.....	7
2.1.6	SPSS	7
2.1.7	PYTHON.....	7
2.1.8	LOGISTIC REGRESSION	8
2.1.9	RANDOM FOREST	8
2.1.10	NEURAL NETWORK.....	8
2.2	Penelitian Terdahulu.....	9
BAB III METODE PENELITIAN.....		28
3.1	Deskripsi Sumber Data.....	28
3.2	Sarana Pendukung	34
3.3	Diagram Alir Penelitian.....	34
BAB IV HASIL DAN PEMBAHASAN.....		38
4.1	Persyaratan Utama.....	38
4.1.1	Hasil.....	38
4.2	Tools Yang Digunakan.....	53
4.3	Pembahasan.....	54
BAB V Kesimpulan dan saran		56
5.1	Kesimpulan.....	56
5.2	Saran	56
DAFTAR PUSTAKA		57
LAMPIRAN		62

DAFTAR TABEL

Tabel 2.1 <i>Literature Review</i>	9
Tabel 3.1 <i>Statistics Total</i>	28
Tabel 3.2 <i>ID</i>	29
Tabel 3.3 <i>Gender</i>	29
Tabel 3.4 <i>Age</i>	29
Tabel 3.5 <i>Hypertension</i>	30
Tabel 3.6 <i>Heart Disease</i>	30
Tabel 3.7 <i>Ever Married</i>	30
Tabel 3.8 <i>Work Type</i>	31
Tabel 3.9 <i>Residence Type</i>	31
Tabel 3.10 <i>Avg Glucose Level</i>	32
Tabel 3.11 <i>BMI</i>	32
Tabel 3.12 <i>Smoking Status</i>	33
Tabel 3.13 <i>Stroke</i>	33



DAFTAR GAMBAR

Gambar 3.3.1 Diagram Alir Penelitian.....	35
Gambar 4.1.1 Import Package	38
Gambar 4.1.2 Read Dataset	38
Gambar 4.1.3 Informasi Data	39
Gambar 4.1.4 Menampilkan Statistik.....	39
Gambar 4.1.5 Missing Value Detection	40
Gambar 4.1.6 Handling Missing Value BMI	40
Gambar 4.1.7 Dendogram	41
Gambar 4.1.8 Missing Value Smoking Status.....	41
Gambar 4.1.9 Memeriksa Missing Value.....	42
Gambar 4.1.10 Memeriksa Duplikat	42
Gambar 4.1.11 Drop Column	42
Gambar 4.1.12 Mendefinisikan Tipe Data	43
Gambar 4.1.13 Memeriksa Outliers	43
Gambar 4.1.14 Mendefinisikan Outliers	44
Gambar 4.1.15 Heatmap Korelasi Analisis	44
Gambar 4.1.16 Melakukan Encoding.....	45
Gambar 4.1.17 Mendefinisikan Variable Fitur dan Target	45
Gambar 4.1.18 SMOTE.....	46
Gambar 4.1.19 Train dan Test Split Data.....	46
Gambar 4.1.20 Import Package Modelling	47
Gambar 4.1.21 K-Nearest Neighbor.....	48
Gambar 4.1.22 Neural Network	49
Gambar 4.1.23 Logistic Regression	50
Gambar 4.1.24 Random Forest.....	51
Gambar 4.1.25 Hasil Komparasi Model.....	52

LAMPIRAN

Lampiran 1. Data Penelitian	62
Lampiran 2. Source Code	62
Lampiran 3. Kartu Bimbingan.....	72
Lampiran 4. Biodata	73

