

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



**UNIVERSITAS
MERCU BUANA**

Muhammad Yan Silmi Kaaffah 41819010104
Muhammad Bima Aditya 41819010040

**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 Yan Silmi Kaaffah 41819010104

Muhammad Bima Aditya 41819010040

MERCU BUANA

PROGRAM STUDI SISTEM INFORMASI

FAKULTAS ILMU KOMPUTER

UNIVERSITAS MERCU BUANA

JAKARTA

2022

SURAT PERNYATAAN ORISINALITAS

Yang bertanda tangan dibawah ini:

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

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.

UNIVERSITAS
MERCU BUANA

Jakarta, 02 Febuari 2022



Muhammad Yan Silmi Kaaffah

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 Kaaffah
NIM (41819010104)
Judul Tugas Akhir : DATA MINING UNTUK PREDIKSI PENYAKIT
STROKE PADA SEMUA USIA
MENGUNAKAN METODE K-NEAREST
NEIGHBOUR

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 2022




Muhammad Yan Silmi Kaaffah

LEMBAR PERSETUJUAN

Nama Mahasiswa : Muhammad Yan Silmi Kaaffah
NIM : 41819010104
Nama Mahasiswa : Muhammad Bima Aditya
NIM : 41819010040
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

UNIVERSITAS
MERCU BUANA
Menyetujui, AS


(Sarwati Rahayu, ST., MMSI)

Dosen Pembimbing

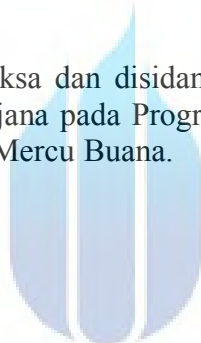
LEMBAR PENGESAHAN

Nama Mahasiswa (1) : Muhammad Yan Silmi Kaaffah
NIM (41819010104)
Nama Mahasiswa (2) : Muhammad Bima Aditya
NIM (41819010040)

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

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 kehadirat 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 Yan Silmi Kaaffah

DAFTAR ISI

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

UNIVERSITAS
MERCU BUANA

DAFTAR GAMBAR

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

DAFTAR LAMPIRAN

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

