

DAFTAR ISI

HALAMAN JUDUL.....	i
HALAMAN PENGESAHAN PEMBIMBING.....	ii
HALAMAN PENGESAHAN PENGUJI.....	iii
SURAT PERNYATAAN KARYA ASLI TUGAS AKHIR.....	iv
SURAT PERNYATAAN BEBAS PLAGIAT.....	v
HALAMAN PERSEMBAHAN.....	vi
ABSTRAK.....	vii
ABSTRACT.....	viii
KATA PENGANTAR.....	ix
DAFTAR ISI.....	x
DAFTAR TABEL.....	xiii
DAFTAR GAMBAR.....	xiv
DAFTAR MODUL PROGRAM.....	xvi
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang Masalah.....	1
1.2 Rumusan Masalah.....	2
1.3 Batasan Masalah.....	2
1.4 Tujuan Penelitian.....	3
1.5 Manfaat Penelitian.....	3
1.6 Tahapan Penelitian.....	3
1.7 Sistematika Penulisan.....	4
BAB II TINJAUAN LITERATUR.....	5
2.1 Metode <i>Ensemble</i>	5
2.2 <i>Imbalanced Data</i>	5
2.3 <i>Synthetic Minority Over-Sampling Technique (SMOTE)</i>	6
2.4 Dimensionalitas Data.....	7
2.5 <i>Principle Component Analysis (PCA)</i>	7
2.6 Klasifikasi.....	9
2.6.1 <i>Decision Tree</i>	9
2.6.2 <i>K-Nearest Neighbors</i>	10
2.6.3 <i>Naïve Bayes</i>	10
2.7 Evaluasi Kinerja Model Klasifikasi.....	10
2.7.1 Akurasi.....	11

2.7.2	<i>AUC</i>	11
2.7.3	G-Mean	12
2.8	Penelitian Terdahulu	12
BAB III METODOLOGI PENELITIAN		14
3.1	Metodologi Penelitian.....	14
3.2	Pengumpulan Informasi.....	14
3.2.1	Studi Pustaka.....	15
3.2.2	Observasi.....	15
3.2.3	Pengumpulan Data.....	15
3.3	<i>Data Pre-processing</i>	16
3.3.1	Sub-Proses <i>Data Cleansing</i>	17
3.3.2	Sub-Proses <i>Label Encoding</i>	18
3.3.3	Sub-Proses <i>Train Test Split</i>	19
3.4	Pemodelan.....	20
3.4.1	Skenario 1	20
3.4.1.1	Pipeline Skenario 1	21
3.4.1.2	<i>Hyperparameter Tuning</i>	27
3.4.1.3	<i>10-Fold Cross Validation</i>	27
3.4.2	Skenario 2	28
3.5	Evaluasi.....	29
3.5.1	Kinerja Model Klasifikasi.....	29
3.5.2	Rancangan Pengujian.....	31
3.6	Kebutuhan Sistem	35
3.7	Tahapan Pengembangan Model.....	35
BAB IV HASIL DAN PEMBAHASAN.....		36
4.1	Modul <i>Data Pre-processing</i>	36
4.1.1	Implementasi <i>Data Cleansing</i>	38
4.1.2	Implementasi <i>Label Encoding</i>	38
4.1.3	Implementasi <i>Train Test Split</i>	39
4.2	Modul Pemodelan	41
4.2.1	Implementasi Skenario 1.....	42
4.2.1.1	Urutan Tahap <i>Pipeline</i> Skenario 1	42
4.2.1.2	Mendefinisikan <i>Hyperparameter</i> Skenario 1.....	43
4.2.1.3	<i>Hyperparameter Tuning</i> dengan <i>10-Fold Cross Validation</i> Skenario 1.....	43
4.2.2	Implementasi Skenario 2.....	46

4.2.2.1	Urutan Tahap <i>Pipeline</i> Skenario 2.....	46
4.2.2.2	Mendefinisikan <i>Hyperparameter</i> Skenario 2.....	46
4.2.2.3	<i>Hyperparameter Tuning</i> dengan <i>10-Fold Cross Validation</i> Skenario 2.....	47
4.3	Modul Evaluasi	48
4.3.1	Kinerja Model Klasifikasi.....	48
4.3.2	Hasil Pengujian	53
4.4	Pembahasan	59
BAB V PENUTUP		60
5.1	Kesimpulan	60
5.2	Saran	60
DAFTAR PUSTAKA.....		61

DAFTAR TABEL

Tabel 2.1 <i>Confusion Matrix</i>	10
Tabel 2.2 Perhitungan Kinerja Klasifikasi	11
Tabel 2.3 Kriteria AUC	12
Tabel 2.4 Penelitian Terdahulu	12
Tabel 2.5 Lanjutan Penelitian Terdahulu	13
Tabel 3.1 Dataset Medis Penderita Stroke	15
Tabel 3.2 Hasil <i>Label Encoding</i>	19
Tabel 3.3 Hasil <i>Train Test Split</i>	20
Tabel 3.4 <i>Subset</i> Dataset Medis Penderita Stroke	22
Tabel 3.5 Tetangga Terdekat Sampel 8	23
Tabel 3.6 <i>Subset</i> Dataset Medis Penderita Stroke Setelah SMOTE	24
Tabel 3.7 Perhitungan Matriks Kovarians	25
Tabel 3.8 Hasil Reduksi Dimensi Fitur	26
Tabel 3.9 Rentang <i>Hyperparameter</i> Skenario 1	27
Tabel 3.10 Rentang <i>Hyperparameter</i> Skenario 2	29
Tabel 3.11 Ilustrasi <i>Confusion Matrix</i> Skenario 1	30
Tabel 3.12 Hasil Contoh Perhitungan Akurasi, AUC, dan G-Means Skenario 1	30
Tabel 3.13 Rancangan Pengujian Kinerja Model Klasifikasi Skenario 1	31
Tabel 3.14 Rancangan Pengujian Kinerja Model Klasifikasi Skenario 2	31
Tabel 3.15 Rancangan Pengujian Performa Ujian Skenario 1	31
Tabel 3.16 Rancangan Pengujian Performa Ujian Skenario 2	32
Tabel 3.17 Rancangan Parameter Terbaik	32
Tabel 3.18 Spesifikasi Perangkat Keras	35
Tabel 3.19 Spesifikasi Perangkat Lunak	35
Tabel 4.1 Sebelum Proses <i>Label Encoding</i>	39
Tabel 4.2 Sesudah Proses <i>Label Encoding</i>	39
Tabel 4.3 Pilihan Nilai <i>N_Components</i>	41
Tabel 4.4 Hasil Pengujian Kinerja Model Klasifikasi Skenario 1	53
Tabel 4.5 Hasil Pengujian Kinerja Model Klasifikasi Skenario 2	53
Tabel 4.6 Hasil Pengujian Performa Ujian Skenario 1	54
Tabel 4.7 Hasil Pengujian Performa Ujian Skenario 2	54
Tabel 4.8 Parameter Terbaik Setiap Model Klasifikasi	56

DAFTAR GAMBAR

Gambar 2.1 Ilustrasi <i>Imbalanced Data</i>	5
Gambar 2.2 Ilustrasi <i>Under-sampling</i> dan <i>Over-sampling</i>	6
Gambar 2.3 Ilustrasi SMOTE	7
Gambar 2.4 Ilustrasi <i>Principle Component Analysis</i>	8
Gambar 3.1 Kerangka Penelitian	14
Gambar 3.2 <i>Flowchart Data Pre-processing</i>	16
Gambar 3.3 <i>Flowchart Data Cleansing</i>	17
Gambar 3.4 <i>Flowchart Label Encoding</i>	18
Gambar 3.5 <i>Flowchart Train Test Split</i>	19
Gambar 3.6 <i>Flowchart</i> Skenario 1	21
Gambar 3.7 Ilustrasi kalkulasi SMOTE	23
Gambar 3.8 Ilustrasi <i>10-Fold Cross Validation</i>	28
Gambar 3.9 <i>Flowchart</i> Skenario 2	28
Gambar 3.10 <i>Flowchart</i> Perhitungan Kinerja Model Klasifikasi	29
Gambar 3.11 Rancangan Visualisasi Perbandingan Kinerja Model	32
Gambar 3.12 Rancangan Visualisasi Perbandingan Skor Ujian	33
Gambar 3.13 Rancangan Visualisasi <i>Scatter Plot</i> distribusi kelas SMOTE	33
Gambar 3.14 Rancangan Visualisasi <i>Bar Plot</i> distribusi kelas SMOTE	34
Gambar 3.15 Rancangan Visualisasi Komponen dari 95% Varians PCA	34
Gambar 3.16 Rancangan Visualisasi Jumlah Reduksi PCA	34
Gambar 4.1 Hasil Visualisasi <i>Imbalanced</i>	37
Gambar 4.2 Sebelum dan Sesudah Proses <i>Data Cleansing</i>	38
Gambar 4.3 Proporsi Data Latih dan Data Uji	40
Gambar 4.4 Urutan Tahap <i>Pipeline</i> Skenario 1	43
Gambar 4.4 <i>Ouput</i> Hasil Pembentukan Model Klasifikasi Skenario 1	45
Gambar 4.6 Urutan Tahap <i>Pipeline</i> Skenario 2	46
Gambar 4.5 <i>Ouput</i> Hasil Pembentukan Model Klasifikasi Skenario 2	47
Gambar 4.6 Hasil Evaluasi Kinerja Model <i>Naïve Bayes</i> Skenario 1	50
Gambar 4.7 Hasil Evaluasi Kinerja Model <i>K-Nearest Neighbors</i> Skenario 1	50
Gambar 4.8 Hasil Evaluasi Kinerja Model <i>Decision Tree</i> Skenario 1	51
Gambar 4.9 Hasil Evaluasi Kinerja Model <i>Naïve Bayes</i> Skenario 2	51
Gambar 4.10 Hasil Evaluasi Kinerja Model <i>K-Nearest Neighbors</i> Skenario 2	52
Gambar 4.11 Hasil Evaluasi Kinerja Model <i>Decision Tree</i> Skenario 2	52

Gambar 4.14 Perbandingan Kinerja Model Klasifikasi	53
Gambar 4.15 Perbandingan Skor Ujian Model <i>Naive Bayes</i>	55
Gambar 4.16 Perbandingan Skor Ujian Model <i>K-Nearest Neighbor</i>	55
Gambar 4.17 Perbandingan Skor Ujian Model <i>Decision Tree</i>	55
Gambar 4.18 Visualisasi <i>Scatter Plot</i> distribusi kelas SMOTE.....	57
Gambar 4.19 Visualisasi <i>Bar Plot</i> distribusi kelas SMOTE.....	57
Gambar 4.20 Visualisasi Komponen dari 95% Varians PCA.....	58
Gambar 4.21 Visualisasi Jumlah Reduksi PCA	58

DAFTAR MODUL PROGRAM

Modul Program 4.1 <i>Source Code Import Library dan Fungsi</i>	36
Modul Program 4.2 <i>Source Code Import Dataset</i>	37
Modul Program 4.3 <i>Source Code Visualisasi Imbalanced</i>	37
Modul Program 4.4 <i>Source Code Tahap Data Cleansing</i>	38
Modul Program 4.5 <i>Source Code Tahap Label Encoding</i>	39
Modul Program 4.6 <i>Source Code Tahap Train Test Split</i>	39
Modul Program 4.7 <i>Source Code Visualisasi Proporsi Data Latih dan Data Uji</i>	40
Modul Program 4.8 <i>Source Code Algoritma Yang Digunakan</i>	41
Modul Program 4.9 <i>Source Code Urutan Tahap Pipeline Skenario 1</i>	42
Modul Program 4.10 <i>Source Code Hyperparameter Skenario 1</i>	43
Modul Program 4.11 <i>Source Code Hyperparameter Tuning 10-Fold Cross Validation</i>	44
Modul Program 4.12 <i>Source Code Pembentukan Model Klasifikasi Skenario 1</i>	45
Modul Program 4.13 <i>Source Code Urutan Tahap Pipeline Skenario 2</i>	46
Modul Program 4.14 <i>Source Code Hyperparameter Skenario 2</i>	47
Modul Program 4.15 <i>Source Code Pembentukan Model Klasifikasi Skenario 2</i>	47
Modul Program 4.16 <i>Source Code Fungsi Evaluasi Kinerja Model Klasifikasi</i>	49
Modul Program 4.17 <i>Source Code Evaluasi Kinerja Model Klasifikasi Skenario 1</i>	49
Modul Program 4.18 <i>Source Code Evaluasi Kinerja Model Klasifikasi Skenario 2</i>	49