

**IMPLEMENTASI ALGORITMA LOGISTIC REGRESSION
UNTUK KLASIFIKASI PASIEN PEREMPUAN
YANG TERKENA PENYAKIT DIABETES**



UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

SKRIPSI

Alvin Octavianus
00000042583

**PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA
TANGERANG
2024**

**IMPLEMENTASI ALGORITMA LOGISTIC REGRESSION
UNTUK KLASIFIKASI PASIEN PEREMPUAN
YANG TERKENA PENYAKIT DIABETES**



Diajukan sebagai salah satu syarat untuk memperoleh
Gelar Sarjana Komputer (S.Kom.)

Alvin Octavianus

00000042583

UMN

UNIVERSITAS

MULTIMEDIA

NUSANTARA

**PROGRAM STUDI INFORMATIKA
FAKULTAS TEKNIK DAN INFORMATIKA
UNIVERSITAS MULTIMEDIA NUSANTARA**

TANGERANG

2024

HALAMAN PERNYATAAN TIDAK PLAGIAT

Dengan ini saya,

Nama : Alvin Octavianus
NIM : 00000042583
Program Studi : Informatika

Menyatakan dengan sesungguhnya bahwa Tesis/Skripsi/Tugas Akhir/Laporan Magang/MBKM saya yang berjudul:

Implementasi Algoritma Logistic Regression untuk Klasifikasi Pasien Perempuan yang Terkena Penyakit Diabetes

merupakan hasil karya saya sendiri, bukan merupakan hasil plagiat, dan tidak pula dituliskan oleh orang lain; Semua sumber, baik yang dikutip maupun dirujuk, telah saya cantumkan dan nyatakan dengan benar pada bagian Daftar Pustaka.

Jika di kemudian hari terbukti ditemukan kecurangan/penyimpangan, baik dalam pelaksanaan skripsi maupun dalam penulisan laporan karya ilmiah, saya bersedia menerima konsekuensi untuk dinyatakan TIDAK LULUS. Saya juga bersedia menanggung segala konsekuensi hukum yang berkaitan dengan tindak plagiarisme ini sebagai kesalahan saya pribadi dan bukan tanggung jawab Universitas Multimedia Nusantara.

Tangerang, 13 Juni 2024



(Alvin Octavianus)

HALAMAN PENGESAHAN

Skripsi dengan judul

IMPLEMENTASI ALGORITMA LOGISTIC REGRESSION UNTUK KLASIFIKASI PASIEN PEREMPUAN YANG TERKENA PENYAKIT DIABETES

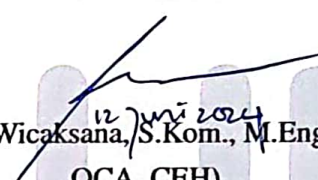
oleh

Nama : Alvin Octavianus
NIM : 00000042583
Program Studi : Informatika
Fakultas : Fakultas Teknik dan Informatika


Telah diujikan pada hari Senin, 03 Juni 2024
Pukul 10.00 s/s 12.00 dan dinyatakan
LULUS

Dengan susunan penguji sebagai berikut


Ketua Sidang



(Arya Wicaksana, S.Kom., M.Eng.Sc.,
OCA, CEH)
NIDN: 015109103

Penguji


(David Agustriawan, S.Kom., M.Sc.,
Ph.D.)
NIDN: 0525088601

Pembimbing


(Moeljono Widjaja, B.Sc., M.Sc., Ph.D.)
NIDN: 0311106903
PJS Ketua Program Studi Informatika,


(Dr. Eng. Niki Prastomo, S.T., M.Sc.)
NIDN: 0419128203

**HALAMAN PERSETUJUAN PUBLIKASI KARYA ILMIAH UNTUK
KEPENTINGAN AKADEMIS**

Yang bertanda tangan di bawah ini:

Nama : Alvin Octavianus
NIM : 00000042583
Program Studi : Informatika
Jenjang : S1
Jenis Karya : Skripsi

Menyatakan dengan sesungguhnya bahwa:

- Saya bersedia memberikan izin sepenuhnya kepada Universitas Multimedia Nusantara untuk mempublikasikan hasil karya ilmiah saya di repositori Knowledge Center, sehingga dapat diakses oleh Civitas Akademika/Publik. Saya menyatakan bahwa karya ilmiah yang saya buat tidak mengandung data yang bersifat konfidensial dan saya juga tidak akan mencabut kembali izin yang telah saya berikan dengan alasan apapun.
- Saya tidak bersedia karena dalam proses pengajuan untuk diterbitkan ke jurnal/konferensi nasional/internasional (dibuktikan dengan *letter of acceptance*)**.

Tangerang, 13 Juni 2024

Yang menyatakan



Alvin Octavianus

U M M N
UNIVERSITAS
MULTIMEDIA
NUSANTARA

** Jika tidak bisa membuktikan LoA jurnal/HKI selama enam bulan ke depan, saya bersedia mengizinkan penuh karya ilmiah saya untuk diunggah ke KC UMN dan menjadi hak institusi UMN.

Halaman Persembahan / Motto

"Believe you can and you're halfway there."

Theodore Roosevelt

UMMN

UNIVERSITAS
MULTIMEDIA
NUSANTARA

KATA PENGANTAR

Puji Syukur atas berkat dan rahmat kepada Tuhan Yang Maha Esa, atas selesainya penulisan laporan Skripsi ini dengan judul: Implementasi Algoritma Logistic Regression untuk Klasifikasi Pasien Perempuan yang Terkena Penyakit Diabetes dilakukan untuk memenuhi salah satu syarat untuk mencapai gelar Sarjana Komputer Jurusan Informatika Pada Fakultas Teknik dan Informatika Universitas Multimedia Nusantara. Saya menyadari bahwa, tanpa bantuan dan bimbingan dari berbagai pihak, dari masa perkuliahan sampai pada penyusunan skripsi ini, sangatlah sulit bagi saya untuk menyelesaikan skripsi ini. Oleh karena itu, saya mengucapkan terima kasih kepada:

1. Bapak Dr. Ninok Leksono, selaku Rektor Universitas Multimedia Nusantara.
2. Bapak Dr. Eng. Niki Prastomo, S.T., M.Sc., selaku Dekan Fakultas Teknik dan Informatika Universitas Multimedia Nusantara serta PJS Ketua Program Studi Informatika Universitas Multimedia Nusantara.
3. Bapak Moeljono Widjaja, B.Sc., M.Sc., Ph.D., sebagai Pembimbing yang telah banyak meluangkan waktu untuk memberikan bimbingan, arahan dan motivasi atas terselesainya skripsi ini.
4. Tuhan Yesus Kristus yang selalu memberikan saya kekuatan dan dukungan, sehingga penulis dapat menyelesaikan skripsi ini.
5. Orang Tua, keluarga saya, dan teman saya yang telah memberikan dukungan, sehingga penulis dapat menyelesaikan skripsi ini.
6. Diri saya sendiri yang telah bekerja keras dalam pembuatan skripsi ini.

Semoga skripsi ini bermanfaat, baik sebagai sumber informasi maupun sumber inspirasi, bagi para pembaca.

Tangerang, 13 Juni 2024



Alvin Octavianus

**IMPLEMENTASI ALGORITMA LOGISTIC REGRESSION
UNTUK KLASIFIKASI PASIEN PEREMPUAN
YANG TERKENA PENYAKIT DIABETES**

Alvin Octavianus

ABSTRAK

Diabetes merupakan sebuah penyakit kronis yang disebabkan ketika pankreas tidak dapat memproduksi cukup insulin secara efektif. Pada Diabetes Atlas Edisi ke-10 tingkat perempuan terkena diabetes lebih tinggi daripada pria yang disebabkan karena faktor perbedaan hormon, penambahan berat badan dan gaya hidup selama masa pubertas. Penelitian ini bertujuan untuk membuat sistem sederhana menggunakan *machine learning* untuk mengetahui pasien perempuan terkena diabetes atau tidak dengan beberapa variabel yang terdapat pada data. Sistem yang dibuat menggunakan klasifikasi dengan algoritma *Logistic Regression*, serta menguji sistem tersebut berdasarkan nilai akurasi dalam setiap uji coba. Dalam percobaannya terdapat beberapa skenario pelatihan yang dilakukan, yaitu pemodelan hanya menggunakan *Logistic Regression*, pemodelan dengan metode *feature extraction*, pemodelan dengan metode *feature selection*, pemodelan dengan *stacking*. Hasil performa evaluasi terbaik pada pengujian dengan data *test* didapatkan nilai akurasi sebesar 91%, *precision* sebesar 87%, *recall* sebesar 87%, dan *f1-score* sebesar 87%.

Kata kunci: *Diabetes, Feature Engineering, GridSearchCV, Logistic Regression, Machine Learning*



Implementation of Logistic Regression Algorithm for Classification of Female Patients Affected by Diabetes Disease

Alvin Octavianus

ABSTRACT

Diabetes is a chronic disease caused when the pancreas does not produce enough insulin effectively. In the 10th Edition of Diabetes Atlas, the rate of women developing diabetes is higher than men due to hormonal differences, weight gain and lifestyle factors during puberty. This study aims to create a simple system using machine learning to determine whether a female patient has diabetes or not with several variables contained in the data. The system created uses classification with the Logistic Regression algorithm, and tests the system based on the accuracy value in each trial. In the experiment, there are several training scenarios performed, namely modeling using only Logistic Regression, modeling with feature extraction, modeling with feature selection, modeling with stacking. The best evaluation performance results on testing with data test obtained an accuracy of 91%, precision of 87%, recall of 87%, and f1-score of 87%.

Keywords: *Diabetes, Feature Engineering, GridSearchCV, Logistic Regression, Machine Learning*



DAFTAR ISI

HALAMAN JUDUL	i
PERNYATAAN TIDAK MELAKUKAN PLAGIAT	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERSETUJUAN PUBLIKASI ILMIAH	iv
HALAMAN PERSEMBAHAN/MOTO	v
KATA PENGANTAR	vi
ABSTRAK	vii
ABSTRACT	viii
DAFTAR ISI	ix
DAFTAR GAMBAR	xi
DAFTAR TABEL	xiv
DAFTAR KODE	xvi
DAFTAR LAMPIRAN	xvii
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang Masalah	1
1.2 Rumusan Masalah	3
1.3 Batasan Permasalahan	4
1.4 Tujuan Penelitian	4
1.5 Manfaat Penelitian	4
1.6 Sistematika Penulisan	5
BAB 2 LANDASAN TEORI	6
2.1 Klasifikasi	6
2.2 Machine Learning	6
2.3 Logistic Regression	7
2.4 Diabetes	9
2.5 Holdout Data	11
2.6 Synthetic Minority Oversampling Technique	12
2.7 Standar Scaller	12
2.8 Principal Component Analysis	13
2.9 Recursive Feature Elimination	15
2.10 Stacking	15
2.11 Gradient Boosting	16
2.12 K-Nearest Neighbors	16
2.13 Grid Search	17
2.14 Evaluasi Model	17
2.14.1 Confusion Matrix	17
2.14.2 Akurasi	19
2.14.3 Recall	19
2.14.4 Precision	19
2.14.5 F1-Score	19
2.14.6 Receiver Operating Characteristic Curve	19
BAB 3 METODOLOGI PENELITIAN	21
3.1 Metodologi Penelitian	21
3.1.1 Studi Literatur	21
3.1.2 Pengumpulan Data	21
3.1.3 Perancangan Sistem	23
3.1.4 Implementasi	23
3.1.5 Validasi	23

3.1.6	Evaluasi	23
3.1.7	Pembuatan Laporan	23
3.2	Perancangan Sistem	24
3.2.1	Flowchart Utama	24
3.2.2	Import Library	25
3.2.3	Import dan Membaca Raw Data	25
3.2.4	Preprocess dan Hold Out Data	27
3.2.5	Data Modeling	34
3.2.6	Evaluasi	36
3.3	Spesifikasi Sistem	36
BAB 4	HASIL DAN DISKUSI	37
4.1	Implementasi Sistem	37
4.1.1	Pemodelan hanya Menggunakan Logistic Regression	37
4.1.2	Pemodelan Menggunakan Logistic Regression dengan Feature Extraction	44
4.1.3	Pemodelan Logistic Regression dengan Menggunakan Feature Selection	72
4.1.4	Pemodelan Logistic Regression dengan Menggunakan Stacking	81
4.1.5	Validasi	88
4.2	Uji Coba	102
4.2.1	Skenario Uji Coba	103
4.2.2	Hasil Uji Coba	103
4.3	Evaluasi Hasil Uji Coba	109
4.4	Klasifikasi	109
4.5	Research Gap	112
BAB 5	SIMPULAN DAN SARAN	114
5.1	Simpulan	114
5.2	Saran	115
	DAFTAR PUSTAKA	116

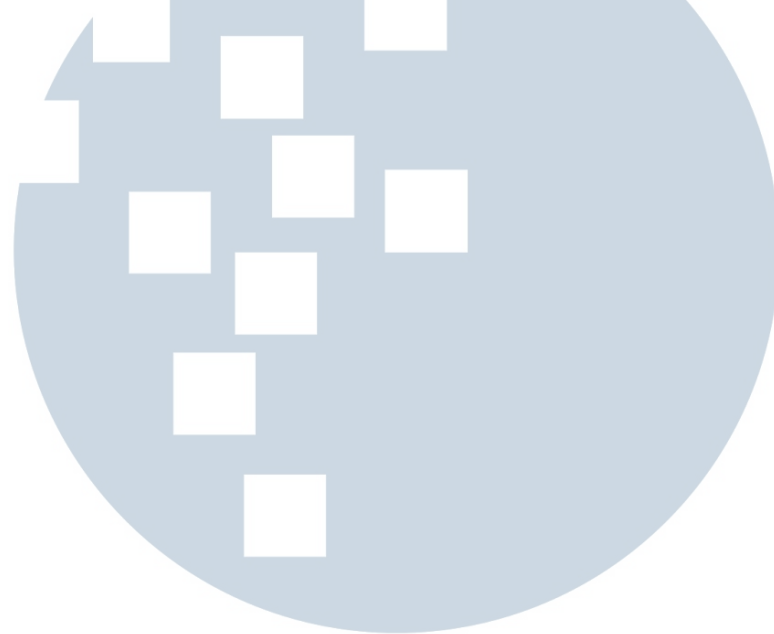


DAFTAR GAMBAR

Gambar 2.1	Kurva <i>Logistic Regression</i>	9
Gambar 2.2	Tahap metode <i>stacking</i>	16
Gambar 2.3	<i>Confusion matrix</i>	18
Gambar 3.1	<i>Flowchart</i> utama	24
Gambar 3.2	Lima baris pertama <i>dataset</i> satu	26
Gambar 3.3	Lima baris pertama <i>dataset</i> dua	26
Gambar 3.4	Lima baris pertama <i>dataset</i> tiga	26
Gambar 3.5	<i>Flowchart preprocess</i> dan <i>hold out</i> data	27
Gambar 3.6	<i>Missing value</i> pada <i>dataset</i> satu	28
Gambar 3.7	<i>Missing value</i> pada <i>dataset</i> dua	29
Gambar 3.8	<i>Missing value</i> pada <i>dataset</i> tiga	29
Gambar 3.9	Hasil dari proses <i>handle missing value</i>	30
Gambar 3.10	Hasil dari pengecekan jumlah <i>dataset</i> satu	31
Gambar 3.11	Hasil dari pengecekan jumlah <i>dataset</i> dua	31
Gambar 3.12	Hasil dari pengecekan jumlah <i>dataset</i> tiga	31
Gambar 3.13	Hasil dari pengecekan jumlah <i>dataset</i> satu setelah SMOTE	32
Gambar 3.14	Hasil dari pengecekan jumlah <i>dataset</i> dua setelah SMOTE	33
Gambar 3.15	Hasil dari pengecekan jumlah <i>dataset</i> tiga setelah SMOTE	33
Gambar 3.16	<i>Flowchart</i> data modeling	34
Gambar 3.17	<i>Flowchart</i> evaluasi	36
Gambar 4.1	<i>Confusion matrix dataset</i> satu	38
Gambar 4.2	<i>Confusion matrix dataset</i> dua	38
Gambar 4.3	<i>Confusion matrix dataset</i> tiga	39
Gambar 4.4	Kurva ROC <i>dataset</i> satu	41
Gambar 4.5	Kurva ROC <i>dataset</i> dua	41
Gambar 4.6	Kurva ROC <i>dataset</i> tiga	42
Gambar 4.7	Hasil <i>confusion matrix</i> untuk pemodelan dengan <i>Logistic Regression</i> menggunakan PCA dua komponen pada <i>dataset</i> satu	51
Gambar 4.8	Hasil <i>confusion matrix</i> untuk pemodelan dengan <i>Logistic Regression</i> menggunakan PCA dua komponen pada <i>dataset</i> dua	52
Gambar 4.9	Hasil <i>confusion matrix</i> untuk pemodelan dengan <i>Logistic Regression</i> menggunakan PCA dua komponen pada <i>dataset</i> tiga	52
Gambar 4.10	Hasil <i>confusion matrix</i> untuk pemodelan dengan <i>Logistic Regression</i> menggunakan PCA tiga komponen pada <i>dataset</i> satu	53
Gambar 4.11	Hasil <i>confusion matrix</i> untuk pemodelan dengan <i>Logistic Regression</i> menggunakan PCA tiga komponen pada <i>dataset</i> dua	54
Gambar 4.12	Hasil <i>confusion matrix</i> untuk pemodelan dengan <i>Logistic Regression</i> menggunakan PCA tiga komponen pada <i>dataset</i> tiga	54
Gambar 4.13	Hasil <i>confusion matrix</i> untuk pemodelan dengan <i>Logistic Regression</i> menggunakan PCA 95% varian pada <i>dataset</i> satu	55

Gambar 4.14	Hasil <i>confusion matrix</i> untuk pemodelan dengan <i>Logistic Regression</i> menggunakan PCA 95% varian pada <i>dataset</i> dua	56
Gambar 4.15	Hasil <i>confusion matrix</i> untuk pemodelan dengan <i>Logistic Regression</i> menggunakan PCA 95% varian pada <i>dataset</i> tiga	56
Gambar 4.16	Hasil <i>confusion matrix</i> untuk pemodelan dengan <i>Logistic Regression</i> menggunakan PCA 99% varian pada <i>dataset</i> satu	57
Gambar 4.17	Hasil <i>confusion matrix</i> untuk pemodelan dengan <i>Logistic Regression</i> menggunakan PCA 99% varian pada <i>dataset</i> dua	58
Gambar 4.18	Hasil <i>confusion matrix</i> untuk pemodelan dengan <i>Logistic Regression</i> menggunakan PCA 99% varian pada <i>dataset</i> tiga	58
Gambar 4.19	Kurva ROC PCA dua komponen <i>dataset</i> satu	60
Gambar 4.20	Kurva ROC PCA dua komponen <i>dataset</i> dua	60
Gambar 4.21	Kurva ROC PCA dua komponen <i>dataset</i> tiga	61
Gambar 4.22	Kurva ROC PCA tiga komponen <i>dataset</i> satu	61
Gambar 4.23	Kurva ROC PCA tiga komponen <i>dataset</i> dua	62
Gambar 4.24	Kurva ROC PCA tiga komponen <i>dataset</i> tiga	63
Gambar 4.25	Kurva ROC PCA 95% varian <i>dataset</i> satu	63
Gambar 4.26	Kurva ROC PCA 95% varian <i>dataset</i> dua	64
Gambar 4.27	Kurva ROC PCA 95% varian <i>dataset</i> tiga	64
Gambar 4.28	Kurva ROC PCA 99% varian <i>dataset</i> satu	65
Gambar 4.29	Kurva ROC PCA 99% varian <i>dataset</i> dua	66
Gambar 4.30	Kurva ROC PCA 99% varian <i>dataset</i> tiga	66
Gambar 4.31	Hasil <i>confusion matrix</i> untuk pemodelan dengan <i>Logistic Regression</i> menggunakan RFE pada <i>dataset</i> satu	75
Gambar 4.32	Hasil <i>confusion matrix</i> untuk pemodelan dengan <i>Logistic Regression</i> menggunakan RFE pada <i>dataset</i> dua	76
Gambar 4.33	Hasil <i>confusion matrix</i> untuk pemodelan dengan <i>Logistic Regression</i> menggunakan RFE pada <i>dataset</i> tiga	76
Gambar 4.34	Kurva ROC RFE <i>dataset</i> satu	78
Gambar 4.35	Kurva ROC RFE <i>dataset</i> dua	78
Gambar 4.36	Kurva ROC RFE <i>dataset</i> dua	79
Gambar 4.37	Hasil <i>confusion matrix</i> untuk pemodelan dengan <i>Logistic Regression</i> menggunakan <i>stacking</i> pada <i>dataset</i> satu	82
Gambar 4.38	Hasil <i>confusion matrix</i> untuk pemodelan dengan <i>Logistic Regression</i> menggunakan <i>stacking</i> pada <i>dataset</i> dua	83
Gambar 4.39	Hasil <i>confusion matrix</i> untuk pemodelan dengan <i>Logistic Regression</i> menggunakan <i>stacking</i> pada <i>dataset</i> tiga	84
Gambar 4.40	Kurva ROC metode <i>stacking</i> <i>dataset</i> satu	85
Gambar 4.41	Kurva ROC metode <i>stacking</i> <i>dataset</i> dua	86
Gambar 4.42	Kurva ROC metode <i>stacking</i> <i>dataset</i> tiga	86
Gambar 4.43	Hasil <i>confusion matrix</i> untuk pemodelan dengan <i>Logistic Regression</i> menggunakan GridSearchCV pada <i>dataset</i> satu	97
Gambar 4.44	Hasil <i>confusion matrix</i> untuk pemodelan dengan <i>Logistic Regression</i> menggunakan GridSearchCV pada <i>dataset</i> dua	98
Gambar 4.45	Hasil <i>confusion matrix</i> untuk pemodelan dengan <i>Logistic Regression</i> menggunakan GridSearchCV pada <i>dataset</i> tiga	99
Gambar 4.46	Kurva ROC GridSearchCV <i>dataset</i> satu	99
Gambar 4.47	Kurva ROC GridSearchCV <i>dataset</i> dua	100
Gambar 4.48	Kurva ROC GridSearchCV <i>dataset</i> tiga	101

Gambar 4.49	Histogram klasifikasi pasien perempuan yang terkena penyakit diabetes dengan <i>dataset</i> satu	110
Gambar 4.50	Histogram klasifikasi pasien perempuan yang terkena penyakit diabetes dengan <i>dataset</i> dua	111
Gambar 4.51	Histogram klasifikasi pasien perempuan yang terkena penyakit diabetes dengan <i>dataset</i> tiga	111



UMMN
 UNIVERSITAS
 MULTIMEDIA
 NUSANTARA

DAFTAR TABEL

Tabel 2.1	Tes diagnosa diabetes	10
Tabel 3.1	Keterangan fitur pada <i>dataset</i>	22
Tabel 4.1	Hasil evaluasi pemodelan dengan <i>Logistic Regression</i> dengan <i>dataset</i> satu	43
Tabel 4.2	Hasil evaluasi pemodelan dengan <i>Logistic Regression</i> dengan <i>dataset</i> dua	43
Tabel 4.3	Hasil evaluasi pemodelan dengan <i>Logistic Regression</i> dengan <i>dataset</i> tiga	44
Tabel 4.4	Varian ratio PCA dua komponen dengan <i>dataset</i> satu	45
Tabel 4.5	Varian ratio PCA dua komponen dengan <i>dataset</i> dua	45
Tabel 4.6	Varian ratio PCA dua komponen dengan <i>dataset</i> tiga	46
Tabel 4.7	Varian ratio PCA tiga komponen dengan <i>dataset</i> satu	46
Tabel 4.8	Varian ratio PCA tiga komponen dengan <i>dataset</i> dua	46
Tabel 4.9	Varian ratio PCA tiga komponen dengan <i>dataset</i> tiga	47
Tabel 4.10	Varian ratio PCA dengan 95% varian dengan <i>dataset</i> satu	47
Tabel 4.11	Varian ratio PCA dengan 95% varian dengan <i>dataset</i> dua	48
Tabel 4.12	Varian ratio PCA dengan 95% varian dengan <i>dataset</i> tiga	48
Tabel 4.13	Varian ratio PCA dengan 99% varian dengan <i>dataset</i> satu	49
Tabel 4.14	Varian ratio PCA dengan 99% varian dengan <i>dataset</i> dua	49
Tabel 4.15	Varian ratio PCA dengan 99% varian dengan <i>dataset</i> tiga	50
Tabel 4.16	Hasil evaluasi untuk pemodelan dengan <i>Logistic Regression</i> menggunakan PCA dengan <i>dataset</i> satu	67
Tabel 4.17	Hasil evaluasi untuk pemodelan dengan <i>Logistic Regression</i> menggunakan PCA dengan <i>dataset</i> satu	68
Tabel 4.18	Hasil evaluasi untuk pemodelan dengan <i>Logistic Regression</i> menggunakan PCA dengan <i>dataset</i> dua	69
Tabel 4.19	Hasil evaluasi untuk pemodelan dengan <i>Logistic Regression</i> menggunakan PCA dengan <i>dataset</i> dua	70
Tabel 4.20	Hasil evaluasi untuk pemodelan dengan <i>Logistic Regression</i> menggunakan PCA dengan <i>dataset</i> tiga	71
Tabel 4.21	Hasil evaluasi untuk pemodelan dengan <i>Logistic Regression</i> menggunakan PCA dengan <i>dataset</i> tiga	72
Tabel 4.22	<i>Ranking</i> fitur yang paling berpengaruh dari hasil seleksi RFE pada <i>dataset</i> satu dan dua	73
Tabel 4.23	<i>Ranking</i> fitur yang paling berpengaruh dari hasil seleksi RFE pada <i>dataset</i> tiga	74
Tabel 4.24	Hasil evaluasi untuk pemodelan dengan <i>Logistic Regression</i> menggunakan RFE	80
Tabel 4.25	Hasil evaluasi untuk pemodelan dengan <i>Logistic Regression</i> menggunakan RFE	80
Tabel 4.26	Hasil evaluasi model <i>Logistic Regression</i> dengan metode <i>stacking</i>	87
Tabel 4.27	Hasil evaluasi model <i>Logistic Regression</i> dengan metode <i>stacking</i> pada <i>dataset</i> tiga	88
Tabel 4.28	Hasil GridSearchCV dengan kombinasi <i>hyperparameter</i> pada <i>dataset</i> satu sebelum SMOTE	90

Tabel 4.28	Hasil GridSearchCV dengan kombinasi <i>hyperparameter</i> pada <i>dataset</i> satu sebelum SMOTE (lanjutan)	91
Tabel 4.29	Hasil GridSearchCV dengan kombinasi <i>hyperparameter</i> pada <i>dataset</i> satu setelah SMOTE	91
Tabel 4.30	Hasil GridSearchCV dengan kombinasi <i>hyperparameter</i> pada <i>dataset</i> dua sebelum SMOTE	92
Tabel 4.31	Hasil GridSearchCV dengan kombinasi <i>hyperparameter</i> pada <i>dataset</i> dua setelah SMOTE	94
Tabel 4.32	Hasil GridSearchCV dengan kombinasi <i>hyperparameter</i> pada <i>dataset</i> tiga sebelum SMOTE	95
Tabel 4.33	Hasil GridSearchCV dengan kombinasi <i>hyperparameter</i> pada <i>dataset</i> tiga setelah SMOTE	96
Tabel 4.34	Hasil evaluasi <i>Logistic Regression</i> dengan GridSearchCV .	101
Tabel 4.35	Hasil evaluasi <i>Logistic Regression</i> dengan GridSearchCV pada <i>dataset</i> tiga	102
Tabel 4.36	Hasil performa pemodelan dengan <i>dataset</i> satu dan dua sebelum SMOTE	104
Tabel 4.37	Hasil performa pemodelan dengan <i>dataset</i> tiga sebelum SMOTE	104
Tabel 4.38	Hasil performa pemodelan dengan <i>dataset</i> satu dan dua setelah SMOTE	105
Tabel 4.39	Hasil performa pemodelan dengan <i>dataset</i> tiga setelah SMOTE	105
Tabel 4.40	Hasil <i>confusion matrix</i> pemodelan dengan <i>dataset</i> sebelum SMOTE	106
Tabel 4.41	Hasil <i>confusion matrix</i> pemodelan dengan <i>dataset</i> setelah SMOTE	107
Tabel 4.42	<i>Research gap</i> penelitian terdahulu	112

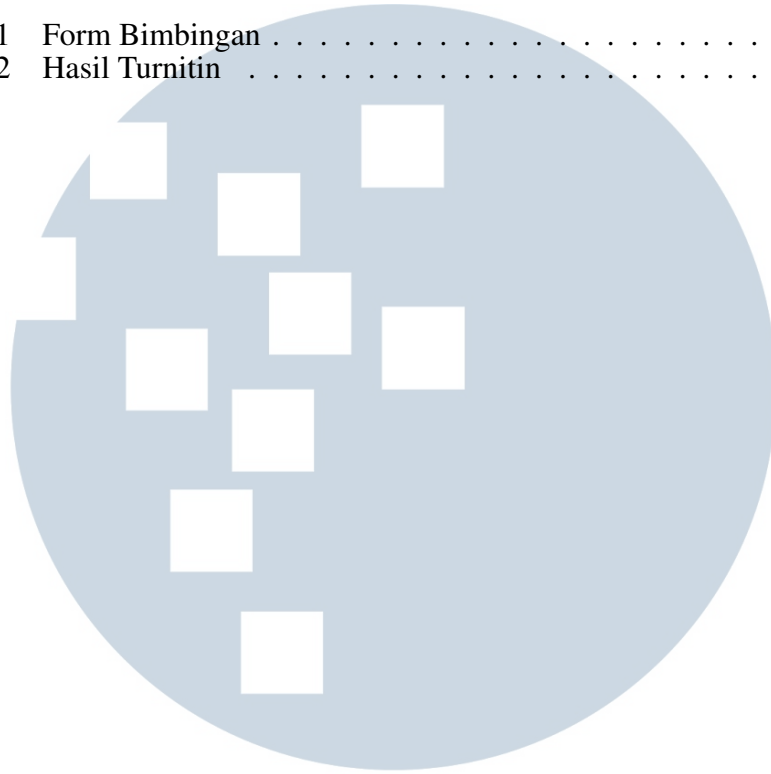


DAFTAR KODE

3.1	<i>Import library</i>	25
3.2	<i>Read dataset</i>	25
3.3	<i>Copy data dan cek missing value</i>	28
3.4	<i>Handle missing value</i>	29
3.5	Cek keseimbangan data	30
3.6	<i>Split dataset</i>	31
3.7	Implementasi SMOTE	32
3.8	Standardisasi data	33
4.1	Implementasi model <i>Logistic Regression</i>	37
4.2	<i>Confusion matrix</i>	37
4.3	Kurva ROC	39
4.4	Evaluasi model	42
4.5	Implementasi PCA dengan dua komponen	44
4.6	Implementasi model <i>Logistic Regression</i> menggunakan PCA	50
4.7	<i>Confusion matrix</i> untuk pemodelan dengan <i>Logistic Regression</i> menggunakan PCA	50
4.8	Kurva ROC untuk pemodelan dengan <i>Logistic Regression</i> menggunakan PCA	59
4.9	Hasil evaluasi untuk pemodelan dengan <i>Logistic Regression</i> menggunakan PCA dua komponen	67
4.10	Pemilihan fitur terbaik oleh RFE	72
4.11	Cetak <i>ranking</i> fitur yang tertinggi	73
4.12	<i>Apply</i> fitur baru pada <i>dataset</i>	74
4.13	<i>Confusion matrix</i> untuk pemodelan dengan <i>Logistic Regression</i> menggunakan RFE	74
4.14	Kurva ROC untuk pemodelan dengan <i>Logistic Regression</i> menggunakan RFE	77
4.15	Hasil evaluasi untuk pemodelan dengan <i>Logistic Regression</i> menggunakan RFE	79
4.16	Definisikan model dasar dan model meta untuk melakukan metode <i>stacking</i>	81
4.17	Pelatihan dan pengujian model dengan metode <i>stacking</i> dengan data <i>train</i> dan data <i>test</i>	81
4.18	<i>Confusion matrix</i> pemodelan <i>Logistic Regression</i> menggunakan metode <i>stacking</i>	82
4.19	Kurva ROC metode <i>stacking</i>	84
4.20	Hasil evaluasi untuk pemodelan dengan <i>Logistic Regression</i> menggunakan <i>stacking</i>	87
4.21	<i>Hyperparameter tuning</i> dengan GridSearchCV	89
4.22	Klasifikasi model	109

DAFTAR LAMPIRAN

Lampiran 1	Form Bimbingan	123
Lampiran 2	Hasil Turnitin	126



UMMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA