

**KLASIFIKASI TRANSAKSI *FRAUD* KARTU KREDIT
MENGUNAKAN *OVERSAMPLING* ADASYN DAN
SELEKSI FITUR SVM-RFECV**

TESIS



**UNIVERSITAS PENDIDIKAN GANESHA
PROGRAM PASCASARJANA
PROGRAM STUDI ILMU KOMPUTER
SINGARAJA
2024**

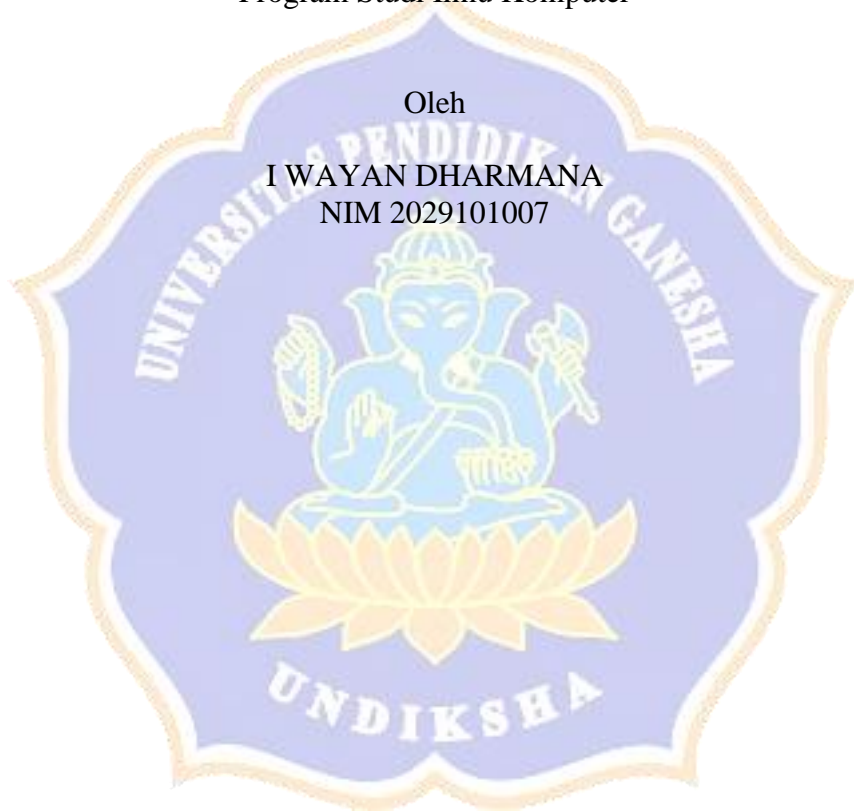
**KLASIFIKASI TRANSAKSI *FRAUD* KARTU KREDIT
MENGUNAKAN *OVERSAMPLING* ADASYN DAN
SELEKSI FITUR SVM-RFECV**

TESIS

Diajukan kepada
Universitas Pendidikan Ganesha
untuk Memenuhi Sebagian Persyaratan
Memperoleh Gelar Magister Komputer
Program Studi Ilmu Komputer

Oleh

I WAYAN DHARMANA
NIM 2029101007

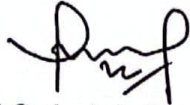


**UNIVERSITAS PENDIDIKAN GANESHA
PROGRAM PASCASARJANA
PROGRAM STUDI ILMU KOMPUTER
SINGARAJA
2024**

LEMBAR PERSETUJUAN PEMBIMBING

Tesis oleh I Wayan Dharmana ini telah diperiksa dan disetujui untuk mengikuti Ujian Tesis

Singaraja, 19 Januari 2024
Pembimbing I



Dr. I Gede Aris Gunadi, S.Si. M.Kom.
NIP. 197703182008121004

Pembimbing II



Dr. Luh Jeni Hrawati Dewi, S.T., M.Pd.
NIP. 197606252001122001

LEMBAR PERSETUJUAN TIM PENGUJI

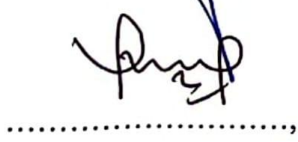
Tesis oleh I Wayan Dharmana ini telah dipertahankan di depan tim penguji dan dinyatakan diterima sebagai salah satu persyaratan untuk memperoleh gelar Magister Komputer di Program Studi Ilmu Komputer, Program Pascasarjana, Universitas Pendidikan Ganesha.

Disetujui pada tanggal: 15 Februari 2024

oleh
Tim Penguji



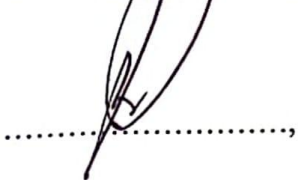
Ketua (Dr. I Made Gede Sunarya, S.Kom., M.Cs.)
NIP. 198307252008011008



Anggota (Dr. I Gede Aris Gunadi, S.Si. M.Kom.)
NIP. 197703182008121004



Anggota (Dr. Luh Joni Erawati Dewi, S.T., M.Pd.)
NIP. 197606252001122001



Anggota (Dr. I Nyoman Sukajaya, M.T.)
NIP. 196711151993031001

Mengetahui Direktur
Program Pascasarjana Undiksha,



Prof. Dr. I Nyoman Jampel, M.Pd.
NIP. 195910101986031003

LEMBAR PERNYATAAN

Saya menyatakan dengan sesungguhnya bahwa tesis yang saya susun sebagai syarat untuk memperoleh gelar Magister Komputer dari Program Pascasarjana Universitas Pendidikan Ganesha seluruhnya merupakan hasil karya sendiri. Bagian-bagian tertentu dalam penulisan tesis yang saya kutip dari hasil karya orang lain telah dituliskan sumbernya dengan jelas dan sesuai dengan norma, kaidah, serta etika akademis.

Apabila dikemudian hari ditemukan seluruh atau sebagian tesis ini bukan hasil karya saya sendiri atau adanya plagiat dalam bagian-bagian tertentu, saya bersedia menerima sanksi pencabutan gelar akademik yang saya sandang dan sanksi-sanksi lainnya sesuai dengan peraturan perundang-undangan yang berlaku di wilayah Negara Kesatuan Republik Indonesia.

Singaraja, 19 Januari 2024
Yang memberi pernyataan,



(I Wayan Dharmana)

PRAKATA

Puji dan syukur penulis panjatkan kepada Tuhan Yang Maha Esa atas anugrah-Nya, sehingga tesis dengan judul “Klasifikasi Transaksi *Fraud* Kartu Kredit Menggunakan *Oversampling* ADASYN dan Seleksi Fitur SVM-RFECV” dapat diselesaikan sesuai dengan yang direncanakan.

Tesis ini ditulis untuk memenuhi salah satu persyaratan untuk memperoleh gelar Magister Komputer Pascasarjana Universitas Pendidikan Ganesha pada Program Studi Ilmu Komputer. terselesaikannya tesis ini telah banyak memperoleh uluran tangan dari berbagai pihak. Untuk itu, izinkan penulis menyampaikan terima kasih dan penghargaan kepada pihak-pihak berikut.

1. Bapak Dr. I Gede Aris Gunadi, S.Si. M.Kom. selaku Dosen Pembimbing I yang telah dengan sabar membimbing, mengarahkan, dan memberikan motivasi yang demikian bermakna, sehingga penulis mampu melewati berbagai hambatan dalam perjalanan studi dan penyelesaian tesis ini;
2. Ibu Dr. Luh Joni Erawati Dewi, S.T., M.Pd. selaku Dosen Pembimbing II yang telah memberikan arahan kepada penulis selama penelitian dan penulisan naskah laporan tesis ini, sehingga tesis ini dapat terwujud dengan baik sesuai harapan;
3. Bapak Dr. I Nyoman Sukajaya, M.T. dan Bapak Dr. I Made Gede Sunarya, S.Kom., M.Cs. selaku penguji yang telah banyak memberikan masukan-masukan yang menunjang penyempurnaan penelitian ini;

4. Koordinator Program Studi Ilmu Komputer dan staf dosen pengajar yang telah banyak membantu dan memotivasi penulis selama penyusunan tesis ini;
5. Direktur Pascasarjana Undiksha dan staf, yang telah banyak membantu selama penulis menyelesaikan tesis ini;
6. Rektor Universitas Pendidikan Ganesha, yang telah memberikan bantuan secara moral dan memfasilitasi berbagai kepentingan penulis dalam menyelesaikan tesis ini;
7. Bapak I Ketut Sanjaya dan Ibu Ni Made Genep selaku orang tua penulis, yang telah banyak memberikan dukungan moral selama penyelesaian tesis ini;
8. Seluruh pihak yang telah membantu dan mendukung penulisan tesis ini, yang tidak dapat disebutkan satu persatu.

Penulis menyadari bahwa tesis ini belum sempurna. Namun, kehadirannya dalam konstelasi masyarakat akademis akan menambah referensi dalam pengembangan ilmu pengetahuan khususnya Ilmu Komputer. Semoga tesis ini bermanfaat bagi semua pihak, khususnya bagi pengembangan pengetahuan dalam mencegah *fraud* dalam sistem pembayaran.

Singaraja, 19 Januari 2024

Penulis

DAFTAR ISI

(halaman)

LEMBAR PERSETUJUAN PEMBIMBING	i
LEMBAR PERSETUJUAN TIM PENGUJI	ii
LEMBAR PERNYATAAN	iii
PRAKATA	iv
ABSTRAK	vi
ABSTRACT	vii
DAFTAR ISI	viii
DAFTAR TABEL	x
DAFTAR GAMBAR	xi
DAFTAR LAMPIRAN	xiii
BAB I PENDAHULUAN	1
1.1 Latar Belakang Penelitian	1
1.2 Identifikasi Masalah	8
1.3 Pembatasan Masalah	8
1.4 Rumusan Masalah	9
1.5 Tujuan Penelitian	9
1.6 Manfaat Penelitian	10
BAB II KAJIAN PUSTAKA	12
2.1 Kajian Teori	12
2.1.1 <i>Fraud</i> Kartu Kredit	12
2.1.2 <i>Machine Learning</i>	14
2.1.3 <i>Feature Selection</i>	15
2.1.4 SVM-RFECV	17
2.1.5 Data Tidak Seimbang	20
2.1.6 <i>Adaptive Synthetic (ADASYN)</i>	22
2.1.7 <i>Support Vector Machine (SVM)</i>	24
2.1.8 <i>Adaptive Boosting (AdaBoost)</i>	24
2.1.9 <i>Light Gradient Boosting Machine (LightGBM)</i>	25
2.1.10 Uji Kinerja Model Klasifikasi	26
2.1.11 <i>Silhouette Index</i>	29

2.2	Kajian Penelitian Yang Relevan.....	29
2.3	Kerangka Berpikir	34
BAB III METODE PENELITIAN.....		38
3.1	Rancangan Penelitian	38
3.2	Teknik Pengumpulan Data	40
3.3	<i>Data Pre-Processing</i>	42
3.3.1	<i>Data Cleaning</i>	43
3.3.2	<i>Encoding Categorical Features</i>	45
3.3.3	<i>Feature Scaling</i>	45
3.3.4	<i>Data Splitting</i>	46
3.3.5	<i>Feature Selection</i>	46
3.4	<i>Data Resampling</i>	47
3.4.1	<i>Oversampling</i> dengan ADASYN.....	48
3.4.2	Analisis Kualitas Data.....	49
3.4.3	Uji Performa Klasifikasi	51
3.5	Analisis dan Evaluasi Hasil	52
3.5.1	Evaluasi Kinerja Prediksi.....	52
3.5.2	Analisis <i>Silhouette Index</i>	53
BAB IV HASIL PENELITIAN DAN PEMBAHASAN.....		55
4.1	Tahapan <i>Data Pre-Processing</i>	55
4.2	Seleksi Fitur dengan SVM-RFECV	63
4.3	Hasil <i>Oversampling</i>	70
4.3.1	Hasil <i>Oversampling</i> dengan Data Latih 70%	70
4.3.2	Hasil <i>Oversampling</i> dengan Data Latih 80%	77
4.3.3	Hasil <i>Oversampling</i> dengan Data Latih 90%	81
4.4	Analisis Strategi <i>Oversampling</i>	86
BAB V PENUTUP.....		96
5.1	Rangkuman.....	96
5.2	Simpulan.....	100
5.3	Saran	101
DAFTAR PUSTAKA		102
LAMPIRAN.....		107

DAFTAR TABEL

(halaman)

Tabel 2.1 Alur SVM-RFE.....	19
Tabel 2.2 Algoritma ADASYN.....	22
Tabel 2.3 Contoh Confusion Matrix	27
Tabel 3.1 Variabel Tabel Transaksi	41
Tabel 3.2 Variabel Tabel Identitas	42
Tabel 3.3 Hyperparameter ADASYN	49
Tabel 3.4 Skenario Kombinasi Strategi Rasio Data Latih dan Oversampling	49
Tabel 4.1 Kelompok Variabel Independen	57
Tabel 4.2 Rincian Jumlah Data Dihapus Sesuai Kelompok Variabel.....	58
Tabel 4.3 Rincian Rasio Data Latih dan Data Uji.....	63
Tabel 4.4 Kinerja AUC Seleksi Fitur dengan Data Latih 70%	64
Tabel 4.5 Variabel Tereliminasi Berdasarkan Seleksi Fitur Data Latih 70%	65
Tabel 4.6 Variabel Tereliminasi Berdasarkan Seleksi Fitur Data Latih 80%	66
Tabel 4.7 Kinerja AUC Seleksi Fitur dengan Data Latih 80%	67
Tabel 4.8 Variabel Tereliminasi Berdasarkan Seleksi Fitur Data Latih 90%	68
Tabel 4.9 Kinerja AUC Seleksi Fitur dengan Data Latih 90%	69
Tabel 4.10 Perbandingan Jumlah Variabel Dengan Rasio Data Latih dan Uji.....	70
Tabel 4.11 Rincian Distribusi Data Oversampling Data Latih 70%	71
Tabel 4.12 Uji Klasifikasi Hasil Oversampling Data Latih 70%	75
Tabel 4.13 Rincian Distribusi Data Oversampling Data Latih 80%	77
Tabel 4.14 Uji Klasifikasi Hasil Oversampling Data Latih 80%	81
Tabel 4.15 Rincian Distribusi Data Oversampling Data Latih 90%	82
Tabel 4.16 Kinerja Uji Klasifikasi Hasil Oversampling Data Latih 90%	86
Tabel 4.17 Rangkuman Kinerja AUC Seluruh Strategi	87
Tabel 4.18 Rangkuman Kinerja Precision Seluruh Strategi.....	88
Tabel 4.19 Rangkuman Kinerja Recall Seluruh Strategi	90
Tabel 4.20 Rangkuman Kinerja F1 Score Seluruh Strategi	92
Tabel 4.21 Rincian Silhouette Index Seluruh Strategi	92

DAFTAR GAMBAR

(halaman)

Gambar 2.1 Alur Umum Teknik RFE.....	18
Gambar 2.2 Contoh Metrik AUC.....	28
Gambar 2.3 Kerangka Berpikir	34
Gambar 3.1 Alur Rancangan Penelitian.....	38
Gambar 3.2 Tampilan Dataset Transaksi Fraud.....	42
Gambar 3.3 Rasio Kolom Memiliki Missing Value	44
Gambar 3.4 Visualisasi Distribusi Data dengan Missing Value	44
Gambar 4.1 Alur Kerja Data Pre-Processing	55
Gambar 4.2 Visualisasi Missing Value Variabel Numerik.....	59
Gambar 4.3 Visualisasi Missing Value Variabel Kategorikal	59
Gambar 4.4 Distribusi Nilai Unik Variabel Kategorikal	60
Gambar 4.5 Hasil Label Encoding Variabel Card4	61
Gambar 4.6 Hasil Normalisasi Menggunakan MinMaxScaler	62
Gambar 4.7 Tampilan Perbandingan Hasil Normalisasi.....	62
Gambar 4.8 Perbandingan Nilai Validasi Silang Data Latih 70%	65
Gambar 4.9 Perbandingan Nilai Validasi Silang Data Latih 80%	67
Gambar 4.10 Perbandingan Nilai Validasi Silang Data Latih 90%	69
Gambar 4.11 Perbandingan Distribusi Hasil Oversampling Data Latih 70%	71
Gambar 4.12 Histogram Distribusi Data Latih 70% (Oversampling 100%)	72
Gambar 4.13 Histogram Distribusi Data Latih 70% (Oversampling 50%)	72
Gambar 4.14 Histogram Distribusi Data Latih 70% (Oversampling 25%)	73
Gambar 4.15 Silhouette Index Data Latih 70%	74
Gambar 4.16 Perbandingan Distribusi Hasil Oversampling Data Latih 80%	78
Gambar 4.17 Histogram Distribusi Data Latih 80% (Oversampling 100%)	79
Gambar 4.18 Histogram Distribusi Data Latih 80% (Oversampling 50%)	79
Gambar 4.19 Histogram Distribusi Data Latih 80% (Oversampling 25%)	79
Gambar 4.20 Silhouette Index Data Latih 80%	80
Gambar 4.21 Perbandingan Distribusi Hasil Oversampling Data Latih 90%	82
Gambar 4.22 Histogram Distribusi Data Latih 90% (Oversampling 100%)	83

Gambar 4.23 Histogram Distribusi Data Latih 90% (Oversampling 50%)	83
Gambar 4.24 Histogram Distribusi Data Latih 90% (Oversampling 25%)	84
Gambar 4.25 Silhouette Index Data Latih 90%	84
Gambar 4.26 Perbandingan Kinerja AUC	87
Gambar 4.27 Perbandingan Kinerja Precision	88
Gambar 4.28 Confusion Matrix LGBM dengan Data Latih 70%	89
Gambar 4.29 Perbandingan Kinerja Recall.....	90
Gambar 4.30 Perbandingan Kinerja F1 Score	91
Gambar 4.31 Perbandingan Silhouette Index	93



DAFTAR LAMPIRAN

(halaman)

Lampiran 1. Kode Sumber	107
Lampiran 2. Dataset IEEE CIS Fraud Detection	107

