

**DETEKSI ANOMALI MENGGUNAKAN *ENSEMBLE
LEARNING* DAN *RANDOM OVERSAMPLING* PADA
PENIPUAN TRANSAKSI KEUANGAN**

SKRIPSI



Oleh:

DEWA RAKA KRISNA SAPUTRA

20081010205

**PROGRAM STUDI INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN"
JAWA TIMUR
2024**

LEMBAR PENGESAHAN SKRIPSI

Judul : DETEKSI ANOMALI MENGGUNAKAN *ENSEMBLE LEARNING* DAN *RANDOM OVERSAMPLING* PADA PENIPUAN TRANSAKSI KEUANGAN
Oleh : Dewa Raka Krisna Saputra
NPM : 20081010205

Telah Diseminarkan Dalam Ujian Skripsi Pada:
Hari Kamis, Tanggal 4 Juli 2024

Mengetahui

1 Dosen Pembimbing



Yisti Vita Via, S.ST. M.Kom.

NIP. 19860425 2021212 001

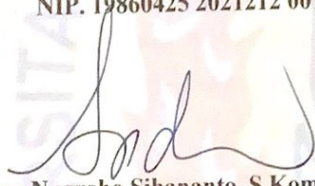
1 Dosen Penguji



Achmad Junaidi, S.Kom, M.Kom

NPT. 1993121 3202203 2010

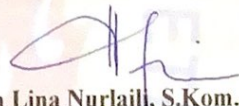
2



Andreas Nugroho Sihananto, S.Kom., M.Kom.

NPT. 211199 00 412271

2



Afina Lina Nurlaili, S.Kom., M.Kom.

NIP. 1993121 3202203 2010

Menyetujui

Dekan

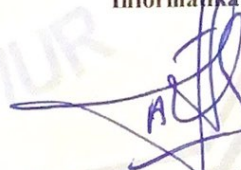
Fakultas Ilmu Komputer



Prof. Dr. Ir. Novrina Hendrasarie, Mt

NIP. 19681126 199403 2 001

Koordinator Program Studi
Informatika



Fetty Tri Anggraeny, S.Kom. M.Kom

NIP. 19820211 2021212 005

SURAT PERNYATAAN BEBAS DARI PLAGIASI

Saya, mahasiswa Program Studi Sarjana Informatika Universitas Pembangunan Nasional “Veteran” Jawa Timur, yang bertanda tangan di bawah ini:

Nama : Dewa Raka Krisna Saputra

NPM : 20081010205

Menyatakan dengan sesungguhnya bahwa Skripsi/Tugas Akhir yang saya kerjakan berjudul:

**“DETEKSI ANOMALI MENGGUNAKAN *ENSEMBLE LEARNING* DAN
RANDOM OVERSAMPLING PADA PENIPUAN TRANSAKSI
KEUANGAN”**

bukan merupakan plagiasi sebagian atau keseluruhan dari Skripsi/Tugas Akhir/Penelitian orang lain dari juga bukan merupakan produk dan software yang saya beli dari pihak lain. Saya juga menyatakan bahwa Skripsi/Tugas Akhir ini secara keseluruhan adalah pekerjaan Saya sendiri, kecuali yang dinyatakan dalam Daftar Pustaka dan tidak pernah diajukan untuk syarat memperoleh gelar di Universitas Pembangunan Nasional “Veteran” Jawa Timur maupun di Institut Pendidikan lain. Bukti hasil pengecekan plagiasi dokumen ini dapat ditelusuri melalui QR Code di bawah.

Apabila di kemudian hari terbukti bahwa dokumen ini merupakan plagiasi karya orang lain, saya sanggup menerima sanksi sesuai aturan yang berlaku.

Demikian atas perhatiannya disampaikan terima kasih.



Surabaya, 14 Juli 2024



Dewa Raka Krisna Saputra

NPM. 20081010205

DETEKSI ANOMALI MENGGUNAKAN *ENSEMBLE LEARNING* DAN *RANDOM OVERSAMPLING* PADA PENIPUAN TRANSAKSI KEUANGAN

Nama Mahasiswa : Dewa Raka Krisna Saputra

NPM : 20081010205

Program Studi : Informatika

Dosen Pembimbing : Yisti Vita Via, S.ST. M.Kom.

Andreas Nugroho Sihananto, S.Kom., M.Kom.

ABSTRAK

Dalam era digital, transaksi keuangan semakin beralih ke metode nontunai, yang didorong oleh kenyamanan dan efisiensi yang ditawarkan. Namun, peningkatan penggunaan kartu kredit dan transaksi online juga meningkatkan risiko penipuan finansial. Penelitian ini mengkaji penggunaan metode *ensemble learning* dan teknik *random oversampling* untuk mendeteksi anomali dalam transaksi keuangan, khususnya penipuan kartu kredit. Algoritma klasifikasi yang digunakan meliputi *Decision Tree (DT)*, *Random Forest (RF)*, *Logistic Regression (LR)*, dan *Naive Bayes (NB)*, dengan pendekatan *ensemble learning* seperti *Bagging*, *Boosting*, dan *Stacking*. Hasil penelitian menunjukkan bahwa metode *ensemble learning* secara signifikan meningkatkan performa deteksi penipuan dibandingkan model dasar (base model). Teknik *stacking* secara khusus menunjukkan peningkatan AUC yang signifikan, dengan beberapa algoritma mencapai AUC sempurna 1.00. *Random Forest (RF)* dengan metode *ensemble learning* menunjukkan performa yang sangat konsisten dan optimal dalam mendeteksi anomali penipuan. Penelitian ini menegaskan bahwa metode *ensemble learning*, terutama *stacking*, efektif dalam membedakan antara transaksi sah dan mencurigakan, sehingga dapat diandalkan untuk deteksi penipuan keuangan.

Kata Kunci: *Financial Transactions, Credit Card Fraud, Ensemble Learning, Random Oversampling, Anomaly Detection, Machine Learning.*

KATA PENGANTAR

Puji syukur kehadiran Tuhan Yang Maha Esa yang telah memberikan nikmat, berkah, dan kesempatan-Nya sehingga penulis dapat menyelesaikan tugas akhir atau skripsi dengan judul “DETEKSI ANOMALI MENGGUNAKAN ENSEMBLE LEARNING DAN RANDOM OVERSAMPLING PADA PENIPUAN TRANSAKSI KEUANGAN.”

Tidak lupa penulis ucapkan terima kasih kepada beberapa pihak yang membantu dan mendukung penulis sampai bisa menyelesaikan tugas akhir atau skripsi dengan lancar. Penulis menyadari bahwa masih terdapat banyak kekurangan pada tugas akhir atau skripsi yang penulis selesaikan ini. Oleh karena itu, penulis mengharapkan banyak saran dan masukan yang dapat membantu penulis untuk memperbaiki tugas akhir atau skripsi ini sehingga menjadi lebih baik.

Surabaya, 2 Juli 2024

Dewa Raka Krisna Saputra

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada para dosen dan dosen pembimbing yang membantu penulis dalam menyelesaikan skripsi ini. Pada kesempatan ini, penulis mengucapkan terimakasih yang sebesar-besarnya kepada:

1. Bapak Prof. Dr. Ir. Akhmad Fauzi, M.MT. selaku Rektor Universitas Pembangunan Nasional Veteran Jawa Timur.
2. Ibu Prof. Dr. Ir. Novirina Hendrasarie, MT, selaku Dekan Fakultas Ilmu Komputer.
3. Ibu Fetty Tri Anggraeny, S.Kom, M.Kom, selaku Koordinator Program Studi Informatika.
4. Ibu Yisti Vita Via, S.ST. M.Kom. selaku Dosen Pembimbing 1 atas kontribusi dan saran yang membangun selama penyusunan skripsi ini.
5. Bapak Andreas Nugroho Sihananto, S.Kom., M.Kom. selaku pembimbing 2 yang juga sudah meluangkan waktu dan memberi arahan kepada penulis dalam menyelesaikan skripsi ini.
6. Orang Tua dan Keluarga yang selalu memberikan dukungan moral, doa restu dan motivasi di setiap proses penulisan ini.
7. Indah Febryana Putri yang telah memberikan saran, semangat dan dukungan selama pengerjaan skripsi.
8. Bayu, Dimas, Thobi, Irnan, Chandra, Aca, dan segenap teman kelas E081, yang selalu memberikan saran, semangat dan dukungan selama pengerjaan skripsi.
9. Semua Teman-teman yang turut serta dalam mendukung penyelesaian skripsi ini, baik secara langsung maupun tidak langsung.

Akhir kata, semoga dengan adanya laporan ini dapat bermanfaat bagi penulis, pembaca serta memberikan ilmu dan pemikiran yang baru bagi pihak yang membutuhkan.

DAFTAR ISI

LEMBAR PENGESAHAN SKRIPSI.....	i
SURAT PERNYATAAN BEBAS DARI PLAGIASI	ii
KATA PENGANTAR.....	iv
UCAPAN TERIMA KASIH.....	v
DAFTAR ISI	vi
DAFTAR GAMBAR.....	ix
DAFTAR TABEL.....	xi
BAB I PENDAHULUAN	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah	3
1.3 Tujuan Penelitian	4
1.4 Manfaat Penelitian.....	4
1.5 Batasan Masalah	4
BAB II TINJAUAN PUSTAKA	6
2.1 Penelitian Terdahulu	6
2.2 Penipuan Transaksi	7
2.3 Deteksi Anomali	8
2.3.1 Anomali Titik	8
2.3.2 Anomali Kontekstual.....	9
2.3.3 Anomali Kolektif.....	9
2.4 <i>Machine learning</i>	10
2.5 <i>Ensemble Learning</i>	10
2.5.1 Bagging	11
2.5.2 Boosting	12
2.5.3 Stacking.....	12
2.6 <i>Base Learner</i>	13
2.6.1 <i>Logistic Regression</i>	13
2.6.2 <i>Decision Tree</i>	14
2.6.3 <i>Naive Bayes</i>	15
2.6.4 <i>Random Forest</i>	16

2.7	<i>Python</i>	18
2.7.1	<i>NumPy</i>	18
2.7.2	<i>Pandas</i>	18
2.7.3	<i>Matplotlib</i>	19
2.7.4	<i>Scikit-Learn</i>	19
2.8	Teknik <i>Resampling</i>	19
2.8.1	<i>Random oversampling</i>	20
2.9	<i>Heatmap Correlation</i>	21
2.10	Pengukuran Performa.....	22
2.10.1	<i>Confusion matrix</i>	22
2.10.2	<i>ROC-AUC Score</i>	23
BAB III METODOLOGI PENELITIAN.....		25
3.1	Alur Penelitian.....	25
3.2	Studi Literatur.....	26
3.3	Metode Pengumpulan Data.....	26
3.4	<i>Data Preprocessing</i>	27
3.4.1	<i>Knowledge Discovery in Database (KDD)</i>	28
3.4.2	Alat Uji.....	32
3.5	<i>Resampling</i>	32
3.6	Implementasi Model.....	32
3.6.1	Implementasi Model Skenario 1.....	33
3.6.2	Implementasi Model Skenario 2.....	52
3.7	Evaluasi Model.....	65
3.8	Analisa Hasil.....	66
BAB IV HASIL DAN PEMBAHASAN.....		68
4.1	Alat Bantu Penelitian.....	68
4.2	Pengumpulan Data.....	68
4.3	<i>Data Preprocessing</i>	69
4.4	<i>Resampling</i>	78
4.5	Implementasi Model.....	80
4.5.1	Implementasi Model Skenario 1.....	80
4.5.2	Implementasi Model Skenario 2.....	90
4.6	Evaluasi Model.....	112
4.6.1	Skenario 1 (<i>Base Learner</i>).....	112

4.6.2 Skenario 2 (<i>Ensemble Learning</i>).....	112
BAB V KESIMPULAN	120
5.1 Kesimpulan.....	120
5.2 Saran	120
DAFTAR PUSTAKA	122

DAFTAR GAMBAR

Gambar 2.1 Visualisasi Anomali Titik	9
Gambar 2.2 Visualisasi Anomali Kolektif.....	10
Gambar 2.3 <i>Bagging</i> dan <i>Boosting Ensemble Learning</i>	11
Gambar 2.4 <i>Stacking Ensemble Learning</i>	12
Gambar 2.5 Konsep Dasar <i>Decision Tree</i>	14
Gambar 2.6 Konsep Dasar <i>Random Forest</i>	17
Gambar 2.7 Visualisasi bentuk Undersampling dan Oversampling.....	20
Gambar 2.8 Proses <i>Random oversampling</i>	21
Gambar 2.9 <i>ROC-AUC Classification Evaluation Metric</i>	24
Gambar 3.1 Diagram Alur Penelitian	25
Gambar 3.2 Contoh Heatmap Korelasi	28
Gambar 3. 3 Implementasi Model Skenario 1.....	33
Gambar 3. 4 Alur Kerja <i>Logistic Regression</i>	34
Gambar 3. 5 Alur Kerja <i>Decision Tree</i>	38
Gambar 3. 6 Alur Kerja Naive Bayes.....	43
Gambar 3.7 Alur Kerja <i>Random Forest</i>	47
Gambar 3. 8 Alur Skenario 2.....	52
Gambar 3. 9 Alur Kerja Bagging	53
Gambar 3. 10 Alur Kerja Boosting.....	58
Gambar 3. 11 Alur Kerja Stacking	63
Gambar 3. 13 Visualisasi ROC AUC Score	67
Gambar 4.1 5 baris pertama pada Dataset.....	69
Gambar 4.2 Informasi tipe data pada Dataset	70
Gambar 4.3 Rangkuman Statistik Dataset.....	71
Gambar 4. 4 Heatmap Korelasi	73
Gambar 4. 5 Transaksi Penipuan Per Jenis Transaksi	74
Gambar 4. 6 Pengecekan Missing Value	75
Gambar 4. 7 Hasil Transformasi Data.....	77
Gambar 4. 8 Hasil Resampling Dataset	79
Gambar 4. 9 Visualisasi Hasil Resampling	80
Gambar 4. 10 Metrik Evaluasi <i>Logistic Regression</i>	81
Gambar 4. 11 <i>Confusion Matrix Logistic Regression</i>	82
Gambar 4. 12 Clasiification Report.....	82
Gambar 4. 13 Metrik Evaluasi <i>Decision Tree</i>	84
Gambar 4. 14 <i>Confusion Matrix Decision Tree</i>	84
Gambar 4. 15 <i>Classification Report Decision Tree</i>	85
Gambar 4. 16 Metrik Evaluasi <i>Gaussian Naive Bayes</i>	87
Gambar 4. 17 Confusion Matrix Decision Tree	87
Gambar 4. 18 <i>Clasiification Report naïve Bayes</i>	87
Gambar 4. 19 Metrik Evaluasi <i>Random Forest</i>	89
Gambar 4. 20 <i>Confusion Matrix Decision Tree</i>	89
Gambar 4. 21 <i>Clasiification Report Random Forest</i>	90
Gambar 4. 22 Evaluasi Metrik Logistic Regression Bagging.....	92
Gambar 4. 23 Confusion Matrix Logitic Regression Bagging.....	92
Gambar 4. 24 Classification Report Logistic Regression Bagging.....	93
Gambar 4. 25 Metrik Evaluasi Decision Tree Bagging	94

Gambar 4. 26 Confusion Matrix Decision Tree Bagging	94
Gambar 4. 27 <i>Classification Report Decisio Tree Bagging</i>	94
Gambar 4. 28 Evaluasi Metrik <i>Naïve Bayes Bagging</i>	95
Gambar 4. 29 <i>Confusion Matrix Naïve Bayes Bagging</i>	95
Gambar 4. 30 <i>Classification Report Gaussian Naïve Bayes Bagging</i>	96
Gambar 4. 31 Evaluasi Metrik Random Forest Bagging	96
Gambar 4. 32 <i>Confusion Matrix Random Forest Bagging</i>	97
Gambar 4. 33 <i>Classification Report Random Forest Bagging</i>	97
Gambar 4. 34 <i>Metrik Evaluasi Logistic Regression Boosting</i>	99
Gambar 4. 35 <i>Confusion Matrix Logistic Regression Boosting</i>	100
Gambar 4. 36 <i>Classification Report Logitic Regression Boosting</i>	100
Gambar 4. 37 <i>Metrik Evaluasi Decision Tree Boosting</i>	101
Gambar 4. 38 <i>Confusion Matrix Decision Tree Boosting</i>	101
Gambar 4. 39 <i>Classification Report Decision Tree Boosting</i>	101
Gambar 4. 40 <i>Metrik Evaluasi Gaussian Naïve Bayes Boosting</i>	102
Gambar 4. 41 <i>Confusion Matrnx Naïve Bayes Boosting</i>	102
Gambar 4. 42 <i>Classification Report Naïve Bayes Boosting</i>	103
Gambar 4. 43 <i>Metrik Evaluasi Random Forest Boosting</i>	103
Gambar 4. 44 <i>Confusion Matrix Random Forest Boosting</i>	104
Gambar 4. 45 <i>Classification Report Random Forest Boosting</i>	104
Gambar 4. 46 <i>Metrik Evaluasi Logistic Regression Stacking</i>	106
Gambar 4. 47 <i>Confusion Matrix Logistic Regression Stacking</i>	107
Gambar 4. 48 <i>Classification Report Logistic Regression Stacking</i>	107
Gambar 4. 49 <i>Evaluasi Metrik Decision Tree Stacking</i>	108
Gambar 4. 50 <i>Confusion Matrix Decision Tree Stacking</i>	108
Gambar 4. 51 <i>Classification Report Decision Tree Stacking</i>	108
Gambar 4. 52 <i>Metrik Evaluasi Gaussian Naïve Bayes Stacking</i>	109
Gambar 4. 53 <i>Confusion Matrix Gaussian Naive Bayes Stacking</i>	109
Gambar 4. 54 <i>Classification Report Naive Bayes Stacking</i>	110
Gambar 4. 55 <i>Metrk Evaluasi Random Forest Stacking</i>	110
Gambar 4. 56 <i>Confusion Matrix Random Forest Stacking</i>	111
Gambar 4. 57 <i>Classification Report Random Forest Stacking</i>	111
Gambar 4.58 Hasil ROC Curve model <i>Decision Tree</i>	114
Gambar 4.59 Hasil ROC Curve model <i>Logistic Regression</i>	115
Gambar 4.60 Hasil ROC Curve Model <i>Naive Bayes</i>	117
Gambar 4.61 Hasil ROC Curve Model <i>Random Forest</i>	118

DAFTAR TABEL

Tabel 3.1 Deskripsi Dataset.....	27
Tabel 3.2 Contoh Sebelum <i>Data Cleaning</i>	29
Tabel 3.3 Contoh Sesudah <i>Data Cleaning</i>	29
Tabel 3.4 Tabel Setelah <i>Data Selection</i>	30
Tabel 3.5 Contoh data sebelum transformasi data.....	30
Tabel 3.6 Contoh data sesudah transformasi data	31
Tabel 3.7 Hasil <i>Splitting Data</i>	31
Tabel 3.8 Data Sampel	34
Tabel 3. 9 Perbandingan Hasil.....	38
Tabel 3.10 <i>Data Training Decision Tree</i>	39
Tabel 3.11 <i>Data Testing Decision Tree</i>	40
Tabel 3.12 Proses Menghitung Nilai <i>Gain</i> pada atribut <i>type_PAYMENT</i>	41
Tabel 3.13 Proses Prediksi <i>Decision Tree</i>	42
Tabel 3.14 <i>Mean</i> dan Standar Deviasi Kelas <i>fraud</i>	45
Tabel 3.15 <i>Mean</i> dan Standar Deviasi Kelas <i>nonfraud</i>	45
Tabel 3.16 Perhitungan <i>Likelihood Probability</i>	45
Tabel 3.17 Perhitungan <i>Likelihood Probability</i> (2).....	45
Tabel 3.18 <i>Data Testing Naïve Bayes</i>	46
Tabel 3.19 <i>Data Training Random Forest</i>	48
Tabel 3.20 <i>Data Testing Random Forest</i>	48
Tabel 3. 21 Bootstrap Sampling 1	49
Tabel 3. 22 Bootstrap Sampling 2	49
Tabel 3. 23 Bootstrap Sampling 3	50
Tabel 3. 24 Perbandingan Hasil Prediksi dengan Data Asli.....	51
Tabel 3. 25 Bootstrap Sample 1.....	54
Tabel 3. 26 Bootstrap Sample 2	54
Tabel 3. 27 Bootstrap Sample 3	54
Tabel 3. 28 Data Uji	56
Tabel 3. 29 Prediksi dari Tree 1.....	56
Tabel 3. 30 Prediksi dari Tree 2.....	56
Tabel 3. 31 Prediksi dari Tree 3.....	56
Tabel 3. 32 Hasil prediksi gabungan	56
Tabel 3. 33 Hasil Prediksi dengan Data Asli	57
Tabel 3. 34 Data Latih <i>Boosting</i>	59
Tabel 3. 35 Data Uji <i>Boosting</i>	59
Tabel 3. 36 Hasil Prediksi Pertama	60
Tabel 3. 37 Bobot baru	61
Tabel 3. 38 Hasil Prediksi Model	61
Tabel 3. 39 Hasil Gabungan Model 1 dan Model 2	62
Tabel 3. 40 Hasil Prediksi	62
Tabel 3. 41 Data Latih <i>Stacking</i>	63

Tabel 3. 42 contoh hasil prediksi.....	64
Tabel 3. 43 Hasil Uji Prediksi <i>Meta Learner</i>	64
Tabel 3. 44 Prediksi Akhir Meta Learner	65
Tabel 3. 45 Skenario <i>Confusion Matrix</i>	65
Tabel 4. 1 Hasil Skenario 1	112
Tabel 4. 2 Hasil Skenario 2	113