

**PENERAPAN METODE *SUPERVISED LEARNING* DAN TEKNIK
RESAMPLING UNTUK PREDIKSI PENIPUAN TRANSAKSI KEUANGAN**

SKRIPSI

Program Studi Sistem Informasi

Jenjang Sarjana



Oleh:

Elven Constancio

NIM. 09031382126169

PROGRAM STUDI SISTEM INFORMASI

FAKULTAS ILMU KOMPUTER

UNIVERSITAS SRIWIJAYA

DESEMBER 2024

LEMBAR PENGESAHAN

SKRIPSI

**PENERAPAN METODE *SUPERVISED LEARNING* DAN TEKNIK
RESAMPLING UNTUK PREDIKSI PENIPUAN TRANSAKSI KEUANGAN**

Sebagai salah satu syarat untuk menyelesaikan studi di Program Studi

Sistem Informasi S1

Oleh:

Elven Constancio

NIM. 09031382126169

Palembang, 30 Desember 2024

Mengetahui,
Ketua Jurusan Sistem Informasi



Ahmad Rifai, S.T., M.T.
NIP. 19791020201021003

Pembimbing I



Ken Ditha Tania, M.Kom., Ph.D.
NIP. 198507182012122003

HALAMAN PERNYATAAN BEBAS PLAGIAT

Saya yang bertanda tangan di bawah ini:

Nama : Elven Constancio

NIM : 09031382126169

Program Studi : Sistem Informasi Bilingual

Judul Skripsi : Penerapan Metode *Supervised Learning* dan Teknik
Resampling untuk Prediksi Penipuan Transaksi Keuangan

Hasil Pengecekan *software iThenticate*/Turnitin : 7%

Menyatakan bahwa laporan skripsi saya merupakan hasil karya saya sendiri dan bukan penjiplakan/plagiat. Apabila ditemukan unsur penjiplakan/plagiat dalam laporan skripsi ini, maka saya bersedia menerima saksi akademik dari Universitas Sriwijaya sesuai dengan ketentuan yang berlaku.

Demikian pernyataan ini saya buat dengan sebenarnya dan tidak ada paksaan oleh siapapun.



Palembang, 30 Desember 2024



Elven Constancio
NIM. 09031382126169

HALAMAN PERNYATAAN INTEGRITAS

Saya yang bertanda tangan di bawah ini:

Nama : Elven Constancio
NIM : 09031382126169
Program Studi : Sistem Informasi Bilingual
Judul Publikasi : Penerapan Metode *Supervised Learning* dan Teknik
Resampling untuk Prediksi Penipuan Transaksi Keuangan
DOI : <https://doi.org/10.47065/bits.v6i3.6110>

Dengan ini menyatakan bahwa publikasi saya dengan judul “**Penerapan Metode *Supervised Learning* dan Teknik *Resampling* untuk Prediksi Penipuan Transaksi Keuangan**” yang diusulkan pada Jurnal *Building of Informatics, Technology and Science* (BITS), Vol. 6, No. 2, Desember 2024, pp. 1427–1439 bersifat orisinal dan saya sendiri yang bertanggung jawab pada setiap proses submisi publikasi tersebut. Bilamana di kemudian hari ditemukan ketidaksesuaian dengan pernyataan ini, maka saya bersedia dituntut dan diproses sesuai dengan ketentuan yang berlaku.

Demikian surat pernyataan ini dibuat dengan sesungguhnya dan sebenarnya.

Palembang, 30 Desember 2024

**Mengetahui,
Dosen Pembimbing**



Ken Ditha Tarria, M.Kom., Ph.D.
NIP. 198507182012122003

Menyatakan,



Elven Constancio
NIM. 09031382126169

HALAMAN PERSETUJUAN

Telah *accepted* di jurnal “*Building of Informatics, Technology and Science (BITS)*”

SINTA 3 pada:

Hari : Minggu
Tanggal : 01 Desember 2024
Nama : Elven Constancio
NIM : 09031382126169
Judul : Penerapan Metode *Supervised Learning* dan Teknik
Resampling untuk Prediksi Penipuan Transaksi Keuangan

Tim Pembimbing : Ken Ditha Tania, M.Kom., Ph.D.



Mengetahui,
Ketua Jurusan Sistem Informasi



Ahmad Rifai, S.T., M.T.
NIP. 19791020201021003

HALAMAN PERSEMBAHAN

“Masalah yang mengeruh, perasaan yang rapuh. Ini belum separuhnya.

Biasa saja, kamu tak apa. (Karena kamu akan baik-baik saja).”

(Hindia - Evaluasi)

Motto:

“Hidup bahagia tidak harus menjadi sempurna. cukup jadi apa adanya.”

Dengan penuh rasa syukur, karya ini penulis persembahkan kepada:

- 🎓 Tuhan Yang Maha Esa
- 🎓 Orang tua, saudara, dan keluarga besar
- 🎓 Dosen Pembimbing Tugas Akhir dan Pembimbing Akademik
- 🎓 Pihak-pihak yang telah memberikan inspirasi dan dukungan yang tak ternilai
- 🎓 Diriku sendiri di masa lalu, yang pernah ragu untuk bermimpi, dan diriku di masa depan, yang akan terus melangkah lebih jauh
- 🎓 Almamaterku Universitas Sriwijaya, Fakultas Ilmu Komputer, tempat yang telah menjadi bagian penting dalam perjalanan hidup ini
- 🎓 Untuk semesta ini sendiri, dengan segala tantangan dan keindahan yang kau hadirkan, terima kasih telah mengajarkanku untuk terus bertumbuh
- 🎓 Terakhir, untuk mereka yang pernah meremehkan dan merendahkan, terima kasih atas luka yang menjadi saksi bahwa segala sesuatu itu mungkin dicapai

KATA PENGANTAR

Dengan segenap rasa syukur kepada Tuhan Yang Maha Esa, atas rahmat, karunia, dan kekuatan yang telah diberikan, penulis akhirnya dapat menyelesaikan skripsi ini dengan judul “**Penerapan Metode *Supervised Learning* dan Teknik *Resampling* untuk Prediksi Penipuan Transaksi Keuangan**”. Skripsi ini merupakan wujud nyata dari usaha, doa, dan dukungan dari berbagai pihak yang telah memberikan kontribusi besar selama proses penyelesaian. Untuk itu, penulis ingin menyampaikan rasa terima kasih yang mendalam kepada:

1. Tuhan Yang Maha Esa, atas rahmat, kekuatan, dan petunjuk-Nya selama penulis menempuh setiap tahap dalam perjalanan ini.
2. Kedua orang tua, saudara, dan keluarga besar yang telah memberikan doa dan dukungan baik secara moral maupun material.
3. Bapak Prof. Dr. Erwin, S.Si., M.Si., selaku Dekan Fakultas Ilmu Komputer Universitas Sriwijaya; Bapak Samsuryadi, M.Kom., Ph.D., selaku Wakil Dekan I Bidang Akademik dan Penelitian; Bapak Dr. Rossi Passarella, M.Eng., selaku Wakil Dekan II Bidang Umum dan Keuangan; serta Bapak Dr. Muhammad Fachrurrozi, M.T., selaku Wakil Dekan III Bidang Kemahasiswaan dan Alumni.
4. Almarhum Bapak Assoc. Prof. Jaidan Jauhari, M.T., yang telah mengemban amanah sebagai Dekan Fakultas Ilmu Komputer Universitas Sriwijaya Periode 2020–2024, atas segala perhatian, nasihat, dukungan, dedikasi, dan sikap baiknya yang senantiasa meninggalkan kesan mendalam bagi penulis dan banyak mahasiswa lainnya.

5. Bapak Julian Supardi, S.Pd., M.T., selaku Wakil Dekan I Bidang Akademik Periode 2020–2024 yang telah memberikan bantuan dalam pengurusan berkas akademik penulis; Bapak Mgs. Afriyan Firdaus, S.Si., M.IT., selaku Wakil Dekan II Bidang Umum dan Keuangan Periode 2020–2024 yang telah banyak berperan dalam proses permohonan dana kegiatan organisasi mahasiswa yang penulis ikuti; Bapak Fathoni, S.T., M.MSI selaku Wakil Dekan III Bidang Kemahasiswaan dan Alumni Periode 2020–2024, yang telah banyak mendukung dan membantu kegiatan organisasi mahasiswa yang penulis ikuti, serta memberikan perhatian, dukungan, dan kedekatannya dengan mahasiswa yang memberikan motivasi selama perkuliahan.
6. Bapak Ahmad Rifai, S.T., M.T., selaku Ketua Jurusan Sistem Informasi Universitas Sriwijaya dan Bapak Dedy Kurniawan, M.Sc., selaku Sekretaris Jurusan Sistem Informasi atas segala perhatian, bantuan, dukungan, serta kedekatannya dengan mahasiswa yang telah memberikan dampak positif selama perkuliahan.
7. Ibu Ken Ditha Tania, M.Kom., Ph.D., selaku Dosen Pembimbing Tugas Akhir, yang telah meluangkan waktu untuk membimbing penulis, memberikan arahan, ilmu, dan motivasi yang tak ternilai selama proses penyusunan skripsi.
8. Ibu Dinda Lestarini, S.Si., M.T., selaku Dosen Pembimbing Akademik yang telah memberikan bimbingan, ilmu, serta arahan selama masa perkuliahan, sehingga menjadi inspirasi dan motivasi bagi penulis dalam menempuh perjalanan akademik ini.

9. Dosen-dosen Jurusan Sistem Informasi Universitas Sriwijaya yang pernah mengajar dan membimbing penulis sejak mahasiswa baru, antara lain:

- Ibu Endang Lestari Ruskan, S.Kom., M.T.,
- Ibu Nabila Rizky Oktadini, M.T.,
- Bapak Bayu Wijaya Putra, S.Kom., M.Kom.,
- Almarhum Bapak Ir. Muhammad Ikhsan Jambak, M.Sc.,
- Bapak Yoppy Sazaki, S.Si., M.T.,
- Bapak Fathoni, S.T., M.MSI
- Bapak Mgs. Afriyan Firdaus, S.Si., M.IT.,
- Ibu Dwi Rosa Indah, S.T., M.T.,
- Ibu Dinna Yunika, S.Si., M.T.,
- Bapak Dedy Kurniawan, M.Sc.,
- Ibu Ken Ditha Tania, M.Kom., Ph.D.,
- Ibu Dinda Lestarini, S.Si., M.T.,
- Bapak Ari Wedhasmara, S.Kom.,
- Bapak Apriansyah Putra, M.Kom., M.T.I.,
- Bapak Dr. Ali Ibrahim, M.T.,
- Bapak Pacu Putra Suarli, M.Sc.,
- Ibu Allsela Meiriza, M.T.,
- Ibu Putri Eka Sevtiyuni, S.Kom., M.T.,
- Bapak Rusdi Effendi, M.T.,
- Bapak Rudi Sanjaya, M.Kom.,
- Ibu Assoc. Prof. Dr. Ermatita, M.Kom.,
- Bapak Assoc. Prof. Deris Stiawan, Ph.D.,

- Ibu Nurul Afifah, S.Kom., M.Kom.,
- Ibu Purwita Sari, S.Si., M.Kom.,
- Bapak Rahmat Izwan Heroza, S.T., M.T.,
- Ibu Mira Afrina, M.Sc., Ph.D.,
- Ibu Sarifah Putri Raflesia, S.Kom., M.T.,
- Ibu Rizka Dhini Kurnia, S.T., M.Sc.,
- Ibu Iin Seprina, S.Kom., M.Kom.,
- Bapak Yadi Utama, S.Kom., M.Kom.,
- Bapak Husni Syahbani, M.T.,
- Ibu Miftahul Falah, S.Si., M.Kom.,
- Bapak Ricy Fernando, M.Kom.,
- Bapak Muhammad Iqbal Tawaqal, M.Ds. (Dosen Multimedia),
- Bapak Prof. Dr. Yusuf Hartono, M.Sc. (Dosen Matematika Diskrit),

dan seluruh dosen lainnya yang telah berperan penting dalam perjalanan akademik penulis.

10. Mbak Rifka Ariza Nurhandini, S.E., selaku Admin Jurusan Sistem Informasi yang telah membantu penulis dalam proses pemberkasan dari mahasiswa baru hingga lulus.
11. Saudari Putri Setiyawati atas segala dukungan, perjuangan, dan bantuan yang sangat berarti selama masa perkuliahan ini sekaligus *partner* dalam perlombaan *web development*.
12. Saudara Arip, Agung, Adib, Alwi, Iqbal selaku teman seperjuangan penulis yang telah bersama-sama melewati perkuliahan ini.

13. Teman-teman sekelas Sistem Informasi Bilingual A 2021, atas kerja sama, kebersamaan, dan keseruan yang mewarnai perjalanan perkuliahan ini.
14. Teman-teman seangkatan dari kelas Sibil B, Sireg A, Sireg B, dan Sireg C 2021 yang telah bersama-sama menjalani perjalanan perkuliahan ini.
15. Teman-teman penulis dari jurusan Teknik Informatika, Sistem Komputer, dan Diploma Komputer, baik angkatan 2021 maupun adik tingkat angkatan 2022 dan 2023.
16. Kakak-kakak Angkatan 2020 yang telah memberikan banyak ilmu dan bantuan kepada penulis.
17. BEM KM Fasilkom Unsri Kabinet Cakra Baswara (2023) dan Adibrata Kriya (2024) yang telah memberikan banyak pengalaman dan kesempatan dalam mengembangkan potensi diri selama menjalani masa perkuliahan, serta mengajarkan arti dari sebuah perjuangan. Terima kasih kepada teman-teman dari BEM KM Fasilkom Unsri yang telah berjuang bersama dalam berbagai pengalaman ini.
18. GDSC Universitas Sriwijaya *batch* 2023, Badan Otonom FASCO tahun 2023, dan Badan Otonom INTEL tahun 2023 yang telah memberikan pengalaman luar biasa dan pembelajaran non-akademik bagi penulis.
19. Staf-staf Fasilkom Unsri yang telah memberikan bantuan dalam pelaksanaan kegiatan akademik dan kemahasiswaan.
20. Bapak Satpam Fasilkom yang senantiasa menjaga keamanan, serta staf kebersihan yang selalu memastikan kebersihan lingkungan Fasilkom.

21. Tanjung Senai, Ogan Ilir dan Perpustakaan Unsri yang telah menjadi saksi bagi penulis mencari inspirasi, serta menyediakan ruang bagi penulis untuk berkarya dan menyelesaikan berbagai tugas.

22. DAMRI dan bus kaleng, yang pernah menjadi kendaraan bagi penulis dalam menempuh jarak antara Indralaya dan Palembang, turut mewarnai perjalanan akademik penulis hingga akhirnya menyelesaikan studi di Universitas Sriwijaya.

Akhir kata, penulis mengharapkan semoga karya ini dapat memberikan manfaat bagi dunia akademik, masyarakat, serta kemajuan ilmu pengetahuan. Penulis menyadari bahwa masih banyak kekurangan dalam karya ini dan dengan rendah hati penulis mengharapkan saran serta kritik yang membangun untuk perbaikan di masa mendatang. Penulis juga berharap karya ini menjadi amal jariyah bagi semua pihak yang telah mendukung, membantu, dan memberikan doa selama perjalanan akademik ini.

Palembang, 30 Desember 2024

Penulis,



Elven Constancio
NIM. 09031382126169

PENERAPAN METODE *SUPERVISED LEARNING* DAN TEKNIK *RESAMPLING* UNTUK PREDIKSI PENIPUAN TRANSAKSI KEUANGAN

Oleh:

Elven Constancio

NIM. 09031382126169

ABSTRAK

Penipuan transaksi keuangan dapat mengakibatkan konsekuensi yang menghancurkan bagi stabilitas perusahaan, serta kerugian besar bagi pemegang saham, industri, dan bahkan pasar secara keseluruhan. Seiring dengan meningkatnya penipuan dalam transaksi keuangan, dibutuhkan metode yang efektif untuk mendeteksi dan mencegah aktivitas penipuan secara akurat. Penelitian ini bertujuan untuk membandingkan kinerja lima model pembelajaran mesin, yaitu *Random Forest*, *K-Nearest Neighbors* (KNN), *Decision Tree*, *XGBoost*, dan *Extra Trees*, dalam mendeteksi penipuan transaksi keuangan yang menggunakan dataset tidak seimbang. Untuk mengatasi masalah ketidakseimbangan data, diterapkan tiga teknik *resampling*, yakni *Synthetic Minority Oversampling Technique* (SMOTE), *Adaptive Synthetic Sampling* (ADASYN), dan *Undersampling*. Eksperimen dilakukan dengan dua rasio pembagian data latih dan data uji, yaitu 70:30 dan 80:20. Hasil evaluasi menunjukkan bahwa model *XGBoost* adalah model yang paling konsisten, dengan nilai *ROC AUC* tertinggi yaitu 99%, terutama setelah penerapan teknik *resampling*. Rasio data 80:20 menghasilkan distribusi yang lebih seimbang serta performa model yang lebih baik dalam mendeteksi kelas minoritas terutama setelah teknik *resampling*. Hasil Penelitian ini menunjukkan bahwa model *XGBoost* dengan teknik *resampling* sangat efektif dalam mengatasi ketidakseimbangan data.

Kata Kunci: Penipuan Transaksi Keuangan, SMOTE, ADASYN, *Undersampling*, *XGBoost*

**APPLICATION OF SUPERVISED LEARNING METHOD AND
RESAMPLING TECHNIQUE FOR FINANCIAL TRANSACTION FRAUD
PREDICTION**

By:

Elven Constancio

NIM. 09031382126169

ABSTRACT

Financial transaction fraud can result in devastating consequences for the stability of companies, as well as huge losses for shareholders, the industry, and even the market as a whole. As fraud in financial transactions increases, there is a need for effective methods to accurately detect and prevent fraudulent activities. This study aims to compare the performance of five machine learning models, namely Random Forest, K-Nearest Neighbors (KNN), Decision Tree, XGBoost, and Extra Trees, in detecting financial transaction fraud using an imbalanced dataset. To overcome the data imbalance problem, three resampling techniques are applied, namely Synthetic Minority Oversampling Technique (SMOTE), Adaptive Synthetic Sampling (ADASYN), and Undersampling. Experiments were conducted with two training and test data sharing ratios, namely 70:30 and 80:20. The evaluation results showed that the XGBoost model was the most consistent, with the highest ROC AUC value of 99%, especially after the application of resampling techniques. The 80:20 data ratio resulted in a more balanced distribution and better model performance in detecting the minority class, particularly after resampling. This study concludes that the XGBoost model with resampling techniques is highly effective in addressing data imbalance.

Keywords: *Financial Transaction Fraud, SMOTE, ADASYN, Undersampling, XGBoost*

DAFTAR ISI

HALAMAN JUDUL	i
LEMBAR PENGESAHAN	ii
HALAMAN PERNYATAAN BEBAS PLAGIAT.....	iii
HALAMAN PERNYATAAN INTEGRITAS.....	iv
HALAMAN PERSETUJUAN	v
HALAMAN PERSEMBAHAN	vi
KATA PENGANTAR.....	vii
ABSTRAK	xiii
ABSTRACT	xiv
DAFTAR ISI.....	xv
DAFTAR GAMBAR	xvii
DAFTAR TABEL	xviii
DAFTAR LAMPIRAN	xix
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.....	5
BAB II TINJAUAN PUSTAKA.....	7
2.1 Penelitian Terdahulu	7
2.2 Dataset	9
2.3 <i>Preprocessing Data</i>	10
2.4 <i>Splitting Data</i>	11
2.5 <i>Resampling Data</i>	12
2.5.1 <i>Synthetic Minority Oversampling Technique Strategy (SMOTE)</i> ...	12
2.5.2 <i>Adaptive Synthetic Sampling (ADASYN)</i>	12
2.5.3 <i>Undersampling</i>	13
2.6 Pembangunan Model.....	13
2.6.1 <i>Random Forest</i>	13
2.6.2 <i>K-Nearest Neighbors (KNN)</i>	13
2.6.3 <i>Decision Tree</i>	14
2.6.4 <i>Extreme Gradient Boosting (XGBoost)</i>	14
2.6.5 <i>Extremely Randomized Trees (Extra Trees)</i>	14
2.7 Evaluasi Model.....	14
2.7.1 <i>Confusion Matrix</i>	15
2.7.2 <i>Accuracy</i>	15
2.7.3 <i>Precision</i>	15
2.7.4 <i>Recall</i>	16
2.7.5 <i>F1-Score</i>	16
2.7.6 <i>Area Under Curve (AUC)</i>	16
BAB III METODE PENELITIAN	17
3.1 Tahapan Penelitian.....	17
3.2 Dataset Penelitian	18

3.3	<i>Preprocessing Data</i>	20
3.4	<i>Splitting Data</i>	21
3.5	<i>Resampling Data</i>	21
3.5.1	<i>Resampling Data</i> Rasio 70:30.....	22
3.5.2	<i>Resampling Data</i> Rasio 80:20.....	22
3.6	Pembangunan Model	23
3.7	Evaluasi Model.....	23
BAB IV HASIL DAN PEMBAHASAN.....		25
4.1	Hasil Klasifikasi Sebelum <i>Resampling</i>	25
4.1.1	Rasio 70:30 Sebelum <i>Resampling</i>	25
4.1.2	Rasio 80:20 Sebelum <i>Resampling</i>	26
4.2	Hasil Klasifikasi Dengan <i>Resampling</i> SMOTE	27
4.2.1	Rasio 70:30 Dengan <i>Resampling</i> SMOTE.....	27
4.2.2	Rasio 80:20 Dengan <i>Resampling</i> SMOTE.....	28
4.3	Hasil Klasifikasi Dengan <i>Resampling</i> ADASYN	29
4.3.1	Rasio 70:30 Dengan <i>Resampling</i> ADASYN.....	30
4.3.2	Rasio 80:20 Dengan <i>Resampling</i> ADASYN.....	31
4.4	Hasil Klasifikasi Dengan <i>Resampling Undersampling</i>	32
4.4.1	Rasio 70:30 Dengan <i>Resampling Undersampling</i>	32
4.4.2	Rasio 80:20 Dengan <i>Resampling Undersampling</i>	33
4.5	Perbandingan Kinerja Model yang Telah Dievaluasi	34
4.5.1	Analisis Performa Model	34
4.5.2	Analisis Performa Berdasarkan Rasio Data	35
4.5.3	<i>Resampling</i> Terbaik.....	35
4.6	Rekapitulasi Kinerja Model.....	36
BAB V KESIMPULAN DAN SARAN		38
5.1	Kesimpulan.....	38
5.2	Saran	39
DAFTAR PUSTAKA.....		40
LAMPIRAN.....		44

DAFTAR GAMBAR

Gambar 3.1 Tahapan Penelitian	17
Gambar 3.2 Grafik Jumlah Data Awal <i>Fraud</i> dan <i>Non-Fraud</i>	18
Gambar 3.3 Grafik Jumlah Data Setelah Proses <i>Preprocessing</i>	20
Gambar 3.4 Grafik Jumlah Data Setelah <i>Resampling</i> (Rasio 70:30)	22
Gambar 3.5 Grafik Jumlah Data Setelah <i>Resampling</i> (Rasio 80:20)	23
Gambar 4.1 Grafik Performa Model Sebelum <i>Resampling</i> (Rasio 70:30)	26
Gambar 4.2 Grafik Performa Model Sebelum <i>Resampling</i> (Rasio 80:20)	27
Gambar 4.3 Grafik Performa Model Dengan SMOTE (Rasio 70:30)	28
Gambar 4.4 Grafik Performa Model Dengan SMOTE (Rasio 80:20)	29
Gambar 4.5 Grafik Performa Model Dengan ADASYN (Rasio 70:30)	30
Gambar 4.6 Grafik Performa Model Dengan ADASYN (Rasio 80:20)	31
Gambar 4.7 Grafik Performa Model Dengan <i>Undersampling</i> (Rasio 70:30)	33
Gambar 4.8 Grafik Performa Model Dengan <i>Undersampling</i> (Rasio 80:20)	34

DAFTAR TABEL

Tabel 2.1 <i>Confusion Matrix</i>	15
Tabel 3.1 Sampel Dataset	19
Tabel 4.1 Hasil Klasifikasi Sebelum <i>Resampling</i> (Rasio 70:30).....	25
Tabel 4.2 Hasil Klasifikasi Sebelum <i>Resampling</i> (Rasio 80:20).....	26
Tabel 4.3 Hasil Klasifikasi Dengan <i>Resampling</i> SMOTE (Rasio 70:30).....	28
Tabel 4.4 Hasil Klasifikasi Dengan <i>Resampling</i> SMOTE (Rasio 80:20).....	29
Tabel 4.5 Hasil Klasifikasi Dengan <i>Resampling</i> ADASYN (Rasio 70:30).....	30
Tabel 4.6 Hasil Klasifikasi Dengan <i>Resampling</i> ADASYN (Rasio 80:20).....	31
Tabel 4.7 Hasil Klasifikasi Dengan <i>Resampling Undersampling</i> (Rasio 70:30) .	32
Tabel 4.8 Hasil Klasifikasi Dengan <i>Resampling Undersampling</i> (Rasio 80:20) .	33
Tabel 4.9 Rekapitulasi Kinerja Model yang Telah Dievaluasi	36

DAFTAR LAMPIRAN

Lampiran 1 Surat Kesediaan Membimbing Tugas Akhir.....	A-1
Lampiran 2 Surat Keputusan Pembimbing Tugas Akhir.....	B-1
Lampiran 3 <i>Form Desk Evaluation</i> Seminar Proposal.....	C-1
Lampiran 4 <i>Form</i> Perbaikan Seminar Proposal	D-1
Lampiran 5 Bukti Publikasi Artikel	E-1
Lampiran 6 Lembar Rekomendasi Ujian Komprehensif	F-1
Lampiran 7 Kartu Konsultasi Tugas Akhir	G-1
Lampiran 8 Hasil Pengecekan <i>Similarity</i>	H-1
Lampiran 9 Surat Keterangan Pengecekan <i>Similarity</i>	I-1

BAB I

PENDAHULUAN

1.1 Latar Belakang

Masalah penipuan dalam transaksi keuangan terus meningkat dalam beberapa tahun terakhir, yang juga berdampak buruk pada kepercayaan di antara para pelaku bisnis dan pelaku pasar (Sun et al., 2023). Penipuan transaksi keuangan dapat mengakibatkan konsekuensi yang menghancurkan bagi stabilitas perusahaan, serta kerugian besar bagi pemegang saham, industri, dan bahkan pasar secara keseluruhan (Wen et al., 2022). Seiring dengan meningkatnya penipuan dalam transaksi keuangan, dibutuhkan metode yang efektif untuk mendeteksi dan mencegah aktivitas penipuan secara akurat. Salah satu masalah utama dalam mendeteksi penipuan transaksi keuangan adalah ketidakseimbangan data, di mana jumlah transaksi non penipuan secara signifikan jauh lebih tinggi dibandingkan dengan jumlah transaksi penipuan. Ketidakseimbangan ini dapat mengakibatkan model lebih cenderung memprediksi transaksi sah dan gagal mendeteksi transaksi penipuan dengan baik.

Pada lingkup pembelajaran mesin, ketidakseimbangan data dapat diatasi dengan teknik *resampling* seperti *oversampling* data menggunakan *Synthetic Minority Oversampling Technique* (SMOTE) yang diterapkan untuk meningkatkan proporsi transaksi penipuan dalam dataset, sehingga memungkinkan model untuk mempelajari pola dari data yang lebih proporsional (Zhao & Bai, 2022). Selain *oversampling* SMOTE, teknik *Adaptive Synthetic Sampling* (ADASYN) juga diterapkan pada penelitian ini. ADASYN memanfaatkan distribusi kepadatan sebagai dasar untuk secara otomatis menetapkan jumlah sampel sintetik yang

diperlukan bagi setiap contoh data minoritas (Xu et al., 2020). Selain *oversampling*, juga digunakan teknik *undersampling* yaitu dengan menghapus sampel dari kelas mayoritas (Zhou & Sun, 2024).

Berdasarkan permasalahan tersebut dan seiring dengan meningkatnya masalah ketidakseimbangan data dalam deteksi penipuan transaksi keuangan, penulis melakukan penelitian dengan menggunakan berbagai model pembelajaran mesin. Beberapa model pembelajaran mesin yang digunakan dalam penelitian ini antara lain *Random Forest*, *K-Nearest Neighbors* (KNN), *Decision Tree*, *XGBoost*, dan *Extra Trees* (*Extremely Randomized Trees*). Model-model ini dipilih karena masing-masing memiliki kemampuan untuk menangani masalah klasifikasi dan ketidakseimbangan data. *Random Forest*, model *ensemble* berbasis pohon keputusan, yang terinspirasi oleh ruang acak dan pemilihan pemisahan acak (Bai et al., 2022). Algoritma KNN menghasilkan prediksi model dengan mengambil suara terbanyak dari K titik data terdekat dengan titik uji sebagai dasar umpan balik (Zhang, 2020). *Decision tree* adalah suatu metode yang digunakan dalam analisis data untuk melakukan klasifikasi dan regresi (Ghiasi & Zendejboudi, 2021). *XGBoost* dipilih karena kemampuannya dalam menangani data yang tidak seimbang dan *overfitting*. *XGBoost* termasuk dalam kategori pembelajaran *ensemble* yang menggabungkan beberapa pohon keputusan (*decision trees*) untuk menghasilkan prediksi yang lebih baik (Amjad et al., 2022). Terakhir, *Extra Trees* dikenal karena kecepatan dan ketepatan, terutama dalam menangani kumpulan data yang sangat besar (Saeed et al., 2021). Hasil dari semua model ini akan dibandingkan untuk mengevaluasi efektivitas masing-masing dalam mendeteksi tindakan penipuan. Hasil ini dapat memberikan wawasan yang berharga untuk

memilih pendekatan yang paling sesuai untuk menangani masalah penipuan transaksi keuangan.

Merujuk pada permasalahan yang telah dijelaskan sebelumnya, penulis memutuskan untuk melakukan penelitian yang akan dibahas dalam laporan skripsi dengan judul **“PENERAPAN METODE *SUPERVISED LEARNING* DAN TEKNIK *RESAMPLING* UNTUK PREDIKSI PENIPUAN TRANSAKSI KEUANGAN”**.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah disampaikan, rumusan masalah untuk penelitian ini dapat dirumuskan sebagai berikut:

1. Model pembelajaran mesin mana yang paling efektif di antara *Random Forest*, *K-Nearest Neighbors*, *Decision Tree*, *Extra Trees*, dan *XGBoost* dalam mendeteksi penipuan transaksi keuangan?
2. Bagaimana pengaruh teknik *resampling* SMOTE, ADASYN, dan *Undersampling* dalam mengatasi ketidakseimbangan data pada dataset transaksi keuangan?
3. Bagaimana perbandingan performa model-model pembelajaran mesin setelah penerapan teknik *resampling* untuk mengatasi ketidakseimbangan pada dataset transaksi keuangan?
4. Di antara rasio pembagian data latih dan uji 70:30 dan 80:20, manakah yang menunjukkan efektivitas tertinggi dalam mendeteksi kelas minoritas pada dataset transaksi keuangan?

1.3 Tujuan Penelitian

Adapun tujuan yang ingin dicapai dalam penelitian ini adalah sebagai berikut:

1. Menguji dan membandingkan kinerja lima model pembelajaran mesin dalam mendeteksi penipuan transaksi keuangan pada dataset tidak seimbang.
2. Menguji dan membandingkan efektivitas teknik *resampling* SMOTE, ADASYN, dan *Undersampling* dalam meningkatkan kinerja model pembelajaran mesin.
3. Menganalisis perbandingan performa model-model pembelajaran mesin setelah penerapan teknik *resampling* untuk mengatasi ketidakseimbangan data.
4. Menentukan rasio data latih dan uji yang paling efektif antara 70:30 dan 80:20 untuk mendeteksi kelas minoritas dalam dataset transaksi keuangan.

1.4 Manfaat Penelitian

Adapun manfaat yang dapat diperoleh dari penelitian ini adalah sebagai berikut:

1. Memberikan wawasan mengenai efektivitas berbagai model pembelajaran mesin dan teknik *resampling* dalam mendeteksi penipuan pada dataset transaksi keuangan yang tidak seimbang.
2. Mengidentifikasi kombinasi model dan teknik *resampling* yang paling efektif untuk meningkatkan akurasi deteksi kelas minoritas dan dalam menangani masalah ketidakseimbangan data.

3. Memberikan pemahaman mendalam tentang bagaimana ketidakseimbangan data memengaruhi kinerja model pembelajaran mesin dan cara mengatasinya.
4. Menjadi referensi untuk penelitian selanjutnya dalam pengembangan metode deteksi penipuan dengan memanfaatkan teknik *resampling* dan evaluasi model yang lebih baik.
5. Membantu organisasi atau institusi keuangan dalam memilih model dan teknik yang tepat untuk meningkatkan akurasi dalam mendeteksi penipuan dan mengurangi risiko finansial.

1.5 Batasan Masalah

Dalam penelitian ini, terdapat beberapa batasan yang ditetapkan untuk memperjelas ruang lingkup penelitian, yaitu sebagai berikut:

1. Dataset yang digunakan dalam penelitian ini adalah *Synthetic Financial Datasets* yang diambil dari *website* Kaggle sehingga hasil penelitian ini mungkin tidak dapat digeneralisasi untuk dataset lain dengan karakteristik yang berbeda.
2. Penelitian ini hanya membandingkan lima model pembelajaran mesin, yaitu *Random Forest*, *K-Nearest Neighbors (KNN)*, *Decision Tree*, *XGBoost*, dan *Extra Trees*.
3. Teknik *resampling* yang digunakan pada penelitian ini adalah *Synthetic Minority Oversampling Technique (SMOTE)*, *Adaptive Synthetic Sampling (ADASYN)*, dan *Undersampling*.
4. Rasio pembagian data latih dan data uji yang dianalisis dalam penelitian ini adalah 70:30 dan 80:20.

5. Evaluasi performa model hanya dilakukan menggunakan beberapa metrik yakni *accuracy*, *precision*, *recall*, *F1-score*, dan *Area Under Curve (AUC)*.
6. Penelitian ini hanya mencakup pengujian model pembelajaran mesin dan teknik *resampling* untuk deteksi penipuan pada dataset transaksi keuangan yang tidak seimbang, tanpa mencakup implementasi nyata, analisis dampak ekonomi atau pengujian pada sistem aktual.

DAFTAR PUSTAKA

- Ahmed, G., Er, M. J., Fareed, M. M. S., Zikria, S., Mahmood, S., He, J., Asad, M., Jilani, S. F., & Aslam, M. (2022). Dad-net: Classification of alzheimer's disease using adasyn oversampling technique and optimized neural network. *Molecules*, 27(20), 7085. <https://www.mdpi.com/1420-3049/27/20/7085>
- Amjad, M., Ahmad, I., Ahmad, M., Wróblewski, P., Kamiński, P., & Amjad, U. (2022). Prediction of pile bearing capacity using XGBoost algorithm: modeling and performance evaluation. *Applied Sciences*, 12(4), 2126. <https://www.mdpi.com/2076-3417/12/4/2126>
- Bai, J., Li, Y., Li, J., Yang, X., Jiang, Y., & Xia, S.-T. (2022). Multinomial random forest. *Pattern Recognition*, 122, 108331. <https://doi.org/https://doi.org/10.1016/j.patcog.2021.108331>
- Camacho, L., Douzas, G., & Bacao, F. (2022). Geometric SMOTE for regression. *Expert Systems with Applications*, 193, 116387. <https://doi.org/https://doi.org/10.1016/j.eswa.2021.116387>
- Charbuty, B., & Abdulazeez, A. (2021). Classification based on decision tree algorithm for machine learning. *Journal of Applied Science and Technology Trends*, 2(01), 20–28. <https://www.jastt.org/index.php/jasttpath/article/view/65/24>
- Dai, Q., Liu, J., & Liu, Y. (2022). Multi-granularity relabeled under-sampling algorithm for imbalanced data. *Applied Soft Computing*, 124, 109083. <https://doi.org/https://doi.org/10.1016/j.asoc.2022.109083>
- Deng, X., Shao, H., Shi, L., Wang, X., & Xie, T. (2020). A classification–detection approach of COVID-19 based on chest X-ray and CT by using keras pre-trained deep learning models. *Computer Modeling in Engineering & Sciences*, 125(2), 579–596. <https://www.ingentaconnect.com/contentone/tsp/cmesc/2020/00000125/0000002/art00006#>
- Febriady, M., Samsuryadi, S., & Rini, D. P. (2022). Klasifikasi Transaksi Penipuan Pada Kartu Kredit Menggunakan Metode Resampling Dan Pembelajaran Mesin. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 6(2), 1010–1016. <https://ejurnal.stmik-budidarma.ac.id/index.php/mib/article/view/3515/2649>
- Ghiasi, M. M., & Zendehboudi, S. (2021). Application of decision tree-based ensemble learning in the classification of breast cancer. *Computers in Biology and Medicine*, 128, 104089. <https://doi.org/https://doi.org/10.1016/j.compbiomed.2020.104089>
- Ghorbani, R., & Ghousi, R. (2020). Comparing Different Resampling Methods in Predicting Students' Performance Using Machine Learning Techniques. *IEEE Access*, 8, 67899–67911. <https://doi.org/10.1109/ACCESS.2020.2986809>

- Hasibuan, L., & Jannah, F. (2024). Deteksi Penipuan Kartu Kredit Menggunakan Support Vector Machine dengan Optimasi Grid Search dan Genetic Algorithm. *Building of Informatics, Technology and Science (BITS)*, 6(1). <https://doi.org/10.47065/bits.v6i1.5355>
- Izotova, A., & Valiullin, A. (2021). Comparison of Poisson process and machine learning algorithms approach for credit card fraud detection. *Procedia Computer Science*, 186, 721–726. <https://doi.org/https://doi.org/10.1016/j.procs.2021.04.214>
- Joseph, V. R. (2022). Optimal ratio for data splitting. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, 15(4), 531–538. <https://doi.org/https://doi.org/10.1002/sam.11583>
- Khushi, M., Shaukat, K., Alam, T. M., Hameed, I. A., Uddin, S., Luo, S., Yang, X., & Reyes, M. C. (2021). A Comparative Performance Analysis of Data Resampling Methods on Imbalance Medical Data. *IEEE Access*, 9, 109960–109975. <https://doi.org/10.1109/ACCESS.2021.3102399>
- Kurniabudi, K., Harris, A., & Veronica, V. (2022). Komparasi Performa Tree-Based Classifier Untuk Deteksi Anomali Pada Data Berdimensi Tinggi dan Tidak Seimbang. *Jurnal Media Informatika Budidarma*, 6(1), 370–377. <https://www.ejurnal.stmik-budidarma.ac.id/index.php/mib/article/view/3473/2431>
- Li, J., Guo, C., Lv, S., Xie, Q., & Zheng, X. (2024). Financial fraud detection for Chinese listed firms: Does managers' abnormal tone matter? *Emerging Markets Review*, 62, 101170. <https://doi.org/https://doi.org/10.1016/j.ememar.2024.101170>
- Liang, W., Luo, S., Zhao, G., & Wu, H. (2020). Predicting hard rock pillar stability using GBDT, XGBoost, and LightGBM algorithms. *Mathematics*, 8(5), 765. <https://www.mdpi.com/2227-7390/8/5/765>
- Maharana, K., Mondal, S., & Nemade, B. (2022). A review: Data pre-processing and data augmentation techniques. *Global Transitions Proceedings*, 3(1), 91–99. <https://doi.org/https://doi.org/10.1016/j.gltp.2022.04.020>
- Maseer, Z. K., Yusof, R., Bahaman, N., Mostafa, S. A., & Foozy, C. F. M. (2021). Benchmarking of Machine Learning for Anomaly Based Intrusion Detection Systems in the CICIDS2017 Dataset. *IEEE Access*, 9, 22351–22370. <https://doi.org/10.1109/ACCESS.2021.3056614>
- Neupane, D., & Seok, J. (2020). Bearing Fault Detection and Diagnosis Using Case Western Reserve University Dataset With Deep Learning Approaches: A Review. *IEEE Access*, 8, 93155–93178. <https://doi.org/10.1109/ACCESS.2020.2990528>

- Priatna, W. (2024). Dampak Pengambilan Sampel Data untuk Optimalisasi Data tidak seimbang pada Klasifikasi Penipuan Transaksi E-Commerce. *Indonesian Journal of Computer Science*, 13(2). <http://ijcs.net/ijcs/index.php/ijcs/article/view/3698>
- Pruneski, J. A., Williams, R. J., Nwachukwu, B. U., Ramkumar, P. N., Kiapour, A. M., Martin, R. K., Karlsson, J., & Pareek, A. (2022). The development and deployment of machine learning models. *Knee Surgery, Sports Traumatology, Arthroscopy*, 30(12), 3917–3923. <https://doi.org/10.1007/s00167-022-07155-4>
- Ramadina, A., & Tania, K. D. (2024). Knowledge Extraction of Gojek Application Review Using Aspect-based Sentiment Analysis. *The Indonesian Journal of Computer Science*, 13(3). <https://ejurnal.seminar-id.com/index.php/bits/article/view/5368/2982>
- Saeed, U., Jan, S. U., Lee, Y.-D., & Koo, I. (2021). Fault diagnosis based on extremely randomized trees in wireless sensor networks. *Reliability Engineering & System Safety*, 205, 107284. <https://doi.org/https://doi.org/10.1016/j.ress.2020.107284>
- Seyyedattar, M., Ghiasi, M. M., Zendehboudi, S., & Butt, S. (2020). Determination of bubble point pressure and oil formation volume factor: Extra trees compared with LSSVM-CSA hybrid and ANFIS models. *Fuel*, 269, 116834. <https://doi.org/https://doi.org/10.1016/j.fuel.2019.116834>
- Shokrzade, A., Ramezani, M., Akhlaghian Tab, F., & Abdulla Mohammad, M. (2021). A novel extreme learning machine based kNN classification method for dealing with big data. *Expert Systems with Applications*, 183, 115293. <https://doi.org/https://doi.org/10.1016/j.eswa.2021.115293>
- Sun, H., Li, J., & Zhu, X. (2023). Financial fraud detection based on the part-of-speech features of textual risk disclosures in financial reports. *Procedia Computer Science*, 221, 57–64. <https://doi.org/10.1016/j.procs.2023.07.009>
- Tang, S., Yuan, S., & Zhu, Y. (2020). Data Preprocessing Techniques in Convolutional Neural Network Based on Fault Diagnosis Towards Rotating Machinery. *IEEE Access*, 8, 149487–149496. <https://doi.org/10.1109/ACCESS.2020.3012182>
- Vaccari, I., Chiola, G., Aiello, M., Mongelli, M., & Cambiaso, E. (2020). MQTTset, a new dataset for machine learning techniques on MQTT. *Sensors*, 20(22), 6578. <https://www.mdpi.com/1424-8220/20/22/6578>
- Vrigazova, B. (2021). The proportion for splitting data into training and test set for the bootstrap in classification problems. *Business Systems Research: International Journal of the Society for Advancing Innovation and Research in Economy*, 12(1), 228–242. <https://hrca.hrca.hr/ojs/index.php/bsr/article/view/17347/9337>

- Wen, S., Li, J., Zhu, X., & Liu, M. (2022). Analysis of financial fraud based on manager knowledge graph. *Procedia Computer Science*, 199, 773–779. <https://doi.org/https://doi.org/10.1016/j.procs.2022.01.096>
- Xu, T., Coco, G., & Neale, M. (2020). A predictive model of recreational water quality based on adaptive synthetic sampling algorithms and machine learning. *Water Research*, 177, 115788. <https://doi.org/https://doi.org/10.1016/j.watres.2020.115788>
- Zamachsari, F., & Puspitasari, N. (2021). Penerapan Deep Learning dalam Deteksi Penipuan Transaksi Keuangan Secara Elektronik. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 5(2), 203–212. <https://jurnal.iaii.or.id/index.php/RESTI/article/view/2952/391>
- Zhang, S. (2020). Cost-sensitive KNN classification. *Neurocomputing*, 391, 234–242. <https://doi.org/https://doi.org/10.1016/j.neucom.2018.11.101>
- Zhao, Z., & Bai, T. (2022). Financial Fraud Detection and Prediction in Listed Companies Using SMOTE and Machine Learning Algorithms. *Entropy*, 24(8), 1157. <https://www.mdpi.com/1099-4300/24/8/1157>
- Zhou, Q., & Sun, B. (2024). Adaptive K-means clustering based under-sampling methods to solve the class imbalance problem. *Data and Information Management*, 8(3), 100064. <https://doi.org/https://doi.org/10.1016/j.dim.2023.100064>