

**KLASIFIKASI PENIPUAN TRANSAKSI E-WALLET
MENGUNAKAN ALGORITMA RANDOM FOREST
DENGAN OPTIMASI RANDOMIZEDSEARCHCV DAN
TEKNIK SMOTEENN**

SKRIPSI

Diajukan untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi Informatika



disusun oleh

NI PUTU SUMERTIANI

22.11.4875

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2025

**KLASIFIKASI PENIPUAN TRANSAKSI E-WALLET
MENGUNAKAN ALGORITMA RANDOM FOREST
DENGAN OPTIMASI RANDOMIZEDSEARCHCV DAN
TEKNIK SMOTEENN**

SKRIPSI

untuk memenuhi salah satu syarat mencapai derajat Sarjana
Program Studi (*Informatika*)



disusun oleh

NIPUTU SUMERTIANI

22.11.4875

Kepada

**FAKULTAS ILMU KOMPUTER
UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2022

HALAMAN PERSETUJUAN

SKRIPSI

**KLASIFIKASI PENIPUAN TRANSAKSI E-WALLET
MENGUNAKAN ALGORITMA RANDOM FOREST
DENGAN OPTIMASI RANDOMIZEDSEARCHCV DAN
TEKNIK SMOTEENN**

yang disusun dan diajukan oleh

Ni Putu Sumertiani

22.11.4875

telah disetujui oleh Dosen Pembimbing Skripsi
pada tanggal <tanggal ujian>

Dosen Pembimbing,

Nama Dosen Pembimbing
NIK. 19030xxxx

HALAMAN PENGESAHAN

SKRIPSI

**KLASIFIKASI PENIPUAN TRANSAKSI E-WALLET MENGGUNAKAN
ALGORITMA RANDOM FOREST DENGAN OPTIMASI
RANDOMIZEDSEARCHCV DAN TEKNIK SMOTEENN**

yang disusun dan diajukan oleh

Ni Putu Sumertiani

22.11.4875

Telah dipertahankan di depan Dewan Penguji
pada tanggal <tanggal ujian>

Susunan Dewan Penguji

Nama Penguji

Tanda Tangan

Nama dan Gelar Penguji 1

NIK. 190302xxx

Nama dan Gelar Penguji 2

NIK. 190302xxx

Nama dan Gelar Penguji 3

NIK. 190302xxx

Skripsi ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Sarjana Komputer
Tanggal < tanggal lulus ujian >

DEKAN FAKULTAS ILMU KOMPUTER

Prof. Dr. Kusrini, M.Kom.

NIK. 190302106

HALAMAN PERNYATAAN KEASLIAN SKRIPSI

Yang bertandatangan di bawah ini,

Nama mahasiswa : Ni Putu Sumertiani
NIM : 22.11.4875

Menyatakan bahwa Skripsi dengan judul berikut:

Tuliskan Judul Skripsi

Dosen Pembimbing : Nama Dosen dan Gelar

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik di Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Dosen Pembimbing.
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini.
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta.
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Perguruan Tinggi.

Yogyakarta, <tanggal lulus ujian skripsi>

Yang Menyatakan,

Meterai Asli
Rp 10.000,-

Nama Mahasiswa

HALAMAN PERSEMBAHAN

(Bila ada) Halaman ini berisi kepada siapa skripsi dipersembahkan. Ditulis dengan singkat, resmi, sederhana, tidak terlalu banyak, serta tidak menjurus ke penulisan informal sehingga mengurangi sifat resmi laporan ilmiah.

KATA PENGANTAR

Bagian ini berisi pernyataan resmi yang ingin disampaikan oleh penulis kepada pihak lain, misalnya ucapan terima kasih kepada Dosen Pembimbing, Tim Dosen Penguji, dan semua pihak yang terkait dalam penyelesaian skripsi termasuk orang tua dan penyandang dana.

Nama harus ditulis secara lengkap termasuk gelar akademik dan harus dihindari ucapan terima kasih kepada pihak yang tidak terkait. Bahasa yang digunakan harus mengikuti kaidah bahasa Indonesia yang baku.

Bagian ini tidak perlu dituliskan hal-hal yang bersifat ilmiah. Kata Pengantar diakhiri dengan mencantumkan kota dan tanggal penulisan diikuti di bawahnya dengan **kata “Penulis” tanpa perlu menyebutkan nama dan tanda tangan.**

Yogyakarta, <tanggal bulan tahun>

Penulis

DAFTAR ISI

(gunakan tools table of content pada menu references di Word)

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN	ii
HALAMAN PENGESAHAN	iii
HALAMAN PERNYATAAN KEASLIAN SKRIPSI	iv
HALAMAN PERSEMBAHAN	v
KATA PENGANTAR	vi
DAFTAR ISI	vii
DAFTAR TABEL	ix
DAFTAR GAMBAR	x
DAFTAR LAMPIRAN	xi
DAFTAR LAMBANG DAN SINGKATAN	xii
DAFTAR ISTILAH	xiii
INTISARI	xiv
ABSTRACT	xv
BAB I PENDAHULUAN	1
1.1	11.2
	Error! Bookmark not defined. 1.3
	31.4
	31.5
	31.6
	4BAB II TINJAUAN PUSTAKA
	4
2.1	62.2
	11BAB III METODE PENELITIAN
	9

3.1	143.2
	143.3
18BAB IV HASIL DAN PEMBAHASAN	
	11
BAB V PENUTUP	13
5.1	325.2
Error! Bookmark not defined. REFERENSI	
	14
LAMPIRAN	15

DAFTAR TABEL

Tabel 2.1. Perbandingan metode	10
Tabel 2.2. Rangkuman Tinjauan Pustaka	11

DAFTAR GAMBAR

Gambar 2.1. Skema Diagram	10
Gambar 2.2. Skema Diagram Alir	11

DAFTAR LAMPIRAN

Lampiran 1. Profil obyek Penelitian	10
Lampiran 2. Dokumentasi Penelitian	11

DAFTAR LAMBANG DAN SINGKATAN

Ω	Tahanan Listrik
μ	Konstanta gesekan
ANFIS	Adaptive Network Fuzzy Inference System
SVM	Support Vector Machines

DAFTAR ISTILAH

Vektor	besaran yang mempunyai arah
Eigen Value	akar akar persamaan

INTISARI

Intisari merupakan outline dari sebuah hasil penelitian/karya ilmiah/naskah/proyek resmi yang memerlukan deskripsi secara singkat. Intisari disusun dengan kalimat yang singkat, jelas, runtut, dan sistematis dan dapat menggambarkan isi laporan secara keseluruhan. Intisari disusun dalam bahasa Indonesia, **disusun menjadi 1 alinea, tidak lebih dari 1 halaman, berkisar antara 150-250 kata, diketik dengan jarak 1 spasi.**

Intisari Skripsi memuat masalah apa yang terjadi dan dampak dari masalah terhadap lingkungan. Metode apa yang dilakukan peneliti dalam menyelesaikan masalah? Bagaimana hasil akhir penelitian, dan siapa yang dapat memanfaatkan hasil penelitian ini. Jika disajikan dalam 3 Alinea (paragraph), maka alinea pertama dalam intisari berisi masalah penelitian dan dampak dari masalah tersebut. Alinea kedua berisi metode penelitian (langkah-langkah penyelesaian masalah). Alinea ketiga mengungkapkan hasil dari penelitian (secara singkat), kontribusi penelitian, dan siapa yang dapat memanfaatkan hasil penelitian tersebut. Jika belum mencapai 250 kata, dapat ditambahkan penelitian lebih lanjut yang dapat direkomendasikan.

Di bagian bawah intisari dituliskan kata-kata kunci, bisa berupa kata-kata penting dalam intisari atau kata yang sering muncul, berjumlah maksimal 5 (lima) kata.

Kata kunci: satu, dua, tiga, empat, lima.

ABSTRACT

Abstract merupakan hasil terjemahan Intisari dalam versi Bahasa Inggris.
Tata cara penulisan dan ketentuan bisa melihat bagian Intisari.

.

Keyword: one, two, three, Four, Five

BAB I

PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi informasi telah mengubah pola transaksi keuangan masyarakat, terutama dengan hadirnya electronic wallet (e-wallet) yang menawarkan kemudahan, kecepatan, dan efisiensi. Fenomena ini didukung oleh meningkatnya penetrasi internet, pertumbuhan pengguna smartphone, serta pergeseran perilaku konsumen yang mengutamakan transaksi praktis. Namun, kemajuan ini juga diiringi oleh meningkatnya kejahatan siber berupa penipuan transaksi e-wallet. Kasus-kasus seperti phishing, social engineering, hingga pemanfaatan celah keamanan aplikasi semakin sering terjadi dan menimbulkan kerugian finansial maupun psikologis bagi pengguna. [1]

Permasalahan utama dalam transaksi e-wallet adalah meningkatnya kompleksitas modus penipuan yang sulit dideteksi dengan metode tradisional. Sistem deteksi yang ada sering kali tidak mampu mengidentifikasi pola penipuan yang terus berkembang, sehingga banyak transaksi penipuan yang lolos dari pengawasan. Selain itu, ketidakseimbangan data antara transaksi normal dan penipuan membuat model deteksi cenderung bias terhadap kelas mayoritas, sehingga mengurangi sensitivitas dalam mendeteksi kasus penipuan. Kondisi ini berdampak pada menurunnya kepercayaan masyarakat terhadap keamanan transaksi digital.[1]

Sebagai upaya mengatasi masalah tersebut, berbagai metode machine learning telah diterapkan untuk mendeteksi transaksi penipuan secara lebih akurat. Salah satunya adalah algoritma Random Forest, yang dikenal efektif dalam mengolah data kompleks dengan menggabungkan hasil prediksi dari banyak pohon keputusan.[2] Agar performa model optimal, dilakukan optimasi hyperparameter menggunakan RandomizedSearchCV yang mampu menemukan parameter terbaik secara efisien. Pendekatan ini dapat meningkatkan akurasi deteksi tanpa membutuhkan waktu komputasi yang terlalu lama dibandingkan

metode pencarian parameter konvensional.

Di sisi lain, untuk mengatasi ketidakseimbangan data, teknik SMOTEENN diterapkan guna menambah jumlah sampel kelas minoritas sekaligus mengurangi noise pada data mayoritas. Kombinasi antara Synthetic Minority Over-sampling Technique (SMOTE) dan Edited Nearest Neighbours (ENN) ini menghasilkan distribusi data yang lebih seimbang, sehingga model dapat belajar secara lebih efektif membedakan antara transaksi normal dan penipuan. Dengan integrasi Random Forest, RandomizedSearchCV, dan SMOTEENN, diharapkan sistem deteksi dapat mencapai tingkat akurasi yang lebih baik serta mengurangi false positive yang merugikan pengguna dan penyedia layanan.[1]

Kesimpulannya, penipuan transaksi e-wallet merupakan permasalahan yang kompleks dan membutuhkan pendekatan teknologi yang adaptif. Penggunaan algoritma Random Forest dengan optimasi RandomizedSearchCV serta penanganan ketidakseimbangan data menggunakan SMOTEENN merupakan solusi yang menjanjikan untuk meningkatkan efektivitas deteksi penipuan. Penelitian ini diharapkan mampu memberikan kontribusi signifikan dalam menciptakan ekosistem keuangan digital yang lebih aman, terpercaya, dan berkelanjutan, sekaligus menjadi acuan strategis bagi pengembang sistem deteksi penipuan di masa mendatang.

1.2 Rumusan Masalah.

1. Bagaimana membangun model klasifikasi untuk mendeteksi penipuan pada transaksi e-wallet menggunakan algoritma Random Forest?
2. Bagaimana pengaruh teknik balancing data SMOTEENN terhadap peningkatan deteksi kelas minoritas (fraud)?
3. Sejauh mana penerapan PCA dan hyperparameter tuning mampu meningkatkan akurasi dan efisiensi model?

1.3 Batasan Masalah

1. Dataset yang digunakan adalah Online Payment Fraud Detection dari platform Kaggle.
2. Model klasifikasi dibatasi pada penggunaan algoritma Random Forest tanpa membandingkan dengan algoritma lain.
3. Teknik yang digunakan terbatas pada SMOTEENN untuk balancing data, PCA untuk reduksi dimensi, dan RandomizedSearchCV untuk tuning parameter.

1.4 Tujuan Penelitian

1. Membangun model prediksi penipuan transaksi e-wallet menggunakan algoritma Random Forest.
2. Mengimplementasikan teknik SMOTEENN untuk mengatasi ketidakseimbangan kelas pada dataset.
3. Meningkatkan performa model melalui reduksi dimensi (PCA) dan optimasi hyperparameter (RandomizedSearchCV).
- 4.

1.5 Manfaat Penelitian.

1. Secara teoritis.

Penelitian ini memberikan kontribusi terhadap pengembangan ilmu pengetahuan di bidang data science dan machine learning, khususnya dalam penerapan algoritma Random Forest untuk deteksi penipuan transaksi digital. Selain itu, studi ini memperkuat pemahaman tentang efektivitas kombinasi teknik balancing data (SMOTEENN), reduksi dimensi (PCA), dan optimasi hyperparameter dalam membangun model klasifikasi yang andal pada kasus data yang tidak seimbang.

2. Secara Praktis.

Secara praktis, hasil penelitian ini dapat dimanfaatkan oleh penyedia layanan e-wallet atau sistem pembayaran digital sebagai dasar pengembangan sistem deteksi penipuan secara otomatis dan real-time. Model yang dihasilkan dapat membantu dalam mengidentifikasi transaksi mencurigakan secara lebih akurat, sehingga meningkatkan keamanan dan kepercayaan pengguna terhadap layanan finansial digital.

1.6 Sistematika Penulisan

Sistematika penulisan dalam laporan skripsi ini adalah sebagai berikut:

- BAB I Pendahuluan.

Berisi latar belakang, rumusan masalah, batasan masalah, tujuan, manfaat penelitian, dan sistematika penulisan.

- BAB II Tinjauan Pustaka

Membahas studi - studi terdahulu yang mendukung penelitian ini. Serta teori-teori yang relevan seperti machine learning, klasifikasi, algoritma random forest, Teknik SMOTEEN, PCA dan hyperparameter tuning dan evaluasi Model guna memperkuat landasan teori dalam analisis deteksi penipuan pada transaksi e-wallet.

- BAB III Metodologi Penelitian.

Menguraikan tahapan penelitian mulai dari Indetifikasi, Studi Literatur, pengumpulan dan persiapan data, preprosesing, balancing data, seleksi fitur, reduksi dimensi, pemodelan hingga evaluasi model.

- BAB IV Hasil dan Pembahasan

Menyajikan hasil implementasi model deteksi penipuan dan analisis kinerja berdasarkan evaluasi matrik yang digunakan.

- BAB V Penutup.

Berisi Kesimpulan dari hasil penelitian serta saran untuk pengembangan penelitian selanjutnya.

BAB II

TINJAUAN PUSTAKA

2.1 Studi Literatur

Terdapat beberapa penelitian yang dapat menjadi acuan yang relevan terhadap topik yang dibahas dalam penelitian ini.

Penelitian yang dilakukan oleh [1] berjudul *Penerapan Deep Learning dalam Deteksi Penipuan Transaksi Keuangan Secara Elektronik* membahas pemanfaatan metode deep learning dalam mengidentifikasi aktivitas penipuan pada transaksi keuangan elektronik. Penelitian ini menggunakan beberapa model seperti Autoencoder, Extra Trees, dan Random Forest, dengan tahapan preprocessing data mencakup SMOTE untuk penyeimbangan data dan encoding pada fitur kategorikal. Hasil dari penelitian ini menunjukkan bahwa model deep learning mampu mencapai nilai AUC tertinggi sebesar 98,7%. Penelitian ini memiliki kesamaan dengan penelitian yang dilakukan, yaitu fokus pada deteksi penipuan transaksi keuangan. Perbedaannya adalah penelitian ini menggunakan pendekatan deep learning, sedangkan penelitian yang dilakukan menggunakan pendekatan machine learning konvensional dengan algoritma Random Forest yang dioptimasi.

Penelitian yang dilakukan oleh [2] yang berjudul *Analisis Sistem Pendeteksi Penipuan Transaksi Kartu Kredit dengan Algoritma Machine Learning* membahas penerapan algoritma klasifikasi seperti Random Forest, Logistic Regression, dan Decision Tree dalam menganalisis data transaksi kartu kredit. Penelitian ini melakukan proses validasi menggunakan k-fold cross-validation serta menerapkan oversampling untuk mengatasi ketidakseimbangan data. Hasil analisis menunjukkan bahwa algoritma Random Forest memberikan performa terbaik dengan akurasi mencapai 97,2%. Penelitian ini memiliki kesamaan dengan penelitian yang dilakukan berupa penggunaan Random Forest dan teknik balancing data. Perbedaannya terletak pada metode optimasi parameter dan tidak adanya integrasi PCA dalam proses reduksi dimensi.

Penelitian yang dilakukan oleh [4] berjudul *Performa Random Forest dan XGBoost pada Deteksi Penipuan E-Commerce Menggunakan Augmentasi Data CGAN* membahas perbandingan performa algoritma Random Forest dan XGBoost setelah dilakukan augmentasi data menggunakan metode CGAN (Conditional Generative Adversarial Network). Penelitian ini bertujuan untuk menangani ketidakseimbangan data pada transaksi e-commerce yang mengandung fraud. Hasil pengujian menunjukkan bahwa augmentasi data mampu meningkatkan akurasi dan recall model secara signifikan, dengan Random Forest mencapai akurasi 94,2% setelah augmentasi. Penelitian ini memiliki kesamaan dengan penelitian yang dilakukan berupa penggunaan Random Forest dan teknik penyeimbangan data. Perbedaannya adalah pada pendekatan augmentasi data yang menggunakan CGAN, sedangkan penelitian ini menggunakan metode SMOTEENN.

Penelitian yang dilakukan oleh [5] dengan judul *Identifying Credit Card Fraud in Illegal Transactions Using Random Forest and Decision Tree Algorithms* membahas implementasi dua algoritma, yaitu Random Forest dan Decision Tree, dalam mendeteksi transaksi ilegal pada kartu kredit. Penelitian ini menggunakan teknik seleksi fitur untuk meningkatkan performa dan efisiensi model. Hasil pengujian menunjukkan bahwa algoritma Decision Tree mampu mencapai akurasi hingga 98%, sedangkan Random Forest mendekati angka tersebut. Penelitian ini memiliki kesamaan berupa penggunaan algoritma Decision Tree dan Random Forest serta fokus pada transaksi keuangan. Perbedaannya, penelitian ini tidak menyertakan proses optimasi hyperparameter ataupun pengurangan dimensi data.

Penelitian yang dilakukan oleh [6] berjudul *Analisis Keamanan Data Terhadap Penggunaan E-Wallet Sebagai Alat Transaksi Digital untuk Mencegah Penipuan Online* menyoroti aspek keamanan dalam penggunaan e-wallet sebagai sarana transaksi digital. Fokus penelitian ini lebih kepada identifikasi faktor-faktor yang menyebabkan kerentanan terhadap penipuan online serta bagaimana

perlindungan data dapat diterapkan untuk mencegah aktivitas penipuan. Meskipun penelitian ini tidak menggunakan algoritma klasifikasi, namun memiliki relevansi dari sisi domain penelitian, yaitu pencegahan penipuan dalam transaksi digital. Perbedaanannya, penelitian ini bersifat deskriptif kualitatif dan tidak mengimplementasikan model prediksi berbasis machine learning.

Tabel 2.1 Keaslian Penelitian

No	Judul penelitian	Nama Penulis	Tahun Publikasi	Hasil Penelitian	Perbandingan Penelitian
1	Penerapan Deep Learning dalam Deteksi Penipuan Transaksi Keuangan Secara Elektronik	Zamachsari & Puspitasari	2021	Menggunakan Autoencoder dan ensemble model (Random Forest, Extra Trees) dengan AUC sebesar 98,7%.	Penelitian ini menggunakan pendekatan deep learning, berbeda dengan penelitian ini yang menggunakan Random Forest, PCA, dan SMOTEENN
2	Analisis Sistem Pendeteksi Penipuan Transaksi Kartu Kredit dengan Algoritma Machine Learning	Ningsih, Gusvarizon & Hermawan	2022	Random Forest memberikan akurasi 97,2% setelah oversampling.	Penelitian ini serupa dari segi algoritma, tetapi tidak mencakup PCA maupun optimasi hyperparameter seperti pada penelitian ini.
3	Deteksi Penipuan Kartu Kredit Menggunakan Metode Random Forest	Billah & Saputra	2024	Random Forest memberikan akurasi 96,5% dan F1-score 96%.	Penelitian ini fokus pada Random Forest, tetapi tidak mengoptimasi parameter atau menggunakan PCA seperti penelitian ini.
4	Performa Random	Tim Peneliti BITS	2024	CGAN digunakan	Berbeda pendekatan

	Forest dan XGBoost pada Deteksi Penipuan E-Commerce Menggunakan Augmentasi Data CGAN			untuk augmentasi data dan meningkatkan akurasi Random Forest dan XGBoost.	pada teknik penyeimbangan data. Penelitian ini menggunakan SMOTEENN, bukan CGAN.
5	Identifying Credit Card Fraud in Illegal Transactions	Werdiningsih et al.	2023	Decision Tree akurasi 98%, Random Forest mendekati 97,5%.	Penelitian ini tidak mencakup optimasi dan PCA seperti pada penelitian ini.
6	Analisis Keamanan Data terhadap Penggunaan E-Wallet untuk Mencegah Penipuan Online	KN, L. K. & Nasution, M. I. P.	2024	Menganalisis faktor keamanan pada e-wallet untuk mencegah penipuan online.	Pendekatan penelitian ini bersifat deskriptif, tidak menggunakan algoritma machine learning seperti penelitian ini.

2.2 Dasar Teori

2.1.1 Mechanice Learning.

Machine Learning (ML) adalah cabang dari kecerdasan buatan (Artificial Intelligence/AI) yang memungkinkan komputer atau sistem untuk belajar dari data tanpa harus diprogram secara eksplisit. Dalam konteks deteksi penipuan, ML digunakan untuk mengembangkan model prediktif yang mampu mengenali pola-pola mencurigakan atau abnormal dalam data transaksi. Sistem ML bekerja melalui proses pelatihan (training) dan pengujian (testing), di mana algoritma diberikan sejumlah data untuk membangun pemahaman terhadap pola historis, dan kemudian diuji untuk mengevaluasi kemampuannya dalam mengklasifikasikan data baru. Pendekatan ini jauh lebih fleksibel dan adaptif dibandingkan metode berbasis aturan (rule-based) tradisional, yang cenderung statis dan tidak mampu menangani variasi baru dari penipuan digital.

2.1.2 Klasifikasi.

Klasifikasi adalah salah satu teknik dalam supervised learning yang digunakan untuk memetakan input ke dalam satu atau beberapa kelas output yang telah ditentukan sebelumnya. Dalam penelitian ini, klasifikasi bertujuan untuk mengidentifikasi apakah suatu transaksi termasuk dalam kategori *penipuan* (fraud) atau *normal* (non-fraud). Algoritma klasifikasi bekerja dengan membangun model dari data latih yang terdiri atas atribut-atribut transaksi sebagai input dan label kelas sebagai target. Keberhasilan klasifikasi sangat bergantung pada kualitas fitur, proporsi distribusi kelas, dan parameter model yang digunakan.

2.1.3 Algoritma Random Forest.

Random Forest merupakan algoritma ensemble learning yang menggabungkan banyak decision tree (pohon keputusan) untuk meningkatkan akurasi prediksi dan mengurangi risiko overfitting. Setiap pohon dibangun dari subset acak data dan subset fitur, dan hasil akhir diputuskan berdasarkan voting mayoritas (untuk klasifikasi). Kelebihan Random Forest adalah kemampuannya

dalam menangani dataset berdimensi tinggi, mendeteksi fitur yang paling berpengaruh, dan memberikan hasil yang relatif stabil. Dalam konteks deteksi fraud, Random Forest sangat efektif karena dapat menangani data tidak seimbang dan relasi non-linear antar fitur.

2.1.4 Teknik SMOTEENN.

Ketidakseimbangan kelas (class imbalance) adalah tantangan besar dalam data fraud, karena data transaksi penipuan biasanya jauh lebih sedikit dibandingkan transaksi normal. Teknik SMOTEENN (Synthetic Minority Over-sampling Technique + Edited Nearest Neighbors) digunakan untuk mengatasi hal ini melalui dua pendekatan:

- **SMOTE** menambahkan data sintetis pada kelas minoritas berdasarkan interpolasi nilai tetangga terdekat, sehingga model memiliki lebih banyak sampel fraud untuk dipelajari.
- **ENN** membersihkan data dengan menghapus sampel dari kelas mayoritas yang sulit diklasifikasikan dan cenderung menghasilkan noise.

Kombinasi SMOTE dan ENN membantu menghasilkan distribusi data yang lebih seimbang dan bersih, meningkatkan akurasi klasifikasi pada kelas minoritas.

2.1.5 Principal Component Analysis(PCA).

Principal Component Analysis (PCA) adalah metode reduksi dimensi yang digunakan untuk mengurangi kompleksitas data dengan mengubah fitur asli menjadi sekumpulan fitur baru yang disebut principal components. Komponen ini dibentuk sedemikian rupa sehingga mempertahankan sebagian besar variansi dalam data, dan saling ortogonal satu sama lain. Dengan menerapkan PCA, dimensi data dapat dikurangi tanpa kehilangan informasi yang penting, sehingga dapat mempercepat proses pelatihan dan mengurangi risiko overfitting. Dalam penelitian ini, PCA digunakan untuk mempertahankan 95% variansi kumulatif, yang berarti sebagian besar informasi tetap terjaga dalam bentuk data yang lebih ringkas.

2.1.6 Hyperparameter Tuning (RandomizedSearchCV)

Dalam algoritma machine learning, *hyperparameter* adalah parameter eksternal model yang nilainya tidak ditentukan secara otomatis dalam proses pelatihan, melainkan harus diatur terlebih dahulu. Contohnya dalam Random Forest, hyperparameter seperti jumlah pohon (*n_estimators*), kedalaman maksimum pohon (*max_depth*), dan jumlah fitur yang dipilih (*max_features*) sangat memengaruhi performa model. RandomizedSearchCV adalah teknik pencarian hyperparameter yang dilakukan secara acak dengan validasi silang (cross-validation) untuk menemukan kombinasi parameter terbaik. Pendekatan ini lebih efisien dibandingkan GridSearchCV, terutama saat ruang pencarian parameter sangat besar.

2.1.7 Evaluasi Model

Evaluasi performa model klasifikasi sangat penting untuk menilai sejauh mana model mampu membedakan antara transaksi penipuan dan non-penipuan. Beberapa metrik evaluasi yang digunakan dalam penelitian ini meliputi:

- **Accuracy:** Persentase prediksi benar terhadap total data.
- **Precision:** Rasio transaksi yang diprediksi fraud dan benar-benar fraud terhadap semua prediksi fraud.
- **Recall (Sensitivity):** Kemampuan model mendeteksi transaksi fraud dari seluruh transaksi fraud yang sebenarnya.
- **F1-Score:** Harmonik rata-rata antara precision dan recall, memberikan ukuran seimbang.
- **ROC-AUC (Receiver Operating Characteristic - Area Under Curve):** Mengukur kemampuan model dalam membedakan kedua kelas.
- **Confusion Matrix:** Menyajikan ringkasan hasil klasifikasi dalam bentuk jumlah prediksi benar dan salah dari masing-masing kelas.

BAB III METODE PENELITIAN

3.1 Objek Penelitian

Penelitian ini menggunakan dataset Online Payment Fraud Detection yang diunduh dari platform Kaggle. Dataset ini berisi 636.262 transaksi dengan beberapa fitur numerik dan kategorikal, di antaranya:

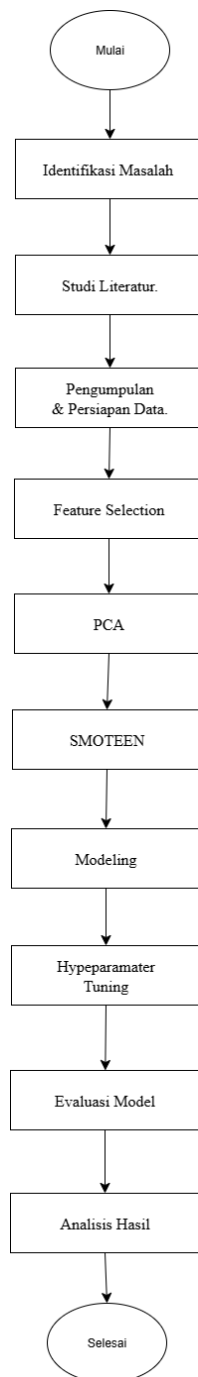
- amount: jumlah transaksi,
- oldbalanceOrg & newbalanceOrg: saldo awal & akhir pengirim,
- oldbalanceDest & newbalanceDest: saldo awal & akhir penerima,
- type: jenis transaksi (transfer, cash-out, payment, dll),
- isFraud: label target (0 = normal, 1 = penipuan).

Dataset bersifat imbalanced, dengan transaksi penipuan hanya sekitar 0,13% dari total data.

3.2 Alur Penelitian

Alur penelitian ini menggambarkan langkah-langkah sistematis yang dilakukan dalam membangun model klasifikasi penipuan transaksi e-wallet berdasarkan data atribut transaksi menggunakan algoritma Random Forest.

Contoh Gambar Alur Penelitian:



Gambar 3.1 Alur Penelitian

3.2.1 Identifikasi Masalah

Menganalisis maraknya penipuan dalam transaksi e-wallet dan keterbatasan metode deteksi berbasis aturan menjadi tantangan utama dalam memastikan keamanan sistem pembayaran digital.

3.2.2 Studi Literatur.

Mengumpulkan teori dan hasil penelitian terdahulu terkait deteksi penipuan menjadi langkah penting dalam membangun fondasi ilmiah penelitian ini. Fokus kajian diarahkan pada penggunaan algoritma machine learning dalam mendeteksi transaksi e-wallet yang mencurigakan, pemanfaatan teknik penyeimbangan data seperti SMOTEENN untuk mengatasi ketimpangan kelas, penerapan Principal Component Analysis (PCA) dalam reduksi dimensi, serta strategi optimasi model melalui hyperparameter tuning. Kajian ini memberikan gambaran menyeluruh terhadap pendekatan yang telah terbukti efektif serta potensi pengembangan metode yang lebih adaptif dan akurat.

3.2.3 Pengumpulan & Persiapan Data.

Dataset diunduh dari Kaggle, kemudian dilakukan:

- Data Cleaning : Menghapus kolom non-informatif(nameOrig, nameDest).
- Encoding: Variabel kategorikal type diubah menggunakan one-hot encoding.
- Scaling: Normalisasi data numerik dengan StandarScaler.
- Spiliting: Membagi data menjadi 80% data latih dan 20% data uji.

3.2.4 Feature Selection.

Tahap ini menggunakan metode *SelectFromModel* berbasis algoritma Random Forest untuk memilih dan mempertahankan fitur yang memiliki kontribusi paling signifikan terhadap prediksi, sehingga meningkatkan efisiensi dan akurasi model.

3.2.5 PCA.

Tahap ini menerapkan Principal Component Analysis (PCA) untuk mereduksi jumlah fitur pada data dengan tetap mempertahankan 95% variansi kumulatif. Tujuannya adalah menyederhanakan struktur data, mengurangi risiko

3.3.6 SMOTEEN

Untuk mengatasi ketidakseimbangan kelas pada data, digunakan metode SMOTEENN yang menggabungkan dua pendekatan: SMOTE menambahkan sampel sintetis pada kelas minoritas (fraud), sementara ENN menghapus sampel mayoritas yang sulit diklasifikasikan guna mengurangi noise dan meningkatkan performa model.

3.2.7 Modeling.

Pada tahap ini, model awal dibangun menggunakan algoritma Random Forest dengan parameter default. Algoritma ini dipilih karena kemampuannya dalam menangani data yang kompleks dan tidak seimbang, serta keunggulannya dalam menghasilkan performa yang stabil melalui teknik ensemble dari beberapa pohon keputusan. Model awal ini berfungsi sebagai baseline untuk dibandingkan dengan model yang telah dioptimasi.

3.2.8 Hyperparameter Tuning

Dilakukan `RandomizedSearchCV` untuk menemukan parameter optimal dengan kombinasi:

- `n_estimators`: [100, 200, 300, 500]
- `max_depth`: [10, 20, 30, None]
- `max_features`: ['sqrt', 'log2']

- min_samples_split: [2, 5, 10]
- min_samples_leaf: [1, 2, 4]

Proses tuning menggunakan 5-fold cross-validation untuk meningkatkan generalisasi model.

3.2.9 Evaluasi Model.

Evaluasi model dilakukan dengan menggunakan beberapa metrik kinerja utama, yaitu Accuracy, Precision, Recall, F1-score, ROC-AUC, dan Confusion Matrix. Metrik-metrik ini digunakan untuk menilai kemampuan model dalam mendeteksi transaksi penipuan secara akurat, khususnya dalam konteks data yang tidak seimbang, guna memastikan bahwa model tidak hanya fokus pada kelas mayoritas (non-fraud) saja.

3.2.10 Analisis Hasil.

Menganalisis perbandingan performa model sebelum & sesudah tuning, termasuk dampak penggunaan balancing SMOTEENN terhadap peningkatan deteksi kelas minoritas.

3.3 Alat dan Bahan.

- **Data Penelitian**

Data yang digunakan adalah Online Payment Fraud Detection Dataset dari Kaggle, dengan jumlah lebih dari 600.000 entri transaksi, termasuk 1-2% transaksi yang dikategorikan sebagai penipuan (isFraud = 1).

- **Alat dan Perangkat lunak.**

- **Bahasa Pemrograman:** Python 3.10.
- **Lingkungan Pengembangan:** Google Colab.
- **Library:** Pandas, NumPy, Scikit-learn, Imbalanced-learn, Matplotlib, Seaborn, Joblib.
- **Dataset:** Online Payment Fraud Detection dari Kaggle

BAB IV

HASIL DAN PEMBAHASAN

4.1 Deskripsi Dataset.

Penelitian ini menggunakan dataset *Online Payment Fraud Detection* yang bersumber dari Kaggle, berisi 636.262 data transaksi keuangan digital. Untuk efisiensi komputasi, digunakan sampel sebanyak 500.000 transaksi. Setiap transaksi dilabeli dengan *isFraud*, yaitu 0 untuk transaksi normal dan 1 untuk transaksi penipuan. Distribusi data sangat tidak seimbang, di mana hanya sekitar 0,13% transaksi tergolong fraud.

Fitur-fitur yang digunakan mencakup jumlah transaksi (*amount*), saldo awal dan akhir dari pengirim (*oldbalanceOrg*, *newbalanceOrg*), saldo awal dan akhir dari penerima (*oldbalanceDest*, *newbalanceDest*), serta jenis transaksi (*type*). Fitur-fitur ini menjadi dasar dalam proses klasifikasi untuk mendeteksi adanya indikasi penipuan.

4.2 Exploratory Data Analysis (EDA)

Pada tahapan ini, dilakukan eksplorasi awal untuk memahami distribusi target dan hubungan antar fitur. Hasil analisis menunjukkan bahwa data sangat tidak seimbang, dengan dominasi transaksi non-fraud. Heatmap korelasi mengidentifikasi fitur seperti *amount*, *oldbalanceOrg*, serta jenis transaksi *TRANSFER* dan *CASH_OUT* memiliki korelasi kuat terhadap fraud. Temuan ini menjadi dasar penting dalam proses seleksi fitur, penyeimbangan data, dan reduksi dimensi di tahap selanjutnya.

4.3 Preprocessing Data.

Tahap preprocessing dilakukan untuk menyiapkan data agar sesuai dengan kebutuhan algoritma machine learning serta meningkatkan kualitas data yang akan digunakan dalam pelatihan model. Proses ini sangat penting untuk memastikan bahwa data bersih, terstruktur, dan dapat diolah secara optimal.

Adapun tahapan preprocessing yang dilakukan meliputi:

1. Data Cleaning : Menghapus kolom non-informatif (nameOrig, nameDest)
2. Encoding: Variabel kategorikal type diubah menjadi one-hot encoding.
3. Scaling : Normalisasi data numerik menggunakan StandardScaler.
4. Split Data: Pembagian dataset menjadi 30% data latih dan 20 % data uji.

4.4. Penyeimbangan data(SMOTEEN)

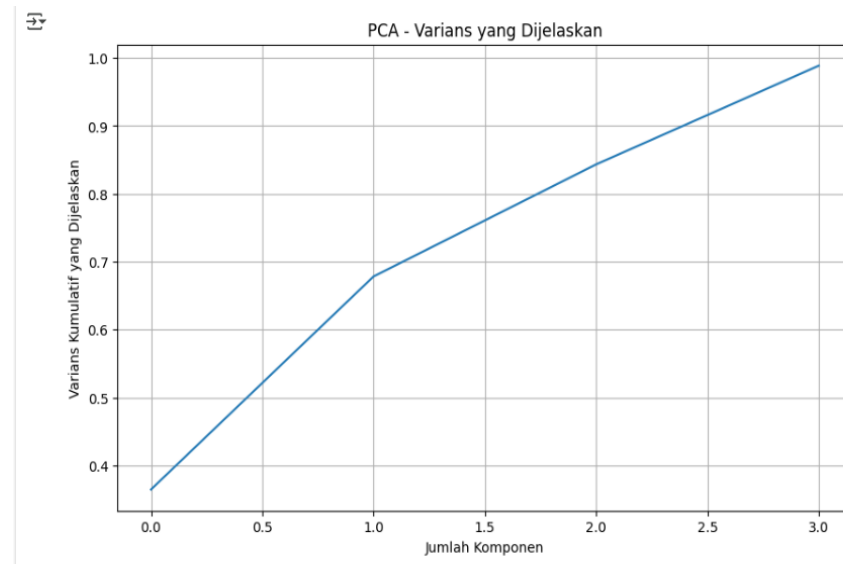
Tahapan ini, penyeimbangan data dilakukan menggunakan teknik SMOTEENN, yaitu gabungan dari Synthetic Minority Over-sampling Technique (SMOTE) dan Edited Nearest Neighbors (ENN). SMOTE digunakan untuk menambahkan data sintetis pada kelas minoritas (penipuan), sementara ENN berfungsi untuk membersihkan data yang salah klasifikasi pada kelas mayoritas (normal). Kombinasi kedua teknik ini mampu menangani permasalahan ketidakseimbangan kelas secara lebih efektif, sehingga model yang dibangun dapat belajar dengan lebih baik dan tidak bias terhadap kelas mayoritas.

4.5. Feature Selection.

Pada tahap ini, dilakukan seleksi fitur menggunakan metode SelectFromModel berbasis algoritma Random Forest. Dari seluruh fitur yang tersedia, terpilih 7 fitur utama yang memiliki kontribusi paling signifikan terhadap prediksi fraud, yaitu: step, amount, oldbalanceOrg, newbalanceOrig, oldbalanceDest, newbalanceDest, dan type_TRANSFER. Fitur-fitur ini akan digunakan pada proses pemodelan selanjutnya untuk meningkatkan efisiensi dan akurasi.

4.6. PCA

Pada tahap ini, dilakukan reduksi dimensi menggunakan Principal Component Analysis (PCA) dengan ambang variansi sebesar 95%. Hasilnya, hanya 4 komponen utama yang diperlukan untuk menjelaskan sebagian besar informasi dalam data, sehingga proses pelatihan model menjadi lebih efisien tanpa kehilangan makna signifikan dari fitur asli.



(Gambar 4.1: Grafik PCA Explained Variance).

4.7 Modeling.

Model awal dibangun menggunakan algoritma Random Forest dengan parameter tertentu dan data latih yang telah diseimbangkan serta direduksi menggunakan PCA. Hasil evaluasi menunjukkan performa yang cukup baik dengan nilai precision sebesar 0,82, recall 0,74, F1-score 0,77, dan ROC AUC sebesar 0,93, yang menandakan kemampuan model dalam mendeteksi transaksi penipuan sudah cukup optimal sebelum dilakukan tuning lebih lanjut.

4.8 Hyperparameter Tuning dengan RandomizedSearchCV.

Untuk meningkatkan performa model, dilakukan proses optimasi hyperparameter menggunakan RandomizedSearchCV. Metode ini dipilih

karena lebih efisien secara komputasi dibanding GridSearchCV, dengan tetap mampu mengeksplorasi kombinasi parameter secara acak dan luas. Proses tuning menggunakan teknik *5-fold cross-validation* untuk memastikan generalisasi model yang baik.

Adapun parameter yang dioptimasi meliputi:

- `n_estimators` : [100, 200, 300, 500].
- `max_depth` : [10, 20, 30, None]
- `max_features` : ['sqrt', 'log2']
- `min_samples_split` : [2, 5, 10]
- `min_samples_leaf` : [1, 2, 4]

Proses tuning menggunakan 5-fold cross-validation. Parameter terbaik menghasilkan kombinasi :

- `n_estimator` = 300
- `max_depth` = 20
- `max_features` = 'sqrt'
- `min_samples_split` = 5
- `min_samples_leaf` = 2

Dari hasil tuning, diperoleh kombinasi parameter terbaik yang digunakan untuk membangun model akhir, yaitu:

- `n_estimators` = 300
- `max_depth` = 20
- `max_features` = 'sqrt'
- `min_samples_split` = 5

- `min_samples_leaf = 2`

Model hasil tuning ini selanjutnya dievaluasi untuk dibandingkan dengan model baseline.

4.9 Evaluasi Model.

Evaluasi model dilakukan untuk mengetahui seberapa baik algoritma Random Forest dalam mengklasifikasikan transaksi e-wallet sebagai aman, warning, atau fraud sebelum dan sesudah proses tuning parameter. Evaluasi dilakukan terhadap data uji menggunakan beberapa metrik utama, yaitu accuracy, precision, recall, F1-score, dan ROC-AUC. Seluruh model dibangun menggunakan fitur-fitur transaksi utama, seperti jumlah transaksi, saldo pengirim dan penerima sebelum dan sesudah transaksi, jenis transaksi, serta urutan transaksi.

Hasil evaluasi ditampilkan pada tabel berikut:

Model	Accuracy	Precision	Recall	F1-score	ROC-AUC
Random Forest (Default)	0.96	0.51	0.85	0.51	0.93
Random Forest (Tuning)	0.99	0.58	0.83	0.62	0.93

Berdasarkan tabel di atas, model Random Forest dengan parameter default memiliki akurasi sebesar 96% dengan precision dan F1-score yang relatif rendah (masing-masing 0,51). Meskipun recall cukup tinggi (0,85), ketidakseimbangan antara precision dan recall ini menunjukkan bahwa model cenderung bias terhadap kelas mayoritas (transaksi aman), sehingga kinerjanya dalam mendeteksi kelas fraud masih kurang optimal.

Setelah dilakukan hyperparameter tuning, model menunjukkan peningkatan signifikan pada precision (0,58) dan F1-score (0,62), sementara akurasi meningkat menjadi 99% dan ROC-AUC tetap stabil di 0,93. Hal ini mengindikasikan bahwa

model hasil tuning lebih seimbang dalam mengklasifikasikan transaksi, mampu mendeteksi transaksi fraud dengan lebih baik, dan mengurangi kesalahan klasifikasi pada kelas minoritas.

Dengan demikian, proses tuning tidak hanya meningkatkan performa prediksi pada kelas minoritas tetapi juga mengurangi potensi bias pada model. Model hasil tuning dianggap lebih stabil, generalizable, dan siap untuk diimplementasikan pada data transaksi nyata dengan variasi yang lebih kompleks.

4.10 Pembahasan Hasil.

Hasil evaluasi menunjukkan bahwa model Random Forest dengan parameter default memiliki akurasi sebesar 96%, dengan precision 0,51, recall 0,85, F1-score 0,51, dan ROC-AUC 0,93. Meskipun akurasinya tinggi, nilai precision dan F1-score yang relatif rendah menunjukkan bahwa model cenderung bias terhadap kelas mayoritas, sehingga kemampuan mendeteksi kelas minoritas (fraud) masih terbatas. Kondisi ini menggambarkan bahwa model default belum optimal dalam mengklasifikasikan transaksi dengan risiko tinggi secara akurat.

Setelah dilakukan hyperparameter tuning, model mengalami peningkatan performa, dengan akurasi mencapai 99%, precision meningkat menjadi 0,58, recall 0,83, F1-score 0,62, dan ROC-AUC tetap stabil di 0,93. Peningkatan precision dan F1-score ini menandakan bahwa model hasil tuning lebih mampu membedakan antara transaksi yang aman dan berisiko, sekaligus mengurangi kesalahan klasifikasi pada kelas fraud. Dengan demikian, proses tuning berhasil membentuk model yang lebih seimbang dan lebih baik dalam mendeteksi potensi kecurangan pada transaksi e-wallet.

4.11 Hasil Deploy.

Setelah proses pembangunan dan evaluasi model selesai, tahap selanjutnya adalah melakukan *deploy* model dalam bentuk aplikasi berbasis web. Tujuan dari

deployment ini adalah untuk memungkinkan pengguna secara langsung menguji dan memprediksi kemungkinan terjadinya penipuan pada transaksi e-wallet dengan memasukkan sejumlah nilai yang merepresentasikan karakteristik transaksi tersebut. Sistem dirancang agar dapat memberikan hasil prediksi secara instan, akurat, dan mudah diakses oleh pengguna. Aplikasi prediksi penipuan transaksi e-wallet yang dikembangkan memiliki tampilan antarmuka yang sederhana dan interaktif. Pada halaman utama aplikasi, pengguna diberikan tujuh kolom input yang mewakili tujuh fitur utama yang telah dipilih dalam proses seleksi fitur, yaitu:

- step,
- amount,
- oldbalanceOrg,
- newbalanceOrig,
- oldbalanceDest,
- newbalanceDest,
- type_TRANSFER.

Model yang telah dilatih dan dituning kemudian dikemas dan diintegrasikan ke dalam sebuah antarmuka berbasis web. Pengguna dapat mengisi masing-masing kolom yang tersedia dengan data transaksi, seperti jumlah transaksi, saldo pengirim dan penerima sebelum dan sesudah transaksi, jenis transaksi, serta urutan transaksi. Setelah semua data diinputkan, pengguna dapat menekan tombol "Prediksi" untuk melihat hasil prediksi tingkat keamanan transaksi e-wallet yang ditampilkan dalam tiga kategori, yaitu Aman, Warning, atau Fraud.

- Aman :

Prediksi Fraud Transaksi E-Wallet

Masukkan detail transaksi untuk memprediksi apakah transaksi **AMAN**, **WARNING**, atau **FRAUD**.

Generate Contoh Data

Contoh Aman

Contoh Warning

Contoh Fraud

Input Data Transaksi

Step (urutan transaksi)

1

- +

Saldo Pengirim Setelah Transaksi

5500,00

- +

Jumlah Transaksi

500,00

- +

Saldo Penerima Sebelum Transaksi

0,00

- +

Saldo Pengirim Sebelum Transaksi

6000,00

- +

Saldo Penerima Setelah Transaksi


500,00

- +

Jenis Transaksi

PAYMENT

▼

 Prediksi

 **AMAN**

Probabilitas Fraud:



- Warning :

Prediksi Fraud Transaksi E-Wallet

Masukkan detail transaksi untuk memprediksi apakah transaksi AMAN, WARNING, atau FRAUD.

Generate Contoh Data

Contoh Aman

Contoh Warning

Contoh Fraud

Input Data Transaksi

Step (urutan transaksi)	Saldo Pengirim Setelah Transaksi
<input type="text" value="2"/>	<input type="text" value="2000,00"/>
Jumlah Transaksi	Saldo Penerima Sebelum Transaksi
<input type="text" value="3000,00"/>	<input type="text" value="0,00"/>
Saldo Pengirim Sebelum Transaksi	Saldo Penerima Setelah Transaksi
<input type="text" value="5000,00"/>	<input type="text" value="3000,00"/>
Jenis Transaksi	
<input type="text" value="CASH_OUT"/>	
<input type="button" value="Prediksi"/>	

Warning

Probabilitas Fraud:



- Fraud.

🏠 Prediksi Fraud Transaksi E-Wallet

Masukkan detail transaksi untuk memprediksi apakah transaksi **AMAN**, **WARNING**, atau **FRAUD**.

🔄 Generate Contoh Data

Contoh Aman
Contoh Warning
Contoh Fraud

📝 Input Data Transaksi

<p>Step (urutan transaksi)</p> <div style="border: 1px solid #ccc; padding: 5px; display: flex; justify-content: space-between;"> 1 - + </div> <p>Jumlah Transaksi</p> <div style="border: 1px solid #ccc; padding: 5px; display: flex; justify-content: space-between;"> 50000,00 - + </div> <p>Saldo Pengirim Sebelum Transaksi</p> <div style="border: 1px solid #ccc; padding: 5px; display: flex; justify-content: space-between;"> 50000,00 - + </div>	<p>Saldo Pengirim Setelah Transaksi</p> <div style="border: 1px solid #ccc; padding: 5px; display: flex; justify-content: space-between;"> 0,00 - + </div> <p>Saldo Penerima Sebelum Transaksi</p> <div style="border: 1px solid #ccc; padding: 5px; display: flex; justify-content: space-between;"> 0,00 - + </div> <p>Saldo Penerima Setelah Transaksi</p> <div style="border: 1px solid #ccc; padding: 5px; display: flex; justify-content: space-between;"> 50000,00 - + </div>
---	---

Jenis Transaksi

TRANSFER
▼

🔍 Prediksi

🚨 Fraud

Probabilitas Fraud:

48.53%

Setelah pengguna memasukkan nilai pada masing-masing kolom input, aplikasi akan memproses data tersebut menggunakan model Random Forest yang telah disimpan (diserialisasi), lalu menampilkan hasil prediksi berupa tiga kemungkinan, yaitu:

- Transaksi Aman (ditampilkan dalam warna hijau)
- Transaksi Warning (ditampilkan dalam warna oranye)
- Transaksi Fraud (ditampilkan dalam warna merah)

Antarmuka aplikasi dilengkapi dengan tombol interaktif seperti:

- Prediksi untuk mengaktifkan fungsi prediksi.

- Generate Contoh Data untuk memuat data contoh yang telah disediakan (Aman, Warning, Fraud).
- Reset Input untuk mengulang pengisian data.

Melalui proses deployment ini, model yang telah dibangun dapat digunakan secara praktis dan memberikan pengalaman interaktif bagi pengguna dalam memprediksi tingkat keamanan transaksi e-wallet berbasis data masukan transaksi. Tampilan aplikasi yang sederhana serta kemudahan penggunaan membuat sistem ini dapat digunakan oleh siapa saja tanpa memerlukan pemahaman teknis yang kompleks.

Bagian ini merupakan halaman kedua dari bab IV. Bagian ini memiliki format penomoran halaman yang berbeda dengan halaman pertama.

BAB V

PENUTUP

5.1 Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa algoritma Random Forest mampu digunakan secara efektif untuk membangun sistem klasifikasi transaksi e-wallet ke dalam tiga kategori, yaitu Aman, Warning, dan Fraud, berdasarkan data atribut transaksi. Pengembangan model dilakukan melalui tahapan yang sistematis, dimulai dari pengumpulan dan pra-pemrosesan data, analisis eksploratif, penyeimbangan kelas menggunakan SMOTEENN, seleksi fitur berbasis feature importance, reduksi dimensi menggunakan PCA, pelatihan model, tuning parameter menggunakan RandomizedSearchCV, hingga evaluasi performa. Model yang dibangun menggunakan tujuh fitur utama, yaitu step, amount, oldbalanceOrg, newbalanceOrig, oldbalanceDest, newbalanceDest, dan type_TRANSFER, yang terbukti cukup mewakili karakteristik data transaksi serta memberikan informasi signifikan untuk mendeteksi indikasi penipuan.

Model awal dengan parameter default menunjukkan akurasi sebesar 96%, namun precision dan F1-score yang rendah mengindikasikan kecenderungan bias terhadap kelas mayoritas. Setelah dilakukan hyperparameter tuning, model menunjukkan peningkatan performa yang signifikan dengan akurasi sebesar 99%, precision 0,58, recall 0,83, F1-score 0,62, dan ROC-AUC 0,93. Peningkatan ini menegaskan bahwa proses tuning mampu menghasilkan model yang lebih seimbang, stabil, dan memiliki kemampuan generalisasi yang lebih baik dalam mendeteksi transaksi berisiko.

Sebagai hasil akhir, model yang telah dibangun dan dituning kemudian dideploy ke dalam aplikasi berbasis web interaktif yang dapat digunakan secara langsung oleh pengguna. Aplikasi ini memungkinkan pengguna untuk memasukkan nilai tujuh fitur transaksi dan memperoleh prediksi tingkat keamanan transaksi e-wallet secara instan dalam tiga kategori, yaitu Aman,

Warning, atau Fraud. Hasil prediksi yang dihasilkan oleh aplikasi telah sesuai dengan pola data, menunjukkan bahwa sistem klasifikasi yang dikembangkan tidak hanya akurat dari sisi akademis, tetapi juga siap diterapkan dalam konteks praktis, seperti sistem keamanan transaksi digital pada layanan keuangan berbasis e-wallet.

5.2 Saran.

Sebagai tindak lanjut dari penelitian ini, terdapat beberapa hal yang dapat dikembangkan untuk meningkatkan kinerja sistem deteksi penipuan transaksi e-wallet. Pertama, penelitian selanjutnya disarankan untuk menambahkan fitur kontekstual seperti informasi lokasi transaksi, perangkat yang digunakan, atau pola perilaku pengguna, yang berpotensi meningkatkan kemampuan model dalam mendeteksi penipuan yang lebih kompleks. Kedua, perlu dilakukan eksperimen dengan algoritma lain seperti Gradient Boosting, XGBoost, atau Deep Learning untuk memperoleh perbandingan kinerja dan menemukan pendekatan yang paling optimal. Ketiga, sistem yang dibangun pada penelitian ini masih berjalan secara batch, sehingga pengembangan model deteksi real-time menjadi langkah penting agar prediksi dapat dilakukan seketika saat transaksi berlangsung. Selain itu, pengujian pada data transaksi dari berbagai platform e-wallet juga diperlukan untuk memastikan model memiliki tingkat generalisasi yang baik pada berbagai skenario nyata. Dari sisi implementasi, penyempurnaan antarmuka aplikasi web agar lebih ramah pengguna serta penambahan fitur seperti riwayat prediksi dan pelaporan dapat menjadi pengembangan yang bermanfaat. Dengan adanya pengembangan lebih lanjut, diharapkan sistem ini mampu memberikan kontribusi yang lebih besar dalam meningkatkan keamanan transaksi digital serta mendukung upaya pencegahan penipuan secara efektif di berbagai layanan keuangan berbasis e-wallet.

REFERENSI

- [1] Kurniawan, G. W. M., & Wisety, U. N. (2025). Pendeteksian Penipuan Menggunakan Pendekatan Metode Klasifikasi Random Forest. *eProceedings of Engineering*, 12(1), 2216-2220.
- [2] Arditha, R., Anugerah, R., & Sutabri, T. (2025). Analisis Penerapan Machine Learning dan Algoritma Anomali untuk Deteksi Penipuan pada Transaksi Digital. *Repeater: Publikasi Teknik Informatika dan Jaringan*, 3(1), 80-90.
- [3] Zamachsari, F., & Puspitasari, N. (2021). Penerapan Deep Learning dalam Deteksi Penipuan Transaksi Keuangan Secara Elektronik. *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, 5(2), 203-212.
- [4] Ningsih, P. T. S., Gusvarizon, M., & Hermawan, R. (2022). Analisis Sistem Pendeteksi Penipuan Transaksi Kartu Kredit dengan Algoritma Machine Learning. *Jurnal Teknologi Informatika Dan Komputer*, 8(2), 386-401.
- [5] Sarmini, S., Sunardi, S., & Fadlil, A. (2024). Performa Random Forest dan XGBoost pada Deteksi Penipuan E-Commerce Menggunakan Augmentasi Data CGAN. *Building of Informatics, Technology and Science (BITS)*, 6(3), 1919–1931.
- [6] Werdiningsih, I., Purwanti, E., Aditya, G. R. W., Hidayat, A. R., Athallah, R. S. R., Sahar, V. A., ... & Somba, D. F. N. (2023). Identifying Credit Card Fraud in Illegal Transactions Using Random Forest and Decision Tree Algorithms. *Jurnal Sisfokom (Sistem Informasi dan Komputer)*, 12(3), 477-484.
- [7] Werdiningsih, I., Purwanti, E., Aditya, G. R. W., Hidayat, A. R., Athallah, R. S. R., Sahar, V. A., ... & Somba, D. F. N. (2023). Identifying Credit Card Fraud in Illegal Transactions Using Random Forest and Decision Tree Algorithms. *Jurnal Sisfokom (Sistem Informasi dan Komputer)*, 12(3), 477-484.
- [8] KN, L. K., & Nasution, M. I. P. (2024). ANALISIS KEAMANAN DATA TERHADAP PENGGUNAAN E-WALLET SEBAGAI ALAT TRANSAKSI DIGITAL UNTUK MENCEGAH PENIPUAN ONLINE. *JOURNAL SAINS STUDENT RESEARCH*, 2(4), 108-116.