

## BAB I PENDAHULUAN

### 1.1 Latar Belakang

Pesatnya transformasi digital di Indonesia telah mendorong peningkatan signifikan pada volume transaksi pembayaran elektronik di berbagai sektor ekonomi. Namun, pertumbuhan ini berbanding lurus dengan meningkatnya ancaman kejahatan siber, khususnya manipulasi transaksi keuangan yang semakin canggih [1]. Pola penipuan dan kecurangan terus berevolusi secara dinamis sehingga metode deteksi konvensional seringkali gagal mengidentifikasi anomali secara *real-time*. Keamanan sistem pembayaran kini menjadi tantangan utama bagi penyedia layanan jasa keuangan untuk menjaga kepercayaan serta loyalitas pengguna [2].

Berdasarkan laporan *Association of Certified Fraud Examiners (ACFE)* pada tahun 2024, kerugian global akibat *fraud* diperkirakan mencapai 5% dari total pendapatan organisasi setiap tahunnya [3]. Di Indonesia, data Badan Siber dan Sandi Negara (BSSN) pada tahun 2023 menunjukkan adanya tren peningkatan anomali trafik yang berkaitan erat dengan upaya pembobolan akun finansial [4]. Tingginya frekuensi serangan ini menciptakan urgensi besar bagi pengembangan sistem proteksi yang jauh lebih adaptif dan responsif. Fenomena tersebut menunjukkan bahwa sektor pembayaran digital masih sangat rentan terhadap ancaman kerugian material.

Pendekatan berbasis aturan (*rule-based*) yang selama ini diterapkan memiliki keterbatasan besar dalam menangani data yang tidak seimbang dan kompleksitas pola *fraud* baru [5]. Sistem saat ini seringkali menghasilkan tingkat *false positive* yang tinggi, sehingga transaksi sah milik pengguna justru terblokir secara keliru. Hal ini mengakibatkan inefisiensi operasional dan menurunkan kualitas pengalaman pengguna dalam bertransaksi secara digital [6]. Diperlukan sebuah solusi cerdas yang mampu melakukan ekstraksi fitur secara mendalam untuk membedakan antara transaksi normal dan mencurigakan [7].

Guna mengatasi kendala tersebut, peneliti mengusulkan pendekatan *hybrid* yang mengombinasikan algoritma *Isolation Forest* dan *XGBoost*, dikarenakan *Isolation Forest* memiliki efisiensi tinggi dalam mengisolasi anomali pada data berdimensi tinggi, sementara *XGBoost* memberikan akurasi klasifikasi yang unggul melalui mekanisme *gradient boosting* [8]. Integrasi kedua metode ini diharapkan mampu meningkatkan performa deteksi serta meminimalisir kesalahan klasifikasi pada transaksi pembayaran digital. Penelitian ini menjadi sangat relevan dalam upaya memperkuat infrastruktur keamanan siber pada ekosistem teknologi finansial di masa depan.

## 1.2 Rumusan Masalah

1. Bagaimana performa penggabungan algoritma *Isolation Forest* dan *XGBoost* dalam mendeteksi anomali serta mengklasifikasikan transaksi *fraud* pada data pembayaran digital?
2. Bagaimana pengaruh penerapan metode *hybrid Isolation Forest* dan *XGBoost* dalam meminimalkan tingkat kesalahan identifikasi transaksi sah (*false positive*) pada transaksi pembayaran digital?

## 1.3 Batasan Masalah

1. **Sumber dan Ukuran Data:** Data yang digunakan dalam penelitian ini merupakan dataset sekunder *IEEE-CIS Fraud Detection* yang berisi riwayat transaksi pembayaran digital. Mengingat ukuran dataset yang sangat besar serta ketidakseimbangan kelas yang tinggi, penelitian ini menggunakan sampel sebesar 20% dari total data yang diambil yaitu 118,108 baris menggunakan teknik *stratified sampling* untuk menjaga proporsi kelas *fraud* dan *non-fraud*. Pembatasan ini dilakukan untuk meningkatkan efisiensi komputasi, mengurangi risiko *overfitting*, serta memastikan proses pelatihan dan evaluasi model dapat dilakukan secara optimal dan reproduktibel, tanpa mengurangi validitas statistik hasil penelitian
2. **Algoritma Penelitian:** Model dikembangkan menggunakan pendekatan

*hybrid* yang menggabungkan *Isolation Forest* untuk ekstraksi fitur anomali dan *XGBoost* sebagai pengklasifikasi utama. Proses penanganan ketidakseimbangan data (*data imbalance*) dibatasi pada penggunaan metode SMOTE (*Synthetic Minority Over-sampling Technique*) di dalam sebuah pipeline pemrosesan.

3. Pra-pemrosesan Data: Tahap pra-pemrosesan dibatasi pada imputasi nilai hilang (*missing values*) dengan nilai khusus (kategorikal: 'MISSING', numerik: -999) dan penggunaan Label Encoding untuk fitur kategorikal. Tidak dilakukan analisis terhadap teknik imputasi tingkat lanjut lainnya.
4. Rekayasa Fitur: Variabel yang digunakan terbatas pada fitur asli dataset serta fitur tambahan hasil rekayasa (*feature engineering*) berupa fitur frekuensi (*count*), rata-rata nominal transaksi per kategori (*mean*), dan pola transaksi berdasarkan jam (*hourly patterns*).
5. Optimasi Model: Proses pencarian parameter terbaik dibatasi pada penggunaan GridSearchCV dengan fokus optimasi pada nilai *F1-Score*. Penentuan klasifikasi akhir menggunakan ambang batas (*threshold*) kustom sebesar 0.3 untuk meningkatkan sensitivitas deteksi (*recall*).
6. Lingkup Implementasi: Luaran penelitian ini terbatas pada pengembangan skema *Inference Pipeline* berbasis Python yang mampu memproses data transaksi baru secara offline, dan tidak mencakup pengembangan antarmuka web atau integrasi ke sistem pembayaran real-time.

#### 1.4 Tujuan Penelitian

1. Membangun dan mengimplementasikan model deteksi *fraud* menggunakan pendekatan *hybrid* yang mengintegrasikan algoritma *Isolation Forest* dan *XGBoost* untuk mengenali pola transaksi yang mencurigakan secara efektif.
2. Menganalisis dan menghitung kinerja model berdasarkan parameter metrik evaluasi yang meliputi *accuracy*, *precision*, *recall*, dan *F1-score* guna memastikan reliabilitas sistem dalam mendeteksi *fraud*.

3. Memvalidasi kemampuan pendekatan *hybrid* dalam meminimalkan tingkat false positive dibandingkan dengan algoritma tunggal, sehingga dapat mengurangi kesalahan identifikasi pada transaksi pembayaran digital yang sah.

#### 1.5 Manfaat Penelitian

- a. Memberikan solusi teknis bagi penyedia layanan pembayaran digital dalam mendeteksi upaya penipuan secara lebih akurat, sehingga dapat menekan kerugian finansial akibat *fraud*.
- b. Memberikan wawasan mendalam mengenai efektivitas integrasi algoritma *unsupervised learning* (*Isolation Forest*) dan *supervised learning* (*XGBoost*) dalam domain deteksi anomali keuangan.
- c. Memperkaya kajian literatur di bidang Sains Data dan Keamanan Siber, khususnya mengenai penerapan teknik *ensemble learning* untuk memperkuat infrastruktur keamanan teknologi finansial.

#### 1.6 Sistematika Penulisan

Sistematika penulisan dalam penelitian ini adalah sebagai berikut:

- a. BAB I PENDAHULUAN, berisi latar belakang masalah terkait urgensi keamanan transaksi digital, rumusan masalah mengenai efektivitas model *hybrid*, batasan masalah untuk menjaga fokus penelitian, tujuan, manfaat penelitian, serta sistematika penulisan.
- b. BAB II TINJAUAN PUSTAKA, berisi studi literatur dari penelitian terdahulu dan landasan teori yang mendukung penelitian, meliputi konsep transaksi pembayaran digital, karakteristik *fraud*, algoritma *unsupervised Isolation Forest*, algoritma *supervised XGBoost*, serta teori pendukung mengenai penanganan imbalanced data.
- c. BAB III METODE PENELITIAN, berisi uraian mengenai objek penelitian, sumber data dataset IEEE-CIS, tahapan penelitian yang meliputi *stratified sampling*, rekayasa fitur (*feature engineering*), tahapan integrasi model *hybrid* (*Isolation Forest* sebagai deteksi

anomali dan *XGBoost* sebagai klasifikasi), serta instrumen evaluasi model.

- d. BAB IV HASIL DAN PEMBAHASAN, berisi laporan hasil implementasi model *hybrid* yang dibangun, pengujian performa berdasarkan metrik Confusion Matrix (analisis pada tingkat *false positive*), perbandingan hasil antara model *hybrid* dengan algoritma tunggal, serta pembahasan hasil eksperimen.
- e. BAB V PENUTUP, berisi kesimpulan akhir yang menjawab rumusan masalah mengenai keunggulan pendekatan *hybrid* dalam mendeteksi *fraud* pembayaran digital, serta saran untuk pengembangan model atau penggunaan dataset yang lebih luas di masa depan.

