

BAB I

PENDAHULUAN

1.1. Latar Belakang

Perkembangan pesat platform e-commerce telah mendorong peningkatan volume transaksi digital secara signifikan. Di balik kemudahan dan efisiensi yang ditawarkan, pertumbuhan ini juga diiringi oleh meningkatnya risiko terjadinya kecurangan transaksi (fraud). Fraud pada sistem e-commerce umumnya melibatkan penyalahgunaan identitas, manipulasi metode pembayaran, maupun eksploitasi celah sistem, yang berpotensi menimbulkan kerugian finansial serta menurunkan tingkat kepercayaan pengguna terhadap layanan digital. Tantangan utama dalam mendeteksi fraud terletak pada sifat pola kecurangan yang dinamis dan terus berkembang, sehingga sulit diidentifikasi menggunakan sistem berbasis aturan yang bersifat statis.

Seiring dengan keterbatasan pendekatan konvensional, berbagai penelitian mulai memanfaatkan teknik machine learning untuk mendeteksi fraud secara lebih adaptif. Model supervised learning seperti Logistic Regression, Random Forest, dan Support Vector Machine banyak digunakan karena kemudahannya dalam implementasi. Namun, pada kondisi data yang sangat tidak seimbang—di mana jumlah transaksi fraud hanya merupakan sebagian kecil dari keseluruhan data—pendekatan supervised tunggal cenderung mengalami penurunan performa, khususnya dalam menjaga keseimbangan antara precision dan recall. Model seringkali bias terhadap kelas mayoritas dan gagal mengenali pola kecurangan yang jarang muncul.

Untuk mengatasi permasalahan tersebut, pendekatan unsupervised seperti Autoencoder mulai digunakan sebagai metode anomaly detection. Autoencoder dilatih menggunakan data transaksi normal untuk mempelajari representasi pola yang wajar, sehingga transaksi dengan karakteristik menyimpang dapat teridentifikasi melalui nilai reconstruction error yang tinggi. Pendekatan ini dinilai efektif dalam menangani data imbalanced karena tidak bergantung pada label kelas minoritas. Namun, sejumlah penelitian menunjukkan bahwa penggunaan Autoencoder saja sering kali belum cukup optimal karena tidak secara langsung memaksimalkan proses klasifikasi akhir antara transaksi fraud dan non-fraud.

Di sisi lain, algoritma XGBoost (Extreme Gradient Boosting) telah banyak digunakan dalam berbagai studi deteksi fraud karena kemampuannya menangani data tabular berskala besar, hubungan fitur yang kompleks, serta ketidakseimbangan kelas melalui mekanisme pembobotan dan regularisasi. Dibandingkan dengan metode machine learning konvensional, XGBoost memiliki keunggulan dalam mempelajari pola non-linear secara lebih stabil dan efektif, terutama pada data dengan distribusi yang tidak seimbang. Meskipun demikian, sebagian besar penelitian masih memanfaatkan XGBoost sebagai model supervised tunggal tanpa mengombinasikannya dengan pendekatan anomaly detection secara terstruktur.

Berdasarkan tinjauan tersebut, masih terdapat celah penelitian (research gap) pada integrasi metode unsupervised dan supervised yang tidak hanya menggabungkan kedua pendekatan, tetapi juga mengevaluasi secara jelas peran masing-masing metode dalam meningkatkan kinerja deteksi fraud. Khususnya, masih terbatas penelitian yang memanfaatkan skor anomali dari Autoencoder sebagai fitur tambahan untuk memperkuat kemampuan klasifikasi XGBoost pada data e-commerce yang sangat imbalanced dan berskala besar.

Oleh karena itu, penelitian ini mengusulkan pendekatan deteksi fraud dua tahap (two-stage) yang mengombinasikan Autoencoder sebagai pendeteksi anomali pada tahap awal dan XGBoost sebagai classifier utama pada tahap lanjutan. Skor anomali yang dihasilkan oleh Autoencoder dimanfaatkan sebagai informasi tambahan dalam proses pembelajaran XGBoost untuk meningkatkan pemisahan antara transaksi fraud dan non-fraud. Pendekatan ini dirancang untuk menghasilkan kinerja deteksi yang lebih seimbang, khususnya dalam meningkatkan recall tanpa mengorbankan precision secara signifikan, serta memberikan kontribusi terhadap pengembangan metode deteksi fraud melalui evaluasi perbandingan dengan pendekatan supervised tunggal pada lingkungan e-commerce yang memiliki tingkat ketidakseimbangan data yang tinggi.

1.2. Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, rumusan masalah penelitian ini adalah:

1. Bagaimana membangun model deteksi fraud dua-tahap (autoencoder + XGBoost) yang efektif pada dataset e-commerce yang sangat imbalance?
2. Seberapa besar perbaikan deteksi (precision, recall, AUC) yang diperoleh

dengan menggunakan anomaly score dari autoencoder dibandingkan model supervised saja?

3. Seberapa robust model terhadap distribusi kelas imbalanced bila diterapkan pada data nyata (split stratified, tidak hanya sampling)?

1.3. Kebaruan dan Kontribusi Penelitian

Kebaruan penelitian ini terletak pada integrasi skor anomali Autoencoder sebagai fitur kuantitatif tambahan dalam proses pembelajaran XGBoost, bukan sekadar digunakan sebagai tahap pra-pemrosesan atau penyaringan data. Berbeda dengan penelitian sebelumnya yang umumnya menerapkan Autoencoder dan XGBoost secara terpisah, penelitian ini secara sistematis mengevaluasi kontribusi skor anomali terhadap peningkatan kinerja XGBoost pada data transaksi e-commerce berskala besar dengan tingkat ketidakseimbangan kelas yang ekstrem.

1.4. Tujuan Penelitian

Tujuan dari penelitian ini adalah:

1. Membangun dan mengimplementasikan pipeline dua-tahap untuk deteksi fraud autoencoder (anomaly detection) + XGBoost (supervised)
2. Mengukur efektivitas anomaly score dari autoencoder dalam meningkatkan performa deteksi fraud (precision, recall, AUC) dibanding model konvensional
3. Menilai performa dan stabilitas model di data nyata (train / val / test) dengan distribusi kelas yang realistis dan imbalanced.

1.5. Manfaat Penelitian

Manfaat penelitian ini meliputi:

1. Manfaat Praktis

Memberikan pendekatan deteksi fraud yang lebih akurat dan adaptif untuk platform e-commerce agar dapat mendeteksi transaksi mencurigakan dengan false positive rendah.

2. Manfaat akademik

Menambah literatur tentang kombinasi autoencoder + gradient boosting di domain fraud detection, terutama pada dataset sangat imbalanced.

3. Manfaat keamanan dan bisnis

Membantu perusahaan e-commerce atau fintech mengurangi risiko kerugian finansial, sekaligus menjaga pengalaman pengguna (tidak memblokir transaksi sah).

4. Manfaat sosial

Meningkatkan kepercayaan pengguna terhadap sistem pembayaran digital,
mengurangi kerugian konsumen akibat fraud.