

ABSTRAK

Pertumbuhan transaksi e-commerce yang semakin pesat diikuti oleh meningkatnya risiko kecurangan digital yang sulit dideteksi secara konvensional, terutama pada data yang sangat tidak seimbang. Untuk mengatasi permasalahan tersebut, penelitian ini mengembangkan model deteksi fraud berbasis pendekatan hybrid yang menggabungkan Autoencoder sebagai pendeteksi anomali tanpa label dan XGBoost sebagai model klasifikasi terawasi. Dataset yang digunakan berasal dari dua versi *Fraudulent E-commerce Transactions* pada platform Kaggle, yang masing-masing memiliki karakteristik berbeda namun sama-sama menunjukkan proporsi fraud yang sangat rendah. Tahapan penelitian mencakup persiapan data, pembersihan, encoding fitur kategori, normalisasi numerik, pelatihan Autoencoder terhadap data non-fraud, serta integrasi skor rekonstruksi sebagai fitur tambahan untuk XGBoost. Evaluasi dilakukan menggunakan metrik yang relevan untuk data imbalanced, seperti precision, recall, F1-score, ROC-AUC, dan PR-AUC. Hasil penelitian menunjukkan bahwa kombinasi Autoencoder dan XGBoost mampu meningkatkan kemampuan model dalam mengenali transaksi fraud, dengan recall dan precision melebihi 80% setelah penyesuaian threshold. Temuan ini menunjukkan bahwa pendekatan hybrid efektif dalam memitigasi ketidakseimbangan data dan dapat diterapkan sebagai sistem pendukung deteksi fraud pada lingkungan e-commerce berskala besar.

Kata kunci: deteksi fraud, e-commerce, *Autoencoder*, *XGBoost*, *anomaly detection*, *data imbalanced*

ABSTRACT

The rapid expansion of e-commerce transactions has led to a rising risk of digital fraud, which is difficult to detect using conventional methods, particularly when dealing with highly imbalanced datasets. To address this problem, this study proposes a hybrid fraud detection model that integrates an Autoencoder for unsupervised anomaly detection with XGBoost as a supervised classification algorithm. The research utilizes two versions of the *Fraudulent E-commerce Transactions* dataset from Kaggle, each presenting different feature structures but similarly exhibiting a very low proportion of fraudulent cases. The research stages include data preparation, cleaning, categorical encoding, numerical scaling, training the Autoencoder exclusively on non-fraud data, and incorporating reconstruction error scores as additional features for XGBoost. Model performance was evaluated using metrics suitable for imbalanced classification, including precision, recall, F1-score, ROC-AUC, and PR-AUC. The results indicate that the hybrid approach significantly improves the detection of fraudulent transactions, achieving precision and recall above 80% after threshold optimization. These findings demonstrate that combining unsupervised anomaly scoring with gradient boosting is effective in handling data imbalance and can serve as a reliable fraud detection component for large-scale e-commerce platforms.

Keywords: fraud detection, e-commerce, *Autoencoder*, *XGBoost*, *anomaly detection*, *imbalanced data*