

ABSTRAK

Penelitian ini bertujuan untuk mengembangkan model deteksi anomali guna membantu auditor dalam mengidentifikasi proyek Pengadaan Barang dan Jasa (PBJ) yang memiliki indikasi red flag atau potensi kecurangan dalam sistem e-procurement di Indonesia. Metode yang digunakan meliputi pendekatan data mining menggunakan machine learning dengan supervised learning dan unsupervised learning. Model supervised learning dikembangkan menggunakan algoritma ensemble Random Forest dan XGBoost, sementara model unsupervised learning menggunakan algoritma Local Outlier Factor, Copula-Based Outlier Detection, dan Autoencoder. Penelitian ini membandingkan performa masing-masing algoritma dengan berbagai proporsi partisi data serta metode resampling untuk menentukan model terbaik dalam mendeteksi anomali. Dataset yang digunakan berasal dari data tender umum PBJ pemerintah pusat periode 2022 hingga 2024. Hasil penelitian menunjukkan bahwa model supervised learning dengan data resampling memberikan performa yang sangat baik, dengan nilai seluruh metrik evaluasi melebihi 99%. Model terbaik yang diperoleh adalah model supervised learning dengan algoritma XGBoost yang menggunakan metode resampling gabungan. Sebaliknya, model unsupervised learning menunjukkan performa yang kurang optimal di seluruh skema pemodelan, baik dalam berbagai proporsi partisi data maupun metode resampling. Temuan ini mengindikasikan bahwa pendekatan supervised learning, khususnya dengan XGBoost dan strategi resampling yang tepat, lebih efektif dalam mendeteksi anomali dalam data e-procurement dibandingkan pendekatan unsupervised learning.

Kata kunci: Data Mining, e-Procurement, Audit

ABSTRACT

This study aims to develop an anomaly detection model to assist auditors in identifying Procurement of Goods and Services (PBJ) projects with red flags or potential fraud in Indonesia's e-procurement system. The methodology involves a data mining approach using machine learning both supervised and unsupervised learning. The supervised learning models are built using ensemble algorithms, namely Random Forest and XGBoost, while the unsupervised learning models utilize Local Outlier Factor, Copula-Based Outlier Detection, and Autoencoder algorithms. This research compares the performance of each algorithm across different data partitioning proportions and resampling methods to determine the most effective anomaly detection model. The dataset consists of general procurement tender data from the central government between 2022 and 2024. The results indicate that supervised learning models with resampled data perform exceptionally well, achieving evaluation metric scores above 99%. The best-performing model is the supervised learning approach using the XGBoost algorithm combined with a hybrid resampling method. In contrast, unsupervised learning models exhibit poor performance across all modeling schemes, regardless of data partitioning proportions or resampling methods. These findings suggest that supervised learning, particularly with XGBoost and appropriate resampling strategies, is more effective in detecting anomalies in e-procurement data compared to unsupervised learning approaches.

Keywords: Data Mining, e-Procurement, Audit