

ABSTRAK

Penyebaran akun palsu di *platform* X (Twitter) telah menjadi tantangan serius bagi keamanan siber dan integritas informasi digital. Akun-akun tersebut sering kali digunakan untuk menyebarkan misinformasi dan memanipulasi opini *public*. Penelitian ini bertujuan untuk membandingkan performa tiga algoritma *machine learning* berbasis pohon (*tree-based*), yaitu *Random Forest*, *XGBoost*, dan *LightGBM*, dalam mendeteksi akun palsu pada *platform* X (Twitter) menggunakan dataset MIB (*Machine Intelligence and Bot detection*). Tahapan penelitian meliputi pra-pemrosesan data, seleksi fitur menggunakan *SelectKBest* dengan *mutual_info_classif*, penanganan ketidakseimbangan kelas menggunakan SMOTE, optimasi *hyperparameter* menggunakan *GridSearchCV*, serta evaluasi menggunakan metrik *Accuracy*, *Precision*, *Recall*, *F1-Score*, dan *ROC-AUC*. Hasil penelitian menunjukkan bahwa *Random Forest* merupakan model dengan performa terbaik secara keseluruhan, diikuti *LightGBM* dan *XGBoost*. Meskipun demikian, selisih performa antar ketiga model sangat kecil sehingga ketiganya dapat dinyatakan memiliki kemampuan yang setara dan optimal dalam mendeteksi akun palsu. Seluruh model tidak mengalami *overfitting* dan termasuk dalam kategori *Excellent* berdasarkan interpretasi nilai *ROC-AUC*, sehingga ketiga algoritma terbukti layak digunakan sebagai sistem deteksi akun palsu pada *platform* X (Twitter).

Kata Kunci: Deteksi Akun Palsu, *Machine Learning*, *Random Forest*, *XGBoost*, *LightGBM*.

ABSTRACT

The proliferation of fake accounts on the X (Twitter) platform has become a serious challenge to cybersecurity and the integrity of digital information. These accounts are frequently used to spread misinformation and manipulate public opinion. This study aims to compare the performance of three tree-based machine learning algorithms, namely Random Forest, XGBoost, and LightGBM, in detecting fake accounts on the X (Twitter) platform using the MIB (Machine Intelligence and Bot detection) dataset. The research stages include data preprocessing, feature selection using SelectKBest with mutual_info_classif, class imbalance handling using SMOTE, hyperparameter optimization using GridSearchCV, and evaluation using Accuracy, Precision, Recall, F1-Score, and ROC-AUC metrics. The results show that Random Forest achieved the best overall performance, followed by LightGBM and XGBoost. Nevertheless, the performance gap among the three models is very small, indicating that all three algorithms have equivalent and optimal capabilities in detecting fake accounts. All models showed no overfitting and were classified in the Excellent category based on ROC-AUC interpretation, confirming that all three algorithms are viable for use as fake account detection systems on the X (Twitter) platform.

Keywords: Fake Account Detection, Machine Learning, Random Forest, XGBoost, LightGBM.