ABSTRAK

Keamanan jaringan telah menjadi tantangan global yang memerlukan solusi yang efektif dan inovatif. Intrusion Detection System (IDS) memainkan peran penting dalam melindungi infrastruktur jaringan dari serangan siber yang terus berkembang. Penggunaan teknik Machine Learning (ML) dalam IDS menawarkan akurasi tinggi dalam mendeteksi dan mengidentifikasi ancaman. Namun, tantangan muncul ketika harus menangani dataset yang tidak seimbang dan berdimensi tinggi. Penelitian ini memperkenalkan pendekatan baru dalam deteksi intrusi jaringan berbasis ML dengan menerapkan Random Oversampling (RO) untuk mengatasi ketidakseimbangan data serta validasi K-fold untuk memastikan bahwa proses pemilihan fitur (Random Forest + PCA) dan pelatihan model dioptimalkan guna menghindari overfitting. Selain itu, setiap model menjalani optimasi maksimum menggunakan Optuna untuk meningkatkan akurasi, presisi, recall, F1-score, parameter lalu lintas NIDS, serta performa kurva ROC. Pendekatan ini dievaluasi pada tiga dataset benchmark: UNSW-NB15, CIC-IDS-2017, dan CIC-IDS-2018. Setiap dataset dimodelkan menggunakan algoritma KNN, Logistic Regression, Decision Tree, Random Forest, GBM, XGBM, Adaboost, Light GBM, CatBoost, dan Extra Tree, sehingga mampu mencapai akurasi tinggi hingga 99%. Metode ini terbukti efektif untuk dataset besar dan tidak seimbang, sebagaimana ditunjukkan pada dataset CIC-IDS-2018 yang berisi lebih dari satu juta data. Hasil penelitian ini melampaui model-model mutakhir sebelumnya, menandai kemajuan signifikan dalam deteksi intrusi jaringan. Kerangka kerja yang fleksibel ini membuka peluang lebih lanjut untuk mengeksplorasi algoritma ML dalam meningkatkan efektivitas IDS.

Keywords: Machine Learning, Deteksi Intrusi Jaringan, Pemilihan fitur dan ekstraksi, Random Oversampling, Principal Component Analysis.