ABSTRAK

Ancaman *insider attack* merupakan salah satu risiko keamanan paling kompleks dalam lingkungan organisasi karena pelakunya berasal dari internal dan memiliki akses sah terhadap sistem. Pola serangan insider threat sering kali bersifat tersembunyi, bertahap, dan sulit dibedakan dari aktivitas normal pengguna, sehingga pendekatan berbasis aturan statis tidak lagi memadai. Selain itu, karakteristik data keamanan yang tidak seimbang serta heterogenitas sumber log, seperti aktivitas *login*, akses *file*, penggunaan perangkat, dan komunikasi jaringan, semakin menambah tantangan dalam mendeteksi serangan *insider* secara akurat. Oleh karena itu, diperlukan pendekatan analitis yang mampu menangkap pola perilaku kompleks serta memiliki kemampuan generalisasi yang baik terhadap data dunia nyata.

Penelitian ini mengusulkan pendekatan deteksi insider threat menggunakan algoritma *Extreme Gradient Boosting* (*XGBoost*) yang dikenal efektif dalam menangani data berdimensi tinggi dan tidak seimbang. Dataset yang digunakan berasal dari *CERT Insider Threat Dataset*, yang mencakup berbagai jenis aktivitas pengguna dalam lingkungan organisasi. Tahapan penelitian meliputi pengumpulan data, *preprocessing*, pembagian data, serta pelatihan dan pengujian model. Evaluasi model dilakukan menggunakan skema *Stratified K-Fold Cross Validation* untuk menjaga proporsi kelas pada setiap *fold*, dengan metrik evaluasi yang berfokus pada kemampuan model dalam menekan tingkat *false positive* dan *false negative*.

Hasil pengujian menunjukkan bahwa model *XGBoost* mampu mencapai performa yang sangat baik ketika dilatih dan diuji pada distribusi data yang sama, khususnya pada dataset r4.2 dan r5.2 dengan nilai *F1-Score* dan *AUC-ROC* yang tinggi. Penerapan *SMOTE* terbukti efektif dalam meningkatkan kemampuan model mendeteksi kelas minoritas melalui peningkatan *recall*, meskipun terdapat *trade-off* berupa penurunan *precision*. Namun, pada dataset r6.2 yang memiliki ketidakseimbangan kelas lebih ekstrem, model gagal menghasilkan prediksi positif pada *threshold* default meskipun nilai AUC-ROC tinggi, yang mengindikasikan masalah pada kalibrasi threshold dan perbedaan distribusi probabilitas. Pengujian lintas versi dataset juga memperlihatkan penurunan performa yang signifikan, menandakan bahwa model masih sensitif terhadap perubahan distribusi data dan belum memiliki kemampuan generalisasi yang memadai untuk skenario deteksi *insider threat* pada lingkungan yang dinamis.

Kata Kunci: XGBoost, SMOTE, *Insider Attack*, CERT *Insider Threat*

ABSTRACT

Insider threats represent one of the most challenging security risks in organizational environments because the attackers originate from within the organization and possess legitimate system access. Insider attack behaviors are often subtle, gradual, and difficult to distinguish from normal user activities, rendering traditional rule-based detection approaches ineffective. Furthermore, the inherent characteristics of security data, such as class imbalance and the heterogeneity of log sources—including login activities, file access, device usage, and network communications—further complicate accurate detection. Consequently, there is a strong need for analytical approaches capable of modeling complex behavioral patterns while maintaining robust generalization performance in real-world scenarios.

This study proposes an insider threat detection approach based on the Extreme Gradient Boosting (XGBoost) algorithm, which has demonstrated effectiveness in handling high-dimensional and imbalanced data. The dataset used in this research is derived from the CERT Insider Threat Dataset, encompassing various types of user activities within an organizational environment. The research methodology includes data collection, preprocessing stages such as data cleaning, feature mapping, and class imbalance handling, followed by data partitioning, model training, and testing. Model evaluation is conducted using a Stratified K-Fold Cross-Validation scheme to preserve class distribution across folds, with evaluation metrics emphasizing the reduction of false positive and false negative rates.

Experimental results indicate that the XGBoost model achieves strong performance when trained and evaluated on datasets with similar distributions, particularly on versions r4.2 and r5.2, where high F1-scores and AUC-ROC values were consistently observed. The application of SMOTE proved effective in improving the detection of minority-class instances by increasing recall, although this was accompanied by a moderate reduction in precision, reflecting a typical trade-off in imbalanced classification problems. However, on dataset version r6.2, which exhibits more extreme class imbalance, the model failed to produce positive predictions at the default decision threshold despite achieving relatively high AUC-ROC values. This discrepancy suggests issues related to threshold calibration and skewed probability distributions rather than an inability of the model to learn discriminative patterns. Cross-version evaluation further revealed significant performance degradation, indicating that the model remains sensitive to distributional shifts and lacks sufficient generalization capability for deployment in dynamic real-world insider threat detection scenarios.

Keywords: *XGBoost, SMOTE, Insider Attack, CERT Insider Threat*