IMPLEMENTASI MODEL GRAPHSAGE DAN GRAPHSVX UNTUK DETEKSI DAN INTERPRETASI POLA ANCAMAN INTERNAL DALAM PROFIL PERILAKU PADA DATASET CERT INSIDER THREAT R6.2

Aaz Zazri Nugraha[1], Arief Ikhwan Wicaksono[2]

INTISARI

Latar Belakang: Ancaman internal merupakan salah satu bentuk serangan siber yang sulit terdeteksi karena dilakukan oleh individu yang memiliki akses sah terhadap sistem organisasi. Banyak organisasi mengalami kerugian signifikan akibat aktivitas pengguna yang menyimpang namun tidak terpantau. Untuk itu, dibutuhkan pendekatan yang mampu mendeteksi pola perilaku mencurigakan secara otomatis dan adaptif, dengan mempertimbangkan hubungan kompleks antar entitas dalam sistem. Representasi graf heterogen dan metode pembelajaran mesin berbasis *Graph Neural Network* (GNN) menjadi solusi yang potensial untuk mendeteksi pola semacam ini.

Tujuan: Penelitian ini bertujuan untuk membangun sistem deteksi ancaman internal berbasis model GraphSAGE dengan interpretasi hasil prediksi menggunakan metode GraphSVX, guna menjelaskan alasan di balik keputusan model secara transparan.

Metode Penelitian: Model GraphSAGE dilatih untuk mengklasifikasikan node user menjadi normal atau insider. Setelah pelatihan, GraphSVX digunakan untuk menafsirkan fitur perilaku yang paling memengaruhi prediksi ancaman.

Hasil: Model GraphSAGE menunjukkan performa akurasi rata-rata sebesar 99,2% dan F1-score sebesar 96%, bahkan mencapai skor sempurna pada sejumlah konfigurasi pengujian. Hasil interpretasi GraphSVX mengungkap bahwa perilaku seperti login di luar jam kerja, akses file, serta kunjungan ke situs cloud merupakan indikator utama yang berkontribusi terhadap deteksi ancaman.

Kesimpulan: Sistem deteksi ancaman internal berbasis GraphSAGE dan GraphSVX berhasil dibangun dan mampu mendeteksi serta menjelaskan pola perilaku mencurigakan dengan baik. Pendekatan ini tidak hanya akurat tetapi juga dapat dijelaskan (*interpretable*), sehingga cocok diterapkan sebagai sistem pendukung keputusan di bidang keamanan siber.

Kata kunci: Ancaman internal, GNN, XAI, GraphSAGE, GraphSVX

[1] Mahasiswa Teknologi Informasi (S-1) Fakultas Teknik dan Teknologi Informasi Universitas Jenderal Achmad Yani Yogyakarta
[2] Dosen Teknologi Informasi Fakultas Teknik dan Teknologi Informasi Universitas Jenderal Achmad Yani Yogyakarta

IMPLEMENTATION OF GRAPHSAGE AND GRAPHSVX MODELS FOR INTERNAL THREAT PATTERN DETECTION AND INTERPRETATION IN BEHAVIORAL PROFILES ON THE CERT INSIDER THREAT R6.2 DATASET

Aaz Zazri Nugraha[1], Arief Ikhwan Wicaksono[2]

ABSTRACT

***Background**: Insider threats are one of the forms of cyberattacks that are difficult to detect because they are carried out by individuals with legitimate access to the organization's systems. Many organizations experience significant losses due to user activities that deviate but go unnoticed. Therefore, an approach is needed that can automatically and adaptively detect suspicious behavioral patterns while considering the complex relationships among entities within the system. Heterogeneous graph representation and machine learning methods based on Graph Neural Networks (GNNs) are a potential solution for detecting such patterns.*

***Objective**: This study aims to develop an insider threat detection system based on the GraphSAGE model, with prediction interpretation using the GraphSVX method to transparently explain the reasons behind the model's decisions.*

***Method**: The GraphSAGE model was trained to classify user nodes as either normal or insider. After training, GraphSVX was used to interpret the behavioral features that most influenced the threat predictions.*

***Result**: The GraphSAGE model achieved an average accuracy of 99.2% and an F1-score of 96%, even reaching perfect scores in several testing configurations. The GraphSVX interpretation revealed that behaviors such as after-hours logins, file access, and visits to cloud-based websites were the main indicators contributing to threat detection.*

***Conclusion**: The insider threat detection system based on GraphSAGE and GraphSVX was successfully developed and is capable of effectively detecting and explaining suspicious behavioral patterns. This approach is not only accurate but also interpretable, making it suitable for implementation as a decision support system in the field of cybersecurity.*

***Keywords**: Insider threat, GNN, XAI, GraphSAGE, GraphSVX*

[1] Undergraduate Student in Information Technology, Faculty of Engineering and Information Technology, Jenderal Achmad Yani University, Yogyakarta
[2] Lecturer in Information Technology, Faculty of Engineering and Information Technology, Jenderal Achmad Yani University, Yogyakarta