BAB 1
PENDAHULUAN

1.1 Latar Belakang

Ancaman internal dalam keamanan informasi menjadi isu yang semakin krusial seiring meningkatnya insiden siber dalam beberapa tahun terakhir. Menurut (World Economic Forum, 2024), ketidakamanan siber berada di peringkat keempat dalam daftar risiko global, dengan 34% pelanggaran data disebabkan oleh aktor internal, yang sering kali menimbulkan kerugian finansial lebih besar dibandingkan serangan eksternal. Di Indonesia, (BSSN, 2024) mencatat 241 dugaan kebocoran data, dengan 183 insiden terjadi di sektor Administrasi Pemerintahan serta 1.814 aduan siber yang diterima sepanjang tahun. Secara global, Laporan dari (Verizon, 2024) menunjukkan bahwa pelanggaran data akibat orang dalam terjadi karena tindakan berbahaya (15%), kelalaian (14%), dan penyalahgunaan hak istimewa (5%). Dampak finansial dari insiden ini juga signifikan. Menurut laporan dari (IBM, 2024), rata-rata biaya pemulihan akibat pelanggaran data internal mencapai USD 4,2 juta (sekitar Rp 67,9 miliar), lebih tinggi dibandingkan insiden yang disebabkan oleh faktor eksternal. Temuan ini menegaskan betapa pentingnya strategi deteksi dan mitigasi ancaman internal dalam meningkatkan keamanan informasi di tingkat nasional maupun global.

Dalam beberapa tahun terakhir, pendekatan berbasis *Artificial Intelligence* (AI), dan *Graph Neural Network* (GNN) telah berkembang pesat dalam mendeteksi ancaman internal. Algoritma berbasis AI mampu menganalisis pola perilaku pengguna secara otomatis, sehingga meningkatkan potensi dalam deteksi dini. Salah satu pendekatan yang menjanjikan dalam deteksi ancaman siber adalah GNN, yang terbukti efisien dalam menganalisis pola hubungan kompleks dalam sistem keamanan informasi (Jegelka, 2023). Meskipun AI dan GNN mampu meningkatkan akurasi deteksi, tantangan utama dalam penerapannya adalah kurangnya transparansi atau sering kali bersifat "*black-box*". Laporan dari (IBM, 2024) mengungkapkan bahwa 70% organisasi yang mengalami pelanggaran data

menganggap kurangnya transparansi dalam sistem AI sebagai penghalang utama penerapan teknologi ini. Oleh karena itu, XAI menjadi kebutuhan mendesak dalam pengembangan solusi keamanan siber berbasis AI.

Salah satu model GNN yang banyak digunakan dalam analisis data graf adalah *Graph Sample and Aggregate* (GraphSAGE). Model ini memungkinkan pembelajaran induktif pada grafik besar dengan menggabungkan informasi dari tetangga setiap *node*, sehingga sangat cocok untuk menganalisis pola perilaku pengguna dalam sistem keamanan siber (Song dkk., 2021). Namun, tantangan utama dalam pemanfaatannya adalah interpretabilitas model yang masih rendah. Untuk mengatasi hal ini, pendekatan *Graph Shapley Value Explanations* (GraphSVX) dikembangkan sebagai solusi untuk meningkatkan interpretabilitas model berbasis grafik. GraphSVX menggunakan konsep nilai *Shapley* dari teori permainan untuk mengukur kontribusi setiap fitur dalam keputusan model GNN, sehingga memberikan transparansi yang lebih baik dalam deteksi ancaman (Duval & Malliaros, 2021).

Selain pendekatan teknis, profil perilaku juga menjadi kunci dalam mendeteksi ancaman internal, karena perilaku yang menyimpang dari norma dapat menjadi indikator potensi risiko. Penelitian oleh (Wang dkk., 2023) menunjukkan bahwa dengan menganalisis pola perilaku pengguna dari berbagai sumber data, seperti log aktivitas, interaksi dengan sistem, dan perilaku komunikasi, organisasi dapat mengidentifikasi aktivitas yang mencurigakan lebih awal. Selain itu, pendekatan teknologi seperti GNN, dapat digunakan untuk memodelkan interaksi kompleks antara pengguna dan sistem, sehingga meningkatkan kemampuan deteksi ancaman.

Untuk mendukung penelitian ini, digunakan dataset *CERT Insider Threat* yang dikembangkan oleh CERT Coordination Center (CERT-CC). Dataset ini telah menjadi standar *de facto* untuk penelitian deteksi ancaman internal. Dataset CERT Insider Threat r6.2, rilis terbaru, menyediakan kumpulan dataset sintetis komprehensif yang menstimulasikan perilaku pengguna normal dan berbahaya dalam lingkungan organisasi (Lindauer, 2020). Dataset ini dirancang untuk

membantu peneliti dan praktisi dalam mengembangkan dan menguji model deteksi ancaman internal yang lebih efektif.

Penelitian ini bertujuan untuk mengatasi gap dalam deteksi dan interpretasi ancaman internal dengan mengimplementasikan GraphSAGE dan GraphSVX pada data representatif dari dataset CERT Insider Threat r6.2. Kombinasi kedua model ini diharapkan mampu meningkatkan akurasi serta transparansi dalam mengidentifikasi pola perilaku mencurigakan, sehingga memberikan wawasan berharga bagi pengembangan solusi keamanan siber yang dapat dijelaskan. Dengan pendekatan implementasi dan evaluasi, penelitian ini berkontribusi dalam mengembangkan solusi keamanan siber yang tidak hanya akurat tetapi juga transparan dan dapat diinterpretasi, yang dapat memenuhi kebutuhan mendesak dalam lanskap keamanan informasi modern.

1.2 Perumusan Masalah

Keterbatasan model GraphSAGE dalam mengidentifikasi potensi ancaman internal yang tersembunyi dalam profil perilaku pengguna, khususnya pada individu dengan tingkat risiko tinggi (*user risk*), serta sejauh mana penerapan GraphSVX dapat meningkatkan interpretabilitas representasi graf untuk mengungkap struktur hubungan kompleks dan pola perilaku yang relevan terhadap deteksi ancaman internal.

1.3 Pertanyaan Penelitian

1. Bagaimana kemampuan model GraphSAGE dalam mendeteksi pola ancaman internal berdasarkan profil perilaku pengguna dalam dataset CERT Insider Threat r6.2?
2. Bagaimana pendekatan GraphSVX digunakan untuk menginterpretasi profil perilaku pengguna yang mengarah pada ancaman internal dalam model GraphSAGE?

1.4 Tujuan Penelitian

Penelitian ini bertujuan untuk menerapkan model GraphSAGE dalam mendeteksi pola perilaku tidak biasa pada dataset CERT Insider Threat r6.2 melalui

representasi graf yang memetakan hubungan antar aktivitas pengguna. Selain itu, penelitian ini juga bertujuan untuk mengintegrasikan pendekatan interpretatif GraphSVX guna menjelaskan faktor-faktor yang memengaruhi hasil deteksi. Secara keseluruhan, penelitian ini ditujukan untuk memahami karakteristik perilaku pengguna yang menyimpang sebagai upaya mengidentifikasi potensi ancaman internal dalam sistem organisasi.

1.5 Manfaat Hasil Penelitian

Manfaat yang diharapkan dari penelitian ini antara lain:

1. Memberikan pengetahuan mendalam mengenai penerapan GNN, khususnya implementasi GraphSAGE dan GraphSVX, untuk mendeteksi dan menginterpretasi ancaman internal, dengan fokus pada pemodelan hubungan antar entitas dalam profil perilaku pengguna.
2. Mengevaluasi efektivitas pendekatan berbasis graf dalam meningkatkan akurasi deteksi ancaman internal serta transparansi model, sehingga dapat mendukung pengambilan keputusan berbasis data dalam sistem keamanan siber.
3. Menyediakan hasil penelitian yang dapat menjadi acuan dalam pengembangan sistem deteksi ancaman internal yang lebih transparan dan dapat diandalkan di masa depan.

1.6 Batasan Penelitian

Penelitian ini memiliki batasan ruang lingkup untuk menjaga fokus analisis dan kedalaman kajian. Dataset yang digunakan berasal dari *CERT Insider Threat Dataset* versi r6.2, yang merupakan data simulasi dan belum sepenuhnya merepresentasikan kondisi nyata dalam organisasi dunia industri. Dari lima skenario yang tersedia dalam dataset tersebut, penelitian ini hanya memfokuskan analisis pada skenario 2 dan skenario 5, karena keduanya menggambarkan pola perilaku pengguna yang relevan dengan pendekatan deteksi berbasis pemodelan profil perilaku.

Skenario 2 menggambarkan seorang karyawan yang mulai mencari pekerjaan di tempat lain, melakukan kontak dengan kompetitor, dan sebelum

mengundurkan diri, mencuri data dengan menggunakan perangkat penyimpanan eksternal. Sementara itu, skenario 5 menggambarkan karyawan yang terdampak pemutusan hubungan kerja dan mengunggah dokumen ke layanan penyimpanan daring untuk kepentingan pribadi. Kedua skenario ini dipilih karena memperlihatkan transisi perilaku yang berpotensi mengarah pada ancaman internal dan memberikan konteks yang sesuai untuk pendekatan deteksi menggunakan model graf.