BAB 1
PENDAHULUAN

1.1 Latar Belakang Masalah

Serangan jaringan berbasis *network-level* seperti DoS (*Denial of Service*), *brute-force*, dan *reconnaissance* tetap menjadi ancaman yang persisten, dengan data telemetri global menunjukkan lebih dari 8 juta serangan DoS dalam paruh kedua tahun 2025 dan sekitar 67,65 miliar kejadian *brute-force*, serta sekitar 640 miliar kejadian *reconnaissance* per tahun [1, 2]. Ketiga kategori serangan ini secara konsisten menjadi fokus pengujian sistem deteksi intrusi pada infrastruktur jaringan nyata, termasuk pada perangkat *router* dan *server* yang rentan terhadap eksploitasi protokol, penjenuhan kapasitas layanan, dan pengintaian sistematis terhadap kerentanan [3]. Kondisi ini mendorong kebutuhan akan sistem deteksi anomali berbasis *machine learning* yang mampu mempelajari pola *traffic* secara adaptif dan mengidentifikasi perilaku menyimpang sebagai indikasi potensi serangan secara lebih akurat [4].

Namun, sebagian besar model deteksi anomali berbasis *machine learning* masih bersifat *black-box*, sehingga proses pengambilan keputusan sulit dipahami secara intuitif oleh pengguna [5]. Kondisi ini menyebabkan analis keamanan sering kali mengalami keterbatasan dalam memahami dasar klasifikasi yang dihasilkan oleh sistem, sehingga proses validasi hasil deteksi dan investigasi insiden masih memerlukan interpretasi manual yang cukup intensif [6]. Keterbatasan ini berimplikasi langsung pada kecepatan respons terhadap insiden, karena analis tidak dapat segera menentukan tindakan mitigasi yang tepat tanpa memahami fitur jaringan mana yang menjadi dasar suatu deteksi dianggap anomali [7].

Meskipun berbagai penelitian telah mengembangkan pendekatan XAI (*Explainable Artificial Intelligence*) untuk meningkatkan transparansi sistem deteksi, sebagian besar pendekatan tersebut masih berfokus pada evaluasi teknis dan karakteristik data. Pemanfaatan hasil interpretasi dalam mendukung proses analisis dan pengambilan keputusan analis keamanan secara sistematis masih relatif terbatas [6]. Akibatnya, masih terdapat ruang pengembangan dalam mengintegrasikan hasil *explainability* ke dalam konteks operasional keamanan siber yang dapat ditindaklanjuti secara langsung [7, 8]. Hal ini menunjukkan bahwa sistem deteksi anomali yang ideal tidak hanya perlu akurat dalam mengklasifikasikan serangan,

tetapi juga mampu menghasilkan penjelasan kontekstual dan rekomendasi mitigasi yang dapat langsung digunakan oleh praktisi keamanan jaringan [9].

Berdasarkan kondisi tersebut, penelitian ini mengusulkan pengembangan sistem deteksi anomali jaringan yang mengintegrasikan XGBoost (*Extreme Gradient Boosting*) sebagai *classification engine*, serta metode XAI melalui SHAP (*SHapley Additive exPlanations*) dan LIME (*Local Interpretable Model-agnostic Explanations*) untuk meningkatkan transparansi keputusan model. XGBoost dipilih karena kemampuannya dalam menangani data berdimensi tinggi dengan performa yang konsisten, sementara SHAP dan LIME digunakan secara komplementer untuk menyediakan interpretasi global dan lokal terhadap hasil prediksi [10]. Sebagai penguatan pada lapisan interpretasi, penelitian ini menambahkan komponen RAG (*Retrieval-Augmented Generation*) yang berfungsi mentransformasikan atribusi fitur yang bersifat teknis menjadi penjelasan naratif berbasis pengetahuan keamanan siber, sehingga analis keamanan dapat memperoleh konteks ancaman dan rekomendasi mitigasi yang dapat ditindaklanjuti secara langsung.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, maka rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana kinerja model XGBoost dalam mendeteksi anomali jaringan pada *dataset* GeNIS yang mencakup berbagai jenis serangan modern?
2. Bagaimana kemampuan metode XAI menggunakan SHAP dan LIME dalam menghasilkan konteks ancaman yang jelas dan akurat terhadap hasil deteksi anomali jaringan yang dihasilkan model XGBoost?
3. Bagaimana integrasi RAG dapat menghasilkan rekomendasi mitigasi yang *actionable* dan sesuai dengan kondisi infrastruktur jaringan nyata berdasarkan penilaian praktisi keamanan jaringan?

1.3 Batasan Permasalahan

Agar penelitian ini tetap terfokus dan sesuai dengan tujuan yang telah ditetapkan, maka batasan permasalahan yang diterapkan adalah sebagai berikut:

1. Pengumpulan data jaringan dilakukan melalui PCAP (*Packet Capture*) dengan analisis terbatas pada *Layer* 2 (*Data Link*), *Layer* 3 (*Network*), dan

Layer 4 (*Transport*) dari model OSI. *Layer* 5–7 tidak dicakup karena analisis *flow-based* pada ketiga lapisan tersebut sudah cukup untuk mendeteksi anomali tanpa memerlukan inspeksi *payload* aplikasi.

2. Ekstraksi fitur dari data PCAP dilakukan menggunakan alat HERA (*Holistic nEtwork featuRes Aggregator*) dengan pendekatan *5-tuple flow aggregation* berbasis kombinasi *source IP*, *destination IP*, *source port*, *destination port*, dan *protocol*.

3. Penelitian difokuskan pada deteksi anomali jaringan berdasarkan *flow-level features*, tidak mencakup *packet-level deep inspection* atau *payload analysis*.

4. Penelitian menggunakan *multi-class classification* dengan 13 kelas berdasarkan *SubCategoryLabel* dari *dataset* GeNIS (*GECAD Network Intrusion Signatures*), mencakup kategori serangan DoS, *brute force*, dan *reconnaissance*.

5. Algoritma *machine learning* yang digunakan adalah XGBoost (*Extreme Gradient Boosting*).

6. Teknik *explainability* yang digunakan dibatasi pada SHAP dan LIME.

7. Penelitian menggunakan *dataset* GeNIS sebagai *primary dataset* pelatihan model.

8. Data yang digunakan bersifat historis. Penelitian tidak mencakup *real-time streaming detection* atau *online learning scenarios*.

9. Evaluasi penelitian berfokus pada performa deteksi model, kualitas interpretabilitas hasil XAI (*Explainable Artificial Intelligent*), serta kualitas narasi RAG (*Retrieval Augmented Generation*) berdasarkan penilaian pakar *cybersecurity*.

10. Simulasi serangan *recon-dns* tidak dilakukan karena teknik eksekusinya memerlukan konfigurasi infrastruktur DNS khusus yang tidak tersedia pada lingkungan pengujian, sehingga kelas ini hanya direpresentasikan melalui label pada *dataset* GeNIS.

11. Sistem berpotensi tidak dapat mendeteksi serangan di luar 10 kelas serangan yang sudah disimulasikan, karena model XGBoost bersifat *supervised* dan

tidak memiliki mekanisme generalisasi terhadap pola serangan yang belum pernah dipelajari selama pelatihan.

1.4 Tujuan Penelitian

Tujuan yang ingin dicapai dalam penelitian ini adalah sebagai berikut:

1. Menganalisis dan mengevaluasi kinerja model XGBoost (*Extreme Gradient Boosting*) dalam mendeteksi anomali jaringan pada *dataset* GeNIS.
2. Mengkaji kemampuan metode XAI menggunakan SHAP dan LIME dalam menghasilkan konteks ancaman yang jelas dan akurat terhadap hasil deteksi anomali.
3. Mengimplementasikan dan mengevaluasi sistem RAG berbasis Haystack yang mampu menghasilkan rekomendasi mitigasi yang *actionable* dan sesuai dengan kondisi infrastruktur jaringan nyata berdasarkan penilaian praktisi keamanan jaringan.

1.5 Manfaat Penelitian

Penelitian ini diharapkan memberikan manfaat bagi beberapa pihak, yaitu peneliti, praktisi keamanan siber, dan akademisi. Adapun manfaat yang diharapkan adalah sebagai berikut:

1. Bagi peneliti, penelitian ini menambah wawasan mengenai deteksi anomali jaringan menggunakan XGBoost, penerapan XAI dalam interpretasi model, serta pengembangan kemampuan integrasi RAG untuk menghasilkan penjelasan yang *actionable*.
2. Bagi praktisi keamanan siber dan *security analyst*, penelitian ini membantu memahami faktor-faktor yang memengaruhi keputusan model dalam mendeteksi anomali, mendukung proses validasi dan investigasi insiden keamanan siber secara lebih sistematis, serta mengurangi ketergantungan pada interpretasi manual dalam analisis *traffic* jaringan.
3. Bagi institusi dan organisasi, penelitian ini mendukung peningkatan pengawasan keamanan jaringan berbasis *machine learning*, membantu identifikasi pola serangan dan potensi ancaman secara lebih cepat dan

terstruktur, serta memberikan gambaran mengenai penerapan sistem deteksi anomali yang *interpretable* dan *actionable*.

4. Bagi akademisi dan peneliti selanjutnya, penelitian ini menjadi referensi dalam pengembangan penelitian terkait deteksi anomali jaringan berbasis XAI, serta menjadi rujukan dalam pengembangan metode analisis interpretabilitas model dan integrasi RAG.

1.6 Sistematika Penulisan

Berisikan uraian singkat mengenai struktur isi penulisan laporan penelitian, dimulai dari Pendahuluan hingga Simpulan dan Saran.

Sistematika penulisan laporan adalah sebagai berikut:

- **Bab 1 PENDAHULUAN**
  Bab ini menguraikan latar belakang masalah yang melatarbelakangi penelitian, meliputi persistensi ancaman serangan jaringan berbasis DoS, *brute-force*, dan *reconnaissance* pada infrastruktur jaringan nyata, keterbatasan model deteksi anomali berbasis *machine learning* yang bersifat *black-box* dalam mendukung proses investigasi dan respons insiden, kesenjangan antara hasil *explainability* teknis dengan kebutuhan operasional analis keamanan, serta kebutuhan akan sistem yang mampu menghasilkan penjelasan kontekstual dan rekomendasi mitigasi yang *actionable*. Bab ini juga memuat rumusan masalah, batasan permasalahan, tujuan penelitian, manfaat penelitian, dan sistematika penulisan.

- **Bab 2 LANDASAN TEORI**
  Bab ini memaparkan dasar teori dan kajian literatur yang menjadi landasan penelitian, mencakup konsep deteksi anomali jaringan, pengumpulan data berbasis PCAP dan agregasi *5-tuple flow*, alat ekstraksi fitur HERA, algoritma XGBoost, metode XAI, serta pendekatan RAG.

- **Bab 3 METODOLOGI PENELITIAN**
  Bab ini menjelaskan metodologi penelitian secara terstruktur, meliputi alur penelitian dari pengumpulan data PCAP hingga evaluasi sistem, tahapan *preprocessing* dan *feature engineering*, pengembangan model XGBoost, integrasi XAI, integrasi sistem RAG dengan *knowledge base MITRE*

ATT&CK dan *CVE*, serta rancangan evaluasi model, kualitas penjelasan XAI, dan kualitas narasi RAG berdasarkan penilaian pakar.

- **Bab 4 HASIL DAN DISKUSI**
  Bab ini menyajikan hasil implementasi dan evaluasi sistem yang dikembangkan, mencakup hasil deteksi anomali oleh model XGBoost, analisis *domain shift* pada evaluasi data *traffic* kampus, analisis interpretasi dari SHAP dan LIME, serta kualitas narasi konteks ancaman dan rekomendasi mitigasi yang dihasilkan oleh sistem RAG berdasarkan penilaian praktisi keamanan jaringan.
- **Bab 5 SIMPULAN DAN SARAN**
  Bab ini menyampaikan simpulan dari keseluruhan penelitian berdasarkan hasil dan analisis yang telah dilakukan, serta saran untuk pengembangan penelitian selanjutnya.