BAB 1
PENDAHULUAN

1.1 Latar Belakang

Dalam beberapa tahun terakhir, perkembangan teknologi informasi dan komunikasi semakin pesat. Hal ini membuka peluang bagi kemajuan dalam berbagai sektor, termasuk perbankan, pemerintahan, pendidikan, dan industri. Namun, seiring dengan manfaat yang diberikan, ancaman di dunia maya atau serangan siber juga semakin meningkat. Data dari lembaga penelitian menunjukkan bahwa serangan siber meningkat setiap tahunnya, baik dari segi jumlah maupun tingkat kompleksitasnya. (Hapsari & Pambayun, 2023)

Bentuk serangan siber yang paling umum meliputi serangan DDoS (*Distributed Denial of Service*), *malware*, *phishing*, serta berbagai jenis eksploitasi jaringan dapat mengakibatkan kerugian baik secara finansial maupun reputasi suatu organisasi. Oleh karena itu, keamanan siber telah menjadi prioritas utama dalam dunia teknologi, dengan tujuan untuk melindungi data dan infrastruktur dari serangan yang dapat berdampak signifikan bagi individu maupun organisasi.

Salah satu teknologi yang dikembangkan untuk menjaga keamanan jaringan adalah *Intrusion Detection System* (IDS). IDS berfungsi untuk memantau lalu lintas jaringan, mendeteksi aktivitas yang mencurigakan, dan memberikan peringatan dini terhadap potensi ancaman. IDS dapat diklasifikasikan menjadi dua kategori utama, yaitu berbasis tanda tangan (*signature-based*) dan berbasis anomali (*anomaly based*). Metode berbasis tanda tangan dikembangkan untuk mendeteksi pola serangan yang sudah dikenali melalui basis data tanda tangan. Di sisi lain, metode anomali didesain untuk mendeteksi aktivitas yang menyimpang dari perilaku normal yang sudah didefinisikan, sehingga lebih mampu mendeteksi serangan *zero day*. (Yang & Shami, 2022)

Namun, kedua metode konvensional ini memiliki keterbatasan: metode berbasis tanda tangan gagal mendeteksi serangan baru (*zero-day*), sementara metode berbasis anomali rentan menghasilkan kesalahan deteksi positif. Untuk mengatasi keterbatasan dari kedua metode ini, pendekatan berbasis *Machine Learning* (ML) dikembangkan sebagai solusi adaptif dan akurat dalam mendeteksi

berbagai jenis serangan. (Yang & Shami, 2022)

Algoritma yang digunakan dalam penelitian ini adalah algoritma *Random Forest*. *Random Forest* merupakan algoritma berbasis *ensemble learning* yang menggabungkan hasil dari banyak pohon keputusan (*decision trees*) untuk mencapai prediksi yang lebih akurat dan stabil. Penelitian menunjukkan bahwa algoritma ini memiliki keunggulan dalam menangani data yang kompleks dan berukuran besar, serta mampu menangani ketidakseimbangan data, yang merupakan karakteristik umum dalam lalu lintas jaringan. (Rastogi dkk., 2022)

Penelitian ini bertujuan untuk merancang dan mengimplementasikan sistem deteksi intrusi berbasis *Machine Learning* menggunakan algoritma *Random Forest*. Melalui penerapan algoritma *Random Forest*, sistem ini diharapkan mampu mengenali pola-pola serangan baru dan mengurangi tingkat kesalahan deteksi, sehingga meningkatkan keamanan jaringan secara keseluruhan. Adapun *dataset* yang digunakan adalah *dataset* CIC-IDS 2017 yang memuat berbagai jenis serangan siber. Dengan pendekatan ini, diharapkan sistem yang dihasilkan dapat menjadi solusi yang adaptif dan berdaya guna dalam mendeteksi dan mencegah serangan siber pada jaringan, sehingga dapat mendukung upaya perlindungan terhadap aset dan informasi penting yang dimiliki oleh organisasi maupun individu.

1.2 Perumusan Masalah

Berdasarkan latar belakang di atas, rumusan masalah dalam penelitian ini adalah:

1. Bagaimana mengimplementasikan *machine learning*, khususnya algoritma *Random Forest*, dalam sistem deteksi intrusi untuk mendeteksi serangan *cyber* pada jaringan?
2. Seberapa baik algoritma *Random Forest* dalam mendeteksi serangan *cyber* tertentu, seperti DOS/DDOS, *Heartbleed*, *Brute Force*, XSS, SQL, *Injection*, *Infiltration*, *Port Scan*?

1.3 Batasan Masalah

Agar ruang lingkup penelitian tidak terlalu luas, batasan masalah yang diterapkan adalah sebagai berikut:

1. Sistem hanya melakukan deteksi intrusi, tanpa pencegahan atau respons otomatis.
2. Sistem akan diimplementasikan pada jaringan terisolasi yang terpapar internet.
3. Algoritma yang menjadi fokus utama adalah *Random Forest*, namun algoritma *machine learning* lainnya (misalnya SVM, *Logistic Regression*, XGBoost, LightGBM, dan *Neural Network*) juga akan diujikan.
4. Data serangan *cyber* yang digunakan untuk pelatihan dan pengujian bersumber dari *dataset* yang tersedia secara publik, yaitu CIC-IDS 2017.
5. Jenis serangan *cyber* yang dideteksi hanya terbatas pada jenis serangan yang tersedia di dalam *dataset* CIC-IDS 2017

1.4 Tujuan dan Manfaat

1.4.1 *Tujuan*

Adapun tujuan dari penelitian ini adalah sebagai berikut:

1) Mengimplementasikan sistem deteksi intrusi yang berbasis *machine learning* menggunakan algoritma *Random Forest*.
2) Menghasilkan prototipe sistem deteksi intrusi yang dapat diaplikasikan pada lingkungan jaringan nyata.

1.4.2 *Manfaat*

Hasil dari penelitian ini diharapkan dapat memberikan manfaat sebagai berikut:

1. Menyediakan model IDS berbasis *Random Forest* yang mampu mendeteksi serangan pada jaringan skala kecil-menengah.
2. Meningkatkan deteksi serangan dinamis (seperti DDoS, *Injection*, dan *zero-day attack*) yang tidak teridentifikasi oleh sistem berbasis *signature* tradisional.