BAB I

PENDAHULUAN

1.1 Latar Belakang

Kemajuan teknologi digital yang pesat telah mendorong integrasi *Internet of Things* (*IoT*) dalam berbagai sektor seperti industri, kesehatan, transportasi, hingga pertahanan. *IoT* memungkinkan berbagai perangkat cerdas untuk saling terhubung dan bertukar data secara otomatis melalui jaringan internet. Namun, pertumbuhan eksosistem ini juga membawa tantangan serius dalam hal keamanan jaringan, karena setiap titik koneksi menjadi potensi pintu masuk bagi serangan siber (Babovic, Protic dan Milutinovic, 2016).

Seiring meningkatnya jumlah perangkat *IoT*, frekuensi dan kompleksitas serangan terhadap jaringan *IoT* turut ikut meningkat seperti serangan *Distributed Denial of Service* (*DDoS*), *brute-force login*, dan infiltrasi jaringan (Thamilarasu dan Chawla, 2019). Laporan terbaru menunjukkan bahwa banyak perangkat *IoT* dibuat menjadi *botnet* atau target serangan karena lemahnya mekanisme autentikasi, enkripsi, dan *update firmware* (Bagaa dkk., 2020). Pendekatan baru dalam deteksi dini ancaman terhadap lalu lintas jaringan yang tidak dapat ditangani secara manual atau dengan sistem *signature-based* konvensional perlu ditemukan.

Proses forensik digital memiliki salah satu *file* yang umum digunakan yang bernama *file Packet Capture* (PCAP), yang berisi rekaman lalu lintas jaringan mentah dalam format *.pcap* (Hoelz dkk., 2008). *File* PCAP digunakan untuk menelusuri aktivitas jaringan secara rinci dan membantu dalam rekonstruksi

kejadian pasca-serangan tetapi klasifikasi manual terhadap ribuan hingga jutaan baris lalu lintas dalam *file* PCAP memerlukan waktu, tenaga ahli dan pengetahuan mendalam terhadap protokol serta pola serangan (Mukherjee dan Haque, 2018).

Machine Learning telah terbukti efektif dalam melakukan deteksi anomali dan klasifikasi pola serangan dalam lalu lintas jaringan secara cepat dan adaptif (Mohammed, dkk., 2016; Mahesh, 2020). Kebanyakan penelitian hanya fokus pada pelatihan model klasifikasi tanpa mengembangkan sistem lengkap yang mampu memproses *file* PCAP dan menyajikan hasil deteksi secara langsung dalam antarmuka pengguna (Ennaji, dkk., 2021).

Berdasarkan itu, penelitian ini bertujuan untuk menjembatani kesenjangan tersebut dengan cara mengembangkan aplikasi PCAP *Analyzer* berbasis web yang mampu menganalisis *file* PCAP, melakukan praproses data otomatis, serta mendeteksi dan mengklasifikasikan jenis serangan dengan menggunakan algoritma *ensemble Machine Learning* seperti *Random Forest*, *AdaBoost*, dan *Gradient Boosting* (Pedregosa dkk., 2011). *Dataset* yang digunakan adalah CICIDS2018, yang merupakan salah satu dataset *benchmark* paling kaya dan realistis dalam domain keamanan jaringan, mencakup berbagai jenis serangan seperti *DDoS*, *Brute-force*, *Botnet*, dan lainnya. Alasan lain dipilihnya *dataset* ini dibandingkan dengan *dataset* lain karena tersedianya aplikasi berbasis terminal bernama *CICFlowMeter-v3* yang tersedia secara publik untuk digunakan.

Aplikasi yang dikembangkan pada penelitian ini diharapkan dapat membantu analis forensik, *administrator* sistem, dan praktisi keamanan siber dalam melakukan deteksi yang cepat dan akurat. Aplikasi yang dibangun memiliki

kemampuan mengolah data pada *file* PCAP secara otomatis dan menghasilkan klasifikasi yang dilakukan oleh model *Machine Learning* yang telah divalidasi, penelitian ini memberikan kontribusi nyata terhadap otomatisasi forensik jaringan dan integrasi *Machine Learning* dalam praktik keamanan siber.

1.2 Rumusan Masalah

Berdasarkan latar belakang di atas, maka dapat dirumuskan masalah dalam penelitian ini sebagai berikut:

1. Bagaimana proses *feature selection* dan *feature engineering* dapat dilakukan terhadap suatu dataset hanya melalui pendekatan analisis data eksploratif (EDA), sehingga mampu menghasilkan dataset yang optimal untuk melatih model klasifikasi serangan siber dengan performa yang baik?
2. Bagaimana proses pengaplikasian model *Machine Learning* yang telah dilatih untuk membuat sebuah aplikasi web REST API?

1.3 Tujuan Penelitian

Berdasarkan latar belakang dan rumusan masalah di atas, adapun tujuan yang didapatkan untuk penelitian ini adalah sebagai berikut:

1. Menerapkan proses *feature selection* dan *feature engineering* berdasarkan EDA dan analisis distribusi data terhadap dataset untuk melatih model *Machine Learning* yang mampu melakukan klasifikasi jenis serangan dengan performa yang optimal.
2. Menjelaskan tahapan-tahapan dari pengembangan sebuah aplikasi web REST API yang dimulai dari pengkonversian *file* PCAP hingga hasil klasifikasi yang bisa digunakan untuk tujuan forensik digital.

1.4 Manfaat Penelitian

Adapun manfaat yang dapat diperoleh dari penelitian ini adalah sebagai berikut:

1. Memberikan wawasan baru dalam melakukan tahapan praproses data untuk melatih sebuah model *Machine Learning*, yang dapat digunakan sebagai referensi untuk melakukan tahapan yang lebih variatif dalam melakukan tahapan praproses data.
2. Mengetahui tahapan pengaplikasian model *Machine Learning* yang telah dilatih kedalam sebuah aplikasi dan juga memberikan solusi yang efisien dalam melakukan proses forensik digital. Implementasi *Machine Learning* dapat membantu dalam menganalisis data yang lebih banyak dan lebih cepat melalui aplikasi berbasis web dibandingkan dengan proses manual.

1.5 Batasan Masalah

Adapun batasan masalah yang terdapat dalam penelitian ini adalah sebagai berikut :

1. Tahapan praproses diterapkan secara seragam pada seluruh tahapan pelatihan model, tanpa adanya penyesuaian khusus untuk masing-masing algoritma.
2. *Random Forest*, *AdaBoosting* dan *Gradient* Boosting yang merupakan tiga algoritma ensemble learning yang akan digunakan untuk melatih model, penelitian ini juga akan memanfaatkan *library* Scikit-learn yang menyediakan tiga algoritma tersebut untuk digunakan secara langsung.
3. Hanya melatih model dengan sebuah *dataset* bernama CICIDS2018 dan menggunakan 3 dari 7 jenis serangan atau label yang tersedia pada *dataset*.

4. Operasi yang digunakan dalam praproses data berfokus pada teknik analisa data atau dengan EDA.