BAB I
PENDAHULUAN

1.1 Latar Belakang

Penggunaan teknologi jaringan internet telah menjadi kunci untuk membantu aktivitas individu atau kelompok seperti masyarakat dan perusahaaan [1]. Sekarang, bisnis, perbankan, pendidikan, dan kesehatan sangat bergantung pada koneksi internet yang stabil dan aman. Namun, seiring dengan pesatnya perkembangan ini, muncul masalah keamanan jaringan yang sering terjadi [2]. Masalah keamanan itu bisa merupakan serangan *DoS*, namun itu bukan satu-satunya. Jenis lainnya juga mencakup berbagai jenis serangan, seperti *Reconnaissance*, *Backdoor*, *Exploits*, dan serangan lainnya [3]. Serangan-serangan ini dapat menyebabkan ketidakstabilan jaringan, penurunan kinerja sistem, dan kebocoran data [4]. Oleh karena itu, perlunya meningkatkan kemampuan untuk mengidentifikasi aktivitas jaringan yang tidak normal atau kita sebut *anomaly* untuk menjaga integritas dan ketersediaan layanan [5].

Ketika sistem jaringan harus menangani lalu lintas data yang sangat beragam, masalah yang semakin kompleks dan juga data yang begitu besar. Dalam situasi seperti ini, pendekatan yang relevan adalah penggunaan algoritma pembelajaran mesin (*Machine Learning*) [6]. Namun demikian, ada algoritma yang memiliki kinerja yang tidak konsisten; ada algoritma yang unggul dalam kecepatan, sementara yang lain lebih akurat tetapi membutuhkan lebih banyak sumber daya [7].

Berdasarkan uraian diatas, pada penelitian ini meninjau bahwa perlu untuk membandingkan kinerja algoritma klasifikasi dalam mendeteksi serangan jaringan dengan membedakan antara kategori normal dan *anomaly* [8]. Dalam penelitian ini menggunakan dataset *UNSW-NB15*, yang dilihat mencakup berbagai jenis serangan seperti *Fuzzers, Exploits, dan DoS*, dianggap lebih menggambarkan kondisi jaringan saat ini.

Penelitian ini berfokus pada empat algoritma klasifikasi berbasis pohon

keputusan: *Random Forest, XGBoost, LightGBM,* dan *CatBoost*. Keempat algoritma tersebut dipilih karena memiliki kemampuan yang setara dalam menangani data berdimensi tinggi, bersifat *non-linear*, dan kompleks seperti pada dataset *UNSW-NB15*. Meskipun seluruh algoritma ini berakar dari konsep decision tree, masing-masing memiliki keunggulan tersendiri. *Random Forest* dikenal stabil dan mudah digunakan tanpa banyak penyesuaian parameter, sementara *XGBoost, LightGBM,* dan *CatBoost* merupakan algoritma berbasis boosting yang cenderung lebih unggul dalam akurasi, namun memerlukan tuning parameter dan sumber daya komputasi lebih besar. *CatBoost* secara khusus dirancang untuk menangani fitur kategorikal secara langsung tanpa perlu encoding manual, yang menjadikannya sangat sesuai untuk dataset jaringan yang memiliki banyak fitur non-numerik. Untuk menjamin kesetaraan dan validitas perbandingan, semua algoritma diterapkan pada data yang telah melalui proses preprocessing yang seragam. Dengan demikian, keempat algoritma ini dapat dibandingkan secara adil dalam hal performa deteksi terhadap lalu lintas jaringan normal dan anomali.

Masing-masing model dilakukan tuning parameter menggunakan *GridSearchCV* agar performanya optimal. Dalam mengevaluasi model, digunakan beberapa metrik seperti *accuracy, precision, recall, F1-score,* dan *ROC AUC*. Karena data yang digunakan cukup tidak seimbang, *F1-score* dipilih sebagai metrik utama untuk menilai seberapa baik model mengenali kedua kelas secara seimbang. Tujuan dari penelitian ini adalah untuk membandingkan seberapa akurat, stabil, dan efisien keempat algoritma tersebut dalam mendeteksi aktivitas serangan pada jaringan komputer.

1.2 Rumusan Masalah

Berdasar latar belakang masalah yang sudah dipaparkan, maka didapatkan rumusan masalah yaitu:

1. Bagaimana kinerja masing-masing algoritma klasifikasi dalam mendeteksi serangan jaringan berbasis data kategori normal dan anomali?
1. Algoritma klasifikasi mana yang memberikan performa paling optimal dalam mendeteksi serangan jaringan berdasarkan hasil evaluasi metrik performa?

Algoritma klasifikasi mana yang paling relevan untuk diimplementasikan pada kasus serupa di dunia nyata berdasarkan tingkat efektivitas performanya?

1.3 Batasan Masalah

1. Penelitian ini menggunakan dataset UNSW-NB15 dan hanya memfokuskan pada klasifikasi biner, yaitu membedakan lalu lintas jaringan normal dan serangan (anomaly).
2. Algoritma yang dibandingkan meliputi empat model berbasis pohon keputusan, yaitu Random Forest, XGBoost, LightGBM, dan CatBoost. Model-model ini dipilih karena kemampuannya dalam menangani data kompleks dan fitur dalam jumlah besar.
3. Evaluasi performa model mencakup metrik Accuracy, Precision, Recall, F1-Score, dan ROC AUC. Setiap algoritma diuji dengan beberapa kombinasi parameter, dan overfitting dianalisis melalui perbedaan hasil pada data latih dan data uji.

1.4 Tujuan Penelitian

1. Mengevaluasi performa empat algoritma klasifikasi berbasis pohon keputusan, yaitu Random Forest, XGBoost, LightGBM, dan CatBoost, dalam mendeteksi serangan jaringan menggunakan data UNSW-NB15 yang diklasifikasikan menjadi dua kategori: normal dan anomali.
2. Menentukan algoritma yang paling optimal dan efisien berdasarkan hasil evaluasi performa menggunakan lima metrik utama, yaitu Accuracy, Precision, Recall, F1-Score, dan ROC AUC, serta melakukan analisis terhadap stabilitas model dengan mengevaluasi kemungkinan terjadinya overfitting melalui perbandingan performa pada data latih dan data uji.

1.5 Manfaat Penelitian

Berdasarkan tujuan penelitian yang ingin dicapai, penelitian ini diharapkan dapat memberikan manfaat sebagai berikut:

1. Memberikan gambaran perbandingan kinerja beberapa algoritma klasifikasi dalam mendeteksi serangan jaringan, sehingga dapat menjadi referensi

teknis dalam pemilihan algoritma yang sesuai untuk kebutuhan keamanan jaringan.

2. Membantu pengguna, pengembang sistem keamanan, dan organisasi dalam memilih algoritma deteksi intrusi yang paling efisien berdasarkan aspek akurasi, waktu komputasi, dan efisiensi sumber daya.
3. Memberikan kontribusi ilmiah dalam pengembangan sistem deteksi intrusi berbasis machine learning, khususnya untuk klasifikasi lalu lintas jaringan antara kategori normal dan anomali.

1.6 Sistematika Penelitian

Dalam penelitian ini terdapat 5 bab dengan beberapa pokok bahasan. Berikut adalah sistematika penulisan yang ada untuk mempermudah pembaca memahami isi skripsi ini secara menyeluruh:

a. BAB I PENDAHULUAN

Memuat latar belakang masalah, rumusan masalah, batasan masalah, maksud dan tujuan penelitian, manfaat penelitian, metode penelitian dan sistematika penulisan.

b. BAB II TINJAUAN PUSTAKA

Menguraikan kajian pustaka dan dasar teori yang berkaitan dengan penelitian, seperti referensi dengan tema yang sama serta definisi yang berhubungan dengan ilmu dan permasalahan yang sedang diteliti.

c. BAB III METODE PENELITIAN

Didalamnya terdapat tinjauan umum tentang objek penelitian, komponen alat dan bahan apa saja yang dibutuhkan dalam penelitian, analisis masalah, solusi yang ditawarkan, serta langkah-langkah yang dijelaskan pada alur penelitian.

d. BAB IV HASIL DAN PEMBAHASAN,

Berisi tentang pembahasan hasil analisis dari evaluasi yang telah dilakukan pada BAB III.

e. BAB V PENUTUP.

Bagian Penutup menyajikan kesimpulan yang diperoleh dari penelitian.