BAB I
PENDAHULUAN

1.1. Latar Belakang

Seiring dengan pesatnya perkembangan teknologi informasi dan komunikasi, jaringan komputer telah menjadi infrastruktur penting dalam berbagai sektor, mulai dari pemerintahan, bisnis, pendidikan, hingga layanan publik. Namun, kemajuan ini juga dibarengi dengan meningkatnya potensi ancaman keamanan, seperti serangan siber dan intrusi terhadap sistem jaringan. Intrusi ini dapat menyebabkan kerugian besar, baik secara finansial maupun reputasi, sehingga mendorong perlunya sistem deteksi intrusi (*Intrusion Detection System*) yang cerdas.

Jaringan yang bersifat tidak seimbang, yaitu jumlah data normal jauh salah satu tantangan utama dalam pengembangan IDS adalah kemampuan untuk mendeteksi berbagai jenis serangan dalam jaringan data yang bersifat tidak seimbang, yaitu jumlah data normal jauh lebih besar dibandingkan dengan data serangan. Kondisi ini dapat mengakibatkan menurunnya akurasi model dalam mengenali serangan yang jarang terjadi. Untuk mengatasi hal tersebut, teknik penyeimbangan data seperti *Adaptive Synthetic Sampling* (*ADASYN*) digunakan untuk meningkatkan representasi kelas minoritas.

Pembelajaran di sisi lain, penggunaan algoritma pembelajaran mesin algoritma telah terbukti efektif dalam meningkatkan kinerja sistem deteksi intrusi. Salah satu algoritma yang menunjukkan kinerja tinggi dalam berbagai kasus klasifikasi adalah *Extreme Gradient Boosting* (*XGBoost*). *XGBoost* mampu mengatasi *overfitting*, memproses data dalam jumlah besar, dan memiliki akurasi yang tinggi, sehingga cocok diterapkan pada sistem *IDS*.

Penelitian ini memanfaatkan kombinasi antara algoritma *ADASYN* untuk penyeimbangan data dan *XGBoost* sebagai model klasifikasi untuk membangun sistem deteksi intrusi yang baik. Dataset yang digunakan adalah CICIDS 2017, yang merupakan salah satu dataset benchmark dalam penelitian IDS karena mencakup berbagai jenis serangan siber yang beragam.

Tujuan dari penelitian ini adalah untuk mengevaluasi efektivitas teknik oversampling *ADASYN* dalam meningkatkan performa *XGBoost* dalam menangani dataset yang tidak seimbang. Evaluasi dilakukan dengan membandingkan performa model sebelum dan sesudah penerapan teknik ini berdasarkan metrik evaluasi seperti *accuracy*, *precision, recall, dan F1-score*. Hasil penelitian ini diharapkan dapat memberikan wawasan yang lebih mendalam mengenai cara terbaik dalam menangani ketidakseimbangan data menggunakan pendekatan machine learning, serta berkontribusi dalam pengembangan metode yang lebih akurat dan efisien untuk berbagai aplikasi yang membutuhkan deteksi kelas minoritas secara lebih optimal.

1.2. Rumusan Masalah

1. Bagaimana menerapkan tahap *preprosessing* untuk pembersihan data *CICIDS2017* menggunakan Teknik *Oversampling* menggunakan *ADASYN*?
2. Bagaimana menerapkan lingkungan uji coba dan data model yang di usulkan?

1.3. Batasan Masalah

Agar penelitian ini lebih terfokus, batasan masalah yang diterapkan adalah sebagai berikut:

1. Dataset yang digunakan dalam penelitian ini adalah dataset yang tidak seimbang yang berkaitan dengan *Network Intrusion Detection*, yaitu *CICIDS 2017.*
2. Teknik *Oversampling* yang digunakan dalam penelitian ini : *ADASYN*.
3. Algortima model yang digunakan dalam penelitian ini, yaitu : *XGBoost.*

1.4. Tujuan

Tujuan dari penelitian ini adalah:

1. Dapat mengetahui cara penggunaan Teknik *Oversampling* pada dataset intrusi jaringan.
2. Dapat mengetahui pengaruh dari penerapan Teknik *Oversampling* pada data pelatihan terhadap performa model deteksi instusi jenis jaringan.

1.5. Manfaat

Manfaat dari penelitian ini yang diharapkan antara lain :

1. Bagi Mahasiswa sebagai peneliti, penelitian ini diharapkan dapat menjadi referensi ilmiah dalam menangani dataset yang tidak seimbang dan meningkatkan performa model *Network Intrusion Detection*.
2. Bagi *Network Security engineer*, diharapkan dapat membantu dalam pengembangan sistem keamanan untuk mendeteksi serangan pada lalu lintas jaringan.