BAB I
PENDAHULUAN

1.1 Latar Belakang

Di era digital saat ini, internet telah menjadi bagian yang tidak terpisahkan dari kehidupan manusia. Berbagai sektor, seperti pendidikan, kesehatan, pemerintahan, dan bisnis, sangat bergantung pada internet untuk menjalankan aktivitas sehari-hari. Perkembangan teknologi informasi ini membawa kemudahan dalam berkomunikasi dan mengakses informasi secara global. Namun, di sisi lain, meningkatnya penggunaan internet juga menghadirkan tantangan baru dalam aspek keamanan siber, termasuk ancaman serangan siber yang semakin kompleks dan canggih.

Untuk mengatasi tantangan tersebut, berbagai teknologi berbasis kecerdasan buatan *(AI)* telah dikembangkan guna meningkatkan sistem keamanan siber. *AI* dan pembelajaran mesin (*Machine Learning)* memiliki kemampuan untuk menganalisis pola serangan siber dan mengidentifikasi ancaman dengan lebih cepat serta akurat. Sistem deteksi intrusi (*Intrusion Detection System/IDS*) menjadi salah satu metode yang semakin berkembang dengan memanfaatkan teknologi *AI* guna mengidentifikasi aktivitas mencurigakan dalam suatu jaringan komputer.

IDS berperan penting dalam mendeteksi dan mencegah ancaman keamanan jaringan dengan menganalisis lalu lintas data secara *real-time. IDS* dapat dikategorikan menjadi dua jenis utama, yaitu berbasis tanda tangan *(signature-based)* dan berbasis anomali *(anomaly-based).* Metode *anomaly-based* yang menggunakan teknik pembelajaran mesin sangat bergantung pada data pelatihan yang berkualitas agar dapat mendeteksi serangan dengan akurasi tinggi. Salah satu tantangan utama dalam implementasi *IDS* adalah ketidakseimbangan *(imbalance*) data, di mana jumlah sampel data serangan jauh lebih sedikit dibandingkan dengan data normal, yang dapat menyebabkan model pembelajaran mesin mengalami bias dalam mendeteksi serangan.

Untuk mengatasi masalah ketidakseimbangan data ini, beberapa teknik telah dikembangkan, termasuk undersampling, oversampling, dan metode hibrida. Oversampling bekerja dengan menambahkan jumlah sampel data minoritas agar

seimbang dengan data mayoritas dengan teknik seperti *Synthetic Minority Over-sampling Technique (SMOTE)* dengan cara menciptakan sampel sintetis yang menyerupai data asli. Dengan demikian, model pembelajaran mesin dapat mempelajari pola serangan secara lebih efektif dan meningkatkan sensitivitas terhadap serangan yang jarang terjadi. Selain SMOTE, terdapat berbagai teknik oversampling lainnya, seperti *Adaptive Synthetic Sampling (ADASYN)* yang berfokus pada pembuatan sampel baru berdasarkan distribusi data minoritas yang ada. Penelitian ini berfokus pada penerapan teknik oversampling dalam IDS guna meningkatkan akurasi deteksi serangan dalam lingkungan data yang tidak seimbang.

1.2 Rumusan Masalah

Berdasarkan latar belakang di atas, perumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana cara menerapkan model pada tahap *preprocessing* untuk cleaning data cicids2017 menggunakan Teknik Oversampling dengan metode SMOTE?
2. Bagaimana melakukan penerapan Teknik *Imbalance* Data menggunakan SMOTE dengan menggunakan *Machine Learning Random Forest?*
3. Bagaimana cara penerapan dan mengevaluasi mode*l klasifikasi biner* antara trafik *BENIGN* dan *ATTACK* pada dataset CICIDS2017 agar model dapat mendeteksi adanya serangan.

1.3 Batasan Masalah

Agar penelitian ini lebih terarah, beberapa batasan masalah yang diterapkan adalah:

1. Dataset yang digunakan dalam penelitian ini berasal dari dataset IDS yang umum digunakan, yaitu CIC-IDS2017.
2. Teknik *oversampling* yang digunakan adalah SMOTE tanpa kombinasi dengan metode *undersampling* lainnya.
3. Algoritma klasifikasi yang digunakan adalah *Random Forest*.

4. Evaluasi model dilakukan dengan *confusion metrik accuracy, precision, recall,* dan *F1-score.*

5. Penelitian ini tidak membahas aspek implementasi IDS dalam lingkungan nyata, tetapi hanya dalam simulasi berbasis dataset.

1.4 Tujuan

Penelitian ini bertujuan untuk:

1. Menganalisis pengaruh teknik SMOTE dalam penerapan sistem deteksi intrusi berbasis *Random Forest.*
2. Menentukan sejauh mana penerapan SMOTE dapat mengurangi bias akibat ketidakseimbangan kelas dalam dataset IDS.
3. Membandingkan performa model *Random Forest* sebelum dan sesudah diterapkan teknik SMOTE.

1.5 Manfaat

Hasil penelitian ini diharapkan dapat memberikan manfaat sebagai berikut:

1. Bagi Peneliti dan Akademisi: Menjadi referensi dalam pengembangan metode deteksi intrusi berbasis machine learning, khususnya dalam mengatasi ketidakseimbangan data.

2. Bagi Pengembang Keamanan Jaringan: Memberikan solusi praktis dalam meningkatkan akurasi IDS menggunakan teknik SMOTE.

3. Bagi Industri dan Pemerintah: Membantu dalam meningkatkan keamanan sistem jaringan dengan model deteksi intrusi yang lebih akurat dan dapat diandalkan.