BAB I

PENDAHULUAN

1.1 Latar Belakang

Keamanan jaringan adalah aspek yang berguna dari teknologi informasi saat ini. Semakin banyak pengguna dan semakin luas area komunikasi, semakin banyak peluang untuk serangan. Serangan terhadap keamanan jaringan lalu lintas informasi adalah *Distributed Denial of Service* (DDoS) yang melumpuhkan *server* dan sistem jaringan dengan membanjiri jaringan menggunakan paket atau permintaan. Misalnya, dalam sebuah studi oleh Kaspersky Lab pada tahun 2017, 33% organisasi menghadapi serangan DDoS pada tahun 2017, dibandingkan dengan 17% pada tahun 2016. Dari organisasi yang terkena serangan DDoS, 20% adalah bisnis yang sangat kecil, 33% UKM, dan 41% industri perusahaan (Afifaturahman et al., 2021; Zidane, 2022).

Machine Learning digunakan untuk membangun *Intrucion Detection System* (IDS) yang berfungsi mendeteksi dan mengklasifikasikan serangan secara otomatis terhadap jaringan lalu lintas serangan jahat terus berubah dan terjadi dalam skala besar sehingga menimbulkan banyak tantangan yang membutuhkan solusi terukur (Vinayakumar et al., 2019). Terdapat dua teknik *Intrusion Detection System* (IDS) biasanya digunakan yakni *signature based intrusion detection system* dan *anomaly based intrusion detection system* (Niko Suwaryo[1], Ismasari Nawangsih[2], 2021; Winanto, 2016). *Anomaly based intrusion system* bekerja dengan mengacu pada pola serangan yang ada dalam lalu lintas, tetapi bermasalah apabila lalu lintas

tersebut berperilaku tidak normal sehingga tidak bisa mengirimkan peringatan adanya serangan kepada sistem. Lalu lintas data dikatakan anomali, apabila terjadi peristiwa yang mencurigakan dari perspektif keamanan informasi (Agarwal & Mittal, 2012; Winanto, 2016)

Penelitian tentang klasifikasi *anomaly network traffic* dilakukan perbandingan metode yang dihasilkan dari proses klasifikasi berdasarkan nilai akurasi *confusion matrix, precision, recall*, dan f1 *score*. *Naive Bayes*, *Support Vector Machine* (SVM) *Linear*, *SVM Polynomial* dan *SVM Sigmoid* menghasilkan persentase akurasi berturut-turut sebesar 85,055%, 99,995%, 99,999%, dan 99,995%. Persentase akurasi tertinggi diperoleh SVM *Polynomial*, sedangkan *Naive Bayes* menghasilkan persentase akurasi terendah (Fluorida Fibrianda & Bhawiyuga, 2018).

Algoritma *J48* memiliki akurasi yang cukup tinggi dibandingkan algoritma *Naive Bayes* dengan pengaturan *testing*, *atributte* dan *intances* yang sama, algoritma *J48* mendapatkan nilai 81,85% sedangkan *Naive Bayes* 80,17% (Cendana & Permana, 2019).

Penelitian tentang nilai akurasi yang cukup baik dilakukan menggunakan algoritma *Random Forest* dalam mendeteksi serangan DDoS dengan rata-rata akurasi yang dihasilkan adalah 92,8% dan akurasi maksimum bisa mencapai 100%. Nilai *precision* dan *recall* memiliki rata-rata hasil 0.90 nilai *precision* dan 0.93 nilai *recall*. Nilai rata-rata *f1_score* nya didapatkan 0.9 dengan *false rate* 7.2%. Waktu

pengambilan keputusannya juga terbilang singkat dengan rata-rata 282.4 ms atau sekitar 0.3 detik (Harto & Basuki, 2021)

Penelitian ini akan berfokus pada klasifikasi dataset *anomaly network traffic (alldays ddos)* dengan membandingkan algoritma *Random Forest*, *Naive Bayes* dan *J48* dengan parameter *metric accuracy, precision, recall*, *sensitivity*, *spesificity* dan *error rate.*

1.2 Rumusan Masalah

Berdasarkan latar belakang masalah yang telah dijelaskan, maka rumusan masalahnya adalah membandingkan nilai *metric accuracy, precision, recall, sensitivity* dan *spesificity* pada dataset *anomaly network traffic (alldays ddos)* menggunakan algoritma *Random Forest, Naive Bayes* dan *J48.*

1.3 Batasan Masalah

Mengingat banyaknya perkembangan yang bisa ditemuskan dalam permasalahan yang telah didefinisikan pada rumusan masalah, maka perlu adanya batasan-batasan masalah yang jelas. Adapun batasan-batasan permasalahannya adalah sebagai berikut:

1. Dataset dengan nama *anomaly network traffic (alldays ddos)* yang digunakan diperoleh dari *kaggle.com/datasets.*.
2. *Accuracy, Precision, Recall, Sensitivity, Specificity* dan *Error Rate* dipilih sebagai parameter pengujian untuk menguji performa dari algoritma *Random Forest*, *Naive Bayes* dan *J48.*

1.4 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah dipaparkan, maka tujuan dari penelitian ini adalah membandingan tingkat *accuracy, precision, recall, spesificity* dan *sensitivity* dari algoritma *Random Forest, Naive Bayes* dan *J48* pada dataset *anomaly network traffic (alldays ddos)*.

1.5 Manfaat Penelitian

Berikut merupakan Manfaat dalam penelitian yang dapat digunakan dan dimanfaatkan.

1. Penelitian ini diharapkan dapat bermanfaat bagi ilmu perkembangan di bidang teknologi informasi khsususnya mengenai pengolahan dataset *anomaly network traffic (alldays ddos)* dengan menggunakan *data mining*.
2. Bagi perkembangan IPTEK, menambah terobosan terkait pendetesian lalu lintas jaringan yang bersifat anomali dengan menggunakan *data mining*.
3. Menambah pengatahuan dan wawasan yang dapat dijadikan acuan untuk meningkatkan profesionalitas dalam mendeteksi anomali pada lalu lintas jaringan agar lebih meningkatkan keamanan khususnya untuk *IT security engineer*.

1.6 Metodologi Penelitian

Metodologi penelitian ini berisikan tentang tahapan-tahapan penelitian pengolahan data, diantaranya :

1 Pengumpulan Data

Pengumpulan data ini merupakan proses yang dilakukan melalui studi literatur dan observasi data dengan pengamatan langsung.

2 Analisis Permasalahan

Tahapan analisis permasalahan dilakukan setelah data hasil studi lieratur dan observasi didapatkan. Masalah yang ditemukan adalah klasifikasi *anomaly network traffic* pada *dataset anomaly* network traffic (alldays ddos). Solusi yang dipilih adalah membandingkan algoritma *Random Forest, Naive Bayes* dan *J48.*

3 Penerapan Algoritma

Tahapan ini dilakukan dengan cara mengubah format *dataset anomaly network traffic (alldays ddos)* dari "csv" menjadi "arff" sebelum diterapkan pada algoritma *Random Forest, Naive Bayes* dan *J48* dengan *tools Weka*.

4 Penarikan Kesimpulan

Tahapan ini didapatkan hasil nilai *accuracy, precision, recall, sensitivity, specificity* dan *error rate* dari uji coba penerapan algoritma *Random Forest, Naive Bayes* dan *J48* menggunakan tools Weka terhadap *dataset anomaly network traffic (alldays ddos)*.

1.7 Sistematika Penulisan

Sistematika penulisan yang digunakan dalam penulisan tugas akhir ini dapat diurakan sebagai berikut :

BAB I PENDAHULUAN

Bab ini akan dibahas tentang latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penelitian.

BAB II LANDASAN TEORI

Bab ini akan dibahas tentang teori-teori dan konsep-konsep yang berhubungan dengan penelitian yang dilakukan dan mendukung dalam pemecahan masalahnya. Selain itu, bab ini juga memuat teori-teori dalam pelaksanaan pengumpulan dan pengolahan data serta melakukan penganalisaan.

BAB III METODOLOGI

Bab ini akan dibahas tentang metodologi dan langkah-langkah selama mengerjakan tugas akhir.

BAB IV HASIL DAN PEMBAHASAN

Bab ini akan dibahas mengenai analisa yang dilakukan terhadap hasil pengumpulan, pengolahan dan analisa data yang diperoleh dari hasil penelitian.

BAB V KESIMPULAN DAN SARAN

Bab ini akan dibahas mengenai kesimpulan yang diperoleh dari hasil penelitian dan analisa data yang telah dilakukan serta saran-saran yang dapat diterapkan dari hasil pengolahan data yang dapat menjadi masukan yang berguna kedepannya.