BAB I

PENDAHULUAN

1.1 Latar Belakang

Keamanan jaringan menjadi aspek penting dalam bidang teknologi informasi saat ini. Semakin banyak pengguna dan semakin luas jangkauan komunikasi, maka semakin banyak pula peluang serangan. Sebagai gambaran, pada penelitian yang dilakukan oleh Lab Kaspersky 2017, 33% organisasi mengalami serangan *Distributed Denial of Service* (DDoS) pada tahun 2017, dibandingkan dengan 17% di tahun 2016. Dari organisasi yang terkena serangan DDoS, 20% adalah bisnis yang sangat kecil, 33% adalah bisnis usaha menengah dan 41% adalah perusahaan (Fibrianda & Bhawiyuga, 2018).

Teknik *machine learning* banyak digunakan untuk mengembangkan *Intrusion Detection System* (IDS) untuk mendeteksi dan mengklasifikasikan serangan dunia maya di tingkat jaringan dan tingkat *host* secara tepat waktu dan cara otomatis. Banyak tantangan muncul karena serangan jahat terus berubah dan terjadi dalam volume yang sangat besar yang membutuhkan solusi yang dapat diskalakan (Vinayakumar et al., 2019).

Intrusion Detection System (IDS) biasanya menggunakan dua jenis teknik yakni *signature based intrusion detection system* dan *anomaly based intrusion detection system*. (Candra Adi Winanto, 2016) *Deteksi Serangan Denial of Service* Menggunakan *Artificial Immune System* mengemukakan bahwa mekanisme kinerja dari *Intrusion Detection System* (IDS) dengan menggunakan teknik *signature based*

dapat mendeteksi serangan yang telah diketahui dengan efektif, tetapi belum mampu memprediksi serangan lama dengan pola yang baru. *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 (Agarwal & Mittal, 2012). Lalu lintas data dikatakan anomali, apabila terjadi peristiwa yang mencurigakan dari perspektif keamanan informasi (García-Teodoro, Díaz-Verdejo, Maciá-Fernández, & Vázquez, 2009).

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 (Fibrianda & Bhawiyuga, 2018).

Algoritma Naive Bayes dapat menghasilkan akurasi yang maksimal dengan data latih yang sedikit. Metode *K-Nearest Neighbour* dipilih karena metode tersebut tangguh terhadap data noise. Hasil yang didapatkan menunjukkan metode Naive Bayes memiliki kinerja yang lebih baik dengan tingkat akurasi 70%, sedangkan metode K-Nearest Neighbor memiliki tingkat akurasi yang cukup rendah yaitu 40% (Devita, Herwanto, & Wibawa, 2018)

Algoritma *K-Nearest Neighbour* memiliki akurasi yang cukup tinggi dibandingkan dengan algoritma *Support Vector Machine* (SVM) dan *Neural Network* (NN) untuk kategori *accuracy*, *precison* dan *recall*. Hasil tersebut menunjukkan bahwa algoritma *K-Nearest Neighbour* dapat memecah data dalam keadaan *higher-feature space* sehingga dua kelas yang berbeda dapat dikelompokkan dengan baik (Doshi, Apthorpe, & Feamster, 2018).

Penelitian ini akan berfokus pada klasifikasi dataset *anomaly network traffic* pada *Intrusion Detection System* (IDS) dengan membandingkan algoritma *K-Nearest Neighbour* KNN dan Naïve Bayes dengan parameter *metric accuracy*, *sensitivity* dan *specificity* sehingga akan dihasilkan nilai g-means yang pada penelitian sebelumnya belum dijabarkan pada parameter *metric* tersebut.

1.2 Rumusan Masalah

Berdasarkan latar belakang masalah yang telah dijelaskan, maka rumusan masalahnya adalah bagaimana membandingkan algoritma *K-Nearst Neighbour* (KNN) dan algoritma Naive Bayes untuk menguji parameter *metric accuracy*, *sensitivity* dan *specificity* klasifikasi pada dataset *Intrusion Detection System* (IDS).

1.3 Batasan Masalah

Mengingat banyaknya perkembangan yang bisa ditemukan 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 lalu lintas jaringan yang digunakan adalah *alldays_ddos* yang diperoleh dari *kaggle.com/datasets*.

2. *Accuracy, Precision, Recall, Specificity, Sensitivty* dan *Error Rate* dipilih sebagai parameter pengujian untuk menguji performa dari algoritma *K-Nearest Neighbour* (KNN) dan Naive Bayes.

1.4 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah dipaparkan, maka tujuan dari penelitian ini adalah membandingkan paramameter *metric accuracy*, *sensitivity* dan *specificity* dari algoritma *K-Nearest Neighbour* (KNN) dan Naive Bayes untuk melakukan klasifikasi lalu lintas jaringan yang bersifat anomali pada dataset *Intrusion Detection System* (IDS).

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 lalu lintas jaringan yang bersifat anomali 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 menjelaskan mengenai waktu dan tempat penelitian, tahapan atau prosedur penelitian, jenis penelitian, pendekatan penelitian, objek penelitian, serta variabel penelitian. Prosedur penelitian terdiri dari pengumpulan data, analisis permasalahan dan pencarian solusi, implementasi solusi, serta penarikan kesimpulan.

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