BAB I

PENDAHULUAN

1.1 LATAR BELAKANG

Internet of Things (IoT) telah muncul sebagai teknologi inovatif di mana jaringan perangkat fisik yang sering disebut sebagai benda atau objek pintar saling terhubung dalam skala global [1]. Penggunaan *Internet of Things* (IoT) semakin marak dan kini menjadi fitur umum dalam kehidupan sehari-hari [2]. *Internet of Things* (IoT) secara umum mencakup penggabungan perangkat berwujud yang mampu beroperasi, digerakkan, dan berkomunikasi secara mandiri untuk meningkatkan dan memfasilitasi penyediaan layanan baru di berbagai domain [3]. Perangkat IoT diharapkan memiliki pemanfaatan yang semakin meningkat di berbagai sektor, seperti *smart buildings, smart city, intelligent transportation systems*, dan *healthcare* [4].

Terlepas dari evolusi *Internet of Things* (IoT) yang cepat, IoT tetap menjadi target yang menarik dan rentan bagi pengguna dan penyerang yang berniat jahat karena langkah-langkah keamanannya yang rendah dan kebutuhan daya komputasi yang tinggi di tingkat perangkat [5]. Belakangan ini, ada tren yang di mana perangkat IoT telah memainkan peran penting dalam mendorong serangan DDoS. Hal ini tetap menjadi ancaman yang belum berkurang [6].

Distributed Denial-of-Service (DDoS) adalah metode serangan yang membanjiri *server* dan sistem jaringan dengan membanjiri mereka dengan jumlah paket atau permintaan yang berlebihan, sehingga menyebabkan mereka tidak

berfungsi atau *crash*. Berbagai bentuk serangan DDoS termasuk *ICMP Flood, SYN Flood, IP Packet Flood,* dan lain-lain [7].

ICMP adalah singkatan dari *Internet Control Message Protocol*. ICMP terutama digunakan untuk tujuan diagnostik, pelaporan kesalahan atau menanyakan server mana pun. Protokol tingkat yang lebih tinggi, seperti TCP, memiliki kemampuan untuk mengenali ketika paket tidak mencapai tujuan yang dimaksudkan. Namun, ICMP memiliki tujuan untuk mengidentifikasi masalah yang lebih serius, seperti "TTL terlampaui" dan "membutuhkan lebih banyak fragmen". Protokol ICMP digunakan untuk mengirimkan berbagai pesan yang mengomunikasikan status jaringan. Berfungsinya TCP, IP, dan protokol serupa lainnya bergantung pada peran penting yang dimainkan oleh sebagian besar jenis pesan ICMP. ICMP tidak boleh dianggap berbahaya dan tidak boleh diblokir. Dalam serangan *ICMP flood,* penyerang membanjiri sumber daya yang dituju dengan paket *ICMP echo request* (*ping*), serta paket ICMP yang besar dan jenis ICMP lainnya, untuk memenuhi dan menghambat infrastruktur jaringan korban [8],[9].

Maka, dibutuhkan suatu sistem yang dapat digunakan untuk melindungi jaringan. *Intrusion Detection System* (IDS) adalah tindakan keamanan yang digunakan untuk memantau dan menganalisis jaringan dan sistem komputer untuk mengetahui adanya potensi akses yang tidak sah atau kerentanan [10]. *Intrusion Detection System* (IDS) digunakan sebagai alat teknologi untuk mengidentifikasi dan mengenali serangan siber. *Intrusion Detection System* (IDS) umumnya menggunakan pengklasifikasi *machine learning* untuk mengkategorikan berbagai

jenis serangan [1]. Seperti penelitian [11], menggunakan metode KNN, SVM, dan *Random Forest* serta pada *dataset* CIC_IDS_2017 mendapatkan tingkat akurasi yaitu 95%, 92%, dan 96.66% ketika diberikan parameter yang optimal. Kemudian penelitian [12], menggunakan metode algoritma C5.0 untuk mendeteksi DDoS dengan tingkat akurasi yang cukup tinggi sebesar 98.38%. Selanjutnya pada penelitian [13], menggunakan kombinasi antara 3 jenis seleksi fitur yaitu *filter*, *wrapper*, dan *embedded* dengan metode KNN, SVM, ANN, dan NB dengan akurasi tertinggi diraih metode KNN dengan akurasi 98.30% dengan menggunakan jenis seleksi fitur *wrapper* dengan 6 fitur terpilih. Sedangkan pada penelitian [14], menggunakan metode *Naïve Bayes*, *Random Forest*, DT, MLP, dan KNN mendapatkan akurasi 93.1%, 98.9%, 98.2%, 96.1%, dan 97.7% di mana metode *random forest* mendapatkan akurasi tertinggi yaitu 98.9%.

Di sisi lain, kemajuan teknologi, protokol dan strategi serangan pada jaringan data telah memunculkan kompleksitas baru dalam penelitian *Intrusion Detection System*. Jaringan mengalami arus lalu lintas yang besar, sehingga menghasilkan data yang rumit dengan dimensi data yang sangat tinggi. Salah satu solusi yang disarankan untuk mengatasi masalah dimensi data adalah pemanfaatan teknik seleksi fitur [15]. Teknik seleksi fitur melibatkan eliminasi variabel yang membantu dalam meningkatkan pemahaman data, mengurangi kebutuhan komputasi, mengatasi tantangan yang ditimbulkan oleh data berdimensi tinggi, dan pada akhirnya meningkatkan kemampuan prediksi model [16]. Seleksi fitur dapat diklasifikasikan ke dalam tiga metode yang luas, yaitu metode *filter*, *wrapper*, dan *embedded* [17]. *Forward Selection* (FS) merupakan salah satu seleksi fitur metode

wrapper. *Forward Selection* (FS) dipilih karena efektivitasnya yang telah terbukti dalam mengidentifikasi *subset* fitur yang optimal [13].

Berdasarkan penelitian sebelumnya, penulis menyarankan untuk menggunakan pendekatan *hybrid* dari metode seleksi fitur dan algoritma klasifikasi untuk mengembangkan sistem deteksi dengan performa yang optimal. Dari sekian banyak algoritma klasifikasi yang telah dikembangkan, algoritma *Random Forest* menggunakan pendekatan acak untuk membangun ensemble kolektif pohon keputusan. *Random forest* terdiri dari banyak pohon keputusan. Tidak ada korelasi yang teramati antara masing-masing pohon keputusan dalam algoritma *random forest*. Setelah *random forest* dibangun, ketika sampel input baru dimasukkan, pohon-pohon keputusan dalam hutan akan menilai dan mengkategorikan sampel berdasarkan keputusan mereka [18]. Salah satu manfaat menggunakan *Random Forest* adalah kemampuannya untuk memproses *dataset* dengan jumlah fitur yang banyak secara efektif sekaligus mengurangi risiko *overfitting*. *Random Forest* juga telah menunjukkan keefektifannya dalam mengidentifikasi anomali dalam jaringan komputer [19]. *Support Vector Machine* (SVM) adalah model serbaguna yang menggunakan teknik analisis klasifikasi dan regresi untuk mendefinisikan dan menganalisis pola yang diamati dalam data [13]. *Support Vector Machine* (SVM) menonjol dari algoritma klasifikasi lainnya karena penekanannya pada memaksimalkan *margin* [20].

Berdasarkan penjelasan-penjelasan diatas, penulis akan melakukan deteksi serangan menggunakan metode *Random Forest* dan SVM serta menggunakan *Forward Selection* sebagai seleksi fitur untuk mendapatkan akurasi yang lebih

optimal dalam mendeteksi serangan. Maka pada penelitian ini penulis mengajukan "**DETEKSI SERANGAN I*CMP FLOOD* MENGGUNAKAN METODE *RANDOM FOREST* DAN *SUPPORT VECTOR MACHINE* DAN SELEKSI FITUR *FORWARD SELECTION* (Studi Kasus: Jaringan *Internet of Things* (IoT))**".

1.2 RUMUSAN MASALAH

Pada penelitian [11], [12], [13], dan [14] berbagai metode dan teknik baru telah dikembangkan untuk mengidentifikasi serangan siber, termasuk serangan DDoS dengan metode *Random Forest* telah menunjukkan akurasi yang lebih baik dibandingkan metode *machine learning* yang lainnya yaitu sebesar 98.9%. Seiring dengan kemajuan sistem keamanan jaringan siber, ada peningkatan yang sesuai dalam munculnya bentuk-bentuk serangan baru yang berasal dari modifikasi metode serangan yang sudah ada. Lonjakan volume data (fitur) yang akan dianalisis menimbulkan tantangan dalam mendeteksi serangan. Namun, sangat penting untuk memiliki pengetahuan untuk mengidentifikasi karakteristik yang sesuai dan relevan untuk mendeteksi serangan seperti DDoS [21]. Mengacu pada penelitian [14], disebutkan bahwa seleksi fitur seperti menggunakan *forward selection* dapat meningkatkan akurasi algoritma klasifikasi.

Berdasarkan latar belakang yang telah dijabarkan, terdapat beberapa pertanyaan masalah yang dapat dirumuskan pada penelitian ini, diantaranya:

1. Bagaimana penerapan algoritma *Random Forest* dan *Support Vector Machine* dalam mendeteksi serangan *ICMP Flood*?

2. Bagaimana penggunaan pemilihan fitur menggunakan *Forward Selection*?
3. Bagaimana perbandingan performa metode *Random Forest* dan *Support Vector Machine* dalam mendeteksi serangan *ICMP Flood*?

1.3 BATASAN MASALAH

Batasan masalah yang diambil penulis dari permasalahan penelitian ini diantaranya:

1. Menggunakan *dataset CIC_IoT_Dataset2023*.
2. Penelitian ini menggunakan *tools* weka.
3. Menggunakan 47 fitur yaitu, "Lamanya waktu aliran paket bertahan", "Panjang/ukuran header dari paket data dalam aliran", "Jenis protokol yang digunakan dalam aliran paket, seperti IP, ICMP, UDP, dan TCP.", "Durasi total koneksi atau aliran data", "Tingkat transmisi paket dari satu titik ke titik lain dalam jaringan", "Tingkat transmisi paket keluar dalam aliran jaringan.", "Tingkat pengiriman paket masuk dalam sebuah aliran.", "Jumlah paket dengan *flag FIN* (untuk mengakhiri koneksi)", "Jumlah paket dengan *flag SYN* (untuk memulai koneksi)", "Jumlah paket dengan *flag RST* (untuk reset koneksi)", "Jumlah paket dengan *flag PSH* (untuk mengirim data segera)", "Jumlah paket dengan *flag ACK* (untuk mengakui penerimaan data)", "Jumlah paket dengan *flag ECE* (untuk pengalaman kemacetan data)", "Jumlah paket dengan *flag CWR* (untuk mengakui bahwa gema kemacetan telah diterima)", "Total pengakuan penerimaan data", "Total permintaan pembukaan

koneksi", "Total penutupan koneksi", "Total paket dengan segment data yang penting", "Total koneksi yang direset", "Jumlah aktivitas untuk protokol HTTP pada lapisan aplikasi (*application layer*)", "Jumlah aktivitas untuk protokol HTTPS pada lapisan aplikasi (*application layer*)", "Jumlah aktivitas untuk protokol DNS pada lapisan aplikasi (*application layer*)", "Jumlah aktivitas untuk protokol Telnet pada lapisan aplikasi (*application layer*)", "Jumlah aktivitas untuk protokol SMTP pada lapisan aplikasi (*application layer*)", "Jumlah aktivitas untuk protokol SSH pada lapisan aplikasi (*application layer*)", "Jumlah aktivitas untuk protokol IRC pada lapisan aplikasi (*application layer*)", "Jumlah aktivitas untuk protokol TCP pada *transport layer*", "Jumlah aktivitas untuk protokol UDP pada *transport layer*", "Jumlah aktivitas untuk protokol DHCP pada lapisan aplikasi (*application layer*)", "Jumlah aktivitas untuk protokol ARP pada *link layer*", "Jumlah aktivitas untuk protokol ICMP pada lapisan jaringan (*network layer*)", "Jumlah paket yang menggunakan protokol IPv4 atau IPv6", "Jumlah paket dengan protokol LLC (*Logical Link Control*) pada *link layer*", "Total panjang paket dalam aliran", "Nilai minimum dari panjang paket dalam aliran", "Nilai maksimum dari panjang paket dalam aliran", "Rata-rata nilai data dalam aliran", "Deviasi standar data dalam aliran", "Ukuran variasi panjang paket dalam aliran", "*Interval* waktu antar paket dalam aliran", "Jumlah paket dalam aliran", "Besar atau kekuatan sinyal dalam data aliran", "Jarak data dari pusat aliran yang

mengindikasikan variasi", "Korelasi antara ukuran paket yang masuk dan keluar.", "Variasi data dari nilai rata-rata", "Berat atau bobot dari paket dalam aliran", "Kelas atau kategori hasil, seperti *BenignTraffic* atau *DDoS- ICMP_Flood*"

4. Menggunakan 1 dataset dari *CIC_IoT_Dataset2023* dan menggunakan 2 dari 34 *label* yaitu, *BenignTraffic* dan *DDoS-ICMP_Flood*.
5. Pelatihan model menggunakan *use training set* dan 10-*fold cross validation* dan pengujian model menggunakan *supplied test set*.
6. Evaluasi performa model menggunakan *confusion matrix* untuk mendapatkan hasil *accuracy, precision, recall,* dan *f1-score*.

1.4 TUJUAN DAN MANFAAT PENELITIAN

1.4.1 Tujuan

Berdasarkan pada pertanyaan masalah, maka tujuan penelitian ini diantaranya:

a. Mengimplementasikan metode *Random Forest* dan *Support Vector Machine* dalam mendeteksi serangan *ICMP Flood*.
b. Mengimplementasikan *Forward Selection* dalam pemilihan fitur untuk mendeteksi serangan *ICMP Flood*.
c. Membandingkan performa dari algoritma *Random Forest* dan *Support Vector Machine* dalam mendeteksi serangan *ICMP Flood*.

1.4.2 Manfaat

Adapun manfaat yang didapatkan dari penelitian ini, diantaranya:

a. Mendapatkan hasil komparasi performa yang optimal antara metode *Random Forest* dan *Support Vector Machine* untuk mendeteksi serangan *ICMP Flood*.
b. Mendapatkan fitur-fitur yang relevan agar dapat mendeteksi serangan *ICMP Flood* lebih optimal.
c. Mengetahui metode yang lebih optimal dalam mendeteksi serangan *ICMP Flood* yang dapat dilihat dari hasil akurasi yang didapatkan.

1.5 SISTEMATIKA PENULISAN

Berdasarkan aturan penulisan ilmiah yang benar, laporan penelitian ini disusun dalam sistematika yang sesuai dan terbagi dalam bab-bab berikut:

BAB I : PENDAHULUAN

Bagian pendahuluan mencakup penyajian latar belakang masalah, rumusan masalah, batasan masalah, tujuan dan manfaat dari penelitian, serta sistematika penulisan.

BAB II : LANDASAN TEORI

BAB II memuat landasan teori yang berhubungan dengan pembahasan yang dianalisis. Adapun beberapa landasan teori yang digunakan, seperti: sekilas tentang IoT, *Machine Learning*, DDoS, *Random Forest*, SVM, Pemilihan Fitur, dll.

BAB III : METODOLOGI PENELITIAN

Menjelaskan tentang alur kerangka kerja penelitian, alur eksperimen, persiapan *dataset* sekunder dari *CIC_IoT_Dataset 2023*, persiapan melakukan *data pre-processing*, penggunaan seleksi fitur, penggunaan metode random forest dan SVM, serta alat bantu penelitian yang bertujuan untuk mendukung penelitian.

BAB IV : ANALISIS DAN HASIL

Pada bagian analisis akan dijelaskan mengenai profil *dataset*, proses *data preprocessing*, analisis data yang digunakan untuk mendeteksi serangan *ICMP Flood* pada jaringan IoT.

BAB V : PENUTUP

Bagian penutup akan menjelaskan kesimpulan beserta saran yang berhubungan dengan hasil penelitian