BAB I

PENDAHULUAN

1.1 Latar Belakang

Pesatnya pertumbuhan berbagai perangkat elektronik yang terhubung ke Internet merupakan bukti yang cukup baik dari teknologi *Internet of Things*. Perangkat elektronik yang biasa digunakan dalam rumah tangga, otomasi industri, dan infrastruktur *smart city* sekarang semuanya terhubung dengan Internet. Gartner memperkirakan 11,2 miliar benda yang terhubung internet akan digunakan di seluruh dunia pada 2018 dan pada 2020 akan mencapai 20,4 miliar (Bastos, Shackleton, & El-Moussa, 2018). Semakin banyaknya perangkat elektronik yang terhubung ke internet dapat menyebabkan resiko keamanannya menjadi lebih tinggi dan bukan tidak mungkin hal itu akan menjadi masalah yang cukup serius.

Ancaman terhadap *Internet of Things* (IoT) menargetkan berbagai perangkat keras, termasuk kamera IP, *router* rumah, dan perangkat pintar. Ancaman ini umumnya mempengaruhi sistem berbasis *Linux* (Andrews, 2018). Jenis serangan *Distributed Denial of Service* (*DDoS*) adalah salah satu jenis serangan di dunia maya yang paling menarik, dimana *Cracker* memperalat sejumlah perangkat yang berkemampuan mengakses internet yang dikenal sebagai *botnet* dan kemudian membuat permintaan simultan ke *server* atau berbagai *server* untuk layanan tertentu, sehingga membanjiri *server* dan membuatnya mengabaikan permintaan dari pengguna yang sah (Angrishi, 2017).

Botnet adalah program berbahaya yang dikendalikan dari jarak jauh oleh *botmaster* melalui saluran *Command and Control* (C&C) (Alejandre, Cortés,

& Anaya, 2017). *Botnet* dapat dimanfaatkan untuk melakukan berbagai tugas yang mematikan, seperti serangan DDoS (*Distributed Denial of Service*), serangan APT (*Advanced Persistent Threats*), atau serangan *phishing* (Abraham et al., 2018). Banyaknya penggunaan serangan internet yang menggunakan *botnet*, menghentikannya adalah prioritas utama dari sudut pandang keamanan.

Jenis *malware* baru bernama *Mirai* menargetkan perangkat IoT seperti kamera IP dan *router* rumah. *Mirai* merupakan salah satu *malware* paling berbahaya dalam beberapa tahun terakhir, *malware* ini digunakan untuk membuat *botnet* sekitar 500.000 perangkat IoT yang dikompromikan kemudian dieksploitasi untuk melakukan beberapa serangan *DDoS* terbesar yang pernah diketahui, diantaranya: menyalahgunakan layanan Internet Perancis dan penyedia hosting OVH pada tanggal 22 September 2016, serangan ke *blog* Krebs On Security pada tanggal 30 September 2016 dan penghapusan layanan *DNS* Dyn yang terkenal pada tanggal 21 October 2016 dengan tingkat kecepatan 1,2 Tbps dan merupakan serangan *DDoS* terbesar yang pernah tercatat (De Donno, Dragoni, Giaretta, & Spognardi, 2018).

Serangan *DDoS* dengan menggunakan *botnet mirai* dan diluncurkan oleh perangkat IoT cenderung menjadi besar dan mengganggu sehingga mengatasi ancaman *botnet mirai* merupakan masalah yang mendesak, langkah awal untuk mengatasinya yaitu dengan cara mendeteksi *botnet mirai*. Terdapat berbagai metode untuk mendeteksi suatu *malware* ataupun *botnet*, salah satunya dengan menggunakan metode pembelajaran mesin.

Pembelajaran mesin merupakan solusi untuk menciptakan mekanisme mendeteksi dan mengidentifikasi jenis serangan baru. Teknik ini memiliki peran penting dalam menyediakan fungsi deteksi intrusi berbasis anomali dalam jaringan IoT (Nomm & Bahsi, 2018). Penelitian yang berjudul "*A Comparison of Machine Learning Approaches to Detect Botnet Traffic*" yang dilakukan Abraham (2018) menyebutkan bahwa algoritma *Random Forest* adalah model superior yang terbukti lebih kuat daripada model lainnya (*Logistic Regression, Naive Bayes, Support Vector Machine* dan *Neural Networks)* dan memiliki kinerja terbaik untuk deteksi anomali (Abraham et al., 2018). Berlandaskan hal tersebut mendeteksi *mirai* menggunakan metode pembelajaran mesin dengan algoritma *random forest* dirasa cukup dikarenakan mempunyai kesamaan pada metode dan algoritma untuk melakukan pendeteksian berbasis anomali.

Dataset yang digunakan yaitu dataset publik dari *UCI Repository "detection of IoT botnet attacks N BaIoT"*. Dataset ini dikumpulkan dari *log* 9 perangkat IoT komersial yang terinfeksi *Mirai* dan *Bashlite*. Perangkat yang terinfeksi *Mirai* ada 7 perangkat, terdiri dari 4 perangkat berjenis kamera keamanan dan masing-masing 1 perangkat berjenis bel pintu, *monitor* bayi dan *thermostat*.

Penelitian ini bertujuan untuk mengukur tingkat akurasi algoritma *random forest* dalam mendeteksi serangan *botnet mirai* (*Scan, ACK, SYN, UDP, UDPplain*) terhadap arsitektur perangkat *Internet of Things* (*security camera*).

Berdasarkan latar belakang yang telah dipaparkan sebelumnya, maka tugas akhir ini berjudul **"Pengolahan Data *Traffic* pada Perangkat *Internet of Things* dengan menggunakan Algoritma *Random Forest*"**.

1.2 Rumusan Masalah

Berdasarkan latar belakang masalah maka rumusan masalah yaitu:

a. Bagaimana menerapkan Algoritma *Random Forest (RF)* untuk mendeteksi serangan *Malware Mirai* dalam penyerangan perangkat *Internet of Things?*
b. Bagaimana menguji akurasi dari Algoritma *Random Forest (RF)* untuk mendeteksi serangan *Malware Mirai* dalam penyerangan arsitektur perangkat *Internet of Things?*

1.3 Tujuan Penelitian

Berdasarkan rumusan masalah diatas maka tujuan penelitian yaitu:

a. Menerapkan Algoritma *Random Forest* untuk mendeteksi serangan *Malware Mirai* (*Scan, ACK, SYN, UDP* dan *UDPplain*) terhadap perangkat *Internet of Things* (*security camera*).
b. Menguji akurasi dari Algoritma *Random Forest* terhadap dataset serangan *Mirai* pada perangkat *Internet of Things* berdasarkan 5 fitur terbaik hasil dari proses *selection features* berdasarkan kriteria *Gini index*.

1.4 Batasan Masalah

Adapun batasan masalah dari penelitian yaitu:

a. *Malware* yang diteliti terfokus pada *malware "Mirai"*.
b. Pengujian yang dilakukan terfokus pada perangkat *Internet of Things* berjenis *security camera*.

c. Pengujian yang dilakukan menggunakan metode *machine learning* dengan terfokus pada Algoritma *Random Forest.*
d. Pengujian yang dilakukan menggunakan dataset publik dari *UCI Repository*.
e. Penelitian dilakukan pada sistem operasi *Windows* 10.
f. Aplikasi yang digunakan berbasis *freeware* yaitu *Rapidminer Educational.*

1.5 Manfaat Penelitian

Manfaat dari penelitian ini yaitu agar hasil dari penelitian dapat dimanfaatkan dan digunakan oleh sebagai berikut:

a. Bagi Ilmu Pengetahuan
    1. Menambah wawasan tentang cara mengukur tingkat akurasi algotritma *random forest* menggunakan *Rapidminer*
    2. Menambah wawasan tentang cara mendeteksi anomali menggunakan *detect outlier (distanced)*

b. Bagi Masyarakat Umum
    1. Mengetahui tingkat akurasi algoritma random forest dalam mendeteksi serangan *botnet mirai* (*Scan, ACK, SYN, UDP* dan *UDPplain*)
    2. Mengetahui anomali yang terdapat pada perangkat *Internet of Things*
    3. Mengetahui proses melakukan *sampling* dan *selection features*

1.6 Metodologi Penelitian

Metode penelitian ini menjelaskan mengenai proses atau tahapan penelitian untuk menerapkan algoritma *Random Forest* untuk mendeteksi serangan *malware*

Mirai pada perangkat *Internet of Things*. Pengujian deteksi penyerangan *malware* ini menggunakan metode *machine learning*, dataset yang digunakan merupakan dataset publik dari *UCI Repository* yang diolah menggunakan *tools* yang bersifat *freeware* yaitu *Rapidminer Educational*.

Alur penelitian yang akan dilakukan ada 5 tahapan, yaitu:

a. Studi Literatur

   Studi literatur ini berisi uraian tentang teori, temuan, dan bahan penelitian lainnya. Proses studi literatur juga untuk menginformasikan keterbaruan dari penelitian yang sudah dilakukan.

b. Pengumpulan Data

   Data yang digunakan adalah data sekunder, didapat dari dataset publik *UCI Repository* deteksi serangan botnet IoT. Pengumpulan data ini menginformasikan rincian dataset yang digunakan.

c. Pemrosesan Data

   Pemrosesan data yaitu melakukan *sampling* dan *feature selection* supaya dataset yang digunakan untuk pengujian sudah dalam dimensi yang rendah.

d. Pemodelan

   Pemodelan yaitu proses membuat rancangan model untuk melakukan proses *features selection* dan proses pengujian.

e. Pengujian dan Evaluasi

   Pengujian yaitu proses menguji model menggunakan algoritma *random forest*, sedangkan evaluasi yaitu proses untuk menganalisis hasil dari beberapa pengujian yang telah dilakukan.

1.7 Sistematika Penulisan

Sistematika penulisan untuk memahami lebih jelas isi laporan, materi-materi yang tertera pada laporan tugas akhir ini dikelompokkan menjadi beberapa sub bab dengan sistematika penyampaian sebagai berikut:

BAB I PENDAHULUAN

Berisi tentang latar belakang, perumusan masalah, batasan masalah, tujuan dan manfaat penelitian, metode penelitian, dan sistematika penulisan.

BAB II LANDASAN TEORI

Bab ini berisikan kajian dari penelitian terdahulu dan teori yang berupa pengertian dan definisi yang diambil dari kutipan jurnal, web, atupun buku serta beberapa literature *review* yang berkaitan dengan penyusunan laporan skripsi ini.

BAB III METODOLOGI PENELITIAN

Bab ini berisikan metodologi penelitian yang memberikan gambaran dan alur dari peneliian yang dilakukan, menjelaskan dari metodologi penelitian, kajian teori, analisis dinamis, analisis dinamis lanjut dan prosedur analisis data.

BAB IV HASIL DAN PEMBAHASAN

Bab ini menjelaskan proses pengujian klasifikasi *serangan malware mirai* menggunakan metode *Machine Learning* menggunakan Algoritma *Random Forest* dengan teknik *Anomaly Detection.*

BAB V KESIMPULAN DAN SARAN

Bab ini berisi kesimpulan dan saran yang berkaitan dengan analisa berdasarkan yang telah diuraikan pada bab-bab sebelumnya.