BAB I

PENDAHULUAN

1.1 Latar Belakang

Perkembangan *Internet of Things* belakangan ini sangatlah pesat, hal tersebut terlihat dari meningkatnya jumlah pengguna berbagai perangkat IoT dari waktu ke waktu. IoT dapat menghubungkan berbagai *device* dan saling bertukar data melalui jaringan internet (Rafsanjani et al. 2022). Namun dalam pengimplementasian teknologi tersebut, terdapat berbagai macam ancaman. Salah satu ancaman serius pada teknologi IoT yaitu serangan *Botnet Mirai* (Meidan et al. 2018). Proliferasi perangkat IoT yang tidak aman telah mengakibatkan lonjakan serangan botnet IoT terhadap infrastruktur Internet. Pada Oktober 2016, *Botnet Mirai* memerintahkan 100.000 perangkat IoT (terutama kamera CCTV) untuk melakukan serangan *denial of service* (DDoS) terdistribusi terhadap infrastruktur Dyn DNS (Doshi, Apthorpe, and Feamster 2018).

Serangan DDoS menggunakan *botnet mirai* dan diluncurkan oleh perangkat IoT seringkali berskala besar dan merusak, sehingga mengatasi ancaman *botnet mirai* adalah masalah yang mendesak, dan langkah pertama untuk mengatasinya adalah dengan mendeteksi *botnet mirai* (Widiyasono et al. 2021). Ada beberapa cara untuk mendeteksi *malware* atau *botnet*, salah satunya menggunakan metode *machine learning* (Donno et al. 2018).

Mechine learning digunakan untuk mekanisme deteksi dan Identifikasi jenis serangan baru. Teknik ini menjadi sebuah solusi dalam menyediakan fungsi deteksi intrusi berbasis anomali dalam jaringan IoT (Doshi, Apthorpe, and Feamster 2018). Pada penelitian ini, dipilih tiga algoritma yang berfungsi untuk mengklasifikasikan serangan *malware mirai*, yaitu algoritma *Forest* (RF), algorima *K-Nearest Neighbor* (KNN), dan algoritma *Decision Tree* (DT).

Algoritma *Random Forest* dipilih karena mencapai performa optimal dengan rata-rata nilai akurasi 95,01%, *recall* 90,82%, F1 Score 93,85% dan nilai presisi terbaik 99,23% dalam Investigasi Forensik Jaringan untuk melakukan teknik klasifikasi dan deteksi serangan *Malware Mirai* (Widiyasono et al. 2021). Algoritma *K-Nearest Neighbour* memiliki akurasi yang cukup tinggi dibandingkan dengan algoritma *Support Vector Machine* dan *Neural Network* 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, and Feamster 2018). Algoritma *Decision Tree* lebih baik dari segi akurasi dalam mendeteksi serangan DDoS daripada algoritma *K-Nearest Neighbour*. dalam segi *running-time*, *Decision Tree* lebih baik daripada *K-Nearest Neighbour* karena *Decision Tree* mempunyai *running-time* yang lebih kecil (Ramadhan, Sukarno, and Nugroho 2019).

Penelitian ini akan berfokus pada klasifikasi anomali pada dataset *N-BaIoT* dengan membandingkan algortima *Random Forest, K-Nearest Neighbour,* dan

Decision Tree dengan parameter *metric accuracy, precision, recall* dan AUC (*area under the ROC Curve).*

1.2 Rumusan Masalah

Berdasarkan latar belakang masalah yang telah dijelaskan, maka rumusan masalahnya adalah membandingkan nilai *metric accuracy*, *precision*, *recall*, dan AUC dari algoritma *Random Forest*, algoritma *K-Nearst Neighbour (KNN)*, dan algoritma *Decision Tree* pada dataset *N-BaIoT.*

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 diperoleh dari *website* https://www.kaggle.com/datasets/ dengan dataset yang bernama *"detection of IoT botnet attack (N-BaIoT)"*.
2. Nilai *Accuracy*, *Precision*, *Recall* dan AUC (*area under the ROC Curve)* dipilih sebagai parameter pengujian untuk menguji performa dari algoritma *Random Forest, K-Nearst Neighbour (KNN),* dan *Decision Tree*.

1.4 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah dipaparkan, maka tujuan dari penelitian ini adalah membandingkan tingkat *accuracy, precision, recall,* dan AUC

dari algoritma *Random Forest*, algoritma *K-Nearst Neighbour (KNN)*, dan algoritma *Decision Tree* pada dataset *N-BaIoT*.

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 *N-BaIoT* dengan menggunakan *data mining.*

2 Bagi perkembangan IPTEK, menambah terobosan terkait pendeteksian serangan *Botnet Mirai* pada *Internet of Thing (IoT)* dengan menggunakan *data mining*.

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.