1. Pendahuluan

Serangan siber telah mengalami pertumbuhan yang luar biasa karena IoT (Internet of Things) telah banyak digunakan. "Internet of Things" adalah jaringan objek yang luas yang dapat berkomunikasi satu sama lain dan internet untuk berbagi informasi dan layanan tanpa perlu keterlibatan manusia [1]. Serangan yang paling sering, menurut IBM 2022 [2], adalah server access, DDoS, RAT, insiders, dan credential harvesting activities.

Metode serangan yang paling banyak digunakan oleh peretas adalah DDoS, umumnya dikenal sebagai Distributed Denial of Service (DDoS). Cara kerjanya adalah membuat layanan tidak tersedia, sehingga menolak layanan kepada pengguna. Sebagian besar waktu, mereka memblokir sumber daya hingga layanan tidak dapat digunakan [3].

Machine Learning dan Data Mining telah diselidiki secara ekstensif untuk masalah deteksi keamanan. Ini adalah yang paling efisien untuk menangani kasus deteksi serangan DDoS [4]. Beberapa model digunakan untuk mengatasi masalah tersebut: XGBoost (Extreme Gradient Boosting), AdaBoost (Adaptive Boosting), dan model ensemble. Model ensemble dirancang untuk meningkatkan kinerja dari satu atau lebih classifier dengan melatih beberapa pengklasifikasi yang berbeda dan menggabungkan hasilnya untuk membuat keputusan akhir [5].

Algoritma XGBoost membuat banyak model menggunakan decision tree dan gradient descent, yang kemudian terintegrasi secara berurutan sambil mengoreksi model sebelumnya untuk menghasilkan model optimal akhir [11/6]. Adaboost adalah teknik klasifikasi yang membuat komite pengklasifikasi yang lemah dan meningkatkan kinerja algoritma ML dengan menggabungkannya menjadi pengklasifikasi yang kuat [7].

Model ensemble menggunakan beberapa metode standar, termasuk bagging, boosting, dan stacking. Model yang digunakan dalam pembahasan berikut adalah stacking karena didasarkan pada dataset yang tidak berubah dan algoritma pembelajaran mesin yang berbeda untuk setiap anggota ensemble dan termasuk proses penggabungan model prediktif [8]. Voting classifier akan digunakan untuk menggabungkan beberapa algoritma untuk memastikan hasil prediksi memiliki presisi yang lebih baik [9].

Makalah ini menggunakan dataset dari Canadian Institute for Cybersecurity yang disebut CIC-DDoS2019 [10], versi 03-11, untuk melakukan pengujian dan pelatihan data. Isinya beberapa file untuk diproses, misalnya, SYN, UDP, UDP-Lag, MSSQL, LDAP, NetBIOS, dan Portmap. Dengan XGBoost, algoritma AdaBoost digunakan untuk mendeteksi serangan DDoS. Menggabungkan algoritma XGBoost dan AdaBoost akan menghasilkan hasil yang lebih baik. Dengan menggunakan dataset yang sama, hasil skenario model ensembel akan dibandingkan dengan algoritma XGBoost dan AdaBoost. Secara teori, model ensemble akan memberikan kinerja yang lebih baik jika dibandingkan dengan keduanya. Selain itu, hasil yang akurat dari model ensembel juga ditentukan oleh berbagai parameter dari algoritma itu sendiri yang ditambahkan pada parameter model ensembel yang dibuat.

Tahapan penting dalam model ensemble adalah evaluasi model. Yang umum dan efektif adalah menerapkan K-Fold Cross Validation. Dapat diartikan sebagai suatu prosedur untuk memisahkan data training dan data testing yang bertujuan untuk mencari kombinasi data yang terbaik dalam hal accuracy, precision, F1, recall, dan lain sebagainya. Semakin besar parameter K-Fold yang dimasukkan, semakin lebih banyak waktu diperlukan untuk mendapatkan hasil [11]. Kapasitas hardware juga memiliki peran penting dalam mempercepat proses iterasi. Tulisan ini bertujuan untuk menggabungkan algoritma XGBoost dan AdaBoost untuk menghasilkan hasil yang lebih akurat dalam mendeteksi serangan DDoS.