ABSTRACT

The implementation of the internet of things (IoT) can make everything connected to the internet but the IoT system can be a very easy target for attackers to infiltrate using malware, more than 1.6 billion or 1,637,973,022 traffic anomalies or cyber attacks to be exact. occurring throughout Indonesia throughout 2021, machine learning techniques can be used for the traffic anomaly classification process using the k-nearest neighbor (KNN) algorithm so that it can distinguish benign or malicious traffic data. The traffic anomaly data used is the IoT-23 aposemat dataset, in the dataset there are 23 datasets, then divided into 20 datasets for malicious scenarios and 3 datasets for benign scenarios. However, the dataset used is 20 datasets for malicious scenarios. The 20 datasets are then preprocessed so that they can be used for model training or classification processes. The accuracy value obtained after the model training process is 0.94 or 94%, the model that has been trained in the model can predict new data traffic into benign or malicious, the new data that has been prepared is as many as 25 new data. The prediction of the 25 new data resulted in 20 data predicted to be correct or appropriate and 5 data predicted to be incorrect or inappropriate, the 5 data divided into 3 data that should be predicted to be benign and 2 data that should be predicted to be malicious.

Keywords: Internet of Things, Malware, Machine learning, Classification, K-Nearest Neighbour (K-NN)

ABSTRAK

Penerapan *Internet of Things* (IoT) dapat membuat semuanya terhubung ke internet tetapi sistem IoT dapat menjadi sasaran yang sangat mudah untuk disusupi penyerang dengan menggunakan *malware*, lebih dari 1,6 miliar atau tepatnya 1.637.973.022 anomali *traffic* atau serangan siber (*cyberattack*) yang terjadi diseluruh wilayah Indonesia sepanjang tahun 2021, teknik *machine learning* dapat dimanfaatkan untuk proses pengklasifikasian anomali *traffic* dengan menggunakan algoritma *k-nearest neighbour* (KNN) sehingga dapat membedakan data *traffic* yang bersifat *benign* atau *malicious*. Data anomali *traffic* yang digunakan adalah dataset aposemat IoT-23, didalam *dataset* tersebut terdapat 23 *dataset*, lalu terbagi kedalam 20 *dataset scenario malicious* dan 3 *dataset scenario benign*. Namun *dataset* yang layak digunakan adalah 20 *dataset scenario malicious*. 20 *dataset* tersebut selanjutnya dilakukan data *preprocessing* supaya dapat digunakan untuk proses *training* model atau pengklasifikasian. Nilai akurasi yang didapatkan setelah proses *training* model sebesar 0.94 atau 94%, model yang sudah dilakukan *training* model dapat memprediksi *traffic* data baru kedalam *benign* atau *malicious*, data baru yang sudah disiapkan adalah sebanyak 25 data baru. Prediksi 25 data baru tersebut menghasilkan 20 data diprediksi benar atau sesuai dan 5 data diprediksi salah atau tidak sesuai, 5 data tersebut terbagi menjadi 3 data yang harusnya diprediksi *benign* dan 2 data yang harusnya diprediksi *malicious*.

Kata kunci: *Internet of Things*, *Malware*, *Machine learning*, Klasifikasi, *K-Nearest Neighbour* (K-NN)