ABSTRAK

Malware menjadi sebuah ancaman dalam jaringan yang harus dideteksi sejak dini. Analisa *malware* dapat dilakukan dengan menerapkan proses klasifikasi berdasarkan trafik darknet. Trafik *darknet* merupakan jaringan internet terenkripsi yang dapat menyembunyikan file *malicious* dari penjahat *cyber*. Penelitian terhadap analisis *malware* terutama klasifikasi trafik *darknet* dengan menggunakan algoritma machine learning telah banyak dilakukan, namun terbatas pada pengukuran kinerja algoritma pada *machine learning* untuk analisis *malware*. Pembaruan dataset dan penggunaan algoritma berbeda sangat diperlukan agar analisa *malware* dapat mengidentifikasi perkembangan *malware* secara cepat dan akurat, oleh karena hal tersebut dalam penelitian ini akan dibahas tentang proses analisis ancaman *malware* pada trafik darknet dengan menggunakan salah satu algoritma *machine learning* yaitu K-*Nearest Neighbour*. Algoritma KNN memiliki nilai akurasi yang tinggi dibandingkan dengan beberapa algoritma lain. Hasil penelitian menunjukkan bahwa klasifikasi trafik *darknet* pada dataset CICDarknet 2020 menggunakan KNN lebih baik dari CNN dan LSTM. Nilai akurasi pada dataset CICDarknet 2020 yang memiliki fitur lengkap adalah 97%. Penelitian ini juga menerapkan teknik seleksi fitur dengan menggunakan *information gain* yang mampu mempercepat proses analisa trafik *darknet* yang dibuktikan dengan waktu eksekusi yang lebih cepat yaitu 6 menit 42 detik dengan akurasi 96,17%, berdasarkan hasil tersebut algoritma KNN dapat diterapkan pada klasifikasi *darknet* agar dapat mendeteksi adanya ancaman serangan malware pada hasil klasifikasi *darknet* dengan akurasi dan kecepatan lebih baik.

Kata kunci: *Darknet, Information Gain, KNN, Machine Learning.*

ABSTRACT

Malware is a threat in the network that must be detected early. Malware analysis can be done by applying a classification process based on darknet traffic. Darknet traffic is an encrypted internet network that can hide malicious files from cyber criminals. Many researches on malware analysis, especially darknet traffic classification using machine learning algorithms, have been carried out, but are limited to measuring the performance of machine learning algorithms for malware analysis. Dataset updates and the use of different algorithms are necessary so that malware analysis can identify malware developments quickly and accurately, because this research will discuss the process of analyzing malware threats on darknet traffic using one of the machine learning algorithms, namely K-Nearest Neighbor. The KNN algorithm has a high accuracy value compared to several other algorithms. The results show that the classification of darknet traffic in the CICDarknet 2020 dataset using KNN is better than CNN and LSTM. The accuracy value in the full-featured CICDarknet 2020 dataset is 97%. This study also applies a feature selection technique using information gain that is able to speed up the darknet traffic analysis process as evidenced by a faster execution time of 6 minutes 42 seconds with an accuracy of 96.17%, based on these results the KNN algorithm can be applied to darknet classification in order to detect any threat of attack. malware on darknet classification results with better accuracy and speed.

Keyword**:** *Darknet, Information Gain, KNN, Machine Learning.*