INTISARI

Analisis Data Log Dionaea dan *Clustering* dengan k-Modes menggunakan Python

Nia Efriana br S
15/380222/SV/08029

Abstrak – Dimensi tinggi dan masifnya data log yang menyimpan informasi mengenai serangan siber merupakan beberapa dari kendala yang dihadapi dalam era di mana kebijakan dan keamanan siber selalu berubah untuk mengatasi serangan siber. Penelitian ini bertujuan untuk memperoleh informasi yang berhubungan dengan serangan siber dan untuk menemukan pola dari analisis *clustering* dalam data log dionaea yang berukuran besar, sistem deteksi serangan yang penggunaanya sering digunakan dalam bidang sistem keamanan siber. Harapan dari penelitian ini untuk menggunakan informasi yang telah diperoleh dalam evaluasi untuk meningkatkan sistem yang telah ada. Dalam penelitian ini, analisis data menggunakan metode *Exploratory Data Analysis* dan teknik visualisasi diterapkan untuk memproses data dan menyajikan hasil dalam bentuk grafik agar mudah dibaca. Selain itu, analisis *clustering* dengan algoritma k-Modes menggunakan Python juga diimplementasikan. Hasil penilaian dari *Confusion Matrix*, *Rand Index*, dan *Adjusted Rand Index* menunjukkan angka validitas rendah yang disebabkan oleh ketidakseimbangan yang terjadi pada fitur yang berisi elemen dengan redundansi tinggi sehingga penyebaran *cluster* dan nilai akurasi prediksi label cenderung rendah.

Kata kunci: Dionaea, Python, *Exploratory Data Analysis* (EDA), *Clustering k-Modes*, visualisasi data.

ABSTRACT

ANALYSIS OF DIONAEA LOG DATA AND CLUSTERING WITH K-MODES ALGORITHM USING PYTHON

Nia Efriana br S
15/380222/SV/08029

Abstract - High-dimensional log data with its massive size- where it stores up information about cyber attacks, are few of many concerns that we're facing in this era of ever-changing in cyber policy and security to counteract the rapid evolution of cyber attacks. This research aims to obtain useful information related to cyber attacks and to discover patterns out of clustering analysis on the immense log data of dionaea, the detection system which is broadly use in the field of security system for cyber. The purport of this study is also to use the extracted informations in evaluating any possible approach to improve the already existing network security system. In this study, data analysis with Exploratory Data Analysis method and visualization technique was applied to process the data and to present the result in graphical form to make it easier in reading the insights from the data. Furthermore, clustering analysis with k-Modes algorithm using Python was implemented. The assessment scores from Confusion Matrix, Rand Index, and Adjusted Rand Index are presented, which indicates poor recovery whereas an imbalance which is likely to happen because of the high redundancy of elements in the feature which were selected for clustering. Hence, the uneven distribution with low accuracy score in each cluster for predicting the labels.

Keywords: *dionaea, Python, Exploratory Data Analysis (EDA), clustering k-Modes, data visualization*