ABSTRAK

Serangan siber terhadap layanan jaringan terus meningkat, sementara mekanisme pemantauan *honeypot* Cowrie di Laboratorium *Cyber Security and Networking* Fakultas Ilmu Komputer UPN "Veteran" Jakarta masih dilakukan secara *batch* dan manual, sehingga menimbulkan keterlambatan dalam mendeteksi dan merespons pola serangan yang sedang berlangsung. Kondisi ini mendorong kebutuhan akan sistem deteksi yang mampu memproses data *honeypot* secara kontinu dan otomatis. Penelitian ini bertujuan merancang dan mengimplementasikan sistem *streaming analytics* berbasis Apache Flink yang mengintegrasikan model klasifikasi *k-Nearest Neighbors* (kNN) dan *Logistic Regression* untuk mendeteksi aktivitas normal dan anomali pada data *honeypot* Cowrie, serta mengevaluasi efektivitas *(accuracy, precision, recall, F1-score)* dan efisiensi (latensi) kedua model tersebut. Penelitian menggunakan metode *experimental research* dengan pendekatan kuantitatif. Data *honeypot* Cowrie dikumpulkan dari basis data MongoDB laboratorium, kemudian melalui tahap *preprocessing* meliputi pelabelan, ekstraksi 19 fitur, penanganan ketidakseimbangan kelas dengan SMOTE, serta normalisasi. Kedua model dilatih dengan data *training* dan diuji dengan data *testing* (rasio 80:20), lalu di implementasikan melalui *pipeline* PyFlink yang terhubung dengan PostgreSQL sebagai penyimpanan hasil. Pengujian hipotesis dilakukan menggunakan analisis deskriptif terhadap latensi klasifikasi per *event*, dengan hasil dianalisis dan didiskusikan bersama dosen pembimbing untuk memastikan kesesuaian dengan kaidah ilmiah.

Kata Kunci : Apache Flink, *Honeypot* Cowrie, *k-Nearest Neighbors, Logistic Regression, streaming analytics*

ABSTRACT

Cyberattacks against network services continue to increase, while the monitoring mechanism for the Cowrie honeypot at the Cyber Security and Networking Laboratory of the Faculty of Computer Science, UPN "Veteran" Jakarta, is still performed through batch and manual processing, causing delays in detecting and responding to ongoing attack patterns. This condition drives the need for a detection system capable of processing honeypot data continuously and automatically. This research aims to design and implement an Apache Flink-based streaming analytics system that integrates k-Nearest Neighbors (kNN) and Logistic Regression classification models to detect normal and anomalous activities in Cowrie honeypot data, as well as to evaluate the effectiveness (accuracy, precision, recall, F1-score) and efficiency (latency) of both models. This research employs the experimental research method with a quantitative approach. Cowrie honeypot data is collected from the laboratory's MongoDB database, then processed through preprocessing stages including labeling, extraction of 19 features, class imbalance handling with SMOTE, and normalization. Both models were trained with training data and tested with testing data (80:20 ratio), then implemented through a PyFlink pipeline connected to PostgreSQL as the result storage. Hypothesis testing is conducted using descriptive analysis of per-event classification latency, with results analyzed and discussed together with the supervisor to ensure scientific validity.

Keywords: Apache Flink, Cowrie honeypot, k-Nearest Neighbors, Logistic Regression, streaming analytics