ABSTRAK

Adopsi protokol HTTP/3 *over Quick UDP Internet Connections* (QUIC) dengan TLS 1.3 telah menciptakan tantangan signifikan bagi administrator jaringan. Penerapan enkripsi ini menyebabkan metode *Deep Packet Inspection* tidak efektif karena ketidakmampuan membaca *payload* yang terenkripsi sehingga visibilitas terhadap kategori layanan yang melintas di jaringan menjadi terbatas. Permasalahan ini diperparah oleh minimnya kajian klasifikasi pada level kategori layanan di *dataset* QUIC skala ISP yang mayoritas penelitian terdahulu berfokus pada identifikasi aplikasi spesifik dengan *dataset* non-QUIC. Oleh karena itu, diperlukan pendekatan klasifikasi non-intrusif yang mampu memetakan kategori layanan pada trafik QUIC tanpa dekripsi.

Penelitian ini membangun model klasifikasi kategori layanan trafik QUIC menggunakan algoritma LightGBM dengan optimasi *hyperparameter* berbasis Optuna dan evaluasi strategi *resampling* dalam satu kerangka terpadu. *Dataset* yang digunakan adalah CESNET-QUIC22 (W-2022-44) yang mencakup 20 juta *flows* dengan 17 kategori layanan. Prapemrosesan data meliputi *data cleaning*, *label encoding*, dan ekstraksi fitur statistik (*mean*, *standard deviation*, *skewness*, *kurtosis*) dari histogram paket. Optimasi *hyperparameter* dilakukan menggunakan TPESampler dengan mekanisme *pruning* adaptif, sedangkan dampak ketidakseimbangan kelas dianalisis melalui tiga strategi *resampling*, yaitu RUS, SMOTE, dan RUS-SMOTE.

Hasil penelitian menunjukkan bahwa model LightGBM tanpa *resampling* mencapai *accuracy* 84,17%, *precision* 84,14%, *recall* 84,17%, *F1-score* 84,00%, AUROC 98,31%, dan AUPRC 92,14%. Optuna meningkatkan *accuracy* sebesar 15,65% dibandingkan konfigurasi *hyperparameter* bawaan. Penerapan RUS-SMOTE memunculkan fenomena *trade-off* berupa penurunan *accuracy* menjadi 80,77%, tetapi meningkatkan *macro recall* dari 73,64% menjadi 78,19% untuk kategori layanan minoritas. Analisis *feature importance* mengidentifikasi *bytes*_rev (*gain* 21,41%) sebagai fitur trafik paling diskriminatif, sementara *learning_rate* (bobot 43,82%) merupakan *hyperparameter* paling berpengaruh. Temuan ini menunjukkan bahwa klasifikasi trafik QUIC berbasis *flow statistics* layak diintegrasikan sebagai komponen non-intrusif dalam sistem pemantauan keamanan jaringan.

Kata Kunci: Klasifikasi Trafik QUIC, LightGBM, Optuna, RUS-SMOTE, CESNET-QUIC22.

ABSTRACT

The adoption of the HTTP/3 over Quick UDP Internet Connections (QUIC) protocol with TLS 1.3 has created significant challenges for network administrators. The implementation of this encryption renders Deep Packet Inspection methods ineffective due to the inability to read encrypted payloads, thereby limiting visibility into the categories of services traversing the network. This problem is agravated by the lack of classification studies at the service category level in ISP-scale QUIC datasets, as most previous research has focused on identifying specific applications with non-QUIC datasets. Therefore, a non-intrusive classification approach is needed that can map service categories in QUIC traffic without decryption.

This research builds a classification model for QUIC traffic service categories using the LightGBM algorithm with hyperparameter optimization based on Optuna and evaluation of resampling strategies in a unified framework. The dataset used is CESNET-QUIC22 (W-2022-44), which includes 20 million flows with 17 service categories. Data preprocessing includes data cleaning, label encoding, and statistical feature extraction (mean, standard deviation, skewness, kurtosis) from packet histograms. Hyperparameter optimization is performed using TPESampler with an adaptive pruning mechanism, while the impact of class imbalance is analyzed through three resampling strategies, namely RUS, SMOTE, and RUS-SMOTE.

The result of the research shows that the LightGBM model without resampling achieved an accuracy of 84.17%, precision of 84.14%, recall of 84.17%, F1-score of 84.00%, AUROC of 98.31%, and AUPRC of 92.14%. Optuna increased accuracy by 15.65% compared to the default hyperparameter configuration. The application of RUS-SMOTE resulted in a trade-off phenomenon in the form of a decrease in accuracy to 80.77%, but increased macro recall from 73.64% to 78.19% for the minority service category. Feature importance analysis identified bytes_rev (gain 21.41%) as the most discriminative traffic feature, while learning_rate (weight 43.82%) was the most influential hyperparameter. These findings indicate that QUIC traffic classification based on flow statistics is feasible to integrate as a non-intrusive component in network security monitoring systems.

Keywords: QUIC Traffic Classification, LightGBM, Optuna, RUS-SMOTE, CESNET-QUIC22.