ABSTRACT

The Internet of Things (IoT) environment is vulnerable to cyber attacks, while the generated traffic data has high dimensions and large feature complexity. This study analyzes the effect of dimensionality reduction using Principal Component Analysis (PCA), Mutual Information (MI), and a hybrid approach (MI+PCA) on the performance of the Gated Recurrent Unit (GRU) model in an IoT intrusion detection system. Evaluations were conducted on the N-BaIoT, CIC-IoT23, and Edge-IIoTset datasets using metrics such as accuracy, precision, recall, F1-score, training time, and the McNemar statistical test. The results show that the effect of dimensionality reduction depends on the characteristics of the dataset. On CIC-IoT23, PCA produces the highest F1-score of 99.19% with a feature reduction of 34.8%, although the training time increases from 778.5 seconds to 1018 seconds. On N-BaIoT, MI maintained 99.95% accuracy with 20.9% feature reduction, while PCA and Hybrid achieved up to 65.2% reduction while maintaining performance at around 99.95%. On Edge-IIoTset, the baseline provided the highest accuracy of 85.92%, while feature reduction increased precision above 98% and specificity above 94%. Overall, dimensionality reduction effectively reduced feature complexity without necessarily improving model performance, so its implementation needs to be tailored to the data characteristics. Based on these findings, further research is recommended to test the model on additional datasets with more diverse characteristics, evaluate real-time scenarios to approximate real-world implementation conditions, and consider other feature reduction methods and computational efficiency aspects to obtain a more optimal model configuration.

Keywords: *IoT, Malware Detection, Dimensionality Reduction, PCA, Mutual Information, GRU, McNemar Test*

ABSTRAK

Lingkungan Internet of Things (IoT) rentan terhadap serangan siber, sementara data trafik yang dihasilkan memiliki dimensi tinggi dan kompleksitas fitur yang besar. Penelitian ini menganalisis pengaruh reduksi dimensi menggunakan Principal Component Analysis (PCA), Mutual Information (MI), serta pendekatan hybrid (MI+PCA) terhadap kinerja model Gated Recurrent Unit (GRU) pada sistem deteksi intrusi IoT. Evaluasi dilakukan pada dataset N-BaIoT, CIC-IoT23, dan Edge-IIoTset menggunakan metrik akurasi, presisi, recall, F1-score, waktu pelatihan, serta uji statistik McNemar. Hasil menunjukkan bahwa pengaruh reduksi dimensi bergantung pada karakteristik dataset. Pada CIC-IoT23, PCA menghasilkan F1-score tertinggi sebesar 99,19% dengan reduksi fitur 34,8%, meskipun waktu pelatihan meningkat dari 778,5 detik menjadi 1018 detik. Pada N-BaIoT, MI mempertahankan akurasi 99,95% dengan reduksi fitur 20,9%, sedangkan PCA dan Hybrid mampu mereduksi hingga 65,2% dengan performa tetap sekitar 99,95%. Pada Edge-IIoTset, baseline memberikan akurasi tertinggi sebesar 85,92%, sementara reduksi fitur meningkatkan presisi di atas 98% dan spesifisitas di atas 94%. Secara keseluruhan, reduksi dimensi efektif menurunkan kompleksitas fitur tanpa selalu meningkatkan performa model, sehingga penerapannya perlu disesuaikan dengan karakteristik data. Berdasarkan temuan tersebut, penelitian selanjutnya disarankan untuk menguji model pada dataset tambahan dengan karakteristik yang lebih beragam, mengevaluasi skenario real-time untuk mendekati kondisi implementasi nyata, serta mempertimbangkan metode reduksi fitur lain dan aspek efisiensi komputasi guna memperoleh konfigurasi model yang lebih optimal.

Kata Kunci: IoT, Deteksi malware, Reduksi dimensi, PCA, MI, GRU, McNemar