Pengembangan Metode Intrusion Detection System berbasis Machine Learning pada tahap seleksi fitur, penyeimbangan kelas, dan ensemble learning = Development of Machine Learning-Based Intrusion Detection System methods at the feature selection, class balancing, and ensemble learning stages

Diwandaru Rousstia, author

Abstrak

Risiko serangan siber berbanding lurus dengan pertumbuhan aplikasi dan jaringan komputer. Intrusion Detection System (IDS) diimplementasikan agar dapat mendeteksi serangan siber dalam lalu lintas jaringan. Akan tetapi terdapat permasalahan pada pendeteksian serangan yang belum diketahui atau jenis serangan baru. Selain itu juga terdapat masalah kinerja tentang waktu deteksi, akurasi deteksi, dan false alarm. Dibutuhkan deteksi anomali dalam lalu lintas jaringan untuk mengurangi permasalahan tersebut dengan pendekatan machine learning. Pengembangan dan pemanfaatan IDS dengan machine learning telah diterapkan dalam beberapa penelitian sebagai solusi untuk meningkatkan kinerja dan evaluasi prediksi deteksi serangan. Memilih pendekatan machine learning yang tepat diperlukan untuk meningkatkan akurasi deteksi serangan siber. Penelitian ini menggunakan metode homogeneous ensemble learning yang mengoptimalkan algoritma tree khususnya gradient boosting tree - LightGBM. Dataset Communications Security Establishment dan Canadian Institute of Cybersecurity 2018 (CSE-CIC-IDS 2018) digunakan untuk mengevaluasi pendekatan yang diusulkan. Metode Polynom-fit SMOTE (Synthetic Minority Oversampling Technique) digunakan untuk menyelesaikan masalah ketidakseimbangan dataset. Penerapan metode spearman’s rank correlation coefficient pada dataset menghasilkan 24 fitur subset dari 80 fitur dataset yang digunakan untuk mengevaluasi model. Model yang diusulkan mencapai akurasi 99%; presisi 99,2%, recall 97,1%; F1-score 98,1%; ROC-AUC 99,1%; dan average-PR 98,1% serta meningkatkan waktu pelatihan model dari 3 menit 25,10 detik menjadi 2 menit 39,68 detik.

.....The risk of cyberattacks is directly proportional to the growth of applications and computer networks. An Intrusion Detection System (IDS) is implemented to detect cyber attacks in network traffic. However, there are problems detecting unknown attacks or new types of attacks. In addition, there are performance issues regarding detection time, detection accuracy, and false alarms. A machine learning approach takes anomaly detection in network traffic to reduce these problems. The development and utilization of IDS with machine learning have been applied in several studies to improve performance and evaluate attack detection predictions. Choosing the right machine learning approach is necessary to improve the accuracy of cyberattack detection. This research uses a homogeneous ensemble learning method that optimizes tree algorithms, especially gradient boosting tree - LightGBM. The Communications Security Establishment and Canadian Institute of Cybersecurity 2018 (CSE-CIC-IDS 2018) dataset evaluated the proposed approach. The Polynom-fit SMOTE (Synthetic Minority Oversampling Technique) method solved the dataset imbalance problem. The application of spearman's rank correlation coefficient method to the dataset resulted in 24 subset features of the 80 dataset features used to evaluate the model. The proposed model achieves 99% accuracy; precision 99.2%, recall 97.1%; F1-score 98.1%; ROC-AUC 99.1%; and an average-PR of 98.1% and increased the training time of the model from 3 minutes 25.10 seconds to 2 minutes 39.68

seconds.