Evaluasi kinerja penerapan metode ensemble learning dan pemilihan fitur pada Intrusion Detection System (IDS) berbasis machine learning = Performance evaluation of machine learning-based intrusion detection system using ensemble learning method and feature selection

Qusyairi Ridho Saeful Fitni, author

Abstrak

Dalam beberapa tahun terakhir, keamanan data pada sistem informasi organisasi telah menjadi perhatian serius. Banyak serangan menjadi kurang terdeteksi oleh firewall dan perangkat lunak antivirus. Untuk meningkatkan keamanan, intrusion detection systems (IDS) digunakan untuk mendeteksi serangan dalam lalu lintas jaringan. Saat ini, teknologi IDS memiliki masalah kinerja mengenai akurasi deteksi, waktu deteksi, pemberitahuan alarm palsu, dan deteksi jenis serangan baru atau belum diketahui. Beberapa studi telah menerapkan pendekatan pembelajaran mesin (machine learning) sebagai solusi, dan mendapat beberapa peningkatan. Penelitian ini menggunakan pendekatan pembelajaran ensemble (ensemble learning) yang dapat mengintegrasikan manfaat dari setiap algoritma pengklasifikasi tunggal. Pada penelitian ini, dibandingkan tujuh pengklasifikasi tunggal untuk mengidentifikasi pengklasifikasi dasar yang digunakan untuk model ensemble learning. Kemudian dataset IDS terbaru dari Canadian Institute for Cybersecurity yaitu CSE-CIC-IDS2018 digunakan untuk mengevaluasi model ensemble learning. Hasil percobaan menujukan bahwa implementasi metode ensemble learning khususnya majority voting dengan tiga algoritma dasar (gradient boosting, decision tree dan logistic regression) dapat meningkatkan nilai akurasi lebih baik dibandingkan implementasi algoritma klasifikasi tunggal, yaitu 0,988. Selanjutnya, implementasi teknik pemilihan fitur spearman-rank order correlation pada dataset CSE-CIC-IDS2018 menghasilkan 23 dari 80 fitur, dan dapat meningkatkan waktu pelatihan model, yaitu menjadi 11 menit 4 detik dibanding sebelumnya 34 menit 2 detik.

.....In recent years, data security in organizational information systems has become a serious concern. Many attacks are becoming less detectable by firewall and antivirus software. To improve security, intrusion detection systems (IDSs) are used to detect anomalies in network traffic. Currently, IDS technology has performance issues regarding detection accuracy, detection times, false alarm notifications, and unknown attack detection. Several studies have applied machine learning approaches as solutions. This study used an ensemble learning approach that integrates the benefits of each single classifier algorithms. We made comparisons with seven single classifiers to identify the most appropriate basic classifiers for ensemble learning. Then the latest IDS dataset from the Canadian Institute for Cybersecurity, CSE-CIC-IDS2018, was used to evaluate the ensemble learning model. The experimental results show that the implementation of the ensemble learning method, especially majority voting with three basic algorithms (gradient boosting, decision tree and logistic regression) can increase the accuracy rate better than the implementation of a single classification algorithm, which is 0.988. Furthermore, the implementation of the spearman-rank order correlation feature selection technique in the CSE-CIC-IDS2018 dataset produced 23 of the 80 features, and could increase the model training time, which was 11 minutes 4 seconds compared to 34 minutes 2 seconds before.