INTISARI

Popularitas android yang meluas menimbulkan risiko keamanan yang signifikan karena rentannya pengguna terhadap malware. Perkembangan teknologi yang pesat memberikan solusi yang cukup efektif dalam mengurangi dampak negatif yang diakibatkan oleh malware pada perangkat android. Salah satu solusi yang dapat diterapkan adalah penerapan machine learning dalam menganalisis data terkait android malware. Penelitian ini bertujuan untuk menilai performa metode random forest dalam menganalisis malware pada android, dengan menggunakan dataset DREBIN yang terdiri dari 215 atribut dan satu atribut target, dan terdiri dari 15,036 baris data. Hasil awal menunjukkan akurasi, presisi, recall, dan f1-score mencapai 100% pada data pelatihan, sementara pada data pengujian, akurasi mencapai 96%, presisi 97%, recall 89%, dan f1-score 92%, yang mengindikasikan potensi overfitting. Untuk mengatasi potensi overfitting, dilakukan cross validation sebanyak sepuluh kali. Hasil setelah penerapan cross validation menunjukkan bahwa kinerja model pada data pelatihan dan pengujian tetap baik, dengan akurasi mencapai 96%, presisi 96% pada data pelatihan dan 97% pada data pengujian, recall 90% pada data pelatihan dan 86% pada data pengujian, serta f1-score 93% pada data pelatihan dan 91% pada data pengujian. Meskipun demikian, penurunan nilai metrik evaluasi yang lebih realistis menunjukkan bahwa model mampu menggeneralisasikan hasilnya dengan lebih baik ke data yang belum pernah dilihat sebelumnya.

Kata kunci: Android Malware, Machine Learning, Random Forest, Cross Validation

ABSTRACT

Android's widespread popularity poses significant security risks due to users' vulnerability to malware. Rapid technological developments provide quite effective solutions in reducing the negative impact caused by malware on Android devices. One solution that can be implemented is the application of machine learning in analyzing data related to android malware. This research aims to assess the performance of the random forest method in analyzing malware on Android, using the DREBIN dataset which consists of 215 attributes and one target attribute, and consists of 15,036 rows of data. Preliminary results show accuracy, precision, recall, and f1-score reaching 100% on training data, while on testing data, accuracy reaches 96%, precision 97%, recall 89%, and f1-score 92%, which indicates the potential for overfitting. To overcome the potential for overfitting, cross validation was carried out ten times. The results after applying cross validation show that the model performance on training and testing data remains good, with accuracy reaching 96%, precision of 96% on training data and 97% on testing data, recall of 90% on training data and 86% on testing data, and f1-score 93% on training data and 91% on testing data. Nonetheless, deriving more realistic values of the evaluation metrics indicates that the model is able to generalize its results better to never-before-seen data.

Keyword: Android Malware, Machine Learning, Random Forest, Cross Validation