Android Malware Detection Using Ensemble Learning and Feature Selection with Insight from SHAP Explainable AI

Rian Adriansyah[1], Parman Sukarno[2], Aulia Arif Wardana[3]
Fakultas Informatika, Universitas Telkom, Bandung
adriansyahrian@student.telkomuniversity.ac.id,
psukarno@telkomuniversity.ac.id,
aulia.wardana@pwr.edu.pl,

Abstract
The increasing proliferation of Android mobile devices, along with a surge in malware targeting these platforms, underscores the need for advanced malware detection methods. This research introduces a robust approach to Android malware detection by employing ensemble learning techniques. The integration of Random Forest, XGBoost, Extremely Randomized Trees, and Histogram-based Gradient Boosting models forms the core of the proposed method. Enhancing this approach, feature selection is performed using Random Forest Feature Importance, focusing on the most relevant features to achieve high accuracy and efficiency. Additionally, SHAP (SHapley Additive exPlanations) Explainable AI is utilized to provide transparency and a comprehensive understanding of the model's decisions, thereby fostering user trust and comprehension. The proposed method was evaluated on the CICMalDroid2020 dataset, demonstrating superior performance, with XGBoost achieving the highest accuracy at 94.88%. This research not only advances the field of malware detection on Android devices but also contributes significantly to the explanation and interpretation of machine learning models within the context of cybersecurity.

Keywords: **android malware, malware detection, ensemble learning, feature selection, shap explainable ai**