PENDETEKSIAN MALWARE ANDROID DENGAN MEMANFAATKAN FITUR PERMISSIONS MENGGUNAKAN ALGORITMA EXTREME GRADIENT BOOSTING (XGBOOST)

[1]Nafisa Alfi Sa'diya, [2]Arief Ikhwan Wicaksono

INTISARI

Latar Belakang: Ancaman malware terhadap sistem operasi Android semakin meningkat seiring pertumbuhan penggunaan perangkat *mobile*. Salah satu pendekatan yang digunakan dalam deteksi malware adalah melalui analisis fitur *permissions*, yaitu izin akses yang diminta oleh aplikasi. Permasalahan utama yang sering muncul adalah ketidakseimbangan kelas (*imbalanced class*) antara aplikasi non-malware dan malware, yang dapat memengaruhi kemampuan model dalam mengenali aplikasi berbahaya secara akurat.

Tujuan: Penelitian ini bertujuan untuk mengkaji penerapan algoritma Extreme Gradient Boosting (XGBoost) dalam mendeteksi aplikasi malware Android berbasis fitur *permissions*, dengan fokus utama pada strategi penanganan ketidakseimbangan kelas.

Metode Penelitian: Penelitian dilakukan dengan pendekatan eksperimental. Dataset yang digunakan merupakan data *permissions* dari aplikasi Android yang telah dilabeli sebagai malware dan non-malware. Proses dimulai dengan *preprocessing* data dan pemilihan 10 fitur *permissions* terbaik. Selanjutnya, dilakukan pelatihan model XGBoost dengan penyesuaian parameter *scale_pos_weight*, disertai eksperimen lanjutan berupa *threshold tuning* dan penerapan teknik *oversampling* SMOTE. Evaluasi dilakukan menggunakan metrik akurasi, *precision*, *recall*, *F1-Score*, dan *confusion matrix*.

Hasil: Model XGBoost menunjukkan *precision* tinggi namun *recall* rendah pada *baseline* awal. Penerapan *scale_pos_weight* mampu meningkatkan sensitivitas model terhadap malware, yang kemudian diperkuat dengan penyesuaian *threshold* dan penggunaan SMOTE. Eksperimen subset data (500, 1000, dan 2000 sampel) menunjukkan bahwa ukuran data memengaruhi stabilitas performa dan waktu komputasi. Model terbaik mencapai *recall* hingga 65% setelah *threshold tuning* pada nilai 0.4.

Kesimpulan: XGBoost terbukti dapat digunakan untuk menangani ketidakseimbangan kelas dalam deteksi malware Android berbasis fitur *permissions*. Kombinasi strategi *scale_pos_weight*, *threshold tuning*, dan SMOTE berhasil meningkatkan performa model terhadap kelas minoritas.

Kata-kunci: Android, Malware, *Imbalanced class*, XGBoost, *Permissions*, SMOTE, *Threshold Tuning*

---

[1]Mahasiswa Teknologi Informasi (S-1) Fakultas Teknik dan Teknologi Informasi Universitas Jenderal Achmad Yani Yogyakarta
[2]Dosen Teknologi Informasi (S-1) Fakultas Teknik dan Teknologi Informasi Universitas Jenderal Achmad Yani Yogyakarta

PENDETEKSIAN MALWARE ANDROID DENGAN MEMANFAATKAN FITUR PERMISSIONS MENGGUNAKAN ALGORITMA EXTREME GRADIENT BOOSTING (XGBOOST)

[1]Nafisa Alfi Sa'diya, [2]Arief Ikhwan Wicaksono

ABSTRACT

***Background**: The growing use of mobile devices has led to a significant increase in malware threats targeting the Android operating system. One approach to detecting such threats is by analyzing permissions—the access rights requested by applications. A major challenge in this domain is the issue of **class imbalance**, where the number of non-malware applications far exceeds that of malware, potentially reducing the model's ability to accurately detect malicious apps.*

***Objective**: This study aims to investigate the implementation of the Extreme Gradient Boosting (XGBoost) algorithm for detecting Android malware based on application permissions, with a primary focus on addressing the class imbalance problem.*

***Method**: This research follows an experimental approach using a publicly available dataset of Android applications labeled as malware or non-malware. The process begins with data preprocessing and Selection of the 10 most relevant permissions features. The XGBoost model is trained with a focus on the scale_pos_weight parameter to handle class imbalance. Additional experiments include threshold tuning and oversampling using SMOTE. Model evaluation is conducted using accuracy, precision, recall, F1-Score, and confusion matrix.*

***Result**: The initial XGBoost model achieved high precision but relatively low recall. The use of scale_pos_weight improved the model's sensitivity to malware, and this effect was further enhanced through threshold tuning and SMOTE. Experiments on subsets of 500, 1000, and 2000 samples showed that data size influences both performance stability and computational time. The best recall, reaching up to 65%, was achieved with a threshold value of 0.4.*

***Conclusion**: XGBoost proves effective in handling class imbalance in Android malware detection based on permissions. The combined use of scale_pos_weight, threshold tuning, and SMOTE improves model performance on minority (malware) classes without significantly sacrificing precision.*

***Keywords**: Android, Malware, Imbalanced class, XGBoost, Permissions, SMOTE, Threshold Tuning*

---

[1]Information Technology Student, Faculty of Engineering and Information Technology, Universitas Jenderal Achmad Yani Yogyakarta

[2]Information Technology Lecturer, Faculty of Engineering and Information Technology, Universitas Jenderal Achmad Yani Yogyakarta