Nama : Muhammad Nabil Nazhmi Kurniali

NIM : A11.2023.15366

MK / KLPKMK : Machine Learning / A11.4406

UTS Machine Learning

Ringkasan + Kata Kunci

Proyek ini bertujuan untuk mengembangkan sistem deteksi malware menggunakan algoritma pembelajaran mesin. Peningkatan ancaman malware menjadi tantangan besar dalam keamanan siber. Dalam penelitian ini, berbagai metode pembelajaran mesin akan diterapkan untuk mengklasifikasikan file executable sebagai malware atau non-malware. Dataset yang digunakan terdiri dari file-file yang diambil dari sumber publik, termasuk The Zoo dan VirusShare untuk memastikan keberagaman dan relevansi data.

Kata Kunci: Deteksi Malware, Pembelajaran Mesin, Klasifikasi, Keamanan Siber, Dataset Publik.

Pendahuluan

Latar Belakang

Keamanan siber menjadi semakin penting dengan adanya ancaman malware yang terus berkembang. Malware dapat menyebabkan kerusakan signifikan pada sistem komputer, data, dan bahkan reputasi organisasi. Dengan adanya teknik deteksi yang lebih efisien, risiko tersebut dapat diurangi.

Masalah

1. Sulitnya mendeteksi varian baru malware yang belum dikenal oleh sistem keamanan tradisional.

2. Metode deteksi berbasis tanda tangan tidak efektif terhadap serangan yang lebih canggih dan kompleks.

State of The Art

Teknologi deteksi malware terbaru banyak menggunakan algoritma pembelajaran mesin, termasuk Random Forest, SVM (Support Vector Machine), dan Neural Networks. Penelitian menunjukkan bahwa pendekatan ini dapat mengenali pola yang lebih halus dalam perilaku file yang terinfeksi dibandingkan metode tradisional.

Dataset

1. Sumber: Dataset dapat diambil dari The Zoo (https://github.com/ytisf/theZoo) dan VirusShare (https://virusshare.com/).

2. Isi Dataset: File executable yang diklasifikasikan sebagai malware atau non-malware, dengan fitur seperti ukuran file, tipe, dan pola perilaku.

3. Jenis: Publik (gratis diakses).

4. Preprocessing:

- Menghapus data yang duplikat atau tidak lengkap.
- Normalisasi ukuran file untuk konsistensi.
- Ekstraksi fitur relevan dari file (misalnya, karakteristik perilaku).

Metoda yang Diusulkan

1. **Pemilihan Model**

- Algoritma pembelajaran mesin, seperti Random Forest dan SVM, dipilih karena kemampuannya dalam menangkap pola kompleks dalam data.

2. **Preprocessing Data**

- Menghapus entri duplikat atau data yang tidak lengkap.
- Melakukan normalisasi data untuk memastikan keseragaman.
- Ekstraksi fitur dari file executable yang relevan untuk analisis.

3. **Pelatihan Model**

- Model akan dilatih menggunakan data fitur dari file yang sudah diproses.
- Hyperparameter Tuning: Menentukan parameter seperti kedalaman pohon untuk Random Forest atau parameter kernel untuk SVM.

4. **Evaluasi Model**

- Gunakan metrik seperti akurasi, precision, recall, dan F1-score untuk mengukur kinerja model.
- Analisis hasil deteksi agar dapat mengetahui seberapa baik model mengenali malware.

5. **Interpretasi dan Penyajian Hasil**

- Confusion Matrix untuk visualisasi kinerja model.

- Visualisasi probabilitas klasifikasi dalam bentuk grafik untuk analisis lebih lanjut.

Hasil

Yang akan didapatkan dari penelitian:

- Visualisasi:

a) Confusion matrix untuk menunjukkan jumlah true positives, false positives, true negatives, dan false negatives.
b) Grafik probabilitas klasifikasi untuk menunjukkan seberapa yakin model dalam klasifikasi yang dibuat.

- Metrik Evaluasi:

Akurasi, precision, recall, dan F1-score.

- Output:

Daftar file yang terdeteksi sebagai malware dan non-malware lengkap dengan analisis performa.

Literatur

1. Khraisat, A., et al. (2019). "Malware Detection using Machine Learning: A Survey." **Journal of Computing and Security**.

2. Kaspersky Lab. "How to Detect Malware." **Kaspersky IT Encyclopedia**.

3. Yeganeh, S., & Moin, N. (2019). "Detecting Malware Using Machine Learning Algorithms." **Journal of Information Security Research**.

4. Kwon, K. R., et al. (2018). "Malware Detection with Machine Learning: An Overview." **IEEE Access.**

5. Zuev, D., et al. (2020). "A Comprehensive Survey on Malware Detection Techniques." **Computers & Security**.