Malware Detection

NIM : 1304211043

Nama : Hafidh Adani

Pendahuluan

Malware atau malicious software merupakan ancaman keamanan yang terus berkembang dan dapat menyebabkan kerusakan sistem, pencurian data, serta gangguan jaringan. Pendekatan tradisional seperti signature-based detection tidak efektif terhadap malware baru yang tidak dikenal oleh karena itu, ancaman tersebut membutuhkan metode deteksi yang lebih efektif. Machine learning menawarkan solusi yang menjanjikan untuk deteksi malware dengan kemampuannya belajar dari data dan mengidentifikasi pola. Dalam laporan ini, kami akan membahas penggunaan dua algoritma machine learning untuk mendeteksi sebuah malware yaitu Naive Bayes dan Support Vector Machine (SVM).

Penjelasan Algoritma

Naive Bayes adalah algoritma klasifikasi yang didasarkan pada teorema Bayes. Algoritma ini mengasumsikan bahwa setiap fitur adalah independen dan memberikan kontribusi yang sama terhadap probabilitas suatu kelas. Naive Bayes sering digunakan dalam deteksi malware karena kesederhanaan dan kecepatan komputasinya.

SVM adalah algoritma klasifikasi yang berusaha menemukan hyperplane terbaik yang memisahkan antara kelas-kelas yang berbeda. SVM bekerja dengan baik pada dataset yang memiliki banyak fitur dan dapat menangani masalah non-linear dengan menggunakan kernel yang sesuai. SVM telah terbukti efektif dalam banyak kasus deteksi malware.

Hasil Deteksi

Berdasarkan hasil dari program yang telah dijalankan, didapatkan data dari masing-masing algoritma sebagai berikut.

- ~ Akurasi adalah matriks yang mengukur persentase sampel yang diklasifikasikan dengan benar oleh model.
- ~ Presisi adalah matriks yang mengukur seberapa banyak sampel positif yang diprediksi benar-benar positif. Dalam konteks deteksi malware, presisi menunjukkan seberapa banyak sampel yang diprediksi sebagai malware benar-benar merupakan malware.
- ~ Recall adalah matriks yang mengukur seberapa banyak sampel positif yang berhasil diidentifikasi dengan benar. Dalam konteks deteksi malware, recall menunjukkan seberapa banyak malware yang berhasil dideteksi dengan benar oleh model.

Naive Bayes:

- Akurasi: 0.69485 atau 69,49%; Dalam hal ini, model Naive Bayes dapat mengklasifikasikan 69,49% dari keseluruhan sampel dengan benar.
- Presisi: 0.6464661768047875 atau 64,65%; Nilai presisi 64,65% menunjukkan bahwa dari semua sampel yang diprediksi sebagai malware, 64,65% di antaranya benar-benar malware.

- Recall: 0.8559679037111334 atau 85,60%; Nilai recall 85,60% menunjukkan bahwa model Naive Bayes dapat mendeteksi 85,60% dari seluruh malware yang ada dalam dataset.

SVM (Support Vector Machine):

- Akurasi: 0.4985 atau 49,85%; Akurasi model SVM hanya 49,85%, yang berarti hanya sekitar setengah dari keseluruhan sampel yang diklasifikasikan dengan benar.
- Presisi: 0.4985 atau 49,85%; Presisi model SVM juga 49,85%, yang menunjukkan bahwa dari semua sampel yang diprediksi sebagai malware, hanya 49,85% yang benar-benar merupakan malware.
- Recall: 1.0 atau 100%; Recall model SVM mencapai 100%, yang berarti model ini berhasil mendeteksi semua malware yang ada dalam dataset.

Berdasarkan hasil ini, dapat disimpulkan bahwa model Naive Bayes memiliki performa yang lebih baik dibandingkan model SVM dalam hal akurasi dan presisi. Namun, model SVM unggul dalam hal recall, yang berarti dapat mendeteksi semua malware yang ada dalam dataset.

Pemilihan model yang optimal akan bergantung pada prioritas dan kebutuhan spesifik dalam penerapan deteksi malware. Jika prioritas utama adalah mendeteksi semua malware (recall tinggi), maka model SVM dapat menjadi pilihan yang baik. Namun, jika akurasi dan presisi menjadi pertimbangan utama, model Naive Bayes mungkin lebih cocok.

Link

Link Github:
https://github.com/TuyulGanteng02/Week-11-MalwareDetection

Link GDrive:
https://drive.google.com/drive/folders/12nNjS7F0xWGyXGw6EMiwA-bFlVspBNVv?usp=sharing