Analisis Metode *Ensemble* untuk Mendeteksi *Malware* pada *Mobile Devices*

Ary Adhigana Suwandi[1], Parman Sukarno[2], Erwid Musthofa Jadied[3]

[1,2,3]Fakultas Informatika, Universitas Telkom, Bandung
[1]ggaarryy@students.telkomuniversity.ac.id, [2]psukarno@telkomuniversity.ac.id,
[3]jadied@telkomuniversity.ac.id

Abstrak

Tahun 2017, sebanyak 371,4 juta penduduk Indonesia telah aktif menggunakan *mobile device.* Sebuah kenyamanan yang dirasakan penduduk Indonesia dalam menggunakan mobile device untuk aktifitas-aktifitas seperti berkumpul, bermain game, dan hal lainnya. Namun pada tahun 2014 tercatat sebanyak 16 juta mobile devices di dunia terserang oleh *malware*. Dalam menghadapi malware tidak dapat digunakan dengan pendekatan tradisional seperti halnya antivirus, pada penelitian dengan Drebin dataset dimana dataset tersebut berbasis fitur perizinan pada aplikasi *mobile devices* Android yang menggunakan *Support Vector Machine (SVM)* sebagai Machine Learning dalam mendeteksi malware pada Android. Pada saat ini malware dapat melakukan perkembangan, oleh karena itu penggunaan Machine Learning (ML) dapat berguna karena ML adalah metode pembelajaran berbasis kebiasaan malware tersebut. Dalam penelitian ini penulis membangun sebuah penggabungan dari ketiga Machine Learning yaitu *KNN*, *Random Forest*, dan *Naïve Bayes* dengan Metode *Ensemble*. Hasil dari pengujian adalah Metode Ensemble menghasilkan akurasi yang lebih baik, yaitu 98,4%.

-

Kata kunci : Ensemble learning, mobile device, threat, flow-based network, machine learning.

Abstract

In 2017, 371.4 Million people in Indonesia are active using mobile devices. The comfortable when use mobile devices is the reason population in Indonesia using mobile devices for many activities such as gathering, playing games, and etc. However, on record in 2014, 16 Millions Mobile mobile devices has been attacked by malware. Approaching traditional method such as antivirus it does not working well when confort malware in study using Drebin dataset which is the research based on features permissions in Android applications that use Support Vector Machine (SVM) as Machine Learning for detecting malware. Currently malware be able to evolution, therefore Machine Leaning (ML) can be useful because ML were learning method based on malware behavior. In this research, we developing three Machine Learning that we will fusion using Ensemble Method. The result from this research is Ensemble Method has better accuracy, 98.4%.

1. Pendahuluan

Latar Belakang

Pada 2017 berdasarkan data databoks.co.id sebanyak 371,4 juta penduduk di Indonesia yang menggunakan *mobile device*, kemajuan teknologi ini menghasilkan banyak kenyamanan pada setiap pengguna terlebih aktifitas seperti, berkumpul, bermain game, dan hal lain dapat dilakukan dimanapun dan kapan saja dengan *mobile device*[1]. Sementara itu, potensi kejahatan dengan memperoleh informasi semakin menunjukan peningkatannya. Namun banyak pengguna yang tidak mementingkan keamanan pada *device* yang dimilikinya, sehingga mereka dapat menjadi korban ancaman *malware*[2]. Pada tahun 2014 tercatat bahwa sebanyak 16 juta *mobile devices* terserang oleh *malware*, dalam penelitian tersebut tercatat bahwa infeksi *malware* meningkat sekitar 25% di tahun 2014 dibandingkan dengan pada tahun 2013 yang meningkat 20%. Terkait laporan tersebut terdapat enam fungsi *malware* yang digunakan, seperti memata-matai pemilik *mobile device* tersebut, melacaknya, memonitori aktifitas telepon dan SMS, memonitori email, dan melacak aktifitas pengguna saat *web browsing*[3].

Pada umumnya dalam menghadapi malware pada mobile devices dapat menggunakan antivirus yang cara kerjanya dengan pendekatan tradisional, namun menurut [1] dalam menghadapi *malware mobile device*, dengan menggunakan pendekatan tradisional seperti antivirus, bukan hal yang efisien. Karena untuk melawan *malware* pada *mobile device* selalu membutuhkan pembaharuan pada database dan setiap malware selalu berubah demi menghindari berbagai macam pendeteksian. Dengan masalah tersebut maka membutuhkan penambahan keamanan pada sisi jaringan sebagai proteksi untuk user dari *malware* yang lebih maju [2]. Namun pada penelitian

[4] yang menggunakan dataset *Drebin* dimana dataset tersebut berbasis fitur seperti perizinan pada *smartphone,* yang dicek menggunakan metode *Support Vector Machine (SVM)*, namun keterbatasan yaitu pada eksekusi yang masih kurang optimal.

Dengan perkembangan *malware* saat ini menggunakan pendeteksi biasa akan sulit dideteksi, dikarenakan *malware* cenderung memiliki beberapa lapisan polimorfik untuk menghindari pendeteksian atau menggunakan mekanisme untuk memperbaharui dirinya menjadi lebih baru dalam waktu yang singkat [5]. Untuk menghadapi hal tersebut digunakanlah *Machine Learning(ML)*, dikarenakan cara kerja ML yaitu memperoleh sebuah label dari dataset dan menghasilkan sebuah model sebagai hasil, yang bisa menangani data baru. Pengklasifikasiannya dipelajari dari semua data yang dimasukkan dan hasilnya dilabeli untuk membangun sebuah model, menggunakan ML dapat meningkatkan akurasi pendeteksian[1].

Dalam pengujian Drebin dataset yang menggunakan SVM, memiliki akurasi yang kurang optimal, peneliti ingin meningkatkan akurasi pada Drebin, dalam penelitian [2][6] mengatakan, bahwa akurasi dapat ditingkatkan dengan menggunakan metode *Ensemble*. Dalam penelitian tersebut menggunakan tiga ML dengan pendekatan ID3, namun peneliti ingin melakukan penggabungan dari tiga ML seperti penelitian tersebut, tetapi dengan jenis ML yang berbeda. Penelitian tersebut mengatakan bahwa akurasi *Ensemble* lebih tinggi dibandingkan akurasi dari sebuah ML dalam penelitian sebelumnya[7]. *Random Forest*(RF) dipilih karena memiliki akurasi yang optimal dalam penelitian untuk metode ID3 dalam mendeteksi malware[6]. Pada penelitian [1] dibandingkan beberapa ML untuk mendeteksi malware, dengan hasil akurasi yang optimal adalah *KNN*. Penulis menambahkan Naïve Bayes untuk digabungkan dalam pendeteksian untuk Drebin dataset. Karena Drebin dataset berisikan nilai biner, dan Naïve Bayes baik dalam mengolah data-data biner dengan pendekatan probabilistik. Maka metode yang diusulkan untuk mendeteksi malware pada Drebin dataset adalah metode *Ensemble* dengan kombinasi *KNN*, *Random Forest*, dan *Naïve Bayes*, yang diharapkan dapat meningkatkan akurasi dari penelitian sebelumnya [4].

Topik dan Batasannya

Rumusan masalah yang diselesaikan pada tugas akhir ini adalah sebagai berikut:

1. Bagaimana rancangan dan nilai akurasi dari metode Ensemble dengan kombinasi *KNN, Random Forest,* dan *Naïve Bayes* dalam medeteksi *malware* pada *mobile device*?

Batasan Masalah yang digunakan pada tugas akhir ini adalah sebagai berikut:

1. Analisis deteksi malware berdasarkan perilaku menggunakan metode Ensemble dengan kombinasi *KNN, Random Forest,* dan *Naïve Bayes*.
2. Data uji coba yang digunakan untuk mendeteksi malware pada mobile device adalah Drebin Dataset.

Tujuan

Pada penelitian ini dilakukan dengan tujuan merancang Metode Ensemble untuk mendeteksi malware dari kombinasi Machine Learning(ML) sebagai berikut *KNN, Random Forest,* dan *Naïve Bayes* dalam mendeteksi serangan malware pada mobile device. Serta menganalisa hasil akhir pendeteksian berdasarkan parameter akurasi

Organisasi Tulisan

Pada penulisan ini untuk BAB 2 akan dijelaskan tentang Studi Terkait, BAB 3 dijelaskan tentang sistem yang dibangun diantaranya adalah pembuatan model untuk Machine Learning. Pada BAB 4 akan dijelaskan tentang evaluasi, dan BAB 5 dijelaskan kesimpulan dan saran.