

BAB I PENDAHULUAN

1.1 Latar Belakang

Malicious Software (Malware) merupakan program komputer yang di desain dengan maksud jelas untuk melakukan serangan ke sistem komputer dan melakukan berbagai aksi penipuan[1]. Malware dapat berupa virus, worm, trojan, *ransomware* atau program berbahaya yang merusak komputer. Malware yang paling berbahaya dan merugikan adalah ransomware.

Ransomware merupakan *software* yang akan memblokir akses pengguna ke file yang terinfeksi kecuali pengguna menebus senilai uang untuk membukanya. Ketika uang sudah dibayarkan, maka pengguna akan mendapatkan password untuk membuka file yang sudah terinfeksi. *Ransomware* menginfeksi sekitar 300.000 komputer di 150 negara dan merugikan pengguna komputer ribuan dolar dalam bentuk uang tebusan dan miliaran orang kehilangan produktivitasnya, hingga rumah sakit di *National Health Service* Inggris terpaksa membatalkan operasi, ambulans dialihkan dan catatan pasien tidak dapat diakses karena *ransomware*[2].

Untuk menghentikan ransomware sebelum menginfeksi sistem baik itu berupa individu atau institusi diperlukan adanya teknik yang mampu mendeteksi ransomware. Namun untuk mendeteksi ransomware diperlukan teknik malware analisis untuk menganalisis ransomware. Malware analisis merupakan cara menganalisis dan mengidentifikasi ciri dan kelakuan dari malware untuk melihat dampak yang dapat ditimbulkan oleh malware yang telah dianalisis[3]. Terdapat dua metode analisis malware yaitu metode statis dan metode dinamis. Statis analisis merupakan metode analisis malware tanpa harus menjalankan malware tersebut tapi hanya dengan membaca code dari malware tersebut[1]. Kemudian metode analisis dinamis adalah metode yang menjalankan malware di dalam virtual box atau tempat khusus analisis malware. Hasil yang didapat dari analisis tersebut dapat berupa ciri code program, string program atau ciri lainnya yang merupakan karakteristik dari malware. Teknik yang ada saat ini yaitu teknik *signature base*.

Teknik deteksi *signature base* merupakan deteksi yang berdasar pada ciri *string* dan kode program yang ada pada malware. *Signature base* tidak efektif untuk

mendeteksi malware atau ransomware yang tidak diketahui cirinya[4]. Terlebih lagi Kriptografi dalam Ransomware menggunakan teknik enkripsi dan mutasi untuk menghindari mekanisme deteksi *signature base* yang digunakan oleh Perangkat Lunak Antivirus[5]. Jadi ketika ransomware menginfeksi file yang ada pada sistem maka ransomware akan mengenkripsi dirinya sendiri dan mengubah cirinya agar tidak terdeteksi oleh *antivirus* yang menggunakan *signature base*.

Solusi dari permasalahan *signature base* ialah dengan menggunakan machine learning dalam mendeteksi ransomware. Machine learning melibatkan data *training* untuk membuat model. Model kemudian memprediksi hasil bahkan ketika data yang ada itu merupakan data baru[6]. Keunggulan dari machine learning adalah bahwa machine learning dapat secara akurat memprediksi hasil dengan data *training* yang memadai[2]. Namun dari keuntungan yang ada pada machine learning dalam mendeteksi ransomware, sulit untuk menemukan algoritma yang tepat untuk untuk diterapkan mendeteksi ransomware agar mendapat akurasi yang tinggi [2].

Metode yang cukup baik dalam machine learning untuk mendeteksi malware adalah metode *random forest*[7]. Namun untuk mendapatkan nilai akurasi yang lebih baik dari metode *random forest* biasa diperlukan pendekatan agar dapat meningkatkan hasil akurasi yang lebih baik. Pendekatan yang digunakan adalah *feature evaluator* untuk memilih fitur yang relevan dengan frekuensi kemunculan api call sebagai fitur dan pendekatan *resampling* dengan teknik *simple random sampling* untuk mengatasi *imbalance* data. Pendekatan ini dinamakan *improved random forest*[8]. *Improved random forest* digunakan karena pendekatan ini mampu memberikan hasil yang lebih baik dari *random forest* biasa dan dapat mengatasi masalah data yang tidak rata atau *imbalance data*. Tahapan yang digunakan untuk mendeteksi ransomware ialah dengan mengekstrak api calls yang ada pada ransomware, kemudian memilih api call berdasarkan frekuensi kemunculannya yang dijadikan sebagai fitur, melakukan fitur evaluator, resampling untuk mengatasi imbalance data dan yang terakhir ialah melakukan klasifikasi. Oleh karena itu peneliti akan menggunakan metode *Improved Random forest* untuk mendeteksi *ransomware*.

Diharapkan penelitian ini nantinya dapat menghasilkan akurasi yang tinggi dan lebih baik dari metode *random forest* biasa dalam mendeteksi *ransomware* dan

peneliti berharap implementasi dari penelitian ini dapat dipakai untuk keperluan dalam hal mencegah ransomware dalam menginfeksi komputer.

1.2 Perumusan Masalah

1.2.1 Permasalahan

Teknik *signature base* tidak efektif dalam mendeteksi *ransomware* varian baru yang belum dianalisis dan diperlukan akurasi yang lebih baik lagi dari metode *random forest* biasa untuk mendeteksi *ransomware*

1.2.2 Metode Usulan

Diusulkan klasifikasi untuk mendeteksi *ransomware* dengan menggunakan metode *improved random forest* dengan menggunakan pendekatan *fitur evaluator* dan *resampling* dan algoritma *C4.5* sebagai *classifier*.

1.2.3 Pertanyaan Penelitian

Apakah metode *Improved Random forest* dengan algoritma *C4.5* sebagai pembentuk pohon keputusan agar dapat mendeteksi *Ransomware* berdasarkan *api calls*?

1.3 Tujuan & Manfaat

1.3.1 Tujuan Penelitian

Untuk mengetahui efektifitas penerapan dan tingkat akurasi metode *Improved Random forest* untuk mendeteksi *Ransomware* berdasarkan *api calls*.

1.3.2 Manfaat Penelitian

Hasil dari penelitian ini diharapkan dapat menghasilkan software yang dapat mendeteksi *ransomware* dengan tingkat akurasi yang tinggi.

1.4 Batasan Masalah

Batasan masalah atau parameter data yang digunakan ialah:

1. Dataset yang digunakan ialah data *ransomware* file *portable executable* dan file *benign* (jinak) yang tidak berisi malware dengan data yang imbalance.

2. Dataset *ransomware* didapatkan dari <https://virusshare.com>[25] sedangkan dataset *benign* didapat dari crawl file dari komputer sendiri.
3. Mengekstrak api call menggunakan library pefile.
4. Menggunakan metode *improved random forest* dengan algoritma *C4.5* sebagai pembentuk pohon keputusan untuk mengklasifikasikan ransomware.
5. Menggunakan model *K-Fold cross validation* untuk mengevaluasi model.
6. Parameter data yang digunakan ialah *API calls* yang dipilih berdasarkan frekuensi kemunculan.

1.5 Sistematika Proposal

Sistematika Penulisan Proposal Skripsi ini adalah sebagai berikut :

BAB I PENDAHULUAN

Pada bab Pendahuluan ini akan di bahas mengenai Latar Belakang, Rumusan Masalah, Tujuan dan Manfaat, Batasan Masalah dan Sistematika Proposal.

BAB II KAJIAN PUSTAKA

Pada bab Kajian Pustakan akan di bahas mengenai landasan teori meliputi *malware*, *malware analisis*, *ransomware*, *api calls*, *improved random forest* dan *algoritma c4.5*.

BAB III METODE USULAN

Pada bab Metode Usulan ini menjelaskan mengenai algoritma metode usulan, penjelasan arsitektur sistem, dataset yang akan digunakan, metode penelitian dan perkiraan jadwal penelitian.

BAB IV HASIL DAN PEMBAHASAN

Pada bab ini menjelaskan hasil yang didapat dari uji coba, seperti hasil uji coba berdasarkan skenario, model evaluasi dan akurasi yang didapat.

BAB V PENUTUP

Pada bab berisi kesimpulan dari penelitian yang sudah dikerjakan yang telah disesuaikan dengan pertanyaan penelitian dan saran akan metode yang diusulkan.