BAB I
PENDAHULUAN

1.1 Latar Belakang

Perkembangan *internet* kian hari semakin maju, banyak industri yang tubuh dan berkembang yang menyebabkan perputaran ekonomi yang luar bisa baik industri yang baru atau yang lama dampaknya besar bagi kehidupan[1]. Internet menjadi sebuah dunia baru yang dijelajahi sejak awal tahun 2000-an, yang membuat banyak kegiatan yang semula ada di dunia nyata kini beberapa berpindah di *internet* dan menjadi sebuah kebutuhan[2]. Dengan banyaknya aktivitas di *internet* orang lupa dan tidak sadar bahwa di dunia maya tersebut terdapat ancaman yang mengintai yaitu kejahatan siber[1][3].

Salah satu bentuk kejahatan siber adalah *malicious* URL yang memiliki banyak macam bentuk seperti *phishing, spoofing, malware downloader* dan *spam* yang digunakan oleh pelaku kejahatan siber untuk mencuri data pribadi [3] - [4]. Umumnya pelaku menyerang akun korban yang mempunyai *kredensial* didalamnya seperti akun *game, ecommerce, banking, social media* dan lain –lain[2][5]. Banyak korban yang terjebak dan termakan *malicious* URL, ini terjadi karena kurang nya pemahaman tentang *malicious* URL atau dalam hal ini phishing, sehingga perlu dikembangkan model, algoritma atau alat yang dapat membantu pengguna untuk mencegah dan meminimalisir dampak buruk yang ditimbulkan dari kejahatan phishing.

Banyak penelitian yang sudah dilakukan tapi umumnya menggunakan metode konvensional seperti *black list* dan *signature*, pendekatan ini juga dipakai oleh banyak vendor antivirus untuk mendeteksi web phishing, kekurangan dari metode ini adalah tidak bisa mendeteksi web phishing yang baru dan masing - masing vendor mempunyai *black list* nya sendiri, jadi vendor yang satu bisa beda dengan yang lainnya[6][2][4]. Ada juga penelitian yang menggunakan *data mining* serta *machine learning* untuk klasifikasi dan pemodelan statistika menggunakan dataset pada deteksi web phishing untuk memperoleh akurasi yang lebih baik dan

juga performa yang tinggi. Pemodelan dan dataset menjadi pembeda pada setiap penelitian termasuk pada penelitian ini.

Penelitian ini peneliti mengusulkan menggunakan *data mining* untuk mengekstraksi dan menemukan pola sehingga dapat dibuat model prediksi dengan algoritma *decision tree*. Perbedaan disini adalah pada dataset, dataset diperoleh dari memanfaatkan API VirusTotal untuk memperoleh dan mengumpulkan analisis vendor mengenai URL web yang telah diperoleh sebelumnya di situs Kaggle dan ada beberapa yang tersebar di sosial media, kemudian diolah dan dibuat model prediksi menggunakan *decision tree* untuk mengetahui hasil pengukuran evaluasi terbaik dibuat empat percobaan rasio pembagian data yang berbeda – beda dan *decision tree* juga dibandingkan dengan algoritma klasifikasi lain. Untuk mempermudah menggetahui hasil prediksi atau pengujin model *decision tree* dibuat aplikasi web sederhana.

1.2 Rumusan Masalah

Berdasarkan penjelasan pada bagian latar belakang, maka dapat dirumuskan sebuah permasalahan dalam penelitian ini adalah bagaimana membuat model prediksi untuk deteksi web phishing menggunakan algoritma *decision tree* dan bagaimana perbandingan dengan rasio pembagian data yang berbeda serta perbandingan akurasi dengan algoritma lain.

1.3 Batasan Masalah

Untuk mempersempit pembahasan pada skripsi ini, maka dibuat batasan batasan sebagai berikut:

1. Penelitian hanya fokus pada bentuk kejahatan siber yaitu *malicious* URL dan salah satu bentuknya yaitu phishing.
2. Penelitian menggunakan algoritma *decision tree* untuk membuat model prediksi dan hanya dibandingkan dengan tiga algoritma klasifikasi lain.
3. Dataset yang digunakan dalam penelitian diperoleh melalui API VirusTotal.
4. Model prediksi akan diaplikasikan dalam bentuk aplikasi web sederhana.

1.4 Tujuan Penelitian

Tujuan yang akan dicapai oleh peneliti dalam penelitiannya adalah mengembangkan model *machine learning* untuk deteksi *malicious* URL terutama web phishing memanfaatkan API VirusTotal.

1.5 Manfaat Penelitian

Manfaat dari penelitian ini adalah sebagai berikut:

a. Memberikan solusi bagi masalah keamanan siber dan mencegah pencurian data pribadi.
b. Memberikan pemahaman yang lebih baik bagi pengguna internet tentang *malicious* URL terutama phishing dan membantu menghindari penyerangan dan menjadi korban.
c. Menjadi alternatif yang lebih efektif dan efisien dalam deteksi web phishing dibandingkan dengan metode *black list* dan *signature*.
d. Memberikan sumbangan terhadap pengembangan teknologi dan ilmu pengetahuan dalam bidang keamanan siber.

1.6 Sistematika Penulisan

BAB I PENDAHULUAN, berisi latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan skripsi ini.

BAB II TINJAUAN PUSTAKA, berisi tinjauan pustaka dan dasar-dasar teori yang digunakan.

BAB III METODE PENELITIAN, didalamnya terdapat tinjauan umum tentang alur penelitian dalam bentuk alur beserta penjelasannya dan peralatan yang digunakan dalam penelitian.

BAB IV HASIL DAN PEMBAHASAN, bab ini berisi tahapan peneliti dalam melakukan penelitian mulai dari pengguaaan dataset, *preprocessing data*, menjalankan model *decision tree,* mengevaluasi lalu pengujiannya.

BAB V PENUTUP, berisi kesimpulan dan saran dari penelitian yang dilakukan.