BAB I
PENDAHULUAN

1.1 Latar Belakang

Phising adalah tindakan dimana penyerang atau peretas mengambil informasi sensitif dari seorang pengguna dengan cara memalsukan halaman website atau mengarahkan pengguna untuk mengunjungi halaman website palsu dimana data pribadi pengguna diambil akibat kurangnya kewaspadaan dan tidak mengerti terhadap serangan phising [1]. Web-phising adalah salah satu metode phising dimana pengguna yang merupakan target serangan web-phising akan diarahkan untuk mengakses halaman website palsu, dimana website palsu tersebut serupa dengan website asli, sehingga target dari web-phising tidak menyadari dan ketika korban mengisikan informasi di web-phising, informasi dan data dari korban bisa disalahgunakan oleh penyerang atau peretas. Berdasarkan data dari BSSN (Badan Siber dan Sandi Negara) tahun 2020 dari bulan Januari sampai April, serangan berkaitan dengan pencurian data pengguna sebesar 43% dari total serangan siber [2]. Jenis serangan dominan menggunakan latar belakang Coronavirus disease 2019 (COVID-19) dengan jenis malicious email phising. Beberapa faktor korban web-phising tidak mengetahui website yang dikunjungi adalah web-phising, yaitu: pengguna tidak memiliki pengetahuan tentang URL, pengguna tidak mengerti website yang asli yang ingin diakses, pengguna tidak melihat alamat website karena beberapa menggunakan direct atau hidden URL, pengguna tidak memiliki waktu untuk mengkonsultasikan website yang akan diakses atau tidak sengaja mengunjunginya, serta pengguna tidak bisa membedakan antara website asli dengan web-phising [3]. Korban web-phising semakin bertambah terlebih lagi pandemi Coronavirus disease (COVID-19) membuat pekerjaan beralih secara online. Apabila pengguna terkena web-phising, semua data yang diinputkan akan diambil, sehingga menimbulkan kerugian bagi korban.

Penelitian tentang cara mendeteksi web-phising terus dikembangkan untuk meminimalisir kerugian bagi korban. Machine learning adalah salah satu cara untuk melakukan deteksi web-phising dengan menggunakan metode-metode yang ada di

dalamnya. Penelitian ini mengembangkan cara mendeteksi website phising telah dilakukan oleh peneliti menggunakan berbagai macam algoritma dan teknik data mining. Penelitian sebelumnya menggunakan satu algoritma untuk klasifikasi pendeteksian web-phising (single classification). Hasilnya, dengan menggunakan Decision Tree menghasilkan kinerja akurasi sebesar 95.35%. Penelitian web-phising juga dilakukan menggunakan Artificial Neural Network, K-Nearest Neighbour, Support Vector Machine (SVM), Random Forest dan Rotation Forest [4]. Penelitian lain menggunakan model Pruned Decision Tree yang menghasilkan kinerja tertinggi dibandingkan dengan Support Vector Machine, Naïve Bayes Classifier atau Neural Network [5].

Penggunaan satu algoritma untuk pemrosesan data membutuhkan waktu yang lama. Salah satu metode di dalam machine learning untuk mengurangi waktu pemrosesan yakni menggunakan fitur seleksi. Fitur seleksi adalah tahap utama pada pre-processing di machine learning. Fitur seleksi adalah metode di machine learning yang berguna untuk mengetahui fitur paling penting di dalam dataset. Fitur seleksi akan meranking semua fitur di dalam dataset sehingga menambah kinerja dari model yang digunakan. Penelitian sebelumnya menunjukan pengaruh fitur seleksi setelah diuji dengan 3 dataset, yakni: bank marketing dataset, car evaluation dataset dan human activity recognition using smartphone dataset. Hasilnya, terdapat peningkatan signifikan dengan kombinasi beberapa algoritma yakni: Support Vector Machine, K-Nearest Neighbor dan Random Forest [6]. Hasil penelitian lain menunjukan fitur seleksi L1-Based Linear Support Vector Machine dengan Classifier Multi-layer Perceptron menghasilkan kinerja terbaik untuk diferensiasi GBM dan LGG pada cross-validation pada analisa radiomik untuk grading glioma [7]. Oleh karena itu, identifikasi fitur yang dominan dengan memanfaatkan mekanisme fitur seleksi menjadi fokus utama dalam penelitian ini. Fitur seleksi populer seperti Pearson Correlation, Information Gain, Gain Ratio, Chi-Square dan Recursive Feature Elimination (RFE) menjadi metode yang diuji kinerjanya terhadap proses deteksi web-phising.