

BAB I

PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi informasi yang pesat telah memberikan kemudahan dalam mengakses berbagai layanan secara daring. Namun, seiring dengan kemajuan tersebut, kejahatan siber pun semakin meningkat, salah satunya adalah serangan *phishing*. *Phishing* merupakan upaya penipuan yang dilakukan dengan menyamar sebagai entitas terpercaya untuk memperoleh informasi sensitif seperti kata sandi, nomor kartu kredit, atau data pribadi pengguna [1].

Website phishing biasanya dibuat dengan tampilan yang sangat mirip dengan situs resmi, sehingga sulit dibedakan secara kasat mata oleh pengguna awam serta kurangnya sistem berbasis web yang mampu mendeteksi *website phishing* secara akurat [2]. Oleh karena itu, dibutuhkan sistem yang mampu mendeteksi *website phishing* secara otomatis dan akurat. Salah satu pendekatan yang efektif untuk menangani masalah ini adalah dengan menggunakan algoritma *machine learning* [3].

Algoritma *random forest* dan *xgboost* merupakan dua metode klasifikasi yang terbukti memiliki performa tinggi dalam banyak kasus deteksi, termasuk deteksi *phishing* [4]. *Random Forest* bekerja dengan membentuk kumpulan pohon keputusan dan menggabungkan hasilnya untuk meningkatkan akurasi prediksi. *xgboost* atau *extreme gradient boosting* adalah metode pembelajaran mesin yang

digunakan untuk menyelesaikan masalah klasifikasi dengan regresi [4]. Algoritma ini menggunakan teknik *gradient boosting decision tree*.

Algoritma ini menggabungkan beberapa model kecil, biasanya *decision tree*, untuk membuat model yang lebih kuat dan memberikan hasil yang lebih baik dalam akurasi [5]. Sementara itu, *xgboost* menggunakan teknik *boosting* untuk meningkatkan performa model secara bertahap dengan meminimalkan kesalahan prediksi. *XGBoost* atau *extreme gradient boosting* adalah metode pembelajaran mesin yang digunakan untuk menyelesaikan masalah klasifikasi dengan regresi [6]. Algoritma ini menggunakan *decision tree* yang dibangun secara sekuensial dengan kedalaman yang sama sehingga basis utamanya adalah *level-wise* atau *depth-wise*. *XGBoost* memperkenalkan beberapa teknik yang membuat algoritma *gradient boosting* lebih efisien dan efektif, seperti *regularization* dan *loss function* [7].

Penelitian ini bertujuan untuk mengimplementasikan dan membandingkan kinerja algoritma *random forest* dan *xgboost* dalam mendeteksi *website phishing*. Sistem ini akan dikembangkan dalam bentuk aplikasi berbasis web agar dapat digunakan dengan mudah oleh pengguna umum dalam mengidentifikasi potensi ancaman *phishing*.

1.2 Identifikasi Masalah

Berdasarkan Latar belakang yang telah dipaparkan, maka dapat diidentifikasi pokok permasalahan utama sebagai berikut:

1. Meningkatnya jumlah *website phishing* yang sulit dikenali secara manual oleh pengguna.
2. Kurangnya sistem berbasis web yang mampu mendeteksi *website phishing* secara akurat.
3. Tidak ada perbandingan pengujian algoritma klasifikasi *random forest* dan *xgboost* dalam mendeteksi *website phishing*.

1.3 Rumusan Masalah

Berdasarkan pembahasan latar belakang di atas, rumusan masalah dalam penelitian ini, yaitu:

1. Bagaimana merancang sistem berbasis web yang mampu mendeteksi *website phishing* secara akurat ?
2. Bagaimana penerapan algoritma klasifikasi *random forest* dan *xgboost* dalam mendeteksi *website phishing* ?
3. Algoritma mana yang memiliki performa terbaik antara *random forest* dan *xgboost* dalam mendeteksi *website phishing* ?

1.4 Tujuan Penelitian

Tujuan Penelitian sebagai berikut:

1. Mengimplementasikan algoritma *random forest* dan *xgboost* dalam sistem deteksi *website phishing*.
2. Membandingkan kinerja kedua algoritma berdasarkan metrik evaluasi seperti akurasi, *precision*, *recall*, dan *F1-score*.

3. Membuat aplikasi berbasis web yang membandingkan hasil pengujian kedua algoritma tersebut.

1.5 Manfaat Penelitian

Manfaat Penelitian Sebagai Berikut:

1. **Bagi pengguna umum:** Memberikan alat bantu untuk menghindari *website phishing* dan menjaga keamanan data pribadi.
2. **Bagi peneliti dan akademisi:** Memberikan referensi dan studi perbandingan antara algoritma klasifikasi *xgboost* dan *random forest* dalam konteks keamanan siber.
3. **Bagi pengembang sistem keamanan:** Menyediakan dasar untuk pengembangan sistem deteksi *phishing* yang dapat diintegrasikan ke layanan keamanan digital.

1.6 Batasan Masalah

Batasan masalah penelitian dalam hal penulisan maupun penelitian sebagai berikut:

1. Data yang digunakan adalah URL yang secara massal digunakan.
2. Algoritma yang digunakan terbatas pada *random forest* dan *xgboost*.
3. Aplikasi berbasis web hanya dapat mendeteksi status *phishing* dari *dataset* yang digunakan.

1.7 Sistematika Penulisan

Sistematika penulisan pada skripsi ini adalah menjelaskan secara ringkas isi dari masing-masing bab dalam skripsi, sebagai berikut:

BAB I PENDAHULUAN

Berisi latar belakang masalah, identifikasi masalah, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, dan sistematika penulisan. Bab ini memberikan gambaran umum mengenai alasan mengapa topik ini dipilih, serta ruang lingkup dan fokus penelitian yang dilakukan.

BAB II TINJAUAN PUSTAKA

Membahas teori-teori dan konsep-konsep yang mendasari penelitian, seperti pengertian *phishing*, *machine learning*, klasifikasi, serta algoritma *xgboost* dan *random forest*. Bab ini juga mencakup kajian terhadap penelitian-penelitian terdahulu yang relevan.

BAB III METODOLOGI PENELITIAN

Menjelaskan pendekatan penelitian yang digunakan, sumber dan jenis data, tahapan *preprocessing data*, implementasi algoritma, serta metode evaluasi model. Bab ini juga menggambarkan alur kerja penelitian secara sistematis.

BAB IV HASIL DAN PEMBAHASAN

Bab ini menyajikan hasil dari proses implementasi sistem deteksi *website phishing* menggunakan algoritma *XGBoost* dan *Random*

Forest. Data yang telah dianalisis ditampilkan dalam bentuk tabel, grafik, dan visualisasi lain untuk memudahkan interpretasi. Selain itu, dilakukan pembahasan terhadap hasil yang diperoleh, dikaitkan dengan tujuan dan rumusan masalah penelitian.

BAB V PENUTUP

Bab ini berisi kesimpulan dari hasil penelitian yang telah dilakukan serta saran-saran untuk pengembangan penelitian di masa mendatang. Kesimpulan dibuat berdasarkan hasil evaluasi model dan pembahasan sebelumnya, sedangkan saran ditujukan bagi peneliti selanjutnya maupun pihak yang berkepentingan terhadap sistem yang dikembangkan.

