BAB I
PENDAHULUAN

1.1 Latar Belakang

Akselerasi teknologi digital telah memicu transformasi yang signfikan pada berbagai sektor kehidupan, dimulai dengan sistem komunikasi, transaksi keuangan, hingga layanan publik. Internet kini menjadi sarana utama dalam mengakses berbagai layanan tersebut. Namun, di balik kemudahan tersebut, terdapat ancaman keamanan siber yang dapat menimbulkan kerugian finansial maupun kebocoran privasi. Ancaman yang paling sering terjadi adalah penipuan tautan palsu (phising), yaitu upaya dalam melakukan penipuan digital dengan memanipulasi korban agar memberikan data pribadi yang sensitif, kredensial akun hingga data keuangan. [1].

Metode deteksi phishing awam ditemukan saat ini masih banyak bergantung pada daftar hitam (blacklist) dan pemeriksaan berbasis server. Pendekatan tersebut memiliki keterbatasan dalam menghadapi serangan baru yang belum terdaftar serta menimbulkan keterlambatan deteksi karena ketergantungan pada koneksi jaringan dan layanan eksternal. Selain itu, pengiriman data penjelajahan ke server pihak ketiga juga berpotensi menimbulkan risiko privasi bagi pengguna[2], [3].

Seiring berkembangnya pendekatan machine learning[4] dalam keamanan siber, berbagai model klasifikasi telah digunakan untuk mendeteksi phishing secara otomatis. Model kompleks seperti deep learning mampu mencapai tingkat akurasi tinggi[5], [6], namun membutuhkan beban komputasi yang besar dan waktu yang relatif lama dalam mengambil referensi sehingga kurang sesuai untuk diterapkan secara langsung di sisi pengguna. Sebaliknya, model yang lebih ringan seperti Logistic Regression[7] menawarkan kecepatan dan efisiensi komputasi yang tinggi, tetapi pada praktiknya masih berpotensi menghasilkan kesalahan klasifikasi[8], [9].

Berdasarkan hasil eksperimen awal, penggunaan satu model tunggal belum sepenuhnya mampu memberikan tingkat keandalan yang optimal dalam mendeteksi berbagai variasi phising[10], [11]. Oleh karena itu, diperlukan pendekatan yang tetap ringan namun lebih andal dalam menangani variasi pola serangan. Dalam penelitian ini dikembangkan sebuah sistem deteksi phishing berbasis ekstensi Google Chrome yang bekerja secara real-time dan sepenuhnya di sisi pengguna dengan mengadopsi pendekatan pencampuran metode. Sistem memanfaatkan Logistic Regression sebagai model deteksi awal dan Random Forest [12]sebagai lapisan verifikasi tambahan untuk menekan tingkat hasil negatif palsu (false negative) tanpa mengorbankan efisiensi komputasi[13]. Pendekatan ini diharapkan mampu memberikan peringatan instan yang lebih akurat dan adaptif bagi pengguna saat mengakses situs web.

1.2 Rumusan Masalah

Pertanyaan berikut akan dijawab dalam penelitian ini :

1. Bagaimana merancang sistem deteksi phishing berbasis ekstensi Google Chrome yang mampu bekerja secara real-time dan sepenuhnya di sisi pengguna?
2. Bagaimana mengkombinasikan algoritma Logistic Regression dan Random Forest dalam suatu mekanisme penggabungan (hybrid) untuk meningkatkan keandalan deteksi phishing?
3. Bagaimana kinerja sistem hybrid dalam mengklasifikasikan tautan menjadi kategori aman, waspada, dan berbahaya berdasarkan hasil analisis model?
4. Bagaimana kinerja model Logistic Regression, Random Forest, dan Hybrid dalam membedakan tautan palsu dan tautan sah berdasarkan metrik evaluasi seperti Accuracy, Recall, F1 Score, Precision, dan ROC-AUC?

1.3 Batasan Masalah

Batasan masalah yang ada pada penelitian ini, diantara lain:

1. Sistem deteksi difokuskan pada analisis fitur pada tautan dan fitur ringan dari struktur halaman tanpa melakukan analisis visual mendalam.
2. Algoritma yang digunakan terbatas pada kombinasi Logistic Regression dan Random Forest tanpa membandingkan dengan model deep learning yang lebih kompleks.
3. Implementasi sistem hanya ditujukan pada ekstensi Google Chrome dan belum mencakup peramban (browser) lainnya.

1.4 Tujuan Penelitian

Penelitian ini ditujukan untuk membangun sistem deteksi phishing yang efisien secara komputasi, ringan, dan mampu beroperasi secara real-time di sisi pengguna melalui implementasi ekstensi Google Chrome. Guna meningkatkan performa deteksi dibandingkan metode konvensional, penelitian ini menerapkan pendekatan *hybrid* yang melakukan penggabungan algoritma Logistic Regression dan Random Forest, dengan fokus pada analisis laman tautan untuk memaksimalkan akurasi serta meminimalkan tingkat false negative. Selain menguji kelayakan model menggunakan metrik klasifikasi standar, penelitian ini juga

mengevaluasi efektivitas sistem dalam memberikan peringatan dini secara langsung kepada pengguna saat mengakses situs web yang terindikasi berbahaya.

1.5 Manfaat Penelitian

Manfaat yang diharapkan dari penelitian ini antara lain:

1. Memberikan kontribusi praktis dalam pengembangan sistem deteksi phishing yang efisien dan adaptif.
2. Menyediakan solusi nyata berupa ekstensi Chrome yang dapat membantu pengguna mendeteksi potensi phishing secara langsung.
3. Menjadi dasar bagi pengembangan lanjutan sistem keamanan web berbasis pendekatan hybrid machine learning.
4. Mencegah lebih awal atas risiko keamanan siber saat mengakses layanan berbasis web.