BAB I
PENDAHULUAN

1.1 Latar Belakang

Kemajuan teknologi digital telah meningkatkan ketergantungan masyarakat pada layanan berbasis internet seperti perbankan digital, e-commerce, dan komunikasi daring. Namun, perkembangan ini turut disertai oleh peningkatan ancaman siber, salah satunya adalah *phishing*. Berdasarkan laporan Anti-Phishing Working Group (APWG), tercatat 989.123 serangan phishing terjadi secara global pada kuartal keempat tahun 2024 dengan estimasi kerugian mencapai USD 128.980 [1]. Di Indonesia, Badan Siber dan Sandi Negara (BSSN) melaporkan lebih dari 26 juta serangan *phishing* pada tahun 2024, terutama menargetkan sektor pemerintahan dan keuangan [2].

Metode deteksi *phishing* tradisional seperti *blacklisting* dan *signature-based detection* semakin dianggap tidak memadai dalam menghadapi serangan yang semakin kompleks dan dinamis. Deteksi berbasis pembaruan basis data cenderung tertinggal dalam mengantisipasi serangan *phishing* baru seperti *zero-day attacks*. Beberapa fitur URL seperti panjang URL, jumlah simbol, dan *domain-level* dapat dianalisis menggunakan *machine learning* untuk meningkatkan efektivitas deteksi [3].

Seiring berkembangnya penelitian, berbagai algoritma *machine learning* telah digunakan untuk mendeteksi *phishing* berbasis URL. Beberapa di antaranya adalah *Support Vector Machine, K-Nearest Neighbor, serta Decision Tree, Logistic Regression, XGBoost, Convolutional Neural Network, dan Deep Learning*. Hasil perbandingan performa algoritma-algoritma tersebut menunjukkan bahwa model CNN mencapai akurasi sebesar 99%, dengan *precision* 98%, *recall* 98%, dan *F1-score* 99%, sedangkan *Random Forest* memperoleh akurasi 98%, *precision* 98%, *recall* 98%, dan *F1-score* 98%. Sementara itu, algoritma seperti *Logistic Regression* hanya mencapai akurasi 95%, *precision* 94%, *recall* 94%, dan *F1-score* 94%, *serta K-Nearest Neighbor* yang memiliki akurasi 96%, *precision* 96%, *recall* 95%, dan *F1-score* 94% [4].

Sementara itu, pendekatan berbasis *Random Forest* telah digunakan secara luas untuk mengklasifikasikan URL *phishing* karena kemampuannya menangani data berdimensi tinggi dan menghasilkan model yang dapat diinterpretasikan [5]. Sebagian besar pendekatan deteksi *phishing* masih bergantung pada analisis konten situs atau metadata email, yang memiliki tantangan privasi dan latensi. Pendekatan deteksi berbasis struktur URL dianggap penting dalam mengidentifikasi ancaman *phishing* tanpa perlu memuat keseluruhan halaman situs [6]. Dalam konteks jaringan, pendekatan ini memungkinkan sistem keamanan untuk mengidentifikasi potensi serangan sejak tahap awal, sebelum pengguna benar-benar terhubung ke situs *phishing*. Efektivitas metode klasifikasi URL dengan berbagai algoritma *machine learning* terbukti mampu meningkatkan sistem keamanan secara preventif [7].

Penelitian ini bertujuan untuk menjawab kebutuhan akan sistem deteksi *phishing* yang adaptif dan *low latency* melalui pengembangan model AI berbasis *Random Forest* yang menganalisis struktur URL. Model ini dirancang agar dapat beroperasi secara efisien dalam sistem keamanan jaringan seperti *firewall* atau *proxy*. Dengan demikian, pendekatan ini diharapkan dapat memperkuat ketahanan siber organisasi dalam menghadapi ancaman *phishing* yang terus berkembang seiring transformasi digital.

1.2 Rumusan Masalah

Berdasarkan latar belakang masalah yang telah dijelaskan, maka rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Bagaimana cara mengidentifikasi URL *phishing* menggunakan algoritma *Random Forest*?
2. Fitur-fitur apa saja dari struktur *URL* yang relevan dan efektif dalam membedakan URL *phishing* dan *non-phishing* di jaringan komputer?

1.3 Batasan Masalah

Berdasarkan latar belakang masalah yang telah dijelaskan, maka rumusan masalah dalam penelitian ini adalah sebagai berikut:

1. Penelitian ini hanya menganalisis URL tanpa melakukan inspeksi terhadap konten halaman web atau isi email.
2. *Model AI* yang dikembangkan terbatas pada penggunaan algoritma *Random Forest*, dan tidak dilakukan perbandingan menyeluruh dengan algoritma lain seperti *SVM*, *XGBoost*, atau *Neural Network*.
3. *Dataset* yang digunakan terdiri dari URL yang telah dilabeli sebagai *phishing* dan *non-phishing*, dengan asumsi bahwa data tersebut valid dan representatif.
4. Implementasi dan pengujian dilakukan pada lingkungan simulasi, bukan pada sistem produksi skala besar.
5. Variabel seperti kecepatan internet, jenis enkripsi trafik, dan teknik penghindaran *phishing* lanjutan (misal URL *shortening*) tidak dianalisis secara mendalam dan diasumsikan konstan.

1.4 Tujuan Penelitian

Tujuan yang ingin dicapai dalam penelitian ini adalah:

1. Mengembangkan model deteksi *phishing* berbasis algoritma *Random Forest* yang mampu mengklasifikasikan URL *phishing*.
2. Mengidentifikasi dan menganalisis fitur struktural URL yang relevan untuk membedakan URL *phishing* dari URL yang *non-phishing*.

1.5 Manfaat Penelitian

Penelitian ini diharapkan memberikan manfaat sebagai berikut:

Secara Teoritis:

Penelitian ini berkontribusi dalam pengembangan kajian ilmiah terkait penerapan *machine learning* di bidang keamanan siber, khususnya dalam deteksi URL *phishing* berbasis struktur URL. Hasil penelitian ini juga dapat menjadi referensi bagi penelitian selanjutnya dalam pengembangan sistem deteksi *real-time* berbasis data jaringan.

Secara Praktis:

Hasil dari penelitian ini dapat dimanfaatkan oleh organisasi atau institusi untuk meningkatkan sistem keamanan jaringan internal. Model yang dikembangkan dapat diintegrasikan dalam sistem deteksi dini untuk memblokir akses ke URL *phishing* sebelum terjadi kerugian data atau finansial, serta mendukung efisiensi dan kecepatan dalam proses mitigasi serangan *phishing*.

1.6 Sistematika Penulisan

Sistematika penulisan skripsi ini disusun untuk memberikan gambaran umum mengenai isi dari masing-masing bab yang terdapat dalam laporan penelitian. Adapun sistematika penulisan yang digunakan adalah sebagai berikut:

BAB I PENDAHULUAN

Bab ini berisi latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan. Bab ini memberikan dasar dan alasan dilakukannya penelitian serta arah dan ruang lingkup penelitian.

BAB II TINJAUAN PUSTAKA

Bab ini berisi tinjauan pustaka terhadap penelitian-penelitian sebelumnya yang relevan, dasar-dasar teori yang mendukung penelitian, serta kerangka teori dan kerangka pemikiran. Teori-teori yang digunakan mencakup keamanan jaringan, *phishing*, URL analysis, serta algoritma *Random Forest*.

BAB III METODE PENELITIAN

Bab ini menjelaskan metode yang digunakan dalam penelitian, meliputi jenis penelitian, objek penelitian, teknik pengumpulan dan pengolahan data, serta tahapan pembangunan model deteksi *phishing*. Rancangan sistem, pemilihan fitur URL, dan pengujian model AI juga dijelaskan di bab ini.

BAB IV HASIL DAN PEMBAHASAN

Bab ini memaparkan hasil implementasi model deteksi URL *phishing* menggunakan Algoritma Random Forest, serta evaluasi performa sistem melalui pengujian. Pembahasan meliputi interpretasi hasil, kelebihan dan keterbatasan model, serta potensi penerapan dalam sistem keamanan jaringan.

BAB V PENUTUP

Bab ini berisi kesimpulan dari hasil penelitian serta saran yang dapat diberikan untuk pengembangan penelitian selanjutnya. Kesimpulan mencakup pencapaian tujuan dan jawaban atas rumusan masalah, sedangkan saran berisi arahan untuk pengembangan sistem atau penelitian lanjutan.