BAB I
PENDAHULUAN

1.1 Latar Belakang

Internet membawa banyak perubahan dalam kehidupan sehari hari mulai dari yang baik maupun yang kurang baik, salah satunya yaitu *phising*. *Phishing* merupakan salah satu jenis serangan siber yang paling umum dan secara dominan menargetkan manusia dibandingkan perangkat komputer [1]. Berdasarkan laporan statistik dari *Indonesia Anti-Phsising Data Exchange* (Q4. 2024), insiden *phising* di Indonesia menunjukkan tren peningkatan sejak 2018 hingga 2023 dengan total 108.806 kasus tercatat, lalu pada 2024 jumlah laporan *phising* mencapai 85.414 kasus di Indonesia[2][3]. *Phishing* dilakukan dengan cara menipu pengguna agar memberikan informasi pribadi melalui situs palsu yang menyerupai situs resmi, sehingga menimbulkan risiko pencurian data dan kerugian finansial yang signifikan[4].

Sebagai upaya mitigasi, berbagai solusi preventif seperti *Intrusion Detection System* (IDS) dan *firewall* telah digunakan dan dikembangkan, termasuk dengan bantuan algoritma *machine learning*. Model *machine learning* seperti *Decision Tree* dan *Logistic Regression* telah banyak digunakan untuk mendeteksi *phishing* karena memiliki tingkat akurasi yang baik serta kemampuan dalam menangani data kategorikal maupun numerik. *Decision Tree* dipilih karena bersifat *interpretable* secara langsung melalui struktur pohonnya, sehingga pengguna dapat memahami aturan keputusan yang diambil oleh model. Sementara itu, *Logistic Regression* digunakan sebagai pembanding karena sifatnya yang linier dan memberikan pemahaman dasar terhadap hubungan antara fitur dan output klasifikasi.

Namun, meskipun *Decision Tree* relatif lebih dapat dijelaskan dibandingkan model *black-box* seperti *ensemble* atau *deep learning*, tetap diperlukan pendekatan tambahan untuk memberikan penjelasan yang lebih komprehensif dan mendalam terhadap prediksi model, terutama pada dataset yang kompleks. Tantangan

interpretabilitas tetap menjadi isu krusial, terutama ketika model harus digunakan oleh pihak non-teknis atau untuk keperluan audit dan keamanan[5].

Untuk menjawab permasalahan tersebut, pendekatan *Explainable Artificial Intelligence* (XAI) diperkenalkan sebagai solusi guna meningkatkan interpretabilitas model. XAI memberikan penjelasan mengenai kontribusi setiap fitur terhadap hasil prediksi secara lokal maupun global. Teknik XAI seperti SHAP (*SHapley Additive exPlanations*), LIME (*Local Interpretable Model-Agnostic Explanations*), dan PDP (*Partial Dependence Plot*) telah digunakan di berbagai domain, termasuk dalam keamanan siber[6]. Meskipun demikian, efektivitas masing-masing teknik XAI dalam menjelaskan hasil deteksi phishing belum banyak dikaji secara komparatif.

Oleh karena itu, penelitian ini bertujuan untuk membandingkan beberapa teknik XAI dalam menjelaskan model deteksi phishing berbasis machine learning. Dengan membandingkan aspek interpretabilitas dari SHAP, LIME, dan PDP, penelitian ini diharapkan dapat memberikan wawasan yang lebih dalam mengenai transparansi model, serta kontribusi nyata bagi pengembangan sistem keamanan siber yang dapat dipercaya dan mudah dipahami oleh pengguna maupun analis keamanan.

1.2 Rumusan Masalah

1. Bagaimana kinerja algoritma SHAP, PDP, dan LIME dalam prediksi model deteksi phishing berbasis *machine learning*?
2. Bagaimana masing-masing metode *Explainable AI* (XAI) menjelaskan kontribusi fitur, dan mana yang paling efektif dalam meningkatkan interpretabilitas model deteksi phishing?

1.3 Batasan Masalah

1. Penelitian hanya difokuskan pada model *machine learning* seperti *Decision Tree*, *Logistic Regression*, dan *Gradient Boosting Algorithm*.
2. Teknik XAI yang digunakan terbatas pada SHAP, LIME, serta PDP
3. Dataset yang digunakan adalah dataset *phishing* yang bersifat open source

dan tidak mencakup data *phishing* terbaru secara real-time.

4. Evaluasi dilakukan berdasarkan aspek interpretabilitas dan kontribusi fitur, tidak mencakup keamanan terhadap *adversarial attack*.

1.4 Tujuan Penelitian

1. Menganalisis dan membandingkan kinerja teknik *Explainable AI* seperti SHAP, LIME, dan PDP dalam menjelaskan model deteksi *phishing* berbasis *machine learning*.
2. Mengidentifikasi kontribusi fitur-fitur dalam proses prediksi untuk meningkatkan pemahaman terhadap cara kerja model.
3. Memberikan rekomendasi teknik XAI yang paling efektif dalam meningkatkan transparansi dan kepercayaan terhadap sistem deteksi *phishing*.

1.5 Manfaat Penelitian

Secara Teoritis

1. Memberikan kontribusi dalam pengembangan studi terkait *Explainable AI* pada domain keamanan siber, khususnya deteksi *phishing*.
2. Menambah referensi ilmiah mengenai perbandingan metode XAI dalam konteks model deteksi berbasis *machine learning*.

Secara Praktis

1. Memberikan panduan bagi pengembang sistem keamanan untuk memilih teknik XAI yang sesuai dalam membangun sistem deteksi *phishing* yang lebih transparan dan dapat dipercaya.
2. Membantu pengguna dan pemangku kebijakan dalam memahami alasan di balik hasil prediksi sistem deteksi *phishing*, sehingga meningkatkan kepercayaan terhadap teknologi yang digunakan.

1.6 Sistematika Penulisan

BAB I PENDAHULUAN, bab ini berisi latar belakang masalah, rumusan masalah,

batasan masalah, tujuan penelitian, manfaat penelitian, serta sistematika penulisan skripsi

BAB II TINJAUAN PUSTAKA, berisi kajian literatur dari penelitian sebelumnya yang relevan, dasar teori yang mendukung penelitian, serta penjelasan konsep-konsep seperti *phishing*, model *machine learning*, dan teknik *Explainable AI* (XAI) seperti SHAP, LIME, dan PDP.

BAB III METODE PENELITIAN, Bab ini menjelaskan objek penelitian, alur dan langkah-langkah pelaksanaan penelitian, alat dan bahan yang digunakan, serta pendekatan analisis yang dilakukan dalam proses perbandingan teknik XAI pada model deteksi *phishing*.

BAB IV HASIL DAN PEMBAHASAN, Berisi hasil eksperimen dari model yang diuji, visualisasi hasil dari masing-masing teknik XAI, serta pembahasan mengenai efektivitas, kelebihan, dan kekurangan dari tiap metode dalam konteks deteksi *phishing*.

BAB V PENUTUP, berisi kesimpulan dari penelitian yang telah dilakukan serta memberi saran yang dapat peneliti rangkum selama proses penelitian