BAB I
PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi informasi yang pesat memudahkan masyarakat dalam mengakses berbagai layanan *digital*, seperti *digital finance* dan media sosial. Namun, kemajuan ini juga diikuti dengan meningkatnya ancaman siber, salah satunya *Phishing*, yaitu kejahatan yang memanfaatkan rekayasa sosial dan penyamaran teknis untuk mencuri informasi identitas pribadi dan kredensial akun keuangan pengguna dengan cara meniru situs web resmi, yang dapat dilakukan melalui berbagai saluran seperti situs web palsu, pesan teks, email, maupun media sosial. Berdasarkan laporan APWG, *Phishing* berbasis situs web tetap menjadi salah satu bentuk yang paling dominan, terlihat dari peningkatan 13,8 persen jumlah situs *Phishing* unik yang dilaporkan pada awal 2026, dari 853.244 kasus pada Q4 2025 menjadi 971.181 kasus pada Q1 2026 [1]. Sementara itu, laporan *Hack3D: The* Web3 *Security Report 2025* yang diterbitkan oleh CertiK menunjukkan bahwa serangan *Phishing* pada tahun 2025 mengakibatkan kerugian sekitar US$722,9 juta [2], yang setara dengan sekitar Rp11,8 triliun. Nilai kerugian tersebut menunjukkan bahwa *Phishing* masih menjadi salah satu ancaman yang memberikan dampak finansial signifikan. *Chain*alysis juga melaporkan bahwa pelaku *Phishing* semakin banyak menggunakan situs Web3 palsu yang menyamar sebagai proyek atau halaman *airdrop* untuk mengelabui pengguna agar menghubungkan dompet dan menyetujui transaksi berbahaya, sehingga aset korban dapat dikuras melalui teknik *crypto drainer*[3]. Fakta ini menunjukkan bahwa *Phishing* telah berkembang menjadi salah satu ancaman utama dalam ekosistem Web3 atau *Cryptocurrency*.

Pendekatan berbasis *blacklist* merupakan salah satu metode yang paling umum digunakan dalam deteksi *Phishing* dengan cara memblokir akses ke domain yang telah teridentifikasi sebagai bahaya. Namun, efektivitas metode ini terbatas yaitu tidak mampu mendeteksi situs *Phishing* baru (*zero-day Phishing*) yang belum tercatat dalam basis data *blacklist* [4]. Keterbatasan tersebut menunjukkan perlunya pendekatan deteksi yang lebih adaptif dan mampu mengenali karakteristik *Phishing*

secara otomatis, sehingga tetap efektif dalam menghadapi kemunculan ancaman baru yang belum pernah teridentifikasi sebelumnya.

Salah satu pendekatan yang relevan adalah penelitian DomURLs-BERT yang diperkenalkan oleh El Mahdaouy dkk. (2024), yaitu model BERT yang dipra-latih khusus pada korpus multilingual berisi nama domain, URL, dan dataset *Domain Generation Algorithm* (DGA). Hasil eksperimen menunjukkan bahwa model ini mampu mendeteksi domain dan URL berbahaya secara efektif pada berbagai tugas klasifikasi meliputi *Phishing*, *malware*, DGA, dan *DNS tunneling*, dengan performa yang mengungguli model berbasis karakter seperti CNN dan RNN [5]. Namun, pendekatan berbasis URL memiliki keterbatasan dimana teknik analisis URL tidak mempertimbangkan konten halaman web, sehingga berpotensi salah mengklasifikasi situs *Phishing*. Pelaku *Phishing* dapat menggunakan nama domain yang tampak sah secara leksikal namun menyembunyikan konten berbahaya di dalam struktur halaman web [6].

Penelitian selanjutnya mengembangkan analisis konten HTML dengan mengekstraksi elemen struktural seperti *hyperlink*, formulir *login*, dan rasio tautan eksternal. Salah satu penelitian oleh Aljofey *et al*. berhasil menggabungkan fitur URL dan HTML menggunakan algoritma XGBoost mampu mencapai akurasi hingga 98,48%. Namun, pendekatan ini tetap memiliki keterbatasan karena penyerang dapat menyembunyikan konten halaman menggunakan JavaScript, maupun teknik *server-side scripting*, sehingga fitur HTML yang diekstraksi tidak selalu mencerminkan perilaku sebenarnya dari halaman *Phishing* [6].

Kondisi ini mendorong eksplorasi dengan menggabungkan informasi dari berbagai sumber analisis untuk menghasilkan representasi yang lebih kaya dibandingkan analisis tunggal. Penelitian tersebut juga menunjukkan bahwa penggabungan fitur yang diekstraksi dari URL dan konten HTML menghasilkan performa klasifikasi yang lebih baik dibandingkan pendekatan yang hanya memanfaatkan satu jenis fitur [6]. Hal ini dikarenakan analisis domain/URL dan konten HTML menyediakan informasi yang berbeda namun saling melengkapi, sehingga menggabungkan kedua sumber informasi tersebut berpotensi menghasilkan keputusan deteksi yang lebih akurat dan *robust*.

Berdasarkan landasan tersebut, penelitian ini mengusulkan sistem deteksi Phishing yang mengintegrasikan analisis domain dan analisis konten HTML. Analisis domain dilakukan berdasarkan karakteristik struktur URL dan nama domain, sedangkan analisis konten HTML dilakukan berdasarkan karakteristik elemen-elemen struktural pada halaman web. Untuk mengatasi keterbatasan masing-masing analisis apabila digunakan secara terpisah, kedua hasil analisis tersebut kemudian digabungkan sehingga menghasilkan keputusan berdasarkan dua analisis yang berbeda. Penelitian ini secara khusus berfokus pada ancaman Phishing pada situs Web3, seperti kampanye airdrop palsu dan situs yang menyamar sebagai *decentralized application* (dApp), karena dinilai sangat relevan dengan kondisi saat ini. Sistem ini diimplementasikan sebagai ekstensi browser pada desktop untuk memberikan perlindungan secara real-time kepada pengguna.

1.2 Rumusan Masalah

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

1. Bagaimana mengembangkan model deteksi Phishing pada situs Web3 yang memanfaatkan analisis domain dan analisis konten HTML untuk meningkatkan akurasi deteksi?
2. Bagaimana mengimplementasikan model deteksi Phishing tersebut sebagai ekstensi browser yang bekerja secara real-time?

1.3 Batasan Masalah

Untuk menjaga fokus penelitian agar lebih terarah, penelitian ini dibatasi pada poin berikut:

1. Sistem hanya melakukan deteksi berdasarkan dua analisis berbeda yaitu analisis domain dan content html.
2. Pada deteksi berbasis analisis domain, sistem hanya mencakup deteksi domain saja, tidak beserta *path* dan *query string*-nya.

3. Pada sistem deteksi berbasis konten HTML, analisis hanya mencakup elemen-elemen pada HTML, tidak mencakup analisis SSL, eksekusi JavaScript, maupun elemen media seperti gambar dan video.
4. Implementasi dilakukan dalam bentuk ekstensi *browser desktop* tidak mencangkup pada perangkat mobile.

1.4 Tujuan

Tujuan penelitian ini adalah:

1. Mengembangkan model deteksi Phishing pada situs Web3 yang memanfaatkan analisis domain dan analisis konten HTML untuk meningkatkan akurasi deteksi.
2. Mengimplementasikan model deteksi Phishing tersebut dalam bentuk ekstensi browser yang bekerja secara real-time.

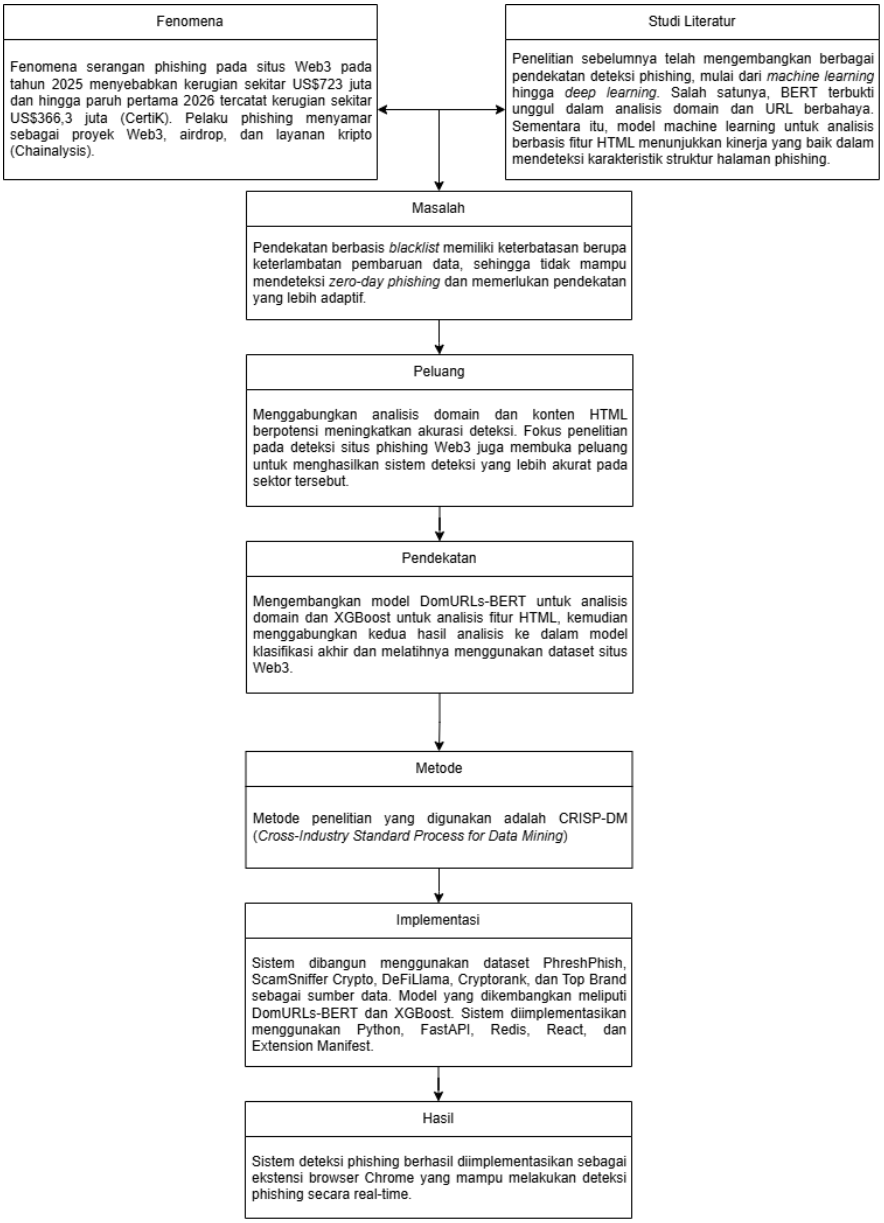
1.5 Manfaat

Penelitian ini diharapkan memberikan manfaat sebagai berikut:

1. Berkontribusi pada pengembangan metode deteksi Phishing melalui integrasi analisis domain dan analisis konten HTML.
2. Berkontribusi pada keamanan lingkungan Web3 melalui deteksi *Phishing* pada proyek *airdrop* dan *decentralized application* (dApp).
3. Menyediakan sistem deteksi *Phishing* dalam bentuk ekstensi *browser* yang mampu memberikan perlindungan secara *real-time* serta menjadi referensi bagi penelitian selanjutnya.

1.6 Kerangka Pemikiran

Kerangka pemikiran ini menggambarkan alur logis penelitian mulai dari fenomena, studi literatur, masalah, peluang, pendekatan, metode, implementasi, hingga hasil yang diharapkan. Kerangka tersebut menjadi dasar dalam merancang metode dan implementasi sistem deteksi yang dikembangkan dalam penelitian ini.



Gambar 1. 1 Kerangka Pemikiran

1.7 Sistematis Penulisan

Skripsi ini disusun dalam lima bab yang saling berkaitan, mulai dari pendahuluan hingga simpulan. Sistematika penulisan ini bertujuan untuk memudahkan pembaca dalam memahami alur penelitian. Adapun sistematika penulisan skripsi ini adalah sebagai berikut.

BAB I PENDAHULUAN

Bab ini berisi uraian mengenai latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, serta kerangka pemikiran yang digunakan dalam penelitian ini.

BAB II TINJAUAN PUSTAKA

Bab ini berisi kajian teori dan konsep yang mendukung penelitian, serta tinjauan terhadap penelitian-penelitian sebelumnya yang berkaitan dengan topik yang dibahas.

BAB III METODOLOGI PENELITIAN

Bab ini menguraikan metode dan tahapan yang digunakan dalam penelitian, mulai dari perancangan hingga proses yang dilakukan untuk mencapai tujuan penelitian.

BAB IV HASIL DAN PEMBAHASAN

Bab ini memaparkan hasil implementasi dari tahapan yang telah diuraikan pada bab metodologi, beserta pembahasan dan analisis terhadap hasil tersebut.

BAB V SIMPULAN

Bab ini berisi simpulan dari keseluruhan penelitian yang telah dilakukan, serta saran yang dapat digunakan untuk pengembangan penelitian selanjutnya.

UNIVERSITAS ISLAM NEGERI
SUNAN GUNUNG DJATI
BANDUNG