BAB 1
PENDAHULUAN

1.1 Latar Belakang Masalah

Phishing merupakan salah satu bentuk ancaman siber yang sering digunakan untuk menyerang pengguna melalui media email dan tautan berbahaya. Pelaku *phishing* umumnya menyamar sebagai pihak tepercaya, seperti institusi keuangan, layanan digital, perusahaan, atau organisasi resmi, untuk memperoleh informasi sensitif dari korban. Informasi yang menjadi target dapat berupa kata sandi, data pribadi, informasi akun, maupun informasi finansial [1].

Serangan *phishing* tidak hanya bergantung pada aspek teknis, tetapi juga memanfaatkan rekayasa sosial. *Email phishing* dapat dibuat menyerupai email resmi melalui penggunaan nama pengirim, gaya bahasa, subjek, isi pesan, serta tautan yang terlihat meyakinkan. Pada sisi lain, *link phishing* sering kali dibuat dengan struktur URL yang menyerupai domain asli, menggunakan tanda hubung, angka, subdomain, path tertentu, atau kata-kata seperti *login*, *verify*, *secure*, dan *account*. Oleh karena itu, pendeteksian *phishing* perlu mempertimbangkan isi pesan email sekaligus karakteristik struktur URL.

Di Indonesia, kasus *phishing* melalui email masih menjadi salah satu ancaman yang perlu diperhatikan. Serangan ini dapat menyerang pengguna individu maupun organisasi karena email masih menjadi media komunikasi digital yang banyak digunakan. Tingginya penggunaan email membuat kanal ini rentan dimanfaatkan oleh pelaku kejahatan siber untuk mengirimkan pesan palsu, mengarahkan pengguna ke halaman tiruan, atau membujuk korban agar memberikan informasi sensitif. Kondisi tersebut menunjukkan bahwa diperlukan sistem deteksi otomatis yang mampu membantu mengidentifikasi *email phishing* dan *link phishing* secara lebih cepat.

Email *phishing* dan *link phishing* memiliki hubungan yang erat dalam pelaksanaan serangan *phishing*. Email sering digunakan sebagai media awal untuk membangun kepercayaan, menciptakan rasa urgensi, atau menyampaikan informasi palsu kepada calon korban. Tautan yang disisipkan di dalam email kemudian digunakan untuk mengarahkan korban ke halaman palsu yang menyerupai situs resmi. Pada halaman tersebut, korban dapat diminta memasukkan informasi sensitif, seperti nama pengguna, kata sandi, kode verifikasi, atau informasi keuangan.

Dengan demikian, email berperan sebagai media penyampaian dan rekayasa sosial, sedangkan tautan berperan sebagai sarana untuk mengarahkan korban menuju tujuan serangan.

Meskipun demikian, tidak seluruh email *phishing* selalu mengandung tautan. Beberapa serangan dapat menggunakan lampiran berbahaya, meminta korban membalas email dengan informasi pribadi, meminta pengiriman uang, atau meminta kode verifikasi secara langsung. Sebaliknya, *link phishing* juga dapat disebarkan melalui media lain, seperti pesan singkat, media sosial, aplikasi percakapan, dan iklan digital. Oleh karena itu, email *phishing* dan *link phishing* merupakan dua objek yang saling berkaitan, tetapi masing-masing tetap dapat muncul dan dianalisis secara terpisah.

Penelitian-penelitian sebelumnya menunjukkan bahwa deteksi *phishing* dapat dilakukan menggunakan berbagai pendekatan, seperti *machine learning*, *deep learning*, *natural language processing*, model berbasis *Transformer*, dan pendekatan berbasis fitur [2]. Pada aspek email, beberapa penelitian telah menggunakan teknik pemrosesan bahasa alami untuk menganalisis isi pesan email. Pada aspek URL, penelitian lain memanfaatkan fitur struktural seperti panjang URL, jumlah subdomain, penggunaan angka, tanda hubung, dan karakter khusus untuk membedakan URL *legitimate* dan URL *phishing* [3, 4]. Selain itu, BERT telah digunakan dalam berbagai tugas klasifikasi teks karena mampu menghasilkan representasi berbasis konteks [5].

Meskipun demikian, masih terdapat ruang penelitian dalam penerapan pendekatan *hybrid* mBERT untuk dua objek *phishing* yang berbeda, yaitu *email phishing* dan *link phishing*. Beberapa penelitian berfokus pada email saja, sedangkan penelitian lain berfokus pada URL saja. Selain itu, pendekatan berbasis BERT umumnya menekankan representasi teks, sedangkan pendekatan berbasis fitur URL lebih menekankan karakteristik struktural. Berdasarkan hal tersebut, penelitian ini menerapkan pendekatan *hybrid* mBERT yang menggabungkan representasi teks dan fitur numerik untuk mendeteksi *email phishing* dan *link phishing*.

Berdasarkan latar belakang di atas, penelitian ini diharapkan dapat menghasilkan model deteksi *email phishing* dan *link phishing* yang mampu memanfaatkan informasi teks sekaligus pola numerik atau struktural dari objek yang dianalisis.

1.2 Rumusan Masalah

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

(a) Bagaimana mengembangkan model klasifikasi *email phishing* menggunakan pendekatan *hybrid* mBERT yang menggabungkan representasi teks dan fitur numerik email?

(b) Bagaimana mengembangkan model klasifikasi *link phishing* menggunakan pendekatan *hybrid* mBERT yang menggabungkan representasi teks URL dan fitur numerik struktur URL?

(c) Bagaimana performa model *hybrid* mBERT dalam mendeteksi *email phishing* dan *link phishing* berdasarkan metrik *accuracy*, *precision*, *recall*, *F1-score*, dan *confusion matrix*?

1.3 Batasan Permasalahan

Batasan permasalahan dalam penelitian ini adalah sebagai berikut:

(a) Fitur email dibatasi pada *sender*, *receiver*, *subject*, *body*, URL, serta fitur numerik yang diekstraksi dari komponen tersebut.

(b) Fitur URL dibatasi pada teks URL dan fitur struktur URL, seperti panjang URL, panjang domain, penggunaan HTTPS, IP address, karakter @, tanda hubung, subdomain, angka, karakter khusus, kata sensitif, *domain entropy*, dan URL *entropy*.

(c) Klasifikasi yang dilakukan bersifat biner, yaitu membedakan antara kelas *legitimate* dan *phishing*.

(d) Evaluasi model dilakukan menggunakan *accuracy*, *precision*, *recall*, *F1-score*, dan *confusion matrix*.

(e) Penelitian ini tidak mencakup implementasi ke dalam sistem *email gateway*, *browser extension*, aplikasi web produksi, maupun sistem keamanan *real-time*.

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut:

(a) Mengembangkan model klasifikasi *email phishing* menggunakan pendekatan *hybrid* mBERT yang menggabungkan representasi teks dan fitur numerik email.

(b) Mengembangkan model klasifikasi *link phishing* menggunakan pendekatan *hybrid* mBERT yang menggabungkan representasi teks URL dan fitur numerik struktur URL.

(c) Mengevaluasi performa model *hybrid* mBERT dalam mendeteksi *email phishing* dan *link phishing* berdasarkan metrik *accuracy*, *precision*, *recall*, *F1-score*, dan *confusion matrix*.

1.5 Manfaat Penelitian

Manfaat yang diharapkan dari penelitian ini adalah sebagai berikut:

(a) Secara teoretis, penelitian ini dapat memberikan kontribusi terhadap pengembangan literatur di bidang keamanan siber, khususnya mengenai penerapan model berbasis *Transformer* dan fitur numerik untuk deteksi *email phishing* dan *link phishing*.

(b) Secara praktis, penelitian ini dapat menjadi dasar pengembangan sistem deteksi yang membantu pengguna atau organisasi dalam mengidentifikasi *email phishing* dan *link phishing* secara otomatis.

(c) Secara metodologis, penelitian ini menyediakan alur pengembangan model deteksi *phishing* berbasis *hybrid* mBERT yang mencakup prapemrosesan data, ekstraksi fitur numerik, *data augmentation* pada *training set*, pencegahan *data leakage*, *early stopping*, dan evaluasi model.

1.6 Sistematika Penulisan

Sistematika penulisan laporan penelitian ini adalah sebagai berikut:

(a) Bab 1 PENDAHULUAN
Berisi latar belakang masalah, rumusan masalah, batasan permasalahan, tujuan penelitian, manfaat penelitian, dan sistematika penulisan.

(b) Bab 2 LANDASAN TEORI

Berisi teori-teori yang mendukung penelitian, seperti *phishing*, *email phishing*, URL *phishing*, pembelajaran mesin, BERT, mBERT, pendekatan *hybrid*, *data augmentation*, *data leakage*, *early stopping*, dan metrik evaluasi.

(c) Bab 3 METODOLOGI PENELITIAN

Berisi tahapan penelitian, pengumpulan data, prapemrosesan data, ekstraksi fitur numerik, pembagian dataset, *data augmentation*, normalisasi fitur, implementasi model *hybrid* mBERT, dan metode evaluasi.

(d) Bab 4 HASIL DAN DISKUSI

Berisi hasil eksperimen, hasil *training*, evaluasi pada *validation set*, evaluasi pada *testing set*, *confusion matrix*, pengujian manual, dan pembahasan hasil penelitian.

(e) Bab 5 SIMPULAN DAN SARAN

Berisi kesimpulan dari hasil penelitian serta saran untuk pengembangan penelitian selanjutnya.