BAB I
PENDAHULUAN

1.1 Latar Belakang

Di era digital 4.0 ini, transformasi teknologi telah mengubah paradigma interaksi manusia secara fundamental. Kehadiran internet tidak lagi sekadar media informasi, melainkan ekosistem hidup yang mengubah cara bekerja, bertransaksi, dan berkomunikasi. Kelimpahan akses digital ini memang menawarkan kemudahan tak terbatas, namun kemudahan ini dapat menciptakan celah keamanan bagi para pengguna internet, di mana keamanan data personal menjadi taruhan utama. Salah satu celah keamanan yang di gunakan oleh para penjahat internet untuk mendapatkan informasi rahasia yaitu melalui *phishing*, mereka bisa mendapatkan informasi pribadi seperti *username,* kata sandi *e-mail*, bahkan data informasi perbankan seperti *mobile banking* dan data kartu kredit dengan memperdaya pengguna menggunakan *link* palsu yang dikirimkan melalui media sosial, pesan instan maupun *e-mail* (Alfawaid & Prianggono, 2023).

Phishing adalah teknik penipuan siber yang melibatkan pembuatan *website* tiruan yang secara seksama meniru situs aslinya. Melalui peniruan tampilan, struktur konten, dan domain URL, pelaku kejahatan merancang jebakan digital yang memancing pengguna internet untuk percaya bahwa mereka sedang mengakses halaman situs asli, padahal sebenarnya mereka telah masuk ke dalam perangkap untuk pencurian data sensitif (Putri & Wijayanto, 2022). Berdasarkan laporan dari *Anti-Phishing Working Group (APWG)* pada kuartal keempat tahun 2024 terdapat sebanyak 989.123 serangan *phishing* yang telah dilakukan oleh para penjahat siber dan menunjukkan peningkatan dari kuartal ketiga yaitu sebanyak 932.923 serangan serta telah terjadi sebanyak 3.763.603 selama tahun 2024. Hal ini menimbulkan banyak kerugian pada pengguna internet, mulai dari kerugian finansial sampai dengan kebocoran data pribadi maupun organisasi (Mahmud & Wirawan, 2024).

Pada Maret 2025, Indonesia mengalami insiden serangan *phishing* yang memanfaatkan celah keamanan dalam protokol *Signalling System* 7 (SS7) yang digunakan oleh operator jaringan telekomunikasi. Para pelaku kejahatan siber berhasil mengeksploitasi kerentanan tersebut menggunakan *fake* BTS (*Base Transceiver Station*) untuk mengirimkan pesan SMS yang tampak berasal dari nomor resmi institusi perbankan, padahal kontennya telah dimodifikasi dan disisipi tautan *phishing* berbahaya. Faktor yang membuat serangan ini sangat efektif adalah penggunaan nomor pengirim yang identik dengan nomor resmi yang biasa digunakan bank untuk mengirimkan kode OTP sehingga para korban tidak menaruh kecurigaan dan cenderung mempercayai keaslian pesan tersebut (Meilina, 2025).

Mengingat semakin meningkatnya serangan *phishing* seiring perkembangan teknologi, berbagai penelitian telah dilakukan untuk mengatasi masalah ini, salah satunya mengembangkan sistem berbasis website untuk mendeteksi *phishing*. Kompleksitas website *phishing* yang terus berkembang memicu para peneliti untuk merancang sistem pendeteksian yang mampu beradaptasi dengan teknik manipulasi digital terbaru. Pendekatan mutakhir dalam mendeteksi *phishing* memanfaatkan kecerdasan buatan, khususnya algoritma *machine learning* dan *deep learning*. Metode ini memungkinkan sistem untuk secara otomatis menganalisis karakteristik website *phishing*, mengenali pola-pola mencurigakan, dan membedakan antara situs asli dengan situs palsu secara akurat (Mahmud & Wirawan, 2024).

Terdapat berbagai algoritma *machine learning* dan *deep learning* yang bisa digunakan untuk mendeteksi website *phishing,* beberapa penelitian sebelumnya telah melakukan dengan beberapa pendekatan, namun lebih condong menggunakan pendekatan berbasis URL dan fitur konten dalam mendeteksi *phishing*. Salah satu penelitian yang telah di lakukan oleh Fatiha at al., (2024) yaitu dengan judul optimasi sistem deteksi *phishing* berbasis web menggunakan algoritma *Decision Tree*. Pada penelitian ini menggunakan salah satu algoritma *machine learning* yaitu *Decision Tree* dimana hasil dari penerapan algoritma ini dalam mendeteksi *website phishing* yaitu dengan akurasi sebesar 95,07% dengan

dataset sebesar 11.055. Penelitian ini menunjukkan bahwa algoritma *machine learning Decision Tree* memberikan akurasi yang sangat tinggi dalam mendeteksi *website phishing*.

Selain itu Egigogo at al., (2024) telah melakukan penelitian dalam mendeteksi *website phishing* dengan melakukan perbandingan penerapan algoritma *deep learning*, adapun algoritma *deep learning* yang digunakan yaitu CNN dan BiGRU. Hasil dari penelitian ini menunjukkan bahwa model *deep learning* yang memiliki Tingkat akurasi paling tinggi dalam mendeteksi *phishing* yaitu model BiGRU dengan akurasi 99,5% yang berarti model ini sangat akurat dalam mendeteksi *website phishing*. Karena penelitian ini ingin melakukan perbandingan, maka peneliti memilih algoritma *Support Vector Machine* (SVM) dan *Long Short-Term Memory* (LSTM) dalam deteksi *website phishing*.

SVM merupakan salah satu algoritma *machine learning* yang sering digunakan dalam melakukan klasifikasi, dimana algoritma ini memiliki beberapa keunggulan seperti performa yang baik ketika digunakan dengan jumlah data yang kecil maupun besar serta pada data yang memiliki atribut yang banyak (Rahman at al., 2021). Sedangkan *Long Short-Term Memory* atau LSTM merupakan salah satu varian *Recurrent Neural Network* atau RNN, algoritma ini mampu mengatasi masalah *vanishing gradient* pada RNN ketika memproses data sekuensial dengan urutan yang panjang dengan memberi *memory cell* untuk menyimpan informasi dalam waktu lama (Septiani, 2020).

Berdasarkan beberapa fenomena yang telah dijelaskan pada paragraf-paragraf di atas, maka penelitian ini berfokus pada perbandingan algoritma *Support Vector Machine* dan *Long Short-Term Memory* sebagai solusi inovatif dalam mendeteksi *website phishing*. Melalui analisis mendalam terhadap kedua algoritma tersebut, penelitian bertujuan menghasilkan metode deteksi yang mampu mengidentifikasi karakteristik *website* berbahaya secara akurat, responsif, dan *real-time*.

1.2 Rumusan Masalah

Berdasarkan fenomena dan masalah yang telah diuraikan pada latar belakang sebelumnya, rumusan masalah pada penelitian ini adalah "Bagaimana perbandingan performa antara algoritma *Support Vector Machine* (SVM) dan *Long Short-Term Memory* (LSTM) dalam mendeteksi *website phishing*?"

1.3 Batasan Masalah

Agar penelitian lebih terarah dan sesuai dengan tujuan yang ingin dicapai, perlu ditetapkan batasan masalah. Batasan ini bertujuan untuk mempersempit bidang dan cakupan penelitian sehingga fokus penelitian dapat tercapai. Berikut merupakan batasan-batasan masalah pada penelitian ini:

1. Penelitian hanya menggunakan dua algoritma untuk melakukan deteksi *website phishing*, yaitu *Support Vector Machine* (SVM) dan *Long Short-Term Memory* (LSTM).
2. Data yang dianalisis terbatas pada karakteristik URL *website phishing* dengan data yang dikumpulkan tidak melebihi periode 5 tahun terakhir.
3. Penelitian ini akan menggunakan metrik evaluasi yaitu *confusion matrix* untuk membandingkan kinerja kedua algoritma, yaitu berupa akurasi, presisi, *recall*, dan *F1-score*.

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah "Membandingkan performa algoritma *Support Vector Machine* (SVM) dan *Long Short-Term Memory* (LSTM) dalam mendeteksi *website phishing*, untuk memperoleh algoritma yang lebih efektif berdasarkan perbandingan performa kedua algoritma tersebut."

1.5 Manfaat Penelitian

Manfaat penelitian ini memberikan nilai penting baik bagi peneliti maupun pembaca, mencakup aspek teoritis dan praktis. Berikut adalah penjelasan mengenai manfaat yang dapat dihasilkan dari penelitian ini:

1. Manfaat secara Teoritis
    a. Memberikan kontribusi pada pengembangan ilmu pengetahuan dalam bidang *machine learning* dan *deep learning*, khususnya terkait perbandingan efektivitas algoritma SVM dan LSTM dalam deteksi *website phishing*.
    b. Memperkaya literatur akademik Universitas Maritim Raja Ali Haji tentang penerapan teknik-teknik kecerdasan buatan dalam keamanan siber, terutama dalam konteks deteksi *phishing*.
    c. Menyediakan kerangka analitis untuk membandingkan algoritma SVM dan LSTM yang dapat digunakan sebagai referensi untuk penelitian-penelitian sejenis di masa depan.
    d. Mengembangkan pemahaman teoretis tentang karakteristik dan pola *website phishing* yang dapat dideteksi melalui pendekatan pembelajaran mesin.
2. Manfaat secara Praktis
    a. Membantu organisasi dan praktisi keamanan siber dalam memilih algoritma yang paling sesuai untuk implementasi sistem deteksi *website phishing* berdasarkan hasil perbandingan yang komprehensif.
    b. Memberikan panduan praktis tentang kelebihan dan keterbatasan masing-masing algoritma (SVM dan LSTM) dalam konteks deteksi *phishing,* sehingga memudahkan proses pengambilan keputusan dalam pengembangan sistem keamanan.
    c. Mendukung upaya edukasi kepada masyarakat mengenai pentingnya deteksi *phishing,* dengan memanfaatkan hasil penelitian untuk mengembangkan alat pendukung atau sistem keamanan berbasis AI yang dapat digunakan secara luas.

1.6 Sistematika Penulisan

Sistematika penulisan tugas akhir ini disusun dalam beberapa tahapan yang terbagi ke dalam sejumlah bab, dengan rincian sebagai berikut:

BAB I PENDAHULUAN

Pada bab ini membahas mengenai latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan.

BAB II KAJIAN LITERATUR

Pada bab ini, menguraikan penelitian-penelitian terdahulu yang berkaitan dengan judul skripsi yang diangkat, serta teori-teori yang digunakan dalam topik penelitian dengan metode yang sama.

BAB III METODE PENELITIAN

Bab ini membahas tentang fokus dan waktu penelitian, bahan atau materi yang diteliti, jenis data yang digunakan, metode pengumpulan data, instrumen penelitian, kerangka penelitian, proses pengumpulan data, serta tahapan analisis dan perancangan aplikasi.

BAB IV HASIL DAN PEMBAHASAN

Pada bab ini, penulis menyajikan hasil pengujian serta pembahasan terkait judul penelitian yang telah dilakukan, dengan penjelasan lengkap mengenai alasan di balik hasil yang diperoleh.

BAB V PENUTUP

Pada bab ini membahas mengenai kesimpulan dan saran dari hasil penelitian**.**

DAFTAR PUSTAKA

Daftar pustaka berisi sumber-sumber yang digunakan peneliti dalam melakukan penulisan.