BAB I
PENDAHULUAN

1.1 Latar Belakang Penelitian

Era digital yang semakin berkembang memberikan dampak positif karena membantu aktivitas masyarakat menjadi lebih mudah tanpa terbatas ruang dan waktu (Ghifari dkk., 2022). Seiring dengan perkembangan teknologi, celah keamanan terhadap serangan siber juga perlu diperhatikan terutama dalam hal keamanan data. Salah satu bentuk serangan siber yang paling sering terjadi ialah serangan *phishing*. *Phishing* merupakan tindakan kejahatan di dunia maya yang melibatkan usaha untuk mencuri identitas pengguna melalui penggunaan *email* atau situs *web* palsu dengan menyamar sebagai lembaga sah yang dapat dipercaya (Bahnsen dkk., 2017). Serangan ini menjadi semakin serius karena semakin banyak orang yang bergantung pada internet untuk bertransaksi maupun berkomunikasi. Pernyataan ini didukung berdasarkan data yang diperoleh dari laporan (Badan Siber dan Sandi Negara (BSSN), 2023) yang menyatakan bahwa situs *phishing* berada di peringkat kedua dari 10 trafik anomali serangan siber terbesar di Indonesia periode Januari - Desember 2023 dengan jumlah trafik sebanyak 47.231.390. Laporan dari (Anti-Phishing Working Group (APWG), 2023) juga menunjukkan bahwa kuartal kedua tahun 2023 merupakan jumlah triwulanan serangan situs *phishing* tertinggi ketiga dari yang pernah diamati yaitu sebanyak 1.286.208 serangan. Maka dari itu, penulis akan menggali lebih dalam terkait *phishing* dalam bentuk situs *web*.

Website yang merupakan pintu utama di era digital ini untuk mengakses internet telah menjadi bagian integral dari kehidupan masyarakat karena sangat dibutuhkan untuk setiap instansi baik pemerintahan maupun swasta (Perbawa & Nurohim, 2020). Sangat disayangkan dengan adanya kemunculan fenomena pencurian data dapat menjadi masalah besar yang dapat mempengaruhi individu, bisnis, dan aktivitas internet lainnya. Hal ini disebabkan oleh adanya situs *web* yang tidak aman, yang dikenal dengan istilah situs atau *website phishing*. Situs *web phishing* beroperasi dengan membuat situs palsu yang menyerupai aslinya, dengan tujuan menipu individu agar mereka memberikan informasi pribadi yang bersifat penting. Teknik rekayasa sosial ini mencakup pembuatan situs *phishing* yang meniru tampilan situs perbankan *online*, hiburan, pendidikan, atau bahkan situs

media sosial untuk mendapatkan data yang sangat rahasia dari korban dengan mencuri akun korban (Ghifari dkk., 2022) (Nugraha dkk., 2022). Selain mencuri data, situs *phising* juga bisa dijadikan sarana untuk menyebarkan virus atau *malware* dengan menyamar sebagai situs *web* yang sah.

Merebaknya situs *phishing* menjadi fenomena yang semakin meresahkan karena melibatkan berbagai metode dan sarana untuk mengekspansi jaringannya. Salah satu faktor utama penyebaran situs *phishing* adalah kecanggihan teknologi dan keterampilan pelaku kejahatan *cyber* dalam menciptakan tautan dan halaman palsu yang sangat meyakinkan. Dengan kemampuan ini, penyerang dapat membuat situs *phishing* yang sulit untuk dibedakan dari situs web asli, terutama dalam meniru tata letak, desain, dan bahkan alamat URL. Penyebaran situs *phishing* juga semakin banyak dilancarkan oleh penyerang, baik melalui *email*, media sosial, *sms*, iklan juga metode lainnya. Dalam skenario *email phishing*, pengguna menerima *email* palsu yang berpura-pura berasal dari lembaga sah yang dikenal dan terdapat tautan yang mengarahkan mereka ke situs *phishing*. Melalui media sosial, tautan *phishing* dapat menyebar melalui akun palsu atau dengan memanfaatkan akun yang sudah terpengaruh. Pesan instan dan iklan *online* pun dapat dimanipulasi untuk menyematkan tautan *phishing* dan bahkan situs *phishing* dapat diciptakan dengan memanfaatkan pembajakan nama domain atau situs *web* yang sudah ada. Dalam beberapa kasus, penyebaran situs *web phishing* juga melibatkan praktik penyebaran *malware* sehingga meningkatkan resiko serangan yang lebih luas dan serius terhadap pengguna yang tidak curiga.

Oleh karena itu, pemahaman dan kewaspadaan masyarakat terhadap potensi ancaman ini sangat penting. Salah satu bentuk cara untuk meningkatkan kewaspadaan agar dapat menghindari serangan situs *phishing* yang dapat membahayakan keamanan data ialah suatu sistem atau aplikasi yang dapat mendeteksi situs *phising* sehingga dapat meminimalisir kerugian yang diakibatkan dari serangan tersebut. Aplikasi yang dikembangkan dapat menggunakan algoritma klasifikasi seperti *machine learning* atau algoritma *deep learning*. Algoritma *machine learning* merupakan algoritma yang dapat meningkatkan kecerdasan sistem melalui pembelajaran dari data yang tersedia, tanpa definisi eksplisit pada algoritma atau program tersebut (Windarni dkk., 2023). Sedangkan algoritma *deep*

learning merupakan penelitian jaringan saraf yang bertujuan untuk mengidentifikasi informasi yang tersembunyi dalam data yang kompleks dengan menggunakan pembelajaran bertahap (Yang dkk., 2019). Penelitian deteksi situs *web phishing* menggunakan model *machine learning* masih memiliki potensi untuk dimaksimalkan (Nugraha dkk., 2022), sehingga untuk meningkatkan juga memaksimalkan kinerja dari deteksi situs *phishing* digunakan pendekatan algoritma *deep learning* dan metode yang diterapkan dalam penelitian ini adalah algoritma *Long Short-Term Memory* (LSTM). Pada penelitian sebelumnya, LSTM dapat menyimpan informasi dalam jangka waktu yang panjang dan memiliki kinerja yang baik dalam menganalisis data sekuensial seperti URL yang memiliki struktur dimulai dari protokol, domain, direktori, *filename* dan parameter sehingga karakternya saling berkaitan dan informasi yang diolah berkesinambungan (Somesha dkk., 2020) (Do dkk., 2021).

Pada penelitian sebelumnya, (Bahnsen dkk., 2017) menggunakan algoritma LSTM untuk mengklasifikasikan situs *phishing*, dan hasilnya menunjukkan bahwa algoritma LSTM dapat menghasilkan nilai akurasi yang tinggi yaitu 98.76% dalam mendeteksi URL *phishing* tanpa adanya ekstraksi fitur manual dari struktur URL dan dapat mengungguli algoritma *Random Forest* yang hanya mendapat nilai akurasi sebesar 93.47% dengan ekstraksi fitur URL-nya. Dataset yang digunakan dalam penelitian tersebut berjumlah 2.000.000 data dengan pembagian data sebanyak 1.000.000 untuk situs *phishing* dan 1.000.000 data situs sah. Penelitian lainnya dilakukan oleh (Somesha dkk., 2020) yang melakukan perbandingan antara tiga algoritma menggunakan *dataset* yang sama yaitu LSTM, *Deep Neural Network* (DNN) dan *Convolutional Neural Network* (CNN). Pada penelitian tersebut algoritma LSTM unggul dalam pemrosesan sekuensial untuk memahami konteks fitur semantik dalam susunan data URL dan memahami ketergantungan karakter dalam urutan URL pada situs *phishing* sehingga menghasilkan nilai akurasi yang tinggi yaitu 99.57%. Akurasi tersebut lebih tinggi dibandingkan algoritma DNN yang mendapat nilai akurasi sebesar 99.52% dan CNN sebesar 99.43%. Dataset yang digunakan dalam penelitian tersebut berjumlah 3526 URL dengan pembagian data sebanyak 2119 situs *phishing* yang dikumpulkan dari situs *PhishTank* dan 1407 merupakan situs sah yang dikumpulkan dari basis data *Alexa*. Penelitian komparatif

lainnya dilakukan oleh (Almousa dkk., 2022) dengan membandingkan tiga algoritma *deep learning* yaitu LSTM, DNN, dan CNN menggunakan *Tan dataset* yang terdiri dari 5000 situs *phishing* dan 5000 situs sah. Algoritma dengan nilai akurasi yang lebih tinggi pada penelitian tersebut adalah LSTM yaitu 97.37%, sedangkan nilai akurasi algoritma CNN dan DNN secara berurutan yaitu 97.27% dan 96.77%.

Berdasarkan penelitian sebelumnya, belum ada yang berfokus pada analisis interpretabilitas model LSTM dalam memahami pola prediksi yang dibuat oleh model karena penelitian sebelumnya merupakan penelitian komparatif. Evaluasi performa model LSTM dalam mendeteksi URL *phishing* juga belum diterapkan secara langsung terhadap sistem aplikasi sehingga belum ada pengujian data baru secara langsung. Maka dari itu, penelitian ini dilakukan untuk mengimplementasikan algoritma LSTM pada sistem deteksi situs *phishing* untuk memahami bagaimana model dalam membuat keputusan dalam mengklasifikasikan URL sebagai *phishing* atau sah. Untuk memahami cara kerja algoritma LSTM, penelitian ini dilakukan untuk menganalisis bagaimana algoritma LSTM menangkap pola-pola penting dalam data dengan mengingat informasi sepanjang waktu dan mengidentifikasi pola yang tidak terlihat jika hanya dalam satu waktu. Penelitian ini juga dilakukan untuk memastikan keandalan dan kinerja model saat diimplementasikan pada sistem deteksi situs *phishing* yang dihasilkan menjadi suatu aplikasi *web* untuk mendeteksi terjadinya serangan *phishing*. Dengan pengaplikasian tersebut, penulis dapat mengetahui sejauh mana model LSTM dapat mengenali pola-pola *phishing* yang lebih kompleks dan bervariasi pada situs-situs yang tersedia di internet. Berdasarkan pemaparan diatas, penulis melakukan penelitian untuk membuat sistem deteksi situs *phishing* berbasis aplikasi *web* dengan judul penelitian "Implementasi Algoritma *Long Short-Term Memory* pada Sistem Deteksi Situs *Phishing* berbasis Aplikasi *Web*".

1.2 Rumusan Masalah Penelitian

Berdasarkan latar belakang yang telah diuraikan, maka rumusan masalah yang diidentifikasi diantaranya:

1. Bagaimana proses klasifikasi algoritma LSTM sehingga dapat mendeteksi serangan situs *phishing*?

2. Bagaimana cara membangun aplikasi *web* sebagai sistem deteksi situs *phishing* yang berfungsi dengan optimal?
3. Bagaimana penerapan algoritma *Long Short-Term Memory* pada sistem deteksi situs *phishing* berbasis aplikasi *web*?

1.3 Batasan Masalah

Adapun beberapa batasan masalah dalam penelitian ini ialah sebagai berikut:

1. Deteksi situs *phishing* berfokus pada implementasi algoritma LSTM sebagai model untuk mendeteksi URL berdasarkan strukturnya.
2. Penelitian berfokus pada pengembangan sistem deteksi situs *phishing* berbasis aplikasi *web* untuk mengidentifikasi situs yang terindikasi *phishing* atau sah.
3. Aplikasi *web* diintegrasikan dengan model yang dilatih berdasarkan data pelatihan dari *dataset* yang digunakan dalam penelitian.

1.4 Tujuan Penelitian

Berdasarkan rumusan masalah yang telah disajikan, maka terdapat tujuan penelitian sebagai sebagai berikut:

1. Mengidentifikasi dan mengevaluasi proses klasifikasi algoritma LSTM dalam mendeteksi serangan situs *phishing* berdasarkan struktur URL.
2. Merancang dan mengembangkan aplikasi *web* sebagai sistem deteksi situs *phishing* menggunakan bahasa pemrograman *Python*.
3. Menerapkan model LSTM pada sistem deteksi situs *phishing* berbasis aplikasi *web* menggunakan data yang relevan.

1.5 Manfaat Penelitian

Penelitian ini diharapkan dapat memberi manfaat baik secara teoritis maupun dalam segi praktisnya. Adapun manfaatnya dijabarkan sebagai berikut:

1.5.1 Manfaat Teoritis

Hasil dari penelitian ini diharapkan dapat memberikan wawasan terkait kajian ilmu pengembangan keamanan jaringan dalam penerapan ilmu komputer yaitu pengembangan aplikasi *web* untuk mendeteksi situs *phishing* menggunakan algoritma LSTM. Penelitian ini juga memberikan pemahaman yang lebih dalam terkait model LSTM ketika membuat keputusan untuk mengklasifikasikan situs

URL sebagai *phishing* atau sah, dan mengetahui bagaimana model menangkap pola-pola penting dalam data. Hal ini dapat membantu pengguna untuk memahami model dan menginterpretasikan hasil prediksi dengan lebih baik untuk memperbaiki kinerja model.

1.5.2 Manfaat Praktis

1. Penelitian ini diharapkan dapat membantu meningkatkan keamanan *web* perusahaan dari pemalsuan nama situs *web* resmi perusahaan.
2. Penelitian ini dapat digunakan untuk mencegah terjadinya serangan situs *phishing* ketika menggunakan internet sehingga dapat meningkatkan perlindungan data pribadi.
3. Penelitian ini dapat menjadi referensi bagi peneliti lainnya yang akan melakukan penelitian tentang penerapan algoritma LSTM dalam mendeteksi situs *phishing*.

1.6 Struktur Organisasi Skripsi

Dalam penulisan sistematika pada penyusunan laporan penelitian, terdapat urutan rangkaian yang membahas setiap bab tersebut diantaranya:

1. BAB I PENDAHULUAN

   Memberi penjelasan secara sistematis terkait latar belakang masalah, perumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah dan struktur organisasi penulisan skripsi.
2. BAB II KAJIAN PUSTAKA

   Menguraikan hasil kajian dari teori penelitian yaitu mengenai serangan situs *phishing*, algoritma LSTM, juga aplikasi deteksi berbasis *web*. Terdapat juga hasil-hasil penelitian yang relevan guna membantu dalam menyelesaikan permasalahan penelitian.
3. BAB III METODE PENELITIAN

   Menjelaskan tahapan-tahapan penelitian, objek penelitian dan metode penelitian yang akan digunakan, serta cara pengambilan dan cara mengolah data.

4. BAB IV TEMUAN DAN PEMBAHASAN

   Membahas hasil dari penelitian secara detail dan implementasi aplikasi *web* deteksi situs *phishing* yang telah dibuat serta pengujian dan evaluasi terhadap kesesuaian penggunaan sistem.

5. BAB V SIMPULAN, IMPLIKASI DAN REKOMENDASI

   Menjelaskan kesimpulan untuk menjawab perumusan masalah secara terperinci dan rekomendasi perbaikan kedepannya.