BAB I
PENDAHULUAN

1.1 Latar Belakang

Banyak kejahatan digital yang berlangsung di era modern dan satu di antaranya adalah Phishing, yang merupakan pemotongan kata untuk menyatakan mendapatkan data pribadi target. Para penjahat biasanya merancang sebuah scam atau penipuan yang membuat target tertipu untuk mengklik suatu situs, email, atau URL yang menyamar sebagai pihak resmi. [1].

Penjahat digital sering menggunakan teknik manipulasi dalam phishing berbasis URL untuk merekam sugestif URL pengguna agar memberikan kata sandi, kartu kredit atau informasi keuangan lainnya. URL phishing sering kali berpura-pura menjadi pesan dari lembaga keuangan tepercaya, toko daring, atau platform daring lainnya. Serangan phishing semacam itu dapat menyebabkan target kehilangan uang atau bahkan membahayakan data pribadi mereka.

Tingkat keberhasilan serangan *phishing* berbasis *URL* sangat tergantung pada seberapa baik pelaku mampu membuat *URL* yang terlihat meyakinkan dan meyakinkan target untuk melakukan tindakan tertentu seperti mengirimkan informasi pribadi. Laporan Ancaman Keamanan Internet Symantec tahun 2019 menyatakan bahwa 5,9% dari semua URL berbahaya pada tahun 2018 merupakan URL phishing, dan 170 URL merupakan URL phishing setiap tahunnya. Kelompok Kerja Anti Phishing (APWG) mengklasifikasikan 46% situs web sebagai phishing antara Oktober 2018 dan Maret 2019[2].

Oleh karena itu, pentingnya pengembangan metode deteksi dan pencegahan yang dapat mengenali pola-pola dan ciri-ciri khas dari *URL phishing* menjadi semakin krusial. Dengan memahami secara mendalam bagaimana *URL phishing* dibuat dan disebarkan, serta mengidentifikasi karakteristik yang dapat dibedakan antara *URL phishing* dan *URL* yang resmi, diharapkan dapat membantu dalam membangun sistem yang dapat melindungi pengguna internet dari serangan *phishing* yang merugikan.

Dalam rangka meningkatkan akurasi dan efisiensi deteksi *URL phishing*, berbagai metode telah dikembangkan. Salah satu Penelitian terdahulu yang dilaksanakan oleh Anupam dan Kar (2020) berfokus pada deteksi situs phishing menggunakan Support Mesin vektor pendukung (SVM) dioptimalkan menggunakan algoritma pengoptimalan yang terinspirasi dari alam. Support Vector Machine (SVM) adalah algoritma pembelajaran yang didasarkan pada konsep pembentukan hiperbidang optimal untuk memisahkan kelas data. Dengan menggunakan konsep ini, SVM dapat mengklasifikasikan data ke dalam berbagai kategori dengan akurasi tinggi[3]. Dalam studi ini, dataset yang dipakai terdiri dari 1353 sampel yang dibagi menjadi dua kelas, yaitu situs phishing dan situs sah. Penelitian ini menunjukkan bahwa algoritma optimasi seperti Firefly Algorithm (FA) dan Grey Wolf Optimizer (GWO) secara signifikan meningkatkan akurasi dan skor F1 dari model SVM dibandingkan dengan metode optimasi tradisional. Hasilnya menunjukkan bahwa model SVM yang dioptimalkan dengan algoritma terinspirasi dari alam lebih unggul dalam mendeteksi situs phishing, dengan implikasi praktis untuk aplikasi dalam perangkat lunak antivirus dan ekstensi browser. Penelitian ini juga menekankan pentingnya eksplorasi lebih lanjut terhadap aplikasi. Contoh lain algoritma optimasi dalam bidang keamanan siber yang terbukti efektif adalah *Convolutional Neural Networks (CNN)* dan *Decission Tree. Convolutional Neural Networks (CNN)* merupakan salah satu jenis deep learning yang telah menunjukkan keunggulan dalam mengambilsi fitur dari data berstruktur kompleks, termasuk teks. *CNN* mempunyai kemampuan untuk menangkap pola-pola hierarkis pada data teks, menjadikannya pilihan yang tepat untuk analisis yang mendalam.[1]. Di sisi lain, Algoritma *Decission Tree* mudah dipahami dan juga mudah diimplementasikan. *Decission Tree* memulai pekerjaannya dengan memilih pemisah terbaik dari atribut yang tersedia untuk klasifikasi yang dianggap sebagai akar pohon.[4]

Penggunaan metode *SVM*, *CNN*, dan *Decission Tree* secara bersamaan dalam suatu model *hybrid* menjanjikan potensi yang besar dalam deteksi *URL phishing*. *SVM* dapat dipakai untuk menganalisis teks dalam *URL* dan mengenali pola-pola yang mencurigakan, seperti penggunaan kata-kata tertentu atau tautan yang meragukan. *CNN* dapat dipakai untuk mengambilsi fitur tekstual dari isi *URL*, menangkap pola-pola yang relevan dalam struktur teks. Sementara itu, *Decission Tree* menawarkan kemampuan

untuk melakukan klasifikasi berbasis aturan yang mudah diinterpretasikan, yang dapat membantu mengidentifikasi faktor-faktor kunci dalam menentukan apakah sebuah *URL* termasuk *phishing* atau *non-phishing*. Dengan menggabungkan kekuatan ketiga metode ini, kita dapat menciptakan sistem deteksi yang lebih tangguh dan efisien, mampu menghadapi berbagai jenis serangan *phishing* yang semakin kompleks dan terus berkembang [3].

Dengan mengacu pada penelitian-penelitian terdahulu ini, penelitian ini bertujuan untuk mengembangkan model *hybrid* yang menggabungkan *CNN* dengan *SVM* serta *CNN* dengan *Decission Tree* untuk mendeteksi dan mencegah serangan *phishing* berbasis *URL*. Kombinasi kedua model *hybrid* ini diharapkan dapat meningkatkan akurasi dan efisiensi dalam klasifikasi *URL phishing*, sekaligus memberikan perbandingan yang jelas untuk menentukan metode terbaik.

1.2 Rumusan Masalah

Berdasarkan latar belakang penelitian yang disebutkan di atas, pertanyaan-pertanyaan berikut dapat dirumuskan:

Bagaimana kinerja model hibrida CNN-SVM dan CNN-Decision Tree dalam mendeteksi URL phishing?

1. Sejauh mana kedua model, yaitu *CNN-SVM* dan *CNN-Decission Tree*, dapat diimplementasikan untuk deteksi *URL phishing* berdasarkan dataset yang tersedia?

1.3 Identifikasi Masalah

Berdasarkan latar belakang penelitian yang telah di sebutkan sebelumnya, maka dapat dibuat identifikasi masalah sebagai berikut :

1. Penelitian ini bertujuan untuk mengidentifikasi metode klasifikasi yang lebih efektif antara *CNN-SVM* dan *CNN-Decission Tree* dalam mendeteksi *URL phishing*. *CNN-SVM* dipilih karena performanya dalam klasifikasi teks, sedangkan *CNN-Decission Tree* menawarkan pendekatan yang lebih interpretatif.

1.4 Tujuan Penelitian

Sesuai dengan identifikasi masalah yang telah dibuat sebelumnya, maka dapat ditentukan tujuan penelitian ini adalah :

1. Membangun *machine learning* yang dapat membantu orang awam dalam memilah *URL phishing* dan *non-phishing* secara tepat.
2. Mengembangkan dan mengevaluasi dua model hybrid, yaitu *CNN-SVM* dan *CNN-Decission Tree*, untuk menentukan model yang paling akurat dalam mendeteksi *URL phishing.*

1.5 Batasan Masalah

Sangat penting untuk menetapkan batasan masalah untuk penelitian agar penelitian lebih terarah dan lebih mudah untuk diselesaikan. Batasan masalah penelitian ini adalah sebagai berikut:

1. Penelitian ini hanya mencakup *URL phishing* sebagai data analisis dan tidak mencakup serangan *phishing* di media lain seperti media sosial. Perbandingan model terbatas pada *CNN-SVM* dan *CNN-Decission Tree*.
2. Sumber data yang dipakai adalah data opensource yang tersedia pada platform *Kaggle*.com