BAB I
PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi internet telah membawa banyak manfaat bagi kehidupan manusia, namun di sisi lain juga menghadirkan berbagai ancaman keamanan. Salah satu ancaman paling signifikan adalah *phishing*, yaitu bentuk kejahatan siber yang bertujuan untuk menipu korban agar memberikan informasi pribadi atau kredensial penting melalui rekayasa sosial dan manipulasi teknis. Phishing umumnya dilakukan dengan menyebarkan tautan berbahaya (phishing link) melalui email, pesan singkat, atau media sosial, yang kemudian mengarahkan korban ke situs web palsu yang menyerupai situs resmi [1].

Menurut Anti-Phishing Working Group (APWG), pada kuartal IV tahun 2024 bulan Oktober hingga Desember, tercatat lebih dari 989.123 serangan phishing yang berhasil dideteksi. Angka ini meningkat dibandingkan triwulan sebelumnya yang berada pada angka 932.923 serangan. Selain itu, APWG juga mencatat peningkatan tajam pada serangan phishing berbasis SMS (*smishing*), terutama dengan memanfaatkan domain berbiaya rendah seperti *.TOP*, *.CYOU*, dan *.XIN*, yang banyak disalahgunakan oleh pelaku dari Tiongkok untuk mengelabui target di Amerika dan negara lainnya [2].

Teknik phishing telah berkembang secara signifikan, menggabungkan taktik canggih seperti spear-phishing, polymorphic phishing, dan konten yang dihasilkan AI, yang meningkatkan kesulitan deteksi dan mitigasi [3]. Untuk melindungi dari serangan siber, penting untuk mendeteksinya secara akurat dan cepat [4], [5]. Namun, mekanisme keamanan tradisional, termasuk enkripsi dan firewall, semakin tidak memadai terhadap sifat serangan phishing yang dinamis dan canggih, yang semakin menekankan perlunya model deteksi yang lebih canggih dan akurat [6], [7].

URL merupakan dasar dari sebagian besar serangan phishing dan berperan penting dalam pelaksanaannya [8]. Oleh karena itu, memahami dan menganalisis

URL phishing sangat penting bagi upaya keamanan siber. Penyerang sering kali menyebarkan URL ini melalui platform pengiriman pesan, email, dan media sosial untuk menipu korban [9].

Dalam beberapa tahun terakhir, teknologi machine learning (ML) telah menunjukkan potensi besar dalam memproses dan menganalisis data kompleks, termasuk dalam konteks keamanan siber [10]. Algoritma machine learning memungkinkan sistem belajar dari dataset historis untuk mengenali pola atau anomali yang mengindikasikan aktivitas berbahaya, termasuk phishing. Salah satu algoritma yang populer digunakan dalam klasifikasi data adalah Support Vector Machine (SVM) [11].

SVM bekerja dengan membuat hyperplane untuk memisahkan titik data ke kelas yang berbeda. Untuk melakukan klasifikasi, SVM membutuhkan kernel Radial Basis Functions (RBF) [12]. Namun, kinerja model SVM dengan kernel RBF sangat bergantung pada pemilihan hyperparameter, seperti nilai C, dan gamma. Pemilihan parameter yang kurang tepat dapat menyebabkan model menghasilkan akurasi rendah atau overfitting [13]. Oleh karena itu, penting dilakukan optimasi hyperparameter agar performa model SVM dapat dimaksimalkan. Salah satu teknik yang umum digunakan untuk tujuan ini adalah Grid Search, di mana semua kombinasi parameter yang memungkinkan dicoba secara sistematis. Meskipun sederhana, Grid Search mampu menghasilkan kombinasi parameter terbaik bila dikombinasikan dengan evaluasi seperti k-fold cross-validation [14].

Berdasarkan permasalahan tersebut, penelitian ini bertujuan untuk mengoptimalkan kinerja algoritma Support Vector Machine (SVM) dalam mendeteksi tautan phishing dengan menggunakan metode Grid Search sebagai teknik pemilihan hyperparameter. Dataset phishing yang digunakan berisi URL yang diklasifikasikan berdasarkan ciri-ciri yang umum ditemukan dalam link phishing, seperti panjang URL, keberadaan simbol mencurigakan, domain, serta struktur jalur. Dengan menerapkan pendekatan ini, diharapkan model yang

dihasilkan akan mampu bekerja secara lebih akurat dan efisien dalam mendeteksi phishing secara otomatis.

Hasil dari penelitian ini diharapkan dapat memberikan kontribusi nyata bagi pengembangan sistem keamanan berbasis kecerdasan buatan, khususnya dalam bidang deteksi dini. Dengan demikian, sistem deteksi phishing dapat diperluas pemanfaatannya untuk membantu perlindungan terhadap individu, organisasi, maupun infrastruktur digital secara keseluruhan.

1.2 Rumusan Masalah

Berdasarkan latar belakang masalah di atas, rumusan masalah dalam penelitian ini adalah sebagai berikut:.

1. Bagaimana kinerja algoritma SVM dalam mendeteksi tautan phishing?
2. Bagaimana pengaruh optimasi hyperparameter menggunakan Grid Search terhadap kinerja SVM dalam mendeteksi URL phishing?

1.3 Batasan Masalah

Penelitian ini memiliki beberapa batasan yang perlu diperhatikan agar fokus penelitian lebih jelas dan terarah. Batasan-batasan tersebut adalah sebagai berikut:

1. Penelitian ini hanya menggunakan dataset Phishing yang tersedia secara publik di Kaggle dan tidak melakukan pengumpulan data primer. Dataset ini berisi variable data yang berkaitan dengan phishing dan digunakan sebagaimana adanya tanpa modifikasi tambahan.
2. Penelitian ini hanya menggunakan algoritma Support Vector Machine (SVM) sebagai model utama dalam deteksi tautan phishing. Algoritma lain tidak dibandingkan atau diuji dalam penelitian ini.
3. Penelitian ini hanya menggunakan Grid Search sebagai metode optimasi hyperparameter untuk meningkatkan kinerja model Support Vector Machine (SVM). Metode optimasi lainnya tidak diterapkan atau dibandingkan dalam penelitian ini.
4. Evaluasi performa model dilakukan menggunakan metrik akurasi, presisi, recall, dan f1-score yang diperoleh dari confusion matrix. Metrik lain seperti

AUC-ROC tidak dianalisis dalam penelitian ini.

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah untuk menyelesaikan permasalahan yang telah diidentifikasi dan mencapai hasil yang dapat diukur ketercapaiannya. Adapun tujuan spesifik yang ingin dicapai dalam penelitian ini adalah sebagai berikut:

1. Mengembangkan model Support Vector Machine (SVM) yang mampu mendeteksi tautan phishing dengan akurasi tinggi berdasarkan data URL.
2. Menganalisis pengaruh optimasi hyperparameter menggunakan metode Grid Search terhadap performa model Support Vector Machine (SVM) dalam mendeteksi tautan phishing.
3. Mengevaluasi peningkatan performa model setelah optimasi hyperparameter, berdasarkan metrik evaluasi seperti akurasi, presisi, recall, dan f1-score.

1.5 Manfaat Penelitian

Penelitian ini diharapkan memberikan beberapa manfaat, baik secara teoritis maupun praktis.

1. Manfaat Teoritis
   a. Menambah literatur akademik di bidang kecerdasan buatan dan keamanan siber, khususnya pada penerapan algoritma Support Vector Machine (SVM) untuk klasifikasi tautan phishing.
   b. Memberikan pemahaman lebih mendalam mengenai pengaruh hyperparameter terhadap kinerja algoritma SVM dalam konteks klasifikasi URL, serta menunjukkan bagaimana metode optimasi Grid Search dapat digunakan sebagai pendekatan sistematis dalam meningkatkan performa model.
   c. Menjadi acuan untuk penelitian selanjutnya yang berkaitan dengan penerapan machine learning dalam deteksi ancaman siber seperti spam, malware, dan social engineering attack lainnya, terutama yang melibatkan klasifikasi teks berbasis URL.

2. Manfaat Praktis
   a. Menjadi acuan untuk penelitian selanjutnya yang berkaitan dengan penerapan machine learning dalam deteksi ancaman siber seperti spam, malware, dan social engineering attack lainnya, terutama yang melibatkan klasifikasi teks berbasis URL.
   b. Menjadi acuan untuk penelitian selanjutnya yang berkaitan dengan penerapan machine learning dalam deteksi ancaman siber seperti spam, malware, dan social engineering attack lainnya, terutama yang melibatkan klasifikasi teks berbasis URL.
   c. Menjadi acuan untuk penelitian selanjutnya yang berkaitan dengan penerapan machine learning dalam deteksi ancaman siber seperti spam, malware, dan social engineering attack lainnya, terutama yang melibatkan klasifikasi teks berbasis URL.

1.6 Sistematika Penulisan

Berikut adalah sistematika penulisan pada penelitian ini:

BAB I PENDAHULUAN

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

BAB II TINJAUAN PUSTAKA

Bab ini berisi pembahasan dasar teori yang berkaitan dengan penelitian serta penelitian-penelitian sebelumnya yang relevan.

BAB III METODE PENELITIAN

Bab ini menguraikan tahapan yang akan dilaksanakan dalam penelitian. Setiap tahap perencanaan penelitian akan dijabarkan secara terperinci berdasarkan suatu kerangka kerja yang telah ditetapkan. Selain itu, bab ini juga mencakup perancangan manajemen proyek untuk pelaksanaan penelitian, yang bertujuan untuk kelancaran dan keberhasilan seluruh proses.

BAB IV HASIL DAN PEMBAHASAN

Bab ini berisi hasil dari penerapan model yang dijelaskan pada bab sebelumnya. Hasil optimasi metode Grid Search pada algoritma Support Vector Machine (SVM) serta metrik evaluasi model akan dibahas secara rinci. Bab ini juga membandingkan hasil sebelum dan sesudah dilakukan optimasi.

BAB V PENUTUP

Bab ini berisi kesimpulan yang merangkum hasil dari penelitian ini, serta saran-saran yang dapat diberikan untuk penelitian lanjutan. Kesimpulan diambil berdasarkan hasil dan pembahasan yang telah dilakukan, dan saran diberikan terkait pengembangan lebih lanjut dari metode atau pendekatan yang digunakan.