BAB I
PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi yang sangat pesat telah membawa perubahan signifikan dalam berbagai aspek kehidupan, khususnya dalam pertukaran informasi digital melalui internet. Salah satu media komunikasi yang masih dominan digunakan hingga saat ini adalah surat elektronik (*email*), baik untuk kepentingan personal, bisnis, maupun institusional. Namun, tingginya ketergantungan terhadap *email* juga membuka peluang terjadinya kejahatan siber, terutama dalam bentuk penipuan berbasis manipulasi sosial atau yang dikenal sebagai *phishing* [1].

Serangan *phishing* terus mengalami peningkatan seiring dengan berkembangnya teknik rekayasa sosial yang semakin kompleks dan sulit dikenali oleh pengguna awam. *Phishing* tidak hanya menyebabkan kerugian finansial, tetapi juga berdampak pada kebocoran data sensitif seperti kredensial akun, informasi pribadi, dan data perbankan. Studi menunjukkan bahwa *email* masih menjadi vektor utama serangan *phishing* karena sifatnya yang masif, murah, dan mampu menjangkau banyak target dalam waktu singkat [1].

Metode konvensional dalam mendeteksi *phishing*, seperti pemfilteran berbasis aturan (*rule-based*) dan *blacklist*, dinilai tidak lagi efektif dalam menghadapi pola serangan *phishing* yang dinamis. Teknik tersebut cenderung bergantung pada pola statis sehingga sulit beradaptasi dengan variasi konten *email phishing* yang terus berkembang. Kondisi ini menuntut adanya pendekatan yang lebih adaptif dan mampu belajar dari data historis serangan [2].

Machine Learning (ML) sebagai bagian dari kecerdasan buatan menawarkan solusi yang lebih cerdas dalam mendeteksi *email phishing*. Dengan pendekatan *supervised learning*, sistem dapat mempelajari karakteristik *email phishing* dan *non-phishing* berdasarkan data berlabel. Pendekatan ini memungkinkan sistem untuk melakukan klasifikasi secara otomatis dan meningkatkan akurasi deteksi seiring bertambahnya data pelatihan [3].

Metode *supervised machine learning* banyak digunakan dalam klasifikasi

teks karena mampu menghasilkan performa yang baik ketika dataset memiliki label yang jelas. Dalam konteks *email phishing*, *supervised learning* sangat relevan karena data dapat dikategorikan ke dalam kelas *phishing* dan *non-phishing*. Beberapa algoritma populer yang sering digunakan antara lain *Naïve Bayes*, *Decision Tree*, *Random Forest*, dan *Support Vector Machine* (SVM) [1].

Support Vector Machine (SVM) dikenal sebagai algoritma *supervised learning* yang unggul dalam menangani data berdimensi tinggi, seperti data teks hasil ekstraksi fitur TF-IDF. SVM bekerja dengan mencari *hyperplane* optimal yang mampu memisahkan dua kelas data dengan margin maksimum. Penelitian sebelumnya menunjukkan bahwa SVM memiliki tingkat akurasi dan stabilitas yang lebih baik dibandingkan beberapa algoritma klasifikasi lainnya dalam mendeteksi *email phishing* [3].

Penelitian tentang penggunaan SVM dalam prediksi *email phishing* menunjukkan bahwa SVM efektif dalam mengenali pola serangan *phishing* pada *body email*, terutama pada dataset berbahasa alami [1]. Kombinasi metode TF-IDF dan algoritma *supervised machine learning* memberikan hasil yang signifikan dalam meningkatkan akurasi deteksi *phishing* dan mampu mengurangi *noise* pada data teks sehingga meningkatkan kualitas fitur yang digunakan oleh model klasifikasi [2]. Metode SVM dengan *kernel linear* menunjukkan hasil akurasi yang sangat tinggi, bahkan mencapai 100% pada dataset tertentu. Hal ini mengindikasikan bahwa SVM sangat andal dalam memisahkan *email phishing* dan *non-phishing*. Namun, penelitian tersebut masih terbatas pada dataset statis dan belum menguji ketahanan model terhadap variasi serangan *phishing* terbaru.

Tujuan utama dari penelitian ini adalah menganalisis serangan *phishing* pada pesan *email* serta mitigasi berbasis *supervised machine learning* yang mampu meningkatkan akurasi deteksi. Secara teoretis, penelitian ini memperkaya kajian keilmuan di bidang keamanan siber dan *text mining*. Secara praktis, hasil penelitian diharapkan dapat menjadi referensi bagi institusi, perusahaan, dan pengembang sistem keamanan dalam meningkatkan perlindungan terhadap ancaman *phishing* melalui *email*

1.2 Permasalahan

Berdasarkan latar belakang yang telah diuraikan, maka rumusan masalah dalam penelitian ini adalah bagaimana kinerja metode supervised machine learning dalam menganalisis dan mengklasifikasikan email phishing dan non-phishing menggunakan algoritma K-Nearest Neighbor (KNN), Decision Tree, dan Support Vector Machine (SVM)

1.3 Batasan Masalah

1. Penelitian ini hanya berfokus pada serangan *phishing* yang disampaikan melalui media *email*.
2. Metode yang digunakan dalam penelitian ini dibatasi pada pendekatan *supervised machine learning*.
3. Dataset yang digunakan merupakan dataset publik berjudul *Phishing Email* yang bersumber dari repositori *Kaggle*.

1.4 Tujuan Penelitian

Tujuan penelitian ini adalah menganalisis serangan *phishing* pada *email* menggunakan metode *supervised machine learning.* Pengujian dilakukan dengan 3 algoritma klasifikasi, yaitu *K-Nearest Neighbor* (KNN), *Decision Tree*, dan *Support Vector Machine* (SVM). Studi kasus pada *email phishing* dan *non-phishing*.

1.5 Manfaat Penelitian

1. Memberikan kontribusi keilmuan dalam bidang keamanan siber, khususnya terkait penerapan metode *supervised machine learning* dalam mendeteksi dan mengklasifikasikan pesan *email phishing*.
2. Menjadi referensi akademik bagi peneliti selanjutnya yang ingin mengembangkan atau membandingkan metode *machine learning* dalam mitigasi serangan *phishing* pada *email*.
3. Membantu memahami karakteristik dan pola pesan *email phishing* berdasarkan hasil klasifikasi menggunakan pendekatan *supervised machine*

learning.

4. Menghasilkan model deteksi *email phishing* yang dapat digunakan sebagai dasar dalam pengembangan sistem keamanan *email* yang lebih adaptif dan akurat.
5. Mendukung upaya mitigasi serangan *phishing* pada *email* dengan meningkatkan kemampuan sistem dalam membedakan *email phishing* dan *non-phishing* secara otomatis.
6. Mengurangi risiko kebocoran data sensitif pengguna akibat serangan *phishing* melalui peningkatan akurasi deteksi *email phishing*.
7. Memberikan manfaat praktis bagi pengguna, institusi, dan pengembang sistem keamanan dalam meningkatkan perlindungan terhadap ancaman *phishing* pada layanan *email.*