BAB I

PENDAHULUAN

1.1 Latar Belakang

Teknologi informasi kian berkembang pesat seiring dengan perkembangan zaman, terutama di bidang Teknologi dan Informasi, saat ini telah mengalami kemajuan pesat dan memberikan berbagai manfaat. Hal ini mempermudah hampir semua aktivitas khususnya pemanfaatan internet, terutama melalui media *Website*[1]. Di sisi lain, kemudahan ini bisa menjadi peluang risiko keamanan bagi mereka yang masih belum familiar dengan prosedur keamanan transaksi *online.* Hal ini dapat dimanfaatkan oleh pelaku kejahatan internet untuk memperoleh informasi rahasia, seperti data pribadi, kata sandi *e-mail*, dan bahkan informasi finansial seperti data kartu kredit dan rekening perbankan *online* tanpa sepengetahuan pengguna *internet*[2].

Seiring dengan pertumbuhan pengguna internet dan perkembangan teknologi, risiko terhadap keamanannya semakin beragam. Salah satu contohnya adalah praktik *phishing*. *Phishing* sendiri merupakan upaya untuk memperoleh informasi krusial dari suatu individu, seperti *username, password,* dan data sensitif lainnya, dengan memberikan situs *web* palsu yang menyerupai aslinya. *Phishing*, yang dapat diartikan sebagai "memancing informasi penting," adalah jenis kejahatan yang bertujuan untuk meraih data rahasia, seperti *username, password*, dan informasi kartu kredit, dengan menyamar sebagai entitas atau bisnis terpercaya dalam komunikasi elektronik resmi, seperti surat elektronik atau pesan instan. Seiring dengan pertumbuhan penggunaan media elektronik, dan sejalan dengan peningkatan kejahatan siber, serangan *phishing* semakin umum terjadi[3].

Website phising akan dirancang oleh pelaku kejahatan *internet* dengan cermat agar menyerupai situs asli, termasuk tampilan, konten, *URL domain*, dan elemen lainnya, dengan tujuan memperdaya korban (pengguna *internet*). Tujuan utamanya adalah membuat korban percaya bahwa mereka sedang

mengakses halaman situs yang sah. Desain situs *phising* dibuat semirip mungkin dengan situs yang sah untuk menipu korban agar yakin bahwa mereka berada di situs yang benar. Apabila korban berhasil tertipu dan memberikan informasi yang diminta, pelaku kejahatan internet dapat dengan mudah memanfaatkan informasi tersebut pada situs yang sah untuk melaksanakan kegiatan yang tidak diinginkan. Dampak dari adanya *phishing* dapat mencakup kerugian finansial, kehilangan data, serta menyebabkan kerugian yang signifikan bagi korban, maka dari itu perlu dilakukan pendeteksian untuk *link* atau *uniform resource locactor (URL)*. Deteksi *phishing* adalah cara untuk mengetahui apakah sebuah alamat situs *web URL* itu palsu atau asli. Ada beberapa cara untuk menilai seberapa aman sebuah *URL*, seperti menggunakan daftar hitam, daftar putih, statistik, atau teknologi pembelajaran mesin. Di antara semuanya, teknologi pembelajaran mesin lebih bagus karena lebih efisien dan akurat. Teknologi ini menggunakan model algoritma khusus untuk memahami pola *URL* yang berbahaya dan mampu mendetksi jenis *URL*, apakah itu *phishing* atau situs yang aman, sesuai dengan kebutuhan[4].

Menurut detik.com selama 5 tahun terakhir ini, pencurian data menggunakan teknik *phising* mengalami peningkatan. Kasus *phising* di Indonesia pada tahun 2023 ini mencapai 26.675 kasus yang terjadi pada periode kuartal I 2023. Banyaknya insiden *phishing* dapat berpotensi menyebabkan kerugian, termasuk kerugian privasi bagi individu atau perusahaan. APWG (Anti-*Phishing* Working Group) mengemukakan bahwa kesadaran masyarakat Indonesia terhadap situs *web phishing* semakin meningkat dari tahun ke tahun[2], oleh karena itu, salah satu langkah yang dapat diambil adalah melalui identifikasi untuk mendeteksi situs *web* yang dicurigai *phishing*. Langkah yang dapat dilakukan salah satunya dengan cara penerapan klasifikasi dalam data *mining* untuk memahami data dan parameter yang menjadi pedoman dalam pendeteksian *phishing*.

Model data *mining* membuktikan keefektifannya dalam mendeteksi serangan *phishing* berbasis *URL* dengan melakukan analisis mendalam terhadap konten *HTML*, otoritas *domain*, dan kode skrip yang terdapat pada halaman

web. Dengan memanfaatkan metode ini, sistem dapat secara cerdas mengidentifikasi pola-pola khusus yang umumnya terkait dengan upaya *phishing*. Analisis konten *HTML* memungkinkan pemantauan terhadap perubahan struktur halaman *web* yang mencurigakan, sedangkan evaluasi otoritas domain memberikan gambaran tentang keabsahan sumber informasi. Selain itu, pengamatan terhadap kode skrip membantu dalam mendeteksi aktivitas mencurigakan atau upaya manipulasi yang dapat merugikan pengguna. Dengan menggabungkan ketiga aspek ini, model data *mining* membantu menciptakan lapisan perlindungan yang kuat untuk mengidentifikasi dan mencegah serangan *phishing* berbasis *URL* secara efisien[5].

Penelitian ini menggunakan suatu model berbasis data *mining* yang bertujuan untuk meningkatkan akurasi dalam mengklasifikasikan *URL phishing* dalam konteks keamanan *web*. Dengan memanfaatkan model data mining, model ini dirancang untuk secara akurat mendeteksi dan mengklasifikasikan *URL* yang memiliki karakteristik serangan *phishing*. Penerapan data *mining* dalam proses analisis *URL* diharapkan dapat meningkatkan keandalan sistem dalam mengenali pola-pola yang umumnya terkait dengan serangan *phishing*. Melalui peningkatan akurasi deteksi dan klasifikasi, model ini diharapkan dapat memberikan lapisan perlindungan yang lebih efektif bagi pengguna *web*, mengurangi risiko akses ke situs-situs berbahaya, dan pada gilirannya, meningkatkan keamanan secara keseluruhan. Dengan demikian, penelitian ini berkontribusi pada pengembangan solusi yang lebih canggih untuk mengatasi ancaman keamanan yang berkaitan dengan serangan *phishing* melalui pendekatan berbasis data *mining*.

Beberapa fungsi dari data mining mencakup analisis asosiasi antar data, klasifikasi data, *clustering* data, prediksi dan lain-lain. Dalam penelitian ini, fungsionalitas yang diterapkan adalah klasifikasi data. Klasifikasi data adalah suatu proses untuk menemukan model atau fungsi yang dapat menjelaskan dan membedakan kelas-kelas data serta konsepnya[6]. Penelitian ini akan melakukan perbandingan antara beberapa algoritma untuk menentukan tingkat akurasi tertinggi. Algoritma yang digunakan dalam penelitian ini melibatkan 3

algoritma yaitu, *k-nearest neighbors*, *random forest*, dan *support vector machine*.

1.2 Rumusan Masalah

Berdasarkan uraian dari latar belakang yang telah dijelaskan, maka terdapat beberapa rumusan masalah yang diangkat pada penelitian ini, yaitu:

1. Bagaimana langkah-langkah implementasi algoritma *K-Nearest Neighbors*, *Random Forest,* dan *Support Vector Machine* dalam melakukan klasifikasi *website phising*?
2. Bagaimana evaluasi hasil pengujian terkait kinerja algoritma *K-Nearest Neighbors, Random Forest,* dan *Support Vector Machine* dalam melakukan klasifikasi *website phishing*?
3. Bagaimana algoritma *K-Nearest Neighbors, Random Forest,* dan *Support Vector Machine* dapat mendeteksi *phishing* melalui *website* yang sudah dirancang?

1.3 Batasan Masalah

Batasan masalah dalam penelitian ini yaitu sebagai berikut:

1. *Dataset* yang digunakan adalah kumpulan data dari situs *web phishing* yang diperoleh dari repositori publik. *Dataset* ini terdiri dari 30 atribut, dan klasifikasinya terbagi dalam dua kelompok, yaitu *phishing* atau tidak *phishing*.
2. Mencari nilai akurasi, f1 *score*, *recall*, dan presisi.
3. Rentang waktu *dataset* yang digunakan adalah 5 tahun terakhir karena dapat memberikan informasi yang lebih akurat dan terperinci sesuai dengan relevansi terhadap kondisi terkini.
4. algoritma yang digunakan untuk mengklasifikasikan dataset akan difokuskan pada 3 algoritma yaitu, *K-Nearest Neighbor, Random Forest,* dan *Support Vector Machine* karena ketiga algoritma tersebut merupakan hasil terbaik dari perbandingan yang telah dilakukan oleh penelitian terdahulu.

5. Alur penelitian hanya dilakukan sampai tahap evaluasi, dan tidak sampai *deployment.*

1.4 Tujuan dan Manfaat Penelitian

1.4.1 Tujuan Penelitian

Tujuan dari penelitian ini adalah:

1. Mengimplementasikan algoritma *K-Nearest Neighbors, Random Forest,* dan *Support Vector Machine* untuk klasifikasi *website phising.*

2. Mendapatkan hasil komparasi dari evaluasi pengujian terkait performa algoritma *K-Nearest Neighbors, Random Forest,* dan *Support Vector Machine* dalam melakukan klasifikasi *website phishing.*

3. Menghasilkan *website* yang dapat mendeteksi *phishing* melalui algoritma.

1.4.2 Manfaat Penelitian

Manfaat dari penelitian ini adalah:

1. Memahami penerapan algoritma *K-Nearest Neighbor, Random Forest,* dan *Support Vector Machine* dalam mengklasifikasikan situs *web phishing.*

2. Menilai manakah kinerja algoritma antara *K-Nearest Neighbor, Random Forest,* dan *Support Vector Machine* yang lebih baik dalam mengklasifikasikan situs *web phishing.*

3. Menganalisis nilai akurasi, f1 *score, recall*, dan presisi, untuk algoritma *K-Nearest Neighbor, Random Forest,* dan *Support Vector Machine.*

1.5 Sistematika Penulisan

Dokumentasi penelitian ini mengikuti susunan tertentu untuk memberikan struktur yang terorganisir dan mempermudah pemahaman. Struktur penulisan terbagi menjadi lima bab dengan fokus yang berbeda, sebagai berikut:

1. BAB I PENDAHULUAN

   Pendahuluan mencakup latar belakang, perumusan masalah, batasan masalah, tujuan, dan manfaat penelitian. Bagian ini menjelaskan mengapa topik penelitian relevan dan perlu untuk diselidiki.

2. BAB II LANDASAN TEORI

   Bagian ini menjelaskan teori yang mendukung penelitian, termasuk framework, tools, dan algoritma yang digunakan. Teori ini bersumber dari jurnal dan buku yang membahas metode atau penelitian terkait.

3. BAB III METODOLOGI PENELITIAN

   Metodologi penelitian membahas objek penelitian, langkah-langkah yang digunakan dalam penelitian, teknik pengambilan data, dan penjelasan tentang variabel penelitian.

4. BAB IV ANALISIS DAN HASIL PENELITIAN

   Bab ini mencakup implementasi metodologi pada objek penelitian untuk mencapai tujuan penelitian. Di dalamnya terdapat penjelasan tentang proses pengolahan data, analisis data, dan hasil akhir yang diperoleh.

5. BAB V SIMPULAN DAN SARAN

   Bagian ini berisi kesimpulan dari hasil penelitian beserta saran berdasarkan kendala atau analisis yang ditemukan. Saran tersebut diharapkan dapat memberikan kontribusi bagi penelitian dengan topik atau tujuan serupa.