BAB I
PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi informasi dan internet yang semakin cepat telah mengubah banyak hal dalam kehidupan, mulai dari cara berkomunikasi, melakukan transaksi uang, hingga penggunaan layanan publik secara digital. Hampir semua bidang kehidupan tidak lepas dari teknologi informasi dan komunikasi. Meski ada banyak keuntungan, perkembangan ini juga membuat ancaman *cybercrime* (kejahatan siber) semakin besar. Salah satu cara penipuan ini adalah serangan *phishing*. Phising adalah istilah yang mengacu pada berbagai kejahatan siber dengan memberikan *link* kepada pengguna internet yang berupaya mencuri informasi pribadi seperti detail bank, kode PIN, kata sandi, dan informasi media sosial. Menurut A.S Sunge [1], keamanan di komunitas internet sering kali sulit diidentifikasi oleh pengguna, sehingga aktivitas *phishing* dan penipuan sering kali terlewat tanpa dideteksi.

Salah satu bentuk serangan *phishing* yang umum terjadi adalah melalui penggunaan *domain name* atau *Uniform Resource Locator (URL)* palsu yang dirancang sedemikian rupa agar menyerupai domain asli. Kondisi inilah yang menjadi alasan betapa pentingnya sistem deteksi otomatis yang dapat mengenali domain *phishing* berdasarkan karakteristik teks dari URL tersebut.

Kasus Phising di Indonesia semakin meningkat, dengan modus penipuan menggunakan SMS, email, dan WhatsApp. Pelaku memalsukan identitas bank atau perusahaan agar bisa mencuri data pribadi seperti OTP, PIN, dan kata sandi. Beberapa kasus yang terungkap melibatkan orang dari luar negeri, dan seringkali tujuannya adalah mengambil uang. Sektor yang sering menjadi target adalah e-commerce dan media sosial. Menurut Laporan IDADX tahun 2024 mencatat bahwa ada sebanyak 106.806 laporan phising selama kurun waktu 5 tahun terakhir sejak tahun 2018[2]. Hal ini menunjukkan pentingnya pengembangan sistem yang mampu mengidentifikasi domain *phishing* secara otomatis dan akurat terutama di

bidang *Machine Learning*.

Menurut Goodfellow dkk. [3], *Machine Learning* adalah bentuk statistik terapan dengan peningkatan penekanan pada penggunaan komputer untuk memperkirakan fungsi kompleks secara statistik. Dalam beberapa tahun terakhir, pendekatan *machine learning* (pembelajaran mesin) menjadi salah satu metode yang efektif dalam mendeteksi *phishing*. Dengan menganalisis fitur teks dari domain atau URL, model *machine learning* dapat mempelajari pola-pola tertentu yang membedakan antara domain *phishing* dan domain sah (*legitimate*).

Menurut Sismoro [4], Algoritma adalah sekumpulan instruksi atau langkah-langkah yang dituliskan secara sistematis dan digunakan untuk menyelesaikan masalah/persoalan logika dan matematika dengan bantuan komputer. Dalam pemrograman dasar, kita dapat mengatakan bahwa suatu algoritma adalah langkah atau hal pertama yang harus dipersiapkan sebelum membuat program [5].

Dalam proses klasifikasi teks, termasuk pada deteksi domain *phishing*, terdapat berbagai algoritma *machine learning* yang dapat digunakan, seperti *Logistic Regression*, *Support Vector Machine* (SVM), *Random Forest*, dan *Naïve Bayes*. Setiap algoritma memiliki karakteristik, kelebihan, serta keterbatasannya masing-masing. *Logistic Regression* dikenal cukup stabil dan mudah diinterpretasikan, namun performanya dapat menurun ketika jumlah fitur sangat besar seperti pada representasi teks. Algoritma SVM mampu memberikan akurasi tinggi, tetapi memiliki waktu komputasi yang jauh lebih lama, terutama ketika data sudah melalui tokenisasi sehingga menghasilkan ribuan fitur. Sementara itu, *Random Forest* menawarkan kemampuan generalisasi yang baik melalui penggabungan beberapa pohon keputusan, tetapi membutuhkan sumber daya komputasi lebih besar dan proses pelatihannya cenderung lambat pada data berdimensi tinggi.

Berdasarkan karakteristik tersebut, penelitian ini membutuhkan algoritma yang efisien untuk data berdimensi tinggi, memiliki waktu pelatihan cepat, namun tetap kompetitif. Oleh karena itu, algoritma *Multinomial Naïve Bayes* dipilih sebagai metode klasifikasi utama. Sebagai varian dari keluarga *Naïve Bayes* yang

dirancang khusus untuk data dengan distribusi frekuensi, *Multinomial Naïve Bayes* sangat efektif dalam menangani klasifikasi teks berskala besar.

Namun, kinerja algoritma ini sangat bergantung pada tahap *preprocessing*, khususnya pada teknik *tokenization* dan representasi fitur. Oleh karena itu, penelitian ini akan mengimplementasikan *Multinomial Naïve Bayes* dengan membandingkan dua pendekatan representasi fitur, yaitu *Bag-of-Words* (BoW) dan TF-IDF. Hasil perbandingan ini diharapkan dapat menentukan metode representasi yang paling optimal dalam meningkatkan akurasi deteksi *phishing* pada URL.

1.2 Rumusan Masalah

Berdasarkan latar belakang diatas, rumusan masalah yang dapat diidentifikasi dalam penelitian ini adalah sebagai berikut:

1. Seperti apakah pengaruh BoW dan TF-IDF terhadap performa *Multinomial Naïve Bayes* dalam mendeteksi Domain *Phishing* berdasarkan fitur teks dari URL?
2. Metode representasi teks manakah yang menghasilkan kinerja terbaik berdasarkan metrik evaluasi seperti *accuracy*, *precision*, *recall*, dan *F1-score*?

1.3 Batasan Masalah

Untuk memperjelas ruang lingkup penelitian, maka batasan masalah dalam penelitian ini adalah sebagai berikut:

1. Dataset yang digunakan berupa dataset Publik dari platform Kaggle dan hanya mencakup data domain atau URL yang dikategorikan sebagai *phishing* dan *non-phishing*.
2. Penelitian ini hanya menggunakan satu algoritma klasifikasi, yaitu *Multinomial Naïve Bayes*.
3. Perbandingan dilakukan hanya pada dua metode representasi teks, yaitu BoW dan TF-IDF.
4. Evaluasi kinerja model menggunakan metrik *accuracy*, *precision*, *recall*,

dan *F1-score*.

5. Penelitian dilakukan menggunakan bahasa pemrograman Python melalui platform Google Colab.

1.4 Tujuan Penelitian

Adapun tujuan dari penelitian ini adalah sebagai berikut:

1. Menerapkan algoritma *Multinomial Naïve Bayes* untuk mendeteksi domain *phishing* berdasarkan fitur teks.
2. Melakukan perbandingan antara dua metode representasi teks (BoW dan TF-IDF) dalam klasifikasi domain *phishing*.
3. Mengetahui metode representasi teks yang memberikan hasil terbaik terhadap kinerja model *Multinomial Naïve Bayes* berdasarkan metrik evaluasi yang digunakan.

1.5 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan manfaat sebagai berikut:

1. Bagi peneliti dan akademisi, hasil penelitian ini dapat menjadi referensi dalam memahami pengaruh metode representasi teks terhadap kinerja algoritma *Multinomial Naïve Bayes* dalam klasifikasi teks *phishing*.
2. Bagi pengembang sistem keamanan siber, penelitian ini dapat menjadi dasar dalam pengembangan sistem deteksi *phishing* otomatis yang lebih akurat dan efisien.
3. Bagi masyarakat umum, hasil penelitian ini dapat meningkatkan kesadaran akan pentingnya keamanan siber dan cara kerja sistem deteksi *phishing* berbasis kecerdasan buatan.

1.6 Sistematika Penulisan

BAB I PENDAHULUAN

Untuk memberikan gambaran yang lebih baik tentang isi skripsi, bab ini penulis memaparkan latar belakang masalah yang mendasari penelitian, perumusan

masalah, Batasan masalah, tujuan penelitan, manfaat penelitian, dan sistematika penulisan skripsi.

BAB II TINJAUAN PUSTAKA

Berisi tinjauan pustaka, teori-teori yang mendasari penelitian, seperti konsep *phishing*, *machine learning*, *Multinomial Naïve Bayes*, dan metode representasi teks (BoW dan TF-IDF).

BAB III METODE PENELITIAN

Menjelaskan tahapan penelitian, mulai dari pengumpulan data, pra-pemrosesan teks, implementasi model, hingga evaluasi.

BAB IV HASIL DAN PEMBAHASAN

Bab ini memaparkan hasil pengujian dan analisis perbandingan performa kedua metode representasi teks.

BAB V PENUTUP

Berisi kesimpulan hasil penelitian serta saran untuk pengembangan penelitian selanjutnya.