

# BAB 1

## PENDAHULUAN

### 1.1 Latarbelakang

Pada era globalisasi 4.0 saat ini kita di tuntut agar bisa mengenal dan menguasai teknologi yang sangat pesat perkembangannya. Ada banyak cara yang dapat dilakukan untuk mengikuti pesatnya perkembangan teknologi. Salah satunya adalah dengan memanfaatkan internet, dengan adanya internet kita lebih mudah untuk mendapatkan informasi dan berkomunikasi [1].

Menurut data yang dirilis oleh [2] ada sebanyak 4,95 miliar pengguna internet aktif pada tahun 2022 dengan persentase 62,5% dari total seluruh penduduk dunia yang dimana data tersebut meningkat sebesar 4% dari tahun 2021. Hampir semua kegiatan sehari - hari dituangkan ke Internet, dan pengguna berinteraksi dengan memiliki akun pribadi yang ditautkan ke data dunia, dengan memberi akses setiap pengguna untuk melihat berbagai informasi diseluruh dunia melalui *website*[3].

Namun dengan meningkatnya pengguna yang mengakses internet membuat kerahasiaan data para pengguna internet semakin rentan untuk dicuri oleh oknum, atau kelompok yang tidak bertanggung jawab [4]. Salah satunya pencurian data yang marak menggunakan *web phishing* [5]. *Phishing* adalah serangan dimana penyerang mencoba mencuri informasi rahasia dari orang yang dituju dengan mengirimkan tautan palsu. Penyerang mencuri informasi pribadi yang dimasukkan oleh pengguna di situs web palsu [6].

Penelitian terdahulu tentang “Perbandingan Klasifikasi Algoritma K-Nn, Neural Network, Naïve Bayes, C 4.5 Untuk Mendeteksi Web Phising” dilakukan oleh [7] menggunakan 1353 dataset yang diolah dan dibagi menjadi 10 atribut, membuktikan bahwa algoritma Decision Tree C4.5 memiliki kinerja yang lebih tinggi dibandingkan dengan tiga algoritma lainnya yaitu Neural Network, Naïve Bayes dan k-NN dengan Akurasi sebesar 89,66%. Berdasarkan penelitian sebelumnya diperlukan pengembangan metode untuk mendapatkan akurasi yang lebih baik lagi, oleh karena itu peneliti tertarik melakukan sebuah penelitian yang berjudul “Analisis Komparasi Algoritma Klasifikasi Prediksi Website Phising Menggunakan Logistic Regression, Decision Tree, Dan Random Forest”.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dipaparkan diatas peneliti menetapkan rumusan masalah sebagai berikut:

1. Bagaimana proses klasifikasi dalam menentukan sebuah *website phishing*?
2. Bagaimana hasil komparasi akurasi dari algoritma *Logistic Regression, Decision Tree, Dan Random Forest*?

## 1.3 Batasan Masalah

Untuk membantu peneliti dalam melakukan penelitian ditentukan batasan masalah sebagai berikut:

1. Dataset website phishing diambil dari Kaggle.
2. Algoritma yang digunakan *Logistic Regression, Decision Tree, Dan Random Forest*.
3. Klasifikasi yang dilakukan hanya kepada *Web Phising*

## 1.4 Tujuan Penelitian

Tujuan dilakukan penelitian ini adalah untuk membantu pengguna internet yang mengakses website untuk memprediksi apakah sebuah *website* itu *valid* atau *Phising*.

## 1.5 Manfaat Penelitian

Manfaat penelitian ini dapat memberikan pengetahuan tentang *web phishing* kepada seluruh pengguna internet supaya selalu waspada terhadap *web phishing* dan diharapkan jumlah kasus pencurian data menggunakan *web phishing* semakin berkurang.