BAB 1
PENDAHULUAN

1.1 Latar Belakang

Seiring dengan perkembangan teknologi dan internet, ancaman siber semakin meningkat, terutama serangan *phising*. Menurut laporan Anti-*Phising* Working Group (APWG) tahun 2023, ada lebih dari 1,2 juta serangan *phising* yang terdeteksi, meningkat 61% dari tahun sebelumnya (APWG, 2023). Serangan *phising* ini tidak hanya menyerang individu tetapi juga organisasi-organisasi besar, dengan kerugian finansial global mencapai USD 42 miliar pada tahun 2023 (FBI Internet Crime Report, 2023). Di Indonesia sendiri, berdasarkan data Badan Siber dan Sandi Negara (BSSN), kasus *phising* meningkat hingga 65% pada tahun 2023 dengan lebih dari 80.000 kasus dilaporkan (BSSN, 2023). Pada kuartal pertama tahun 2023, terdapat 26.675 laporan serangan *phising* dengan sektor pemerintahan dan individu menjadi target utama(Bankjombang, 2023). Menurut studi terbaru, *phising* telah menjadi metode paling umum dalam pencurian data dan serangan siber lainnya (Ashar Ahmed Fazal & Maryam Daud, 2023). Oleh karena itu, metode deteksi yang tepat dan otomatis sangat diperlukan untuk mengidentifikasi situs *phising* secara efisien.

Salah satu pendekatan yang terbukti efektif dalam mendeteksi *phising* adalah penggunaan algoritma machine learning. Algoritma Random Forest memiliki akurasi tinggi dalam deteksi situs *phising*, dengan tingkat akurasi mencapai 97,14% serta false positive rate yang lebih rendah dibandingkan Decision Tree dan Support Vector Machine (SVM) (Costner dkk., 2023). Meskipun demikian, dalam penelitian ini algoritma Decision Tree CART (Classification and Regression Trees) dipilih karena keunggulannya dalam menghasilkan pohon keputusan yang mudah diinterpretasi. Keunggulan ini sangat penting untuk memudahkan proses klasifikasi situs *phising*, terutama dalam menangani data numerik dan kategorikal secara efektif, sehingga menghasilk

an model yang ringkas dan efisien. Studi oleh (Subarkah & Ikhsan, 2021)menunjukkan bahwa penggunaan Decision Tree CART dalam deteksi *phising* dapat mencapai akurasi 95,28%.

Beberapa penelitian telah membuktikan efektivitas algoritma Decision Tree dalam mendeteksi *phising*. Studi oleh (Ashar Ahmed Fazal & Maryam Daud, 2023)menunjukkan bahwa model Decision Tree dapat mencapai akurasi hingga 95,97%, dengan tingkat true positive sebesar 96,64% dan false positive hanya 3,04%, menggunakan dataset dari UCI *Machine Learning Repository* (Ashar Ahmed Fazal & Maryam Daud, 2023). Namun Sebagian penelitian sebelumnya masih bergantung pada dataset statis, sehingga belum optimal dalam mendeteksi *phising* terbaru yang berkembang dengan cepat.

Untuk mengatasi keterbatasan tersebut, penelitian ini menggunakan teknik data mining dan web crawling untuk mengumpulkan data dari situs web, terutama dengan memanfaatkan dataset dari *PhisTank*, sebuah layanan yang menyediakan daftar situs *phising* terbaru. Data variabel yang dianalisis mencakup fitur *URL*, struktur *HTML*, serta metadata situs web, yang akan diklasifikasikan menggunakan algoritma *Decision Tree CART* (*Classification and Regression Tree*) guna membangun model deteksi *phising* yang presisi.

Decision Tree CART menjadi pilihan utama dalam menangani data dinamis dari *PhisTank* karena kemampuannya dalam menangani berbagai tipe data, baik numerik maupun kategorikal. Model ini sangat fleksibel dalam menyesuaikan diri dengan data yang terus diperbarui secara *real-time*, seperti daftar *phising* terbaru dari *PhisTank*. Selain itu, *CART* mampu membentuk pohon keputusan yang sederhana dan mudah dipahami, sehingga memungkinkan analisis yang lebih transparan dan interpretasi hasil yang lebih jelas dibandingkan metode lainnya. Keunggulan lain dari *Decision Tree CART* adalah kemampuannya dalam mengurangi noise pada data serta menangani dataset dengan jumlah variabel yang besar secara efisien, tanpa mengorbankan performa deteksi.

Meskipun sebagian studi telah membahas deteksi *phising* menggunakan *Decision Tree* terdapat beberapa kesenjangan dengan penggunaan dataset UCI atau sumber lain yang tidak diperbarui secara *real-time*. Minimnya penelitian yang

mengintegrasikan teknik *web crawling* dengan *Decision Tree* untuk mendeteksi situs *phising* secara presisi. Penggunaan dataset lama cenderung memiliki performa yang rendah dalam mengidentifikasi tren *phising* terbaru. Dalam penelitian ini penggunaan data real-time dari *PhisTank* yang diperoleh dari teknik *web crawling* dengan pendekatan model ini dapat menyesuaikan dengan tren *phising* terbaru, menambah efektivitas deteksi dibandingkan dengan metode konvensional.

Mengingat semakin canggihnya serangan *phising* yang dapat mengancam keamanan data pengguna internet. Dengan meningkatnya transaksi online dan aktivitas digital lainnya, deteksi *phising* yang cepat dan relevan menjadi sangat krusial untuk mencegah pencurian data. Penggunaan model *Decision Tree CART* dalam sistem berbasis *web crawling* memungkinkan sistem deteksi *phising* yang lebih otomatis, presisi, dan adaptif terhadap serangan baru.

Berdasarkan permasalahan diatas pengunaan algoritma *Decision Tree CART* memiliki keunggulan dalam interpretabilitas, kecepatan, dan efisiensi klasifikasi, sehingga kompatible untuk deteksi *phising* secara real-time. Selain itu, *PhisTank* digunakan sebagai sumber data karena menyediakan daftar situs *phising* yang memungkinkan model deteksi yang relevan dan terbaru.

1.2 Perumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan diatas, maka dengan penggunaan metode deteksi *phising* yang ada masih memiliki keterbatasan, dalam hal penggunaan dataset statis yang kurang mampu mengakomodasi perkembangan tren *phising* terbaru. Selain itu, minimnya penelitian yang mengintegrasikan teknik *web crawling* dengan algoritma *Decision Tree CART* menjadi tantangan dalam menciptakan sistem deteksi *phising* yang lebih adaptif dan presisi. Untuk itu, penelitian ini dilakukan agar dapat mengembangkan model deteksi *phising* berbasis *Decision Tree CART* yang dikombinasikan dengan *web crawling* menggunakan dataset *real-time* dari *PhisTank*.

1.3 Pertanyaan Penelitian

Dari hasil rumusan masalah yang sudah ditetukan tersebut, maka dapat di identifikasi pertanyaan yang akan melatar belakangi penelitian. Pertanyaan tersebut antara lain:

1. Bagaimana cara mengimplementasikan teknik *web crawling* menggunakan algoritma *Decision Tree CART* dalam mendeteksi *website phising*?
2. Bagaimana evektifitas algoritma *Decision Tree CART* dalam mendeteksi situs *phising* berbasis dataset real-time dari *PhisTank*?
3. Apa fitur yang paling berpengaruh dalam mendeteksi *website phising* berdasarkan analisis menggunakan *Decision Tree CART*?

1.4 Tujuan Penelitian

Tujuan penelitian ini adalah melakukan evaluasi model deteksi *website phising* dengan menggunakna teknik data *mining* dan *web crawling* berbasis algoritma *Decision Tree CART* dengan memanfaatkan data *real-time PhisTank* dan menganalisis efektivitas fitur *URL*, struktur *HTML*, dan metadata situs web dalam meningkatkan akurasi klasifikasi *phising*.

1.5 Manfaat Hasil Penelitian

Manfaat yang dapat diperoleh dari penelitian ini adalah:

1. Memberikan kontribusi dalam pengembangan metode deteksi *phising* berbasis *Decision Tree CART* yang relevan dan adaptif.
2. Menghasilkan sistem deteksi serangan berbasis data *real-time* yang reponsif terhadap pola serangan terbaru.
3. Pemanfaatan *PhisTank* sebagai sumber data *real-time*, sehingga meningkatkan efektivitas sistem deteksi *phising* dibandingkan pendekatan berbasis dataset statis.