BAB I
PENDAHULUAN

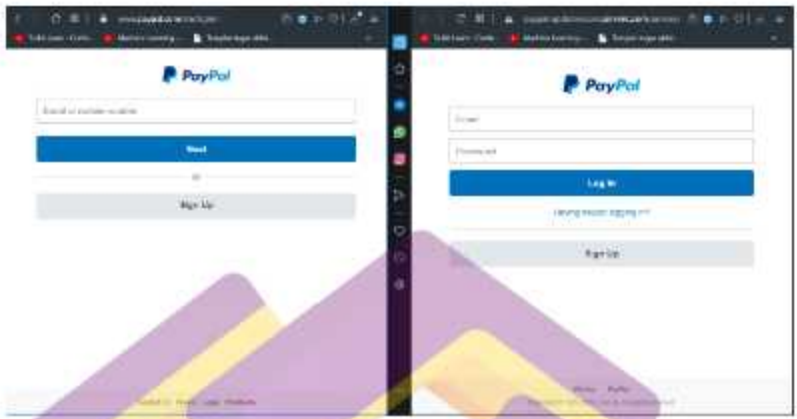
1.1 Latar Belakang

Perkembangan Ilmu Pengetahuan dan Teknologi (IPTEK), terutama pada Teknologi Informasi memberi kemudahan manusia dalam menunjang aktivitasnya diberbagai bidang terutama dalam kehidupan sehari-hari [1]. Perkembangan yang paling mencolok yaitu kemudahan manusia dalam mengakses internet dimanapun dan kapanpun. Berkembangnya Teknologi Informasi juga berdampak semakin meningkatnya pengguna internet di dunia, hal ini membuat internet sudah menjadi kebutuhan *primer* bagi setiap orang di dunia yang menggunakan *smartphone* maupun komputer dalam mengaksesnya [2]. Dengan adanya internet ini, setiap orang yang sebelumnya melakukan aktivitas harus secara *offline* namun sekarang bisa dilakukan dimana saja secara *online* dan *realtime* [3], sehingga menjadi lebih mudah serta efisien dari sisi tenaga dan waktu. Di sisi lain, berbagai kemudahan yang didapat ini menjadi celah keamanan terhadap pemanfaatan yang dilakukan oleh penjahat dunia maya kepada pengguna internet yang masih awan mengenai rawannya transaksi di dunia maya dengan tujuan untuk mencuri berbagai informasi seperti data pribadi, email, kata sandi, bahkan yang paling terburuk yaitu informasi finansial seperti data internet banking dan kartu kredit [3]. Kejahatan ini merupakan bagian dari *cybercrime* [4] yang sering mengincar pada situs *ecommerce* atau *internet banking*. Selain mengincar data pribadi *internet banking* atau *ecommerce*, *phishing* juga marak digunakan untuk mencuri data akun *game online* dengan membagikan tautan yang berisi iming-iming sebuah hadiah dari *game* seperti *skin hero*, *diamond*, atau lainnya secara gratis dengan syarat harus

mengisi sebuah formulir yang telah dibuat oleh penjahat internet untuk mencuri data akun tersebut. Kasus *phishing* pada *game online* ini banyak terjadi di forum grup sosial media facebook dengan mengirimkan komentar spam tautan pada sebuah postingan orang lain. Hal ini membuat para pengguna internet menjadi menjadi takut untuk melakukan transaksi secara online serta mengakses suatu situs yang ada di internet.

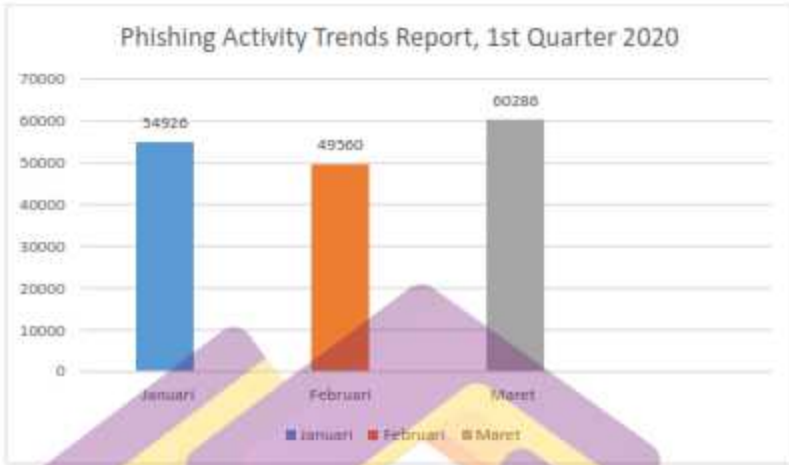
Phishing merupakan suatu tindakan aktivitas kriminal dunia maya dengan tujuan untuk mendapatkan informasi pribadi yang rahasia secara ilegal. Informasi tersebut seperti *username*, *password*, dan data identitas sensitif lainnya dengan memanfaatkan *website phishing* [5]. *Website phishing* merupakan sebuah situs yang dikembangkan dan didesain oleh penjahat internet dengan sedemikian rupa seolah situs tersebut mirip dengan aslinya seperti dari segi tampilan, konten, URL domain atau lainnya. Tujuannya yaitu ngelabuhi korbannya untuk memasukkan suatu data informasi pribadi yang sensitif yang diminta, maka dengan mudah penjahat internet mendapatkan data informasi yang bisa digunakan untuk melakukan aktivitas yang tidak diinginkan dengan berujung kerugian yang diderita korbannya baik data informasi maupun finansial [6]. Di bawah ini merupakan contoh perbandingan halaman *website phishing*[1] dan non-*phishing*[2] yang menyerupai tampilan pada *website online banking* PayPal aslinya seperti pada Gambar 1.1.

[1]http://paypal-updatesecure.serveirc.com/online/customer_center/customer-IDPP00C229/myaccount/signin/, diakses pada tanggal 26 Mei 2020
[2] https://www.paypal.com/us/signin, diakses pada tanggal 26 Mei 2020



Gambar 1.1 Perbandingan *Website Phishing* Dengan Non-*Phishing*

Jika diperhatikan *website phishing* hampir mirip sekali dengan aslinya, bahkan jika orang awam yang mengakses maka dengan mudahnya terlena untuk memasukkan email dan *password*. Menurut APWG *(Anti-Phishing Working Group)* [7], tiap tahunnya kesadaran masyarakat terhadap *website phishing* meningkat, tetapi jumlah *website phishing* dan dampak kerugian yang ditimbulkan tumbuh lebih cepat. Pada laporan APWG kuartal awal 2020, *phishing activity trend* pada bulan Januari 2020 terdapat 54.926 yang terdeteksi *website phishing*, sedangkan pada bulan Februari dan Maret masing-masing terdapat 49.560 dan 60.286 *website* yang terdeteksi sebagai *website phishing*, berikut laporan yang APWG kuartal awal 2020 disajikan dengan grafik pada Gambar 1.2.



Gambar 1.2 Grafik *Phishing Activity Trend* Kuartal Pertama 2020

Untuk mengatasi terhadap maraknya *phishing* yang terjadi di dunia maya, maka diperlukan sistem untuk mendeteksi situs tersebut termasuk dalam kategori *phishing* atau non-*phishing* dengan menerapkan metode dari *machine learning*. Pada penelitian sebelumnya dengan judul "MODEL KLASIFIKASI UNTUK DETEKSI SITUS PHISING DI INDONESIA" oleh Febry Eka Purwiantono & Aris Tjahyanto terdapat beberapa model klasifikasi untuk deteksi situs *phishing*. Penelitian tersebut penulis membuat sebuah model klasifikasi yang mampu untuk deteksi situs *phishing* lalu menguji model klasifikasi yang telah dibuatnya [3]. Beberapa algoritma klasifikasi dalam uji cobanya yaitu SMO (*Sequential Minimal Oprimization*), *Naïve Bayes*, *Bagging*, dan *Multilayer Perceptron*.

Selain itu, penelitian sebelumnya dengan judul "Prediksi Website Pemancing Informasi Penting Phising Menggunakan Support Vector Machine (SVM)" oleh Zuhri Halim tahun 2017 menggunakan beberapa metode klasifikasi

dalam penelitiannya untuk dilakukan perbandingan beberapa algortima antara lain *Support Vector Machine, Naïve Bayes,* dan *Decision Tree* dalam prediksi *website phishing*.

Berdasarkan uraian di atas, terdapat beberapa metode algoritma klasifikasi yang digunakan oleh penelitian sebelumnya dalam klasifikasi dan prediksi *website phishing*. Penelitian selanjutnya yang akan dilakukan oleh peneliti memutuskan untuk menggunakan algoritma klasifikasi *decision tree* (CART) yang mana algoritma klasifikasi *decision tree* (CART) ini memiliki beberapa kelemahan dan kekurangan yaitu kesulitan dalam merancang pohon keputusan yang optimal dan akumulasi jumlah kesalahan dari setiap level dalam pohon keputusan itu besar [8], hal ini akan berdampak pada tingkat akurasi yang kurang optimal, maka diperlukan suatu metode yang dapat meningkatkan akurasi yaitu dengan menggunakan metode *bagging*. *Bagging* dirancang untuk meningkatkan stabilitas serta akurasi dari algoritma *machine learning* dalam klasifikasi statistik dan regresi. Maka dari itu peneliti memutuskan memilih dengan judul penelitian **"Optimasi Algoritma Klasifikasi *Decision Tree* (CART) Menggunakan Metode *Bagging* Untuk Deteksi *Website Phishing*"**.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah disampaikan, rumusan masalah pada penelitian ini adalah "Apakah penerapan metode *bagging* untuk optimasi algoritma klasifikasi *decision tree* (CART) dapat memberikan tingkat akurasi yang lebih baik dalam deteksi *website phishing*?".

1.3 Batasan Masalah

Adapun batasan ruang lingkup dari penelitian ini sebagai berikut.

a. Menggunakan metode optimasi yaitu *bagging*.
b. Optimasi digunakan pada algoritma *decision tree* (CART).
c. Dataset yang digunakan berasal dari UCI *Machine Learning Repository*.
d. Perhitungan menggunakan *library* dari scikit-learn.
e. Sistem dibangun dengan framework Flask menggunakan bahasa pemrograman Python.
f. Sistem digunakan sebagai pembuktian bahwa algoritma yang diimplementasikan berjalan untuk deteksi *website phishing*.

1.4 Maksud dan Tujuan Penelitian

Tujuan dari penelitian ini adalah untuk mengoptimasi algoritma klasifikasi *decision tree* (CART) menggunakan salah satu penggabungan algoritma dari *machine learning* yaitu metode *bagging* untuk deteksi *website phishing* dengan tingkat akurasi yang lebih baik dan akurat.

1.5 Manfaat Penelitian

Manfaat dari penelitian yang dilakukan adalah sebagai berikut

a. Bagi Peneliti

   Memahami dan mengimplementasi dari algoritma penggabungan *machine learning* yaitu *bagging* sebagai pengalaman berharga peneliti dalam meningkatkan kemampuan tentang penelitian yang dilakukan dengan topik "Optimasi Algoritma Klasifikasi *Decision Tree* (CART) Dengan Metode *Bagging* Untuk Deteksi *Website Phishing*".

b. Bagi Masyarakat

Memberikan wawasan terhadap masyarakat awam pengguna internet tentang apa itu *phishing* dan *website phishing*, serta bagaimana cara mengetahui dan mewaspadai.

c. Bagi Ilmu Pengetahuan

Mengimplementasi algoritma penggabungan *machine learning* yaitu *bagging* untuk deteksi *website phishing*, serta memberikan kontribusi terhadap ilmu pengetahuan tentang pengujian algoritma pada penelitian ini.

d. Bagi Peneliti Lanjutan

Hasil dari penelitian ini dapat digunakan sebagai acuan referensi untuk melakukan penelitian yang baru selanjutnya dengan menggunakan metode optimasi algoritma klasifikasi yang lainnya.

1.6 Metode Penelitian

Metode penelitian yang digunakan adalah pendekatan kuantitatif dengan menggunakan metode *bagging* yang merupakan bagian dari algoritma penggabungan *machine learning*. Dataset *website phishing* dan non-*phishing* diperoleh dari UCI *Machine Learning Repository*.

1.7 Sistematika Penulisan

a. **BAB I PENDAHULUAN**

Bagian pendahuluan menjelaskan mengenai usulan penelitian tentang latar belakang, rumusan masalah, maksud tujuan penelitian, dan sistematika penulisan skripsi.

b. **BAB II LANDASAN TEORI**

Bagian ini membahas mengenai landasan teori dari berbagai penelitian yang terkait sebelumnya yang digunakan sebagai dasar untuk melakukan penelitian supaya memahami konsep dan teori terhadap permasalahan yang akan diteliti. Bagian ini meliputi kajian pustaka dan dasar teori dari penelitian yang dilakukan.

c. **BAB III METODOLOGI PENELITIAN**

Bagian ini membahas komponen alat dan bahan apa saja yang dibutuhkan dalam penelitian serta langkah-langkah yang dijelaskan pada alur penelitian.

d. **BAB IV HASIL DAN PEMBAHASAN**

Bagian ini membahas mengenai hasil uji coba terhadap metode yang diimplementasikan serta pembahasan mengenai analisis hasil dari temuan penelitian setelah melakukan uji coba data.

e. **BAB V KESIMPULAN DAN SARAN**

Bagian ini membahas mengenai hasil dari penafsiran seluruh penelitian dan juga membahas mengenai saran bagi penelitian yang akan datang tentang kekurangan dari penelitian ini yang nantinya bisa dikembangkan pada penelitian selanjutnya.

f. **DAFTAR PUSTAKA**

Bagian ini berisikan daftar pustaka yang dipakai sebagai acuan referensi literatur dalam penelitian ini.