BAB I
PENDAHULUAN

1.1 Latar Belakang

Indonesia ialah salah satu negara dengan pengguna internet terbesar di dunia. Menurut laporan *We Are Social*, terdapat 204.7 juta pengguna internet per Januari 2022. Jumlah itu naik sekitar 1,03% dibandingkan tahun 2021. Pada januari 2021, jumlah penggunaan internet di Indonesia tercatat sebanyak 202,6 juta pengguna. Jika dibandingkan dengan tahun 2018, jumlah penggunaan internet pada saat ini sudah naik sebesar 54,25%. Sementara itu tingkat penetrasi pengguna internet di Indonesia sekitar 73,7% dari total penduduk pada awal 2022 [1].

Pengguna internet tersebut kebanyakan digunakan untuk aktifitas sosial di dunia maya. Terbukti dengan jumlah penggunaan sosial media yang begitu besar jumlahnya, mencapai 191,4 juta pengguna pada Januari 2022. Jumlah pengguna sosial media di Indonesia pada awal tahun 2022 setara dengan 68,9% dari total populasi yang ada di Indonesia atau sudah lebih dari setengah populasi di Indonesia telah menggunakan sosial media.

Akan tetapi dengan meningkatnya penggunaan internet di Indonesia, ancaman terhadap penggunanya semakin kian beragam. Salah satunya ialah serangan *phishing*, tujuan dari serangan tersebut ialah membuat penggunanya percaya bahwa mereka berinteraksi dengan situs resmi yang dimaksud. Umumnya informasi yang dicari *phiser*(pelaku Phishing) yaitu informasi tentang kartu kredit korban atau pun password akun, dengan cara dialihkan ke web palsu atau dengan mengirimkan email ke korban, dimana korban akan diminta untuk memberikan infromasi pribadinya. Oleh karena itu perlu adanya pendeteksi *website phishing*.

Phishing termasuk salah sati jenis kejahatan siber yang sering terjadi melalui jaringan internet. *Phishing* sering terjadi pada *social media* seperti *email, Instagram, YouTube, Twitter*, dan *Facebook*. Sebagian besar *Phiser* melakukan aksinya dengan bentuk penipuan membuat *link* yang mengarah pada *website* palsu [2]. *Link* tersebut dapat menjerumuskan pengguna *social media* untuk mengklik dan melakukan *download malware* atau

memasukkan data-data pribadi ke *website* palsu yang memiliki tampilan yang sama dengan *website* aslinya.

Seiring maraknya kejadian *Phishing* tentu dapat menimbulkan kerugian dalam hal informasi pribadi, bahkan mungkin terjadi penyalahgunaan (ekspoilitasi) data yang dapat merugikan korban[2]. Berdasarkan APWG (*Anti-Phishing Working Group*) dalam [3] mengatakan dari tahun ke tahun, masyarakat Indonesia semakin sadar tentang adanya *website phishing*. Akan tetapi jumlah *website phishing* yang merugikan tumbuh lebih cepat. Padahal masyarakat Indonesia sering berbelanja *online* ataupun melakukan transaksi secara *online* melalui *e-wallet* atau *m-banking*.

Dilaporkan dari [4] *website phishing* semakin meningkat selama beberapa tahun terakhir, berdasarkan laporan yang diterbitkan APWG (*Anti-Phishing Working Group*). Dalam laporan tersebut, jumlah paling banyak ditemukannya *website phishing* ialah pada bulan Januari 2021. Sebanyak 245.771 situs *phishing* terdeteksi, pakar APWG mencatat jumlah situs *phishing* menurun pada bulan Februari 2021 dengan 158.898 situs *phishing* terdeteksi. Akan tetapi, pada bulan Maret 2021 jumlah situs *phishing* kembali meningkat hingga mencapai 207.208 situs *phishing*, hal tersebut merupakan bulan terburuk keempat dalam sejarah pelaporan APWG.

Selain *website phishing*, APWG juga melaporkan mengenai penipuan lewat *email*, yakni *Business Email Compromise* (BEC). Bisa juga dikenal sebagai upaya tipu daya perusahaan dengan *email* palsu agar mengirimkan sejumlah uang ke rekening bank. APWG mengatakan jumlah rata-rata nilai transfer dalam serangan BEC mencapai US$ 85.000 atau sekitar Rp 1,2 Miliar. Dari laporan APWG menemukan bahwa Namecheap menjadi tempat para *phiser* mendaftarkan domain yang digunakan untuk melakukan penipuan BEC. Perusahaan Namecheap sempat digugat oleh Facebook pada tahun 2020 lalu, dikarenakan dari 46,3% dari semua domain berbahaya yang digunakan untuk serangan BEC terdaftar pada Namecheap.

Dilaporkan dari [5] pada Q1 2022 Indonesia Anti-Phishing Data Exchange (IDADX) melaporkan adanya 3.180 serangan *phishing*. Dari deretan kasus yang ada, lembaga keuangan menjadi sasaran utama serangan *phishing* dan International Revenue Service (IRS) menjadi organisasi yang paling dijadikan sasaran. Pada kuartil pertama bulan Januari menjadi bulan yang paling banyak kasus serangan *phishing* tercatat 1.267 kasus, lalu pada

bulan Februari mengalami penurunan kasus hanya 1.059 kasus, lalu pada bulan Maret mengalami penurunan lagi hanya 1.037 kasus serangan. Dari laporan tersebut lembaga keuangan lah yang menjadi sasaran utama serangan *phishing* tercatat 50% dari total kasus yang terjadi, lalu disusul dengan *e-commerce* 27%, asset *crypto* 11%, media sosial 5%, *Internet Service Provider* (ISP) 5%, dan *gaming* 2%.

Oleh karena itu salah satu cara yang dapat dilakukan yaitu menerapkan klasifikasi untuk mendeteksi *website phising*, dalam penelitian ini peneliti mencoba memberikan gambaran identifikasi *website Phishing* menggunakan algoritma C4.5 kemudian akan dibandingkan dengan algoritma *Support Vector Machine*, dari perbandingan tersebut diharapkan penelitian ini akan memberikan gambaran algoritma mana yang paling efesien dan akurat dalam mengidentifikasi *website phishing*.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dikemukakan, perumusan masalah pada penelitian ini adalah bagaimana tingkat akurasi, presisi, recall dan skor F1 dari penggunaan algoritma C4.5 dan *Support Vector Machine* dalam mengindentifikasi *website phishing*.

1.3 Batasan Masalah

Untuk membatasi luasnya cakupan ruang lingkup yang akan dibahas dari pokok pembahasan dan permasalahan yang akan diselesaikan, maka permasalahannya dibatasi sebagai berikut :

1. *Dataset* yang digunakan *Phishing Websites Dataset* berasal dari UCI Machine Learning Repository.
2. *Dataset* yang digunakan sebanyak 11055 data.
3. Seleksi fitur yang digunakan adalah *Information Gain*.
4. Algoritma yang digunakan adalah algoritma C4.5 dan SVM.

1.4 Tujuan Penelitian

Tujuan yang diinginkan pada penelitian ini adalah mengetahui tingkat akurasi, presisi, recall dan skor F1 dari algoritma C4.5 dan *Support Vector Machine* dalam melakukan klasifikasi *website phishing*.

1.5 Manfaat Penelitian

Penelitian ini memiliki beberapa manfaat baik secara teoritis maupun secara praktis. Adapun manfaat dari penelitian ini ialah sebagai berikut :

1. Manfaat Teoritis

   Manfaat teoritis yang diharapkan dari penelitian ini adalah dapat menambah ilmu pengetahuan maupun menjadi bahan refrensi yang berkaitan tentang *website phishing*. Selain itu, penelitian ini juga diharapkan dapat menjadi salah satu refrensi agar dilakukannya penelitian lain secara mendalam tentang mengidentifikasi *website phishing*.

2. Manfaat Praktis

   a. Bagi Akademis

      Penelitian ini diharapkan mampu menjadi salah satu sumber refrensi untuk menambah kajian penelitian baru bagi dunia akademis serta sebagai bahan pertimbangan untuk penelitian lebih lanjut khususnya identifikasi *website phishing*.

   b. Bagi Peneliti

      Penelitian ini diharapkan dapat menjadi sarana untuk mengaplikasikan teori serta ilmu pengetahuan yang diperoleh selama masa perkuliahan. Peneliti juga berharap hasil penelitian ini, dapat memperdalam wawasan mengenai identifikasi *website phishing*.

1.6 Sistematika Penulisan

Gambaran secara garis besar mengenai hal-hal yang akan dibahas dalam penulisan skripsi ini

BAB I PENDAHULUAN

Dalam bab ini akan membahas tentang latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan.

BAB II LANDASAN TEORI

Dalam bab ini akan memaparkan tentang tinjauan pustakan dan landasan teori yang menjelaskan dasar-dasar teori yang nantinya akan digunakan maupun diterapkan pada penulisan penelitian.

BAB III METODELOGI PENELITIAN

Dalam bab ini memaparkan tentang langkah-langkah yang dilakukan dalam proses penelitian.

BAB IV HASIL DAN PEMBAHASAN

Dalam bab ini berisi tentang hasil penelitian dan pembahasan yang telah dilakukan.

BAB V PENUTUP

Dalam bab ini memberikan penjelasan mengenai kesimpulan dan saran yang diharapkan berguna untuk penelitian selanjutnya.