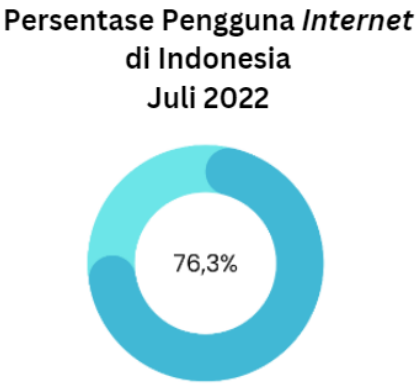
# BAB I
# PENDAHULUAN

## 1.1 Latar Belakang

*Internet* telah menjadi kebutuhan penting bagi masyarakat saat ini dalam mendukung segala aktivitas sehari-hari. *Internet* menjadi salah satu bentuk inovasi dari kemajuan teknologi informasi yang menciptakan transformasi atau perubahan sosial dalam masyarakat (Yenni et al., 2021). Hal tersebut didukung oleh pergeseran aktivitas yang dilakukan secara *online* sebagai akibat dari pandemi COVID-19 menjadi salah satu faktor penyebab meningkatnya jumlah pengguna *internet* dan aktivitas *online* (Nur Latifah et al., 2022).

Data jumlah pengguna *internet* di Indonesia yang ditampilkan oleh *Internetworldstats* pada bulan Juli 2022 menunjukkan bahwa dari populasi Indonesia yang mencapai 278.268.685, sebanyak 212.354.070 orang merupakan pengguna *internet* (*Internet World State*, 2022). Pada Gambar 1.1 merupakan grafik persentase pengguna *internet* pada bulan Juli 2022. Data tersebut membuktikan bahwa sebagian besar aktivitas masyarakat telah bergeser ke dunia *internet,* termasuk dalam aspek komunikasi, pencarian *internet*, pendidikan, pekerjaan, dan transaksi bisnis (Yenni et al., 2021).




Gambar 1. 1 Grafik Lingkaran Persentase Pengguna *Internet* di Indonesia

Sumber: *Internet World State*, 2022

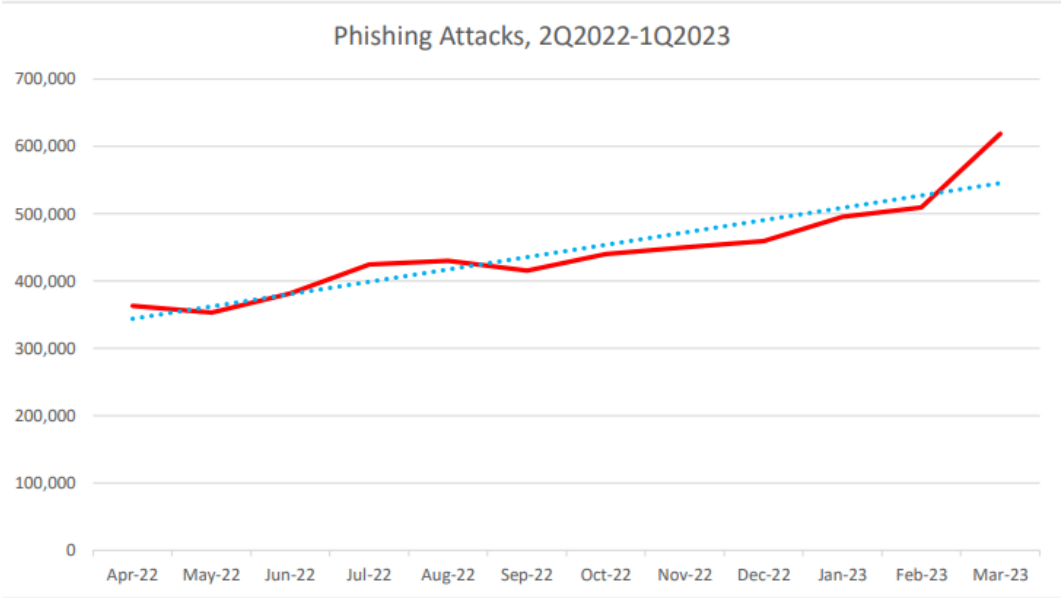
Meskipun *internet* telah banyak memberikan kemudahan bagi manusia, namun teknologi informasi memiliki sifat yang kontradiktif, karena selain berperan dalam kemajuan zaman dan kesejahteraan manusia, teknologi informasi juga memberikan

peluang untuk melanggar hukum. Perkembangan teknologi *internet* memiliki sisi positif, yaitu sebagai bentuk dari kreativitas manusia, tetapi tidak bisa dipungkiri bahwa dengan adanya *internet* dan pengguna *internet* yang massif telah memicu peningkatan tindak kejahatan siber (*cyber-crime*) atau tindakan ilegal melalui jaringan *internet*. Badan Siber dan Sandi Negara (BSSN) melaporkan dalam Lanskap Keamanan Siber 2023, bahwa berdasarkan laporan yang diterima dari *stakeholder* pada layanan aduan siber, kategori aduan terbanyak adalah kejahatan siber sebesar 86%. Kemudian, selama tahun 2023 total trafik anomali di Indonesia adalah 403.990.813 anomali (BSSN, 2024). Adapun bentuk kejahatan di dunia maya yang dimungkinkan oleh teknologi *internet* salah satunya adalah ancaman *phishing* (Gani, 2014; Kharisma Putra et al., 2023).

Tindakan *phishing* diatur dalam Undang-Undang Informasi dan Transaksi Elektronik (UU ITE) yaitu Undang-Undang Nomor 19 Tahun 2016 Tentang Perubahan atas Undang-Undang Nomor 17 Tahun 2008 yang mengatur segala aktivitas masyarakat di dunia siber (Ayu et al., 2023). *Phishing* merupakan tindakan yang melanggar hukum karena melibatkan upaya mencuri informasi akun atau data pribadi seseorang yang bersifat sensitif melalui *website* ataupun *e-mail* tiruan suatu institusi sehingga berpotensi merugikan pihak yang terkena dampak dalam hal finansial (Yuniati & Kresna A., 2020).

Berdasarkan data penanganan insiden BSSN dalam laporan Lanskap Keamanan Siber 2023, telah disampaikan bahwa *phishing* termasuk ke dalam 10 besar insiden yang terjadi pada tahun 2023. Selanjutnya, berdasarkan peningkatan jumlah pada beberapa tahun terakhir, *phishing* diprediksi akan banyak terjadi dan berpotensi menjadi salah satu ancaman siber yang dapat menjadi serangan tingkat lanjut pada tahun 2024. Pada Gambar 1.2 menunjukkan tren peningkatan serangan *phishing* pada kuartal kedua 2022 hingga kuartal kesatu 2023. Serangan *phishing* sering digunakan sebagai metode atau langkah awal untuk serangan yang lebih kompleks seperti *generic* trojan RAT, *ransomware*, *malware*, dan serangan lainnya melalui tautan suatu situs yang tidak sah yang merupakan bagian dari trafik anomali tertinggi berdasarkan BSSN dalam laporan Lanskap Keamanan Siber 2023 (BSSN, 2024). Dengan demikian, adanya peningkatan *phishing* secara tidak langsung dapat berpotensi pula meningkatkan lebih banyak atau menyumbangkan angka pada

beberapa jenis kejahatan siber tersebut. Oleh karena itu, *phishing* merupakan masalah yang penting untuk diatasi.




Gambar 1. 2 Peningkatan *Phishing*

Sumber: *Anti Phishing Working* Group, 2023

Penelitian yang dilakukan oleh *Anti Phishing Working Group* (APWG) yang dimuat dalam laporan yang berjudul "*Phishing Activity Trends Report*" mengungkapkan bahwa pada Januari hingga Maret 2023 merupakan kuartal terburuk untuk *phishing*, yaitu terdapat 1.624.144 total serangan *phishing* yang tercatat. Pada kuartal kesatu 2023, APWG melaporkan bahwa serangan *phishing* paling banyak menyasar pada industri sektor finansial dengan proporsi 23,5% (*Anti Phishing Working Group*, 2023).

Berdasarkan permasalahan *phishing* yang semakin umum dan penting diatasi, terdapat alternatif solusi yang telah dilakukan oleh beberapa penelitian sebelumnya, yaitu berupa deteksi pada suatu URL menggunakan metode *machine learning*. Klasifikasi *malicious website* menggunakan algoritma *K-Nearest Neighbor* berdasarkan *application layers* dan *network characteristic* telah dilakukan oleh Green Arter. S, dkk (Sandag et al., 2018). Algoritma pelatihan model *K-Nearest Neighbor* dalam memprediksi *malicious website* menghasilkan skor akurasi sebesar 93,61% setelah dilakukan 10-*fold cross validation.* Lebih lanjut, penelitian lain yang telah dilakukan Abdi Bimantara dan Tiara Annisa. D (Bimantara & Dina, 2019), yaitu mengenai klasifikasi *web* berbahaya menggunakan algoritma *logistic*

*regression.* Data yang digunakan sebanyak 1000 baris dengan 10 variabel prediksi. Hasil akurasi algoritma pelatihan model *logistic regression* yang didapatkan cukup tinggi yaitu sebesar 94%. Kemudian, penerapan algoritma *logistic regression* pada klasifikasi pesan SMS telah dilakukan oleh Fitran D. Pramakrisna, dkk (Pramakrisna et al., 2022). Algoritma pelatihan model *logistic regression* telah berhasil mengklasifikasikan pesan SMS dengan skor akurasi 97%. Penelitian P. Amba Bhavani, dkk (Madhumitha et al., 2022) menganalisis beberapa metode *machine learning* untuk mendeteksi *link phishing,* yaitu menggunakan algoritma CNN LSTM & CNN Bi-LSTM, *Logistic regression*, dan XGBoost. Algoritma dilatih menggunakan sebanyak 25 ribu lebih *dataset* dengan 50 variabel prediksi. Hasil penelitian ini diperoleh bahwa algoritma *logistic regression* memiliki tingkat akurasi tertinggi kedua, yaitu sebesar 91,98%.

Dalam menghadapi permasalahan *phishing* yang terjadi, berbagai macam algoritma telah digunakan berdasarkan beberapa penelitian terdahulu terkait mendeteksi tautan *phishing* sehingga penelitian ini penting seiring meningkatnya ancaman *phishing*. Penelitian ini terinspirasi dari penelitian sebelumnya, yaitu menerapkan algoritma *logistic regression* dalam klasifikasi *web* berbahaya yang dapat menghasilkan akurasi tinggi seperti penelitian yang dilakukan oleh (Pramakrisna et al., 2022), (Bimantara & Dina, 2019), dan (Madhumitha et al., 2022). Namun, pada penelitian (Bimantara & Dina, 2019) *dataset* yang digunakan untuk melatih algoritma *logistic regression* sebanyak 1000 baris data dan memiliki lebih dari satu variabel prediksi. Penelitian (Madhumitha et al., 2022) menggunakan sekitar 25.000 *dataset* dan 50 variabel prediksi. Perbedaan dengan penelitian (Pramakrisna et al., 2022) adalah menerapkan algoritma *logistic regression* dalam prediksi *link phishing* yang mana pada penelitian (Pramakrisna et al., 2022) menerapkan algoritma *logistic regression* dalam klasifikasi SMS *spam* dan tidak *spam*. Oleh karena itu, Penelitian ini melengkapi penelitian (Madhumitha et al., 2022) dan (Bimantara & Dina, 2019), yaitu menggunakan *Framework AI Project Cycle* dalam membuat sistem deteksi *link phishing*, karena langkahnya yang dapat menyesuaikan dari penggunanya serta menggunakan *dataset* dengan jumlah variabel prediktor yang beragam, yaitu kumpulan URL untuk melatih algoritma klasifikasi sehingga menghasilkan model algoritma yang dapat mendeteksi URL

dengan tepat dan akurasi yang lebih baik. Dengan demikian, penelitian ini bertujuan membuat suatu sistem berbasis aplikasi *website* dengan menerapkan algoritma *logistic regression* dalam konteks deteksi *link phishing* menggunakan *Flask Framework* sehingga diharapkan dapat melindungi pengguna *internet* dari serangan *phishing* dan dapat mengurangi resiko serangan kelanjutan. *Website* aplikasi deteksi *link phishing* pada penelitian ini berbeda dengan produk yang sudah ada, yaitu dalam segi penggunaan algoritma menggunakan *logistic regression*, penggunaan *dataset* dari sumber yang berbeda, variabel prediktor yang berbeda dengan jumlah yang berbeda pula, sehingga dapat menghasilkan *website* aplikasi deteksi *link phishing* yang bisa mengenali potensi URL *phishing* dengan baik.

## 1.2 Rumusan Masalah dan Identifikasi Masalah

Berdasarkan pada latar belakang yang telah diuraikan, terdapat rumusan masalah sebagai berikut:

1. Bagaimana implementasi algoritma *logistic regression* dalam mendeteksi *link phishing*?
2. Berapa tingkat akurasi algoritma *logistic regression* yang dibuat dalam mendeteksi *link phishing* menggunakan jumlah variabel prediktor yang beragam?

## 1.3 Batasan Masalah

Agar pembahasan tetap terfokus pada inti permasalahan maka lingkup permasalahan yang akan dibahas akan dibatasi sebagai berikut:

1. Model deteksi *link phishing* dibangun menggunakan algoritma *logistic regression*.
2. Model deteksi *link phishing* menggunakan data URL untuk melatih model *logistic regression* dalam memprediksi *link phishing*.
3. Pada penelitian ini tidak ada pengujian *website* aplikasi deteksi *link phishing*, penelitian fokus melakukan eksperimen pengembangan model sehingga menghasilkan model yang paling akurat untuk deteksi *link phishing*.
4. Model deteksi *link phishing* dibangun menggunakan bahasa pemrograman *Python*.

## 1.4 Tujuan Penelitian

Berdasarkan rumusan permasalahan diatas, berikut merupakan tujuan dari penelitian yang dilakukan:

1. Mengetahui implementasi algoritma *logistic regression* untuk deteksi *link phishing*.
2. Mengetahui tingkat akurasi algoritma *logistic regression* yang dibuat dalam mendeteksi *link phishing* menggunakan jumlah variabel prediktor yang beragam.

## 1.5 Manfaat Penelitian

Penelitian ini diharapkan memberikan manfaat sebagai berikut:

1. Manfaat Teoritis

   Mengaplikasikan ilmu yang telah diperoleh dan dapat dijadikan dasar untuk melakukan penelitian dalam bidang *Natural Language Processing* khususnya menerapkan algoritma *logistic regression*.

2. Manfaat Praktis

   Membantu memberikan informasi kepada pengguna *internet* dalam meningkatkan kesadaran keamanan siber serta mendorong praktik keamanan yang lebih baik.

## 1.6 Struktur Organisasi Skripsi

Struktur organisasi skripsi memberikan gambaran mengenai sistematika penulisan pada setiap bab yang mengacu pada Peraturan Rektor UPI No. 7867/UN40/HK/2021 tentang Pedoman Penulisan Karya Ilmiah UPI Tahun 2021. Pedoman ini menyusun penulisan skripsi ke dalam lima bab, yaitu: pendahuluan; tinjauan pustaka; metodologi penelitian; hasil dan pembahasan; serta kesimpulan, implikasi, dan rekomendasi, yang dijelaskan secara rinci sebagai berikut:

1. BAB I PENDAHULUAN

   Pada bab I dalam penelitian ini mencakup latar belakang, rumusan masalah dan identifikasi masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan struktur organisasi skripsi.

2. BAB II KAJIAN PUSTAKA

   Pada bab II dalam penelitian ini berisi ulasan literatur dengan topik penelitian, teori-teori yang mendasari, serta penelitian terdahulu yang berhubungan.

Literatur tersebut dijadikan sebagai acuan penelitian, seperti *Phishing*, *Uniform Resource Locator*, Deteksi *Link Phishing*, Algoritma *Logistic Regression*, *Python*, *Jupyter Notebook*, *Google Colaboratory*, *Web*, *Hypertext Markup Language*, *Cascading Style Sheets*, *JavaScript*, Flask *Framework*, Visual Studio Code, dan Ngrok.

3. BAB III METODE PENELITIAN

   Pada bab III dalam penelitian ini menguraikan metode dan teknik yang digunakan dalam penelitian, mencakup jenis penelitian, desain penelitian, prosedur penelitian dengan menerapkan *Framework AI Project Cycle* yang diawali dengan studi literatur, *Problem Scoping*, *Data Acquisition*, *Data Exploration*, *Modelling*, *Evaluation*, dan *Deployment*. Selain itu, pada bab III juga mencakup lingkungan komputasi, teknik pengumpulan data, serta teknik analisis data.

4. BAB IV TEMUAN DAN PEMBAHASAN

   Pada bab IV dalam penelitian ini menyajikan temuan-temuan berupa hasil penelitian serta pembahasannya. Temuan penelitian yang diperoleh pada setiap tahapan dipaparkan sesuai dengan tahapan yang ada pada prosedur penelitian mulai dari temuan pada tahap studi literatur, *Problem Scoping*, *Data Acquisition*, *Data Exploration*, *Modelling*, *Evaluation*, hingga *Deployment*.

5. BAB V SIMPULAN, IMPLIKASI, DAN REKOMENDASI

   Pada bab V dalam penelitian ini menguraikan kesimpulan dari hasil penelitian, implikasi, dan rekomendasi berupa saran-saran penelitian yang konstruktif untuk pengembangan penelitian selanjutnya.