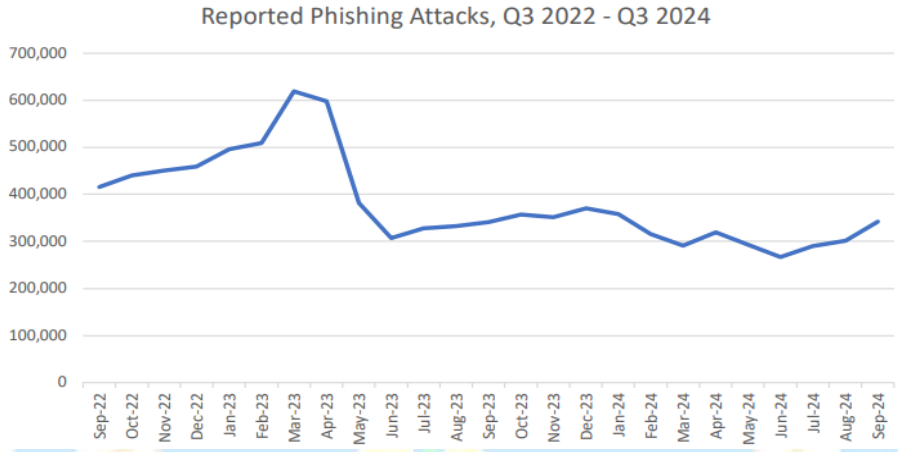
BAB I
PENDAHULUAN

1.1 Latar Belakang

Perkembangan teknologi informasi yang semakin pesat telah mengubah berbagai aspek kehidupan, salah satunya adalah pemanfaatan internet sebagai bagian dari transformasi digital. Dewasa ini, ketergantungan manusia akan internet semakin meningkat setiap tahunnya, karena memberikan banyak kemudahan di dalam berbagai aktivitas, seperti komunikasi dan transaksi. Sebagai media informasi, internet berguna untuk mencari informasi yang terbaru, tetapi tidak jarang dimanfaatkan oleh pelaku kejahatan siber untuk mencuri informasi pribadi pengguna melalui serangan *phishing* (Windarni et al., 2023). *Phishing* merupakan serangan berbasis rekayasa sosial (*social engineering*) yang memungkinkan penjahat siber untuk mencuri informasi kredensial, menyebarkan *ransomware*, hingga melakukan penipuan termasuk pencurian finansial. Selain itu, serangan ini juga memungkinkan pelaku untuk mendapatkan akses strategis ke lingkungan target. Dengan menggunakan situs web palsu yang dirancang dengan baik, *phishing* digunakan untuk memperoleh informasi sensitif yang bersifat pribadi, seperti nomor rekening dan kata sandi dari pengguna yang tidak waspada (Yerima & Alzaylaee, 2020).

Menurut laporan Anti-Phishing Working Group (APWG), serangan *phishing* mengalami peningkatan signifikan selama tahun 2024, dengan lonjakan terbesar pada kuartal ketiga yang diamati sebanyak 932.923 serangan *phishing*, meningkat dari 877.536 pada kuartal kedua. Tren baru yang diamati termasuk penggunaan email yang dipersonalisasi dengan menyertakan gambar Google Street View dari rumah target untuk meningkatkan kredibilitas serangan. *Platform* media sosial kembali menjadi sektor yang paling sering diserang mencakup 30,5% dari total serangan *phishing*, di antaranya X, Facebook, hingga *platform* khusus untuk mencari kerja seperti LinkedIn yang menyumbang sekitar 47% dari keseluruhan serangan *phishing* di *platform* media sosial. Selain itu, *smishing*, serangan *phishing*

yang dilakukan melalui SMS atau pesan teks meningkat lebih dari 22% selama periode tersebut. Untuk serangan berbasis email, akun Gmail juga digunakan paling banyak dalam persentase 83,1% dari semua penipuan Business Email Compromise (BEC), menyoroti risiko yang signifikan terhadap pengguna layanan email berbasis *cloud* (APWG, 2024; Tanti, 2024).

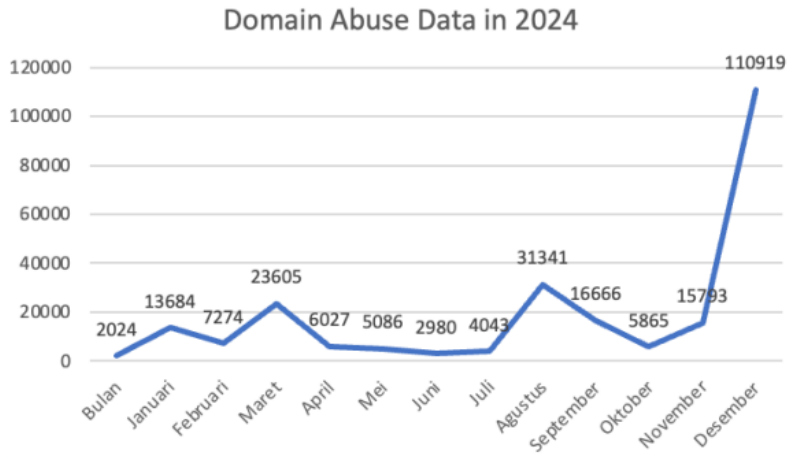


Gambar 1.1 Jumlah Serangan *Phishing* pada Q3 2022 - Q3 2024

(Sumber: APWG, 2024)

Dalam skala nasional, serangan *phishing* juga menjadi masalah yang tidak kalah serius. Berdasarkan laporan Indonesia Domain Abuse Data Exchange (IDADX) pada kuartal keempat tahun 2024, terdapat sebanyak 132.577 serangan dalam bentuk *domain abuse* yang merujuk kepada aktivitas kriminal berdasarkan nama domain pada berbagai tautan (*links*), terdiri dari *phishing*, *malware, spam*, hingga aktivitas berbahaya lainnya seperti SARA. Jumlah serangan pada periode tersebut mengalami peningkatan yang cukup signifikan dengan peningkatan terbesar pada bulan Desember sebanyak 110.919 serangan, jauh melesat dibandingkan dengan bulan sebelumnya sekitar 15.793 serangan. Dalam hal ini, serangan *domain abuse* dalam kategori *phishing* dilaporkan menjadi yang terbanyak di antara kategori lainnya, dengan jumlah sebesar 85.414 serangan. Selain itu, berdasarkan jenis sektor industri yang menjadi target serangan *domain abuse*, sektor finansial mendominasi dengan persentase sebesar 41,70% dari total serangan, diikuti oleh

jejaring sosial 39,86%, *e-commerce* 13,06%, *gaming* 1,83%, dan *cryptocurrencies* 1,15% (IDADX, 2024).



Gambar 1.2 Data *Domain Abuse* Sepanjang Tahun 2024

(Sumber: IDADX, 2024)

Salah satu bentuk serangan *phishing* yang paling umum adalah yang berbasis URL, di mana pelaku menyamarkan tautan palsu agar terlihat mirip dengan tautan resmi yang sah. Dengan memanfaatkan URL palsu ini, penjahat siber dapat mengelabui pengguna untuk mencuri informasi pribadi atau kredensial milik mereka. Pendekatan tradisional dalam mendeteksi URL *phishing* seperti daftar hitam (*blacklist*) memiliki keterbatasan signifikan, karena tidak dapat mendeteksi URL baru yang belum tercantum dalam *database* yang memerlukan pembaruan terus menerus. Selain itu, pendekatan ini memiliki tingkat kesalahan yang cukup tinggi ketika berhadapan dengan URL *phishing* yang dirancang dengan baik (Prabakaran et al., 2023).

Pendekatan untuk mengidentifikasi URL *phishing* kerap memerlukan analisis manual seperti daftar (*blacklist* dan *whitelist*), heuristik, dan kesamaan visual (*visual similarity*), menyebabkan kesulitan untuk mengikuti strategi penyerang yang terus berubah (Do et al., 2022). Dengan begitu, pendekatan menggunakan implementasi *Artificial Intelligence* (AI) secara otomatis dan *real-time* diperlukan untuk mengidentifikasi serangan *phishing* berbasis URL, seperti dengan pembelajaran mesin atau *machine learning* misalnya. *Machine learning* dapat

belajar secara otomatis dari data pelatihan sehingga banyak peneliti sudah menggunakannya untuk deteksi *phishing*, tetapi memiliki kekurangan berupa rekayasa fitur (*feature engineering*) secara manual yang dapat menyebabkan beberapa fitur penting yang bisa dikenali oleh mesin terlewatkan. Untuk mengatasi kekurangan dari metode *machine learning* dalam mendeteksi serangan *phishing* berbasis URL, banyak peneliti memperkenalkan metode deteksi menggunakan cabang dari pembelajaran mesin, yaitu pembelajaran mendalam atau *deep learning* dikarenakan kemampuannya dalam melakukan ekstraksi fitur secara otomatis (Sasi & Balakrishnan, 2024).

Pembelajaran mendalam (*deep learning*) merupakan cabang dari pembelajaran mesin (*machine learning*) yang menghadirkan metode yang lebih maju dengan kemampuan untuk mengekstrak pola kompleks dari data besar secara otomatis. Tidak seperti pembelajaran mesin yang memanfaatkan arsitektur jaringan saraf yang berlapis-lapis untuk mempelajari representasi data secara hierarkis. Dalam konteks deteksi *phishing*, *deep learning* telah membuktikan kemampuannya dalam mengatasi keterbatasan metode tradisional dengan menghasilkan akurasi yang lebih tinggi dan mampu menangani data yang lebih beragam dan dinamis (Alshingiti et al., 2023).

Dalam penerapannya, terdapat banyak model *deep learning* yang dikembangkan untuk mendeteksi serangan *phishing* berbasis URL. Salah satu pendekatan yang banyak digunakan adalah pembelajaran terawasi (*supervised learning*), di mana model dilatih menggunakan data berlabel. Meskipun telah berhasil dikembangkan, pendekatan *supervised learning* yang berfokus pada meminimalkan kesalahan klasifikasi dengan bergantung pada banyak data pengamatan dan serangan yang telah diketahui, masih memiliki keterbatasan dalam mendeteksi serangan *phishing*. Salah satu tantangan utamanya adalah serangan *phishing* tipe *zero-day*, di mana URL *phishing* dibuat dan segera dihapus setelah informasi korban berhasil dicuri. Selain itu, ketika URL tidak seimbang (*imbalanced*), model cenderung mengalami kebingungan atau kesalahan klasifikasi (Bu & Cho, 2021).

Untuk mengatasi keterbatasan tersebut, pendekatan melalui pembelajaran tidak terawasi (*unsupervised learning*) berbasis deteksi anomali (*anomaly detection*) menjadi alternatif yang menjanjikan. Salah satu penelitian yang diusulkan oleh Bu & Cho (2021) menerapkan pendekatan berbasis *Convolutional Autoencoder* (CAE) yang dilatih hanya menggunakan URL yang normal dengan memanfaatkan karakteristiknya dalam merekonstruksi data yang belum pernah dilihat sebelumnya. Pendekatan lain juga diperkenalkan oleh Pandian (2024) melalui *framework Variational Autoencoder* dan *Convolutional Neural Network* (VAE-CNN) untuk ancaman keamanan siber, di mana VAE memiliki keunggulan dalam kompresi data, sedangkan CNN unggul dalam pengenalan pola dan deteksi anomali. Keunggulan ini dibuktikan melalui penelitian terkait deteksi *phishing* berbasis *deep learning* yang dilakukan menggunakan *Convolutional Neural Network* (CNN), berhasil memperoleh presisi sebesar 98%, *recall* sebesar 97%, dan *F1-score* sebesar 98% karena kemampuannya dalam mengekstrak fitur lokal dari setiap URL (Asiri et al., 2024). Selain itu, deteksi menggunakan model *Variational Autoencoder* (VAE) yang dikombinasikan dengan *Transformer* dalam arsitektur *Generative Adversarial Network* (GAN) memperoleh presisi sebesar 96,40%, *recall* sebesar 96,44%, dan *F1-score* sebesar 97,77% sebab mampu menangkap representasi laten dari setiap URL (Sasi & Balakrishnan, 2024).

Berdasarkan pemaparan di atas, maka penelitian yang diusulkan berjudul "Deteksi Serangan *Phishing* berbasis URL dengan *Variational Autoencoder* dan *Convolutional Neural Network* (VAE-CNN)", bertujuan untuk mengimplementasikan model *hybrid* dari algoritma *Variational Autoencoder* dan *Convolutional Neural Network* (VAE-CNN) dalam mendeteksi serangan *phishing* berbasis URL melalui deteksi anomali. Data yang akan digunakan di dalam penelitian, di antaranya URL *legitimate* untuk situs yang terbukti sah dan tervalidasi dan *phishing* untuk situs yang terdeteksi mengandung muatan *phishing*. Selain itu, penelitian ini diharapkan mampu mengurangi ketergantungan pada metode yang memerlukan analisis manual dan memberi peluang terhadap pengembangan lebih lanjut dalam sistem keamanan siber berbasis pembelajaran mendalam atau *deep learning* secara tidak terawasi (*unsupervised*).

1.2 Perumusan Masalah

Berdasarkan pemaparan latar belakang di atas, rumusan masalah yang diperoleh dalam penelitian ini adalah, "Bagaimana performa model *hybrid deep learning Variational Autoencoder* dan *Convolutional Neural Network* (VAE-CNN) dalam mendeteksi serangan *phishing* berbasis URL?"

1.3 Batasan Masalah

Agar penelitian lebih terarah dan berorientasi pada permasalahan utama serta tidak keluar dari ruang lingkup yang telah ditetapkan, batasan masalah yang dirumuskan adalah sebagai berikut:

1. Penelitian ini bertujuan untuk mengetahui performa model *hybrid deep learning*, yaitu *Variational Autoencoder* dan *Convolutional Neural Network* (VAE-CNN) dalam mendeteksi serangan *phishing* berbasis URL melalui pendekatan deteksi anomali (*anomaly detection*).
2. Data yang digunakan dalam penelitian terdiri dari URL berbagai situs web (*websites*). Sumber data diperoleh dari *platform* yang tersedia secara publik dan relevan dengan penelitian, seperti URL *legitimate* diambil dari PhishDataset di GitHub dan URL *phishing* dari media sosial X serta URLScan.io Phishing URL Feed.
3. Data yang digunakan untuk pelatihan model merupakan URL *legitimate* (sah). Meskipun data tetap dilabeli untuk keperluan manajemen dan segmentasi data, model VAE-CNN akan mempelajari pola URL *legitimate* yang normal melalui pendekatan *unsupervised learning*, di mana label tidak digunakan selama pelatihan. Selama pelatihan, model tidak memerlukan label karena fokus pada pembelajaran pola normal dan deteksi anomali berbasis *reconstruction loss* menggunakan nilai ambang batas (*threshold*).
4. Data pengujian mencakup campuran URL *legitimate* dan *phishing*. Label pada data pengujian hanya digunakan untuk evaluasi kinerja model, seperti menghitung metrik presisi, *recall*, dan *F1-score*.

5. Data yang digunakan yaitu sebanyak 10.000 URL, dengan segmentasi data yang digunakan untuk pengembangan model, yakni untuk pelatihan dengan jumlah data sebanyak 7.000 URL *legitimate* tanpa label *phishing*, kemudian untuk pengujian dengan jumlah data sebanyak 3.000 URL yang terdiri dari 1.500 URL *legitimate* dan URL *phishing*.

1.4 Tujuan Penelitian

Tujuan utama yang ingin dicapai dari penelitian ini adalah untuk menguji dan mengukur performa dari model *hybrid deep learning Variational Autoencoder* dan *Convolutional Neural Network* (VAE-CNN) dalam mendeteksi serangan *phishing* berbasis URL.

1.5 Manfaat Penelitian

Adapun manfaat yang dapat diperoleh dari penelitian ini, antara lain:

1. Bagi peneliti dan akademisi, penelitian ini memberikan wawasan baru, baik secara teoritis maupun praktis melalui implementasi model *hybrid deep learning* VAE-CNN dalam mendeteksi serangan *phishing* berbasis URL melalui pendekatan deteksi anomali (*anomaly detection*) secara tidak terawasi (*unsupervised*) sekaligus menguji dan mengevaluasi performa yang diberikan, serta memberikan kontribusi dalam pengembangan metode deteksi serangan *phishing* berbasis URL menggunakan pendekatan berbasis *deep learning* seperti VAE-CNN yang dapat menjadi dasar bagi penelitian selanjutnya.
2. Bagi pembaca dan masyarakat secara umum, penelitian ini ditujukan sebagai urgensi untuk meningkatkan kesadaran akan ancaman keamanan siber (*cybersecurity threats*) selama berselancar di internet ketika mengakses berbagai situs web dan melakukan berbagai aktivitas di dalamnya, sekaligus sebagai langkah preventif dalam mencegah dampak yang ditimbulkan oleh serangan *phishing*.

1.6 Sistematika Penulisan

Penulisan merupakan tahapan krusial dalam penyusunan skripsi. Berikut ini merupakan perincian tahapan penulisan yang dilakukan secara sistematis:

BAB I PENDAHULUAN

Bab ini berisi pemaparan mengenai latar belakang penelitian yang dilakukan, perumusan masalah, batasan masalah, tujuan dan manfaat penelitian, serta sistematika penulisan.

BAB II KAJIAN LITERATUR

Bab ini menjelaskan terkait kajian literatur berupa penelitian terdahulu yang relevan dengan topik penelitian yang dilakukan dan landasan teori sebagai dasar penelitian secara teoritis.

BAB III METODOLOGI PENELITIAN

Bab ini membahas tentang perancangan penelitian yang dilakukan secara seksama.

BAB IV HASIL DAN PEMBAHASAN

Bab ini berisi pembahasan mengenai penelitian yang dilakukan, meliputi lingkungan uji coba yang digunakan di dalam penelitian, penjelasan rinci terkait proses yang terjadi secara bertahap hingga hasil yang diperoleh setelah penelitian dilaksanakan.

BAB V PENUTUP

Bab ini berisi kesimpulan dan saran yang diperoleh dari penelitian yang telah dilakukan.

DAFTAR PUSTAKA

Bab ini memuat daftar referensi yang digunakan sebagai acuan dalam proses penelitian.