BAB I
PENDAHULUAN

1.1 Latar Belakang

Di tengah dominasi media sosial sebagai sarana komunikasi digital saat ini, *platform* X (Twitter) menempati posisi yang cukup unik karena pendekatannya yang berbeda terhadap kebebasan konten. Pada Juni 2024, *platform* ini secara resmi memperbarui kebijakannya untuk mengizinkan peredaran konten dewasa yang diproduksi secara konsensual. Keputusan ini secara tidak langsung mengubah lanskap *platform* tersebut menjadi lebih terbuka, namun sekaligus menciptakan celah keamanan baru. Keterbukaan ini sering kali disalahgunakan oleh pelaku kejahatan siber yang melihat peluang emas untuk menyusupkan ancaman di balik rasa penasaran pengguna melalui manipulasi struktur URL (*URL obfuscation*) untuk menyembunyikan identitas situs asli yang berbahaya [1].

Fenomena ini menjadi sangat mengkhawatirkan, khususnya di Indonesia, di mana kemudahan akses media sosial kini menjangkau hampir seluruh lapisan masyarakat tanpa memandang batasan usia; mulai dari anak-anak usia Sekolah Dasar (SD) hingga orang dewasa. Prevalensi konten dewasa di media sosial dianggap sangat berbahaya bagi nilai moral generasi muda, sehingga penyaringan konten menjadi hal yang esensial **[2]**. Tingginya rasa penasaran dari berbagai kalangan umur ini dieksploitasi secara psikologis oleh pelaku kejahatan siber sebagai titik lemah untuk menyebarkan ancaman.

Eskalasi penyebaran konten berbahaya di *platform* X tidak terlepas dari peran agen otomatis atau bot. Penelitian menunjukkan bahwa bot yang berbahaya (*malicious bots*) di media sosial memiliki peran signifikan dalam menyebarkan tautan penipuan dan disinformasi **[3]**. Bot sering kali memanfaatkan pemendekan URL (*URL shortening*), seperti penggunaan domain t.co, untuk mendistribusikan ancaman secara masif. Hal ini menciptakan lapisan ancaman ganda: konten visual digunakan sebagai penarik perhatian untuk memancing klik dari pengguna segala

umur, sementara manipulasi URL teknis digunakan untuk mengelabui sistem keamanan, menjadikan deteksi manual sangat sulit dilakukan **[3]**.

Di sisi lain, upaya moderasi terhadap konten sensitif sebenarnya telah menjadi perhatian peneliti terdahulu. Studi telah berhasil menerapkan metode *Deep Learning* untuk mengklasifikasikan teks konten dewasa dan non-dewasa secara akurat di media sosial **[4]**, termasuk secara spesifik pada cuitan berbahasa Indonesia **[2]**. Namun, mayoritas penelitian ini masih berfokus pada klasifikasi konten teks semata. Belum banyak penelitian yang secara spesifik mengintegrasikan deteksi konten dewasa tersebut dengan analisis risiko tautan (*link*) yang menyertainya, padahal tautan tersebut sering kali merupakan pintu masuk utama serangan siber.

Guna mengatasi kompleksitas serangan yang memadukan rekayasa sosial konten dewasa dengan manipulasi teknis tersebut, diperlukan metode deteksi yang adaptif. *Phishing* sendiri merupakan teknik penipuan yang menyamar sebagai entitas terpercaya, dan pendekatan berbasis *Machine Learning* terbukti menjadi solusi yang efektif untuk mengenali pola situs *phishing* yang dinamis **[5]**. Algoritma seperti *Decision Tree* telah berhasil diimplementasikan untuk mengoptimisasi sistem deteksi *phishing* **[6]**. Lebih lanjut, penelitian komparatif membuktikan bahwa algoritma XGBoost mampu mencapai tingkat akurasi yang lebih tinggi dan optimal dalam membedakan URL berbahaya dibandingkan algoritma lainnya **[7]**. Berdasarkan landasan tersebut, penelitian ini akan menerapkan algoritma XGBoost untuk mendeteksi tautan *phishing* yang tersebar melalui konten dewasa di *platform* X.

Berangkat dari urgensi permasalahan yang telah dipaparkan sebelumnya, penelitian ini hadir untuk menawarkan pendekatan baru dalam mengisi celah keamanan tersebut. Penulis mengajukan pengembangan model deteksi otomatis yang difokuskan pada tautan *phishing* yang sering kali menyusup melalui konten dewasa di media sosial. Dengan memanfaatkan keunggulan metode *Machine Learning* yang mampu mengenali pola data kompleks dan adaptif, sistem ini dirancang untuk dapat membedakan tautan berbahaya dan aman secara presisi. Sebagai wujud solusi konkret, penelitian ini diajukan dengan judul **"Deteksi**

Malicious Link pada Konten Dewasa di Platform X Menggunakan Metode Machine Learning", dengan harapan dapat memberikan kontribusi nyata bagi keamanan ekosistem digital

Rumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan, rumusan masalah penelitian ini adalah:

Bagaimana metode machine learning dapat digunakan untuk mendeteksi link phishing berdasarkan analisis URL?

1.3 Batasan Masalah

Agar penelitian ini tetap terarah dan tidak melebar dari fokus utama, maka beberapa batasan masalah yang digunakan dalam penelitian ini adalah sebagai berikut:

1. **Sumber Data:** Dataset hanya mencakup tautan yang berasal dari unggahan konten dewasa pada platform X.
2. **Karakteristik Tautan:** Fokus analisis terletak pada tautan pendek (*shortlink*) dengan domain `t.co`.
3. **Metode Pelabelan:** Penentuan label aman atau bahaya dilakukan secara manual melalui observasi perilaku *redirect* dan alamat tujuan akhir.
4. **Ruang Lingkup Analisis:** Analisis dibatasi pada fitur leksikal dan statistik URL tanpa memeriksa isi konten halaman web secara mendalam.
5. **Volume Dataset:** Dataset yang digunakan berjumlah 1.000 data dengan komposisi seimbang (500 aman dan 500 berbahaya).
6. **Algoritma:** Algoritma *Machine Learning* yang digunakan adalah XGBoost.
7. **Metrik Evaluasi:** Pengukuran performa model menggunakan metrik *Accuracy, Precision, Recall,* dan *F1-Score.*

1.4 Tujuan Penelitian

Berdasarkan rumusan masalah yang ada, maka tujuan dari penelitian ini adalah:

1. Mengumpulkan dan mengidentifikasi tautan yang tersebar pada konten dewasa di platform X untuk memetakan pola penyisipan serta karakteristik penyebarannya.
2. Menganalisis karakteristik leksikal dan statistik (seperti entropi) untuk menemukan pola pembeda antara tautan *phishing* dan tautan aman pada ekosistem t.co.
3. Mengimplementasikan algoritma *Machine Learning* XGBoost yang mampu melakukan klasifikasi tautan *phishing* secara otomatis berdasarkan hasil ekstraksi fitur.
4. Mengevaluasi performa model dalam mendeteksi ancaman siber menggunakan metrik *Accuracy*, *Precision*, *Recall*, dan *F1-Score* untuk memastikan keandalan sistem.

1.5 Manfaat Penelitian

Penelitian ini diharapkan dapat memberikan beberapa manfaat sebagai berikut:

1. **Manfaat Akademis**

   Penelitian ini dapat menjadi referensi bagi mahasiswa dan peneliti yang tertarik pada deteksi link phishing berbasis analisis URL, khususnya pada konten dewasa di platform X. Selain itu, penelitian ini menambah literatur terkait pemanfaatan machine learning dalam mendeteksi ancaman siber dari data media sosial.

2. **Manfaat Praktis**

   Hasil penelitian dapat membantu pengguna media sosial lebih waspada terhadap link berbahaya, terutama yang disisipkan dalam konten sensitif.

Model machine learning yang dikembangkan dapat menjadi dasar sistem peringatan otomatis untuk mencegah pencurian data.

3. **Manfaat untuk Keamanan Siber**

   Penelitian ini menjadi acuan bagi pengembang atau praktisi keamanan siber dalam memahami pola penyebaran phishing di platform X, sehingga dapat digunakan untuk meningkatkan mekanisme deteksi dini.

4. **Manfaat untuk Platform dan Komunitas Digital**

   Penelitian ini dapat memberikan masukan bagi platform X terkait pola link berbahaya pada konten dewasa, sehingga dapat memperkuat moderasi konten dan perlindungan pengguna.

1.6 Sistematika Penulisan

Agar pembahasan dalam skripsi ini lebih mudah dipahami, penulisan penelitian disusun ke dalam beberapa bab yang saling berhubungan. Secara garis besar, alur penelitian ini dijelaskan melalui sistematika berikut:

BAB I Pendahuluan

Bab ini berisi gambaran awal mengenai penelitian yang dilakukan. Di dalamnya membahas latar belakang masalah, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, serta sistematika penulisan skripsi. Bagian ini menjadi dasar untuk memahami alasan penelitian dilakukan dan arah penelitian secara keseluruhan.

BAB II Tinjauan Pustaka

Pada bab ini, penulis menguraikan teori-teori dan referensi yang berkaitan dengan penelitian. Mulai dari penjelasan tentang phishing, karakteristik URL, konten dewasa di platform X, konsep dasar machine learning, hingga penelitian terdahulu yang relevan. Semua teori yang disajikan pada bab ini digunakan untuk memperkuat landasan penelitian.

BAB III Metodologi Penelitian

Bab ini menjelaskan langkah-langkah yang ditempuh selama proses penelitian. Mulai dari cara mengumpulkan link dari konten pornografi di X, proses analisis manual untuk menentukan apakah link tersebut phishing atau aman, proses ekstraksi fitur URL, penerapan metode machine learning, hingga cara mengevaluasi model. Bab ini juga memuat alur penelitian dan alat bantu yang digunakan.

BAB IV Hasil dan Pembahasan

Bab ini menyajikan hasil dari proses penelitian yang dilakukan. Mulai dari hasil pengumpulan link, hasil analisis manual, ciri-ciri link phishing yang ditemukan, performa model machine learning setelah dilatih, serta pembahasan hasil secara lebih mendalam. Bagian ini menjadi inti dari penelitian karena menunjukkan apa yang berhasil ditemukan penulis.

BAB V Kesimpulan

Bab terakhir berisi kesimpulan dari penelitian yang telah dilakukan serta saran yang dapat menjadi acuan untuk penelitian selanjutnya. Kesimpulan disusun berdasarkan hasil dan pembahasan pada bab sebelumnya.