BAB 1
PENDAHULUAN

1.1. Latar Belakang

Pusat data nasional merupakan tempat untuk menyimpan sebagian besar data masyarakat Indonesia, sebagai instansi yang penting pusat data nasional harus dijaga dengan ketat agar tidak terjadi peretasan yang membuat kebocoran data (Ramdhan dkk., 2024). Kebocoran data yang terjadi di Pusat Data Nasional Sementara 2 Surabaya menjelaskan pentingnya penerapan keamanan siber yang efektif dalam sebuah organisasi (Adristi & Ramadhani, 2024). Dalam skala yang lebih sederhana, ancaman ini sering ditemui dalam kehidupan sehari-hari melalui penyebaran tautan palsu. Sebagai contoh, masyarakat seringkali menerima pesan singkat melalui WhatsApp atau SMS yang berisi tautan dengan imbalan hadiah, kurir paket, hingga tagihan pajak palsu. Tautan tersebut seringkali menggunakan teknik manipulasi agar terlihat serupa dengan situs resmi, namun sebenarnya mengarahkan pengguna ke situs *phishing* untuk mencuri data atau menyebarkan *malware* pada perangkat korban. Oleh karena itu, deteksi dini terhadap *Uniform Resource Locator* (URL) berbahaya menjadi salah satu upaya untuk mencegah ancaman keamanan siber pada level instansi negara maupun pengguna individu.

Text mining adalah salah satu pendekatan yang dapat digunakan untuk menganalisis sentimen dari teks yang tersedia, seperti pada media sosial (Elfaiz dkk., 2025). *Text mining* merupakan suatu proses penambangan intisari dari suatu dokumen data berupa teks yang bentuknya lebih tidak teratur yang dapat dibuat pola untuk menghasilkan sebuah informasi yang berguna (Astuti dkk., 2024). Text mining digunakan untuk mengkategorikan dan menganalisis sentimen dari data teks yang besar (Rachman dkk., 2025).

Algoritma *Support Vector Machine* merupakan metode yang paling sering digunakan untuk klasifikasi teks (Khaira dkk., 2023). Support Vector Machine (SVM) adalah bagian dari pembelajaran mesin yang bekerja dengan cara menemukan *hyperplane* terbaik yang berguna untuk memisahkan dua buah kelas

pada input (Sembiring dkk., 2024). Keunggulan ini berasal dari kemampuannya memisahkan data non-Linearberdimensi besar secara linear dengan bantuan fungsi *Kernel,* sehingga membuat SVM sangat efektif dalam berbagai jenis analisis data (Hendriyanto dkk., 2022). Selain kemampuan dalam menangani data *non-linear*, alasan utama SVM sering diimplementasikan dalam klasifikasi teks maupun pesan singkat adalah efektivitasnya dalam menangani High Dimensional Feature Space. Data teks yang telah melalui proses TF-IDF biasanya menghasilkan ribuan fitur unik (kata/token), namun hanya sedikit dari fitur tersebut yang benar-benar relevan bagi klasifikasi. Dalam kasus sehari-hari seperti pesan teks atau URL, pola serangan seringkali tersembunyi dalam kombinasi kata-kata yang langka namun spesifik. SVM mampu mengidentifikasi *support vectors* yang dari sekumpulan data yang luas sehingga model tetap memiliki kemampuan klasifikasi yang baik meskipun jumlah sampel data latih mungkin terbatas.

Pada penetian terdahulu akurasi yang dihasilkan Algoritma *Support Vector Machine* menunjukkan kinerja stabil dengan peningkatan akurasi pada data *train* dan *test* seiring bertambahnya jumlah *Fold* dalam *cross-validation*. Seperti pada penelitian terdahulu yang berjudul *Klasifikasi Deteksi Link Phising DANA Kaget Menggunakan Metode Support Vector Machine Berbasis Website* oleh Vebriani & Yustanti (2024) menghasilkan Akurasi *train* pada *Fold 5* dan 85% dan pada *Fold* 10, akurasi mencapai 90% dalam mengklasifikasi link dana *phising*. Penelitian ini bertujuan untuk mengembangkan sebuah sistem deteksi URL berbahaya menggunakan algoritma SVM dalam *text mining*. Sistem ini diharapkan mampu melakukan prediksi dengan tingkat akurasi yang tinggi, sehingga dapat menjadi alat yang efektif dalam melindungi pengguna internet dari tautan berbahaya.

1.2. Perumusan Masalah

Rumusan masalah pada penelitian ini berfokus pada bagaimana hasil algoritma SVM dalam memprediksi URL berbahaya berdasarkan hasil akurasi, presisi, dan recall yang dihasilkan?

1.3. Tujuan Penelitian

Tujuan penelitian ini adalah menghitung nilai akurasi, presisi dan *Recall* pada algoritma SVM dalam memprediksi URL berbahaya.

1.4. Manfaat Penelitian

Diharapkan penelitian ini dapat memberi manfaat bagi :

1. Praktikan

   kegiatan penelitian ini dijadikan sebagai pengalaman berhaga dalam mengembangkan ilmu dan dapat memberikan pengetahuan yang berkelanjutan mengenai *text mining* dan algoritma SVM.
2. Masyarakat

   Dengan adanya klasifikasi keamanan URL dapat mencegah terjadinya *cybercrime* pada masyarakat.

1.5. Batasan Penelitian

Batasan masalah dalam memperkirakan ketepatan algoritma SVM untuk prediksi URL berbahaya sebagai berikut:

1. Dalam penelitian ini, atribut yang digunakan adalah URL dan jenis keamanan sebagai label.
2. Klasifikasi dibagi menjadi 4 label yaitu *benign*, *phishing*, *defacement*, dan *malware*.
3. Pada beberapa tahap pemrosesan data menggunakan *library* sehingga hasil *output* bergantung pada *library* yang dipakai.