

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Perkembangan penggunaan Internet dengan ancaman serangan siber berbanding lurus dimana terdapat tren peningkatan jumlah serangan siber, sebagai contoh, selama bulan Mei 2019 tercatat 7930 serangan siber dan jumlah tersebut meningkat menjadi 11359 serangan pada Mei 2020 (Buil-Gil et al., 2021). Serangan siber pada dasarnya terjadi akibat adanya kerentanan keamanan pada sistem (Allodi & Massacci, 2017) dan informasi mengenai serangan siber tersebut dapat ditemukan pada beberapa media sosial (Rao et al., 2018). Informasi yang tersedia dalam media sosial dapat digunakan untuk melakukan analisis tren serangan siber (You et al., 2019). Kuantitas informasi yang besar seperti yang ditunjukkan oleh R. Williams memunculkan tantangan dalam *retrieval*-nya (Williams et al., 2018). Jika informasi di-*retrieve* dan dianalisis secara manual, banyak waktu dan usaha akan diperlukan.

Web scraping merupakan proses mengumpulkan data dari *website* dengan menggunakan suatu *crawler* yang berjalan secara otomatis (Krotov & Silva, 2018). *Scraping* memiliki kemampuan untuk mengambil data pada *website* dalam jumlah banyak secara cepat. *Web scraping* dapat dibuat dengan memanfaatkan *library* seperti Selenium atau Scrapy (You et al., 2019). *Web scraping* menghasilkan data mentah yang tidak berlabel. *K-Means Clustering* dapat digunakan dalam proses *text mining* data tidak berlabel tersebut agar dapat dilakukan *text clustering* untuk analisis tren serangan siber dengan efisien (Zhu et al., 2019). *K-Means Clustering*

dipilih dalam penelitian ini karena *K-Means Clustering* merupakan salah satu algoritme yang mudah digunakan, cepat dan telah beberapa kali digunakan oleh peneliti lain untuk mendeteksi topik yang sedang *trending* (Sinaga & Yang, 2020).

Berdasarkan paparan tersebut, dalam skripsi ini dibangun suatu sistem yang dapat mengambil data (*scraping*) dari media sosial Twitter (<https://twitter.com>). Twitter di-*scraping* karena Twitter mengandung banyak informasi dengan setidaknya 340 juta *post* dibuat setiap harinya. Selain itu, pengguna Twitter akan berbagi informasi saat adanya kejadian serangan siber baru (V et al., 2018). Data yang telah di-*scraping* akan dianalisis menggunakan *K-Means Clustering* untuk menentukan tren serangan siber. Perancangan dan hasil yang diperoleh dari analisis tersebut dituangkan dalam skripsi dengan judul “Analisis Tren *Cyber Attack* pada Media Sosial Twitter Menggunakan *K-Means Clustering*”.

1.2 Rumusan Masalah

Berdasarkan latar belakang, masalah pada penelitian ini adalah banyaknya informasi yang terdapat pada sosial media Twitter membuat analisis serangan siber menjadi tidak efektif dan efisien jika dilakukan secara manual.

1.3 Tujuan Penelitian

Tujuan penelitian ini adalah memudahkan pengguna dalam melakukan analisis berita mengenai serangan siber yang terdapat pada Twitter.

1.4 Manfaat Penelitian

Manfaat penelitian ini dibagi menjadi dua yaitu:

1. Manfaat Teoritis: Diharapkan dapat memberikan pengetahuan mengenai cara melakukan *scraping* data dari Twitter dan melakukan analisis menggunakan *K-Means Clustering*.
2. Manfaat Praktis: Diharapkan dapat membantu dalam mengenali tren serangan siber dalam suatu periode waktu tertentu sehingga dapat meningkatkan pemahaman serangan siber dan meningkatkan keamanan sistem.

1.5 Batasan Masalah

Batasan masalah dari penelitian ini berupa:

1. Data diambil dari media sosial Twitter (<https://www.twitter.com/>) dengan kata kunci:
 - *Cyberattacks*.
 - *Cyberattack*.
 - *Cybernews*.
2. Data yang digunakan hanya yang berbahasa Inggris.
3. Data yang diambil berada dalam kurun waktu 01 September 2021 sampai dengan 31 November 2021.
4. Hasil analisis berupa informasi serangan siber yang telah terjadi.

1.6 Sistematika Penulisan

Sistematika penulisan dari penelitian ini mencakup:

- BAB I Pendahuluan: Bagian ini terdiri dari latar belakang masalah, rumusan masalah, tujuan penelitian, manfaat penelitian, batasan masalah, dan sistematika penelitian.
- BAB II Landasan Teori: Bagian ini terdiri dari teori yang mendukung penelitian ini dan penelitian terdahulu.
- BAB III Metodologi Penelitian: Bagian ini terdiri dari analisis proses, dan analisis sistem.
- BAB IV Hasil dan Pembahasan: Bagian ini terdiri dari hasil penelitian, dan pembahasan.
- BAB V Kesimpulan dan Saran: bagian ini terdiri dari kesimpulan dan saran.

