BAB I
PENDAHULUAN

1.1 Latar Belakang Masalah

Menurut statista.com, pengguna internet pada tahun 2021 mencapai 4,6 miliar jiwa, dan diantara jumlah tersebut, 4,2 miliar pengguna internet adalah pengguna media sosial, yang diantaranya yaitu Twitter. Media sosail merupakan *platform* digital yang digunakan para penggunanya untuk dapat saling berbagi informasi, berkomunikasi, dan membangun relasi antara sesama pengguna media sosial[1].

Seiring dengan kebebasan penggunaan media sosial, para penggunanya dapat mengunggah berbagai hal, semisal hal-hal informatif seperti *update* berita terkini. Namun ditemukan juga beberapa pengguna yang mengunggah hal-hal yang berbahaya[2]. Sebagai contoh, pengguna yang mengunggah unggahan yang mengandung ancaman (*threat*) dan/atau kerentanan (*vulnerability*) terhadap keamanan suatu sistem di media sosial, dalam hal ini khususnya media sosial Twitter.

Pada era digitalisasi seperti saat ini, *therat* dan *vulnerability* terhadap keamanan suatu sistem menjadi perhatian tersendiri[3]. *Threat* dan *vulnerability* yang dipublikasi di media sosial tentunya dapat merugikan pemilik sistem, karena dikhawatirkan akan disalah gunakan oleh orang-orang yang melihat postingan tersebut. Sehingga, dalam rangka mendeteksi publikasi *threat* dan *vulnerability* terhadap keamanan suatu sistem di media sosial, dibuat sistem *text mining* sebagai metode pendeteksi *threat* dan *vulnerability* pada Twitter menggunakan algortima naïve bayes dan TF-IDF (Term Frequency – Inverse Document Frequency). Dengan harapan postingan yang mengandung *threat* dan *vulnerability* terhadap keamanan suatu sistem dapat terdeteksi untuk bisa diambil tindakan lebih lanjut.

Menurut penelitian yang dilakukan oleh Dinda Ayu Muthia[4]. (2018) dengan judul "Komparasi Algoritma Klasifikasi *Text Mining* Untuk Analisis Sentimen Pada Review Restoran", mendapati hasil perbandingan antara algoritma Naïve Bayes dan Support Vector Machine menunjukkan hasil akurasi sebesar 87% oleh

algoritma Naïve Bayes, sedangkan algoritma Support Vector Machine menunjukkan hasil akurasi sebesar 56%. Dari penelitian tersebut didapati bahwa implementasi algoritma Naïve Bayes untuk melakukan *text mining* menunjukkan hasil baik dan akurasi yang tinggi. Oleh karena itu, sistem pendeteksi *threat* dan *vulnerability* pada twitter dibuat menggunakan algoritma Naïve Bayes.

Tugas akhir ini bertujuan untuk membuat sistem yang dapat membedakan *tweet* yang mengnadung unsur *threat* atau *vulnerability*, dan yang bukan, dengan menggunakan algoritma pembobotan TF-IDF dan algoritma klasifikasi Naïve Bayes.

1.2 Rumusan Masalah

Berdasarkan latar belakang masalah yang telah dijabarkan sebelumnya, maka dirumuskan masalah pada tugas akhir ini adalah sebagai berikut:

1. Bagaimana cara membuat sistem pendeteksi postingan mengandung *threat* dan *vulnerability* pada Twitter?
2. Bagaimana performa algoritma klasifikasi Naïve Bayes dalam mendeteksi postingan mengandung *threat* dan *vulnerability* pada Twitter?

1.3 Tujuan dan Manfaat

Tujuan dan manfaat pada tugas akhir ini adalah sebagai berikut:

1. Membuat pendeteksi *threat* dan *vulnerability* pada unggahan Twitter menggunakan algoritma Naïve Bayes.
2. Melakukan uji performasi pada sistem pendeteksi *threat* dan *vulnerability* pada unggahan Twitter menggunakan algoritma Naïve Bayes.

1.4 Batasan Masalah

Adapun batasan masalah pada tugas akhir ini adalah sebagai berikut:

1. Output yang dihasilkan merupakan klasifikasi tweet menjadi kelas *threat* atau *vulnerability*, atau tidak keduanya.

2. Sistem tidak melakukan tindakan lebih lanjut terhadap temuan unggahan yang mengandung *threat* dan *vulnerability* terhadap suatu sistem.
3. Algoritma klasifikasi yang digunakan adalah algoritma Naïve Bayes.
4. Algoritma pembobotan yang digunakan adalah algoritma TF-IDF (Term Frequency – Inverse Document Frequency).
5. Seluruh dataset yang digunakan pada proses pembuatan sistem menggunakan dataset berbahasa Inggris

1.5 Metode Penelitian

Adapun metode penelitian yang dilakukan dalam pengerjaan tugas akhir ini adalah sebagai berikut:

1. Studi Literatur

   Pada langkah awal dilakukan analisa masalah dan studi literatur, dimana penulis mencari berbagai literatur mempelajari berbagai referensi yang berkaitan dengan *text mining*, *text pre-processing*, algoritma pembobotan TF-IDF, dan algoritma klasifikasi Naïve Bayes.

2. Pengumpulan Data

   Setelah melakukan studi literatur, dilakukan pengumpulan data yang akan digunakan sebagai dataset, dengan metode *web scraping*, untuk mengunduh data dari situs web Twitter.

3. Implementasi dan Pengujian.

   Pada tahap ini dilakukan pengimplementasian dan pengujian hasil pembuatan sistem dan pengolahan data, dan kemudian dilakukan pengujian untuk mengetahui jika sistem sudah berjalan sesuai dengan target yang ingin dicapai, serta mendapatkan nilai akurasi terbaik yang bisa didapat.

4. Evaluasi

   Tahap evaluasi dilakukan untuk memperbaiki kekurangan dan error yang ada pada sistem, sehingga hasil akhir sistem yang diharapkan bisa tercapai.

5. Penulisan Tugas Akhir

   Pada tahap ini, dilakukan penyusunan buku Tugas Akhir yang mencakup teori, tahapan, dan hasil dari penelitian Tugas Akhir yang telah dibuat.

1.6 Sistematika Penulisan

Pada penulisan laporan Tugas Akhir ini, laporan dibagi menjadi lima bab, dimana setiap bab secara singkat dijelaskan sebagai berikut:

BAB I PENDAHULUAN

Bab ini berisi latar belakang, rumusan masalah, batasan masalah, dan tujuan dalam pembuatan penelitian Tugas Akhir.

BAB II DASAR TEORI

Pada bab ini dijelaskan seluruh metode dan tori yang berkaitan dan diimplementasikan dalam penelitian Tugas Akhir

BAB III PERANCANGAN SISTEM

Dalam bab ini dijelaskan gambaran umum, analisis, perancangan, serta alur pengujian dari sistem yang dibuat.

BAB IV SKENARIO PENGUJIAN

Bab ini menjelaskan tentang parameter dan skenario yang dijalankan dalam melakukan pengujian sistem yang telah dibuat.

BAB V KESIMPULAN DAN SARAN

Pada bab terakhir ini, menjelaskan kesimpulan akhir yang diketahui dari penelitian Tugas Akhir yang telah selesai dibuat, dan juga saran bagi penelitian lebih lanjut.