

BAB I

PENDAHULUAN

1.1 Latar Belakang

Twitter merupakan media sosial yang memungkinkan pengguna untuk membuat cuitan berbentuk text, foto dan video. Pada bulan juli 2018, twitter menduduki peringkat ke dua belas dalam hal jumlah pengguna aktif terbanyak di dunia.

Indonesia menduduki peringkat ke 9 dalam hal jumlah pengguna yang menggunakan twitter di dunia. Rata-rata pengguna aktif harian di twitter mencapai 126 juta pengguna. Pengguna twitter semakin bertambah setiap harinya di Indonesia. Dengan twitter pengguna dapat mengikuti *trend* terkini dengan topik tertentu menggunakan tagar. Para pengguna juga dapat mengunggah cuitan dengan panjang maksimal 280 karakter disertai hastag, tagging dan foto. Aplikasi twitter dapat digunakan menggunakan *smartphone* baik platform Android, IOS atau Windows Phone serta dapat di akses menggunakan web.

Alasan twitter banyak digunakan adalah karena pengguna dapat dengan mudah mengunggah cuitan melalui *smartphone* atau website. Namun twitter juga mempunyai kekurangan, yaitu banyaknya cuitan yang mengandung unsur ujaran kebencian. Pada bulan maret 2018 POLRI menemukan 642 konten provokatif. Faktor utama penyebab cuitan ujaran kebencian adalah prasangka buruk terhadap orang atau kelompok tertentu. Ujaran kebencian biasa ditemukan saat menjelang Pilkada dan Pilpres. Cuitan ujaran kebencian selain dilakukan oleh pengguna tertentu seringkali juga dilakukan oleh akun-akun palsu yang digunakan untuk

menyerang seseorang atau kelompok tertentu dan tentunya ini sangat mengganggu dan berpotensi untuk membuat kegaduhan.

Untuk mengatas hal ini, twitter telah memberikan solusi yang memungkinkan pengguna untuk melaporkan tweet yang dianggap mengandung ujaran kebencian. namun tindakannya membutuhkan waktu yang lama. Maka dari itu perlu dibangun suatu sistem untuk mengklasifikasikan cuitan ujaran kebencian yang berbahasa Indonesia. Ada berbagai metode untuk klasifikasi seperti *Naive Bayes*, *Decision Trees*, *Support Vector Machines* atau *K-Nearest Neighbor*. Metode yang digunakan dalam penelitian ini adalah *Naive Bayes*. Metode *Naive Bayes* menggunakan konsep probabilitas setiap kelas dalam proses pembelajaran klasifikasinya. Metode ini dipilih karena mudah di implementasikan dan tidak terlalu membutuhkan sumber daya komputer yang besar.

Berdasarkan uraian diatas maka penulis akan melakukan penelitian sebagai skripsi dengan judul **"Implementasi Algoritma *Naive Bayes* untuk Klasifikasi Ujaran Kebencian pada Twitter"**.

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, maka dapat dirumuskan rumusan masalah sebagai berikut:

1. Bagaimana implementasi algoritma *Naive Bayes* untuk klasifikasi cuitan ujaran kebencian pada Twitter?

2. Berapa presentase tingkat akurasi algoritma *Naive Bayes* pada pengklasifikasian cuitan ujaran kebencian pada Twitter?

1.3 Batasan Masalah

Agar penelitian dapat terfokus, maka perlu adanya batasan untuk membatasi lingkup penelitian ini. Adapun batasan penelitian yang dilakukan antara lain:

1. Metode yang akan digunakan pada penelitian ini adalah *Naive Bayes*.
2. Sistem menggunakan media *website*.
3. Proses *stemming* menggunakan *library* nalapa.
4. Data di dapatkan dari twitter tahun 2017 - 2019.
5. Cuitan yang akan di klasifikasikan adalah cuitan yang berbahasa Indonesia.
6. Cuitan akan dikategorikan ke dalam dua kelas yaitu *hate speech* dan *not hate speech*.
7. Metode yang digunakan untuk mengevaluasi model adalah *10-Fold Cross Validation* dan *Confusion Matrix*.

1.4 Maksud dan Tujuan Penelitian

Maksud dan tujuan penelitian ini adalah sebagai berikut:

1. Membangun model yang dapat mengklasifikasikan cuitan *hate speech* atau *not hate speech*.
2. Mengevaluasi hasil klasifikasi dari model yang di usulkan.

1.5 Metodologi Penelitian

Pada penelitian ini, penulis menggunakan beberapa metode penelitian, adapun metode penelitian yang digunakan adalah sebagai berikut:

1.5.1 Metode Pengumpulan Data

Metode pengumpulan data diperlukan untuk agar mendapatkan informasi yang relevan dengan topik penelitian yang diambil. Adapun metode pengumpulan data yang digunakan adalah studi pustaka. Studi kepustakaan adalah proses untuk mengumpulkan bahan-bahan untuk penelitian, antara lain buku, data cuitan berbahasa Indonesia dari twitter, jurnal-jurnal, skripsi, maupun situs internet mengenai twitter dan algoritma *Naive Bayes* untuk menunjang proses penelitian.

1.5.2 Metode Eksperimen

Evaluasi *10-Fold Cross-Validation* dan *Confusion Matrix* digunakan untuk mengevaluasi akurasi dari metode *Naive Bayes*.

1.5.3 Metode Analisis

Data dan informasi yang didapatkan pada proses pengumpulan data kemudian di analisis untuk mengembangkan model yang dapat mengklasifikasikan cuitan *hate speech* atau *not hate speech*.

Sedangkan untuk mengevaluasi akurasi dari metode *Naive Bayes* menggunakan *10-Fold Cross-Validation* dan *Confusion Matrix*.

1.5.4 Metode Perancangan

Pemodelan proses ini menggambarkan bagaimana bisnis beroperasi. Menggambarkan perintah yang dilakukan dan bagaimana cara data berpindah diantara perintah-perintah tersebut. Dalam penelitian ini penulis menggunakan DFD (*Data Flow Diagram*).

1.5.5 Metode Pengembangan

Mengumpulkan dataset baru kemudian mengolah kembali sehingga menghasilkan pengembangan sistem dengan dataset yang baru untuk memaksimalkan prediksi kelas.

1.5.6 Metode Pengujian

Pengujian menjadi salah satu elemen penting dalam pengembangan perangkat lunak. Fungsi pengujian adalah untuk mengetahui nilai akurasi. Metode yang penulis gunakan untuk pengujian adalah *10-Fold Cross Validation* dan *Confusion Matrix*.

1.6 Sistematika Penulisan

Penulisan skripsi ini di tulis secara sistematis dalam beberapa bab, antara lain sebagai berikut:

BAB I PENDAHULUAN

Bab ini membahas tentang latar belakang penelitian, rumusan masalah, batasan masalah, maksud dan tujuan, metode penelitian dan sistematika penulisan.

BAB II LANDASAN TEORI

Bab ini membahas tentang teori-teori dan tinjauan pustaka yang menjadi rujukan yang dijadikan dasar pada proses penelitian.

BAB III ANALISIS DAN PERANCANGAN

Bab ini menjabarkan tentang analisis dari model yang diusulkan dan perancangan program yang akan digunakan untuk menguji model yang akan diusulkan dengan menggunakan algoritma *Naive Bayes*.

BAB IV IMPLEMENTASI DAN PEMBAHASAN

Bab ini menjelaskan secara rinci tentang program yang dibangun dan model yang diusulkan.

BAB V PENUTUP

Bab ini berisi kesimpulan dan saran dari penulis mengenai penelitian dan pengembangan lebih lanjut yang dapat dilakukan berkaitan dengan klasifikasi menggunakan algoritma *Naive Bayes*.

DAFTAR PUSTAKA

Berisi sumber rujukan penulis dalam melakukan penelitian.