

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Pada era milenial ini, penggunaan internet di Indonesia dapat dikatakan berkembang pesat. Adanya koneksi internet dapat dimanfaatkan dengan baik, sebagai sarana untuk menyalurkan hobi, baik itu jual beli secara *online*, menulis atau menyampaikan pendapat terhadap sesuatu melalui sosial media. Sosial media yang cukup populer di Indonesia salah satunya adalah Twitter. Melalui Twitter pengguna dapat menyampaikan pendapatnya secara bebas. Yang sedang marak pada twitter adalah orang-orang berlomba untuk membuat forum *base* yang dapat dinikmati oleh pengguna.

Forum *base* atau sering disebut grup adalah tempat untuk berbagi cerita, promosi jualan, maupun tanya jawab. Aktivitas pada *base* Twitter itu sendiri, dikendalikan oleh admin atau bot Twitter. Pada *base* sering kali disalahgunakan, dengan mengeluarkan kata-kata tidak senonoh. Mayoritas pengguna twitter, banyak yang dikategorikan masih berusia dibawah umur. Hal ini, tidak pantas untuk dibaca maupun dicontoh oleh siapapun. Oleh karena itu, perlu diterapkan pengklasifikasian *tweet* yang tidak senonoh yang terdapat di Twitter.

Penelitian tersebut memanfaatkan *tweet* dengan cara mengumpulkan dataset teks yang telah diambil pada *base* twitter. Dengan tujuan untuk mencari tahu

polaritas dari sentimen positif dan negatif. Maka dari itu *text mining* merupakan suatu solusi yang dapat dilakukan untuk mengetahui adanya *sentiment* konten pada Twitter, sebagai dasar dari dilakukannya pemilihan konten dengan dataset teks.

Sehubungan dengan diberlakukannya pengklasifikasian tweet tidak senonoh, pengguna dapat mengelompokkan tweet yang bermakna baik maupun buruk untuk dikirim ke *base*. Agar pengguna dapat lebih bijak dalam menggunakan Twitter.

Dalam Pengklasifikasian Teks dapat menggunakan berbagai algoritma, antara lain ada algoritma *Naïve Bayes*, *C4.5*, *ID3* dan lain-lain. Setiap algoritma yang digunakan memiliki kelebihannya masing-masing.

Penelitian ini menggunakan algoritma *Naive Bayes* yaitu sebuah model yang dapat bekerja dengan baik pada proses pembagian kategori untuk teks [1]. *Naive bayes* melakukan klasifikasi dengan menggunakan dua proses yaitu proses *training* dan proses *testing* [2].

## 1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dijabarkan diatas,maka permasalahan yang dapat dirumuskan adalah sebagai berikut :

1. Bagaimana cara algoritma *Naïve Bayes Classifier* dapat membantu memisahkan tweet yang berisikan kata-kata tidak senonoh?
2. Berapa tingkat akurasi yang diperoleh menggunakan algoritma *Naïve Bayes Classifier* untuk mengklasifikasi tweet tidak senonoh?

## 1.3 Batasan Masalah

Berdasarkan latar belakang dan rumusan masalah yang telah dijabarkan,peneliti membatasi cakupan ruang lingkup yang dibahas pada skripsi ini agar tidak melebar ke topik pembahasan yang lain,maka digunakan batasan masalah sebagai berikut :

1. Data yang diperoleh dalam penelitian ini hanya tweet yang ada pada *base twitter*.
2. Tweet yang diambil hanya menggunakan bahasa secara umum.
3. Algoritma yang akan digunakan untuk pengklasifikasian ini hanya *Naïve Bayes Classifier*.
4. Tweet yang diambil hanya menggunakan bahasa Indonesia.

## 1.4 Maksud dan Tujuan Penelitian

Adapun tujuan penelitian dalam penyusunan skripsi ini sebagai berikut :

1. Untuk mengetahui cara dan kinerja algoritma *Naïve Bayes Classifier* dalam memisahkan tweet yang berisi informasi kata-kata tidak senonoh.

- Mengetahui tingkat akurasi yang diperoleh dari Algoritma *Naïve Bayes Classifier* untuk melakukan klasifikasi tweet dengan baik.

### 1.5 Manfaat Penelitian

Adapun manfaat penelitian dalam penyusunan skripsi ini sebagai berikut: Peneliti :

- Penelitian ini diharapkan bagi pengguna twitter agar lebih bijak dalam mengeluarkan kata-kata.
- Penelitian ini dapat mengetahui kata-kata mana saja, yang termasuk kategori tidak senonoh agar tidak dicontoh.

Pengguna :

- Menambah wawasan tentang penggunaan tweet yang baik atau tidak menyimpang dalam peraturan yang dibuat oleh tiap *base* pada Twitter.

Twitter :

- Menambah *detail* informasi yang diberikan kepada Twitter, supaya memisahkan *tweet* terlebih dahulu ke dalam label positif dan label negatif.

### 1.6 Metode Penelitian

Metode penelitian adalah langkah yang dilakukan oleh peneliti dalam melakukan pengumpulan data, studi pustaka, *crawling data*, analisis kebutuhan, analisis data, perancangan sistem, dan *testing*.

#### 1.6.1 Metode Pengumpulan Data

Proses pertama yang dilakukan dalam melakukan klasifikasi teks adalah proses pengumpulan data. Pengumpulan data dilakukan dengan menggunakan

Twitter API yang diimplementasikan pada bahasa pemrograman Python.

### 1.6.2 Metode Studi Pustaka

Tahapan penelitian dimana pengumpulan data dengan membaca serta mengumpulkan berbagai jurnal, referensi *paper text classification, crawling data*, dan Algoritma *Naïve Bayes Classifier* yang saling berhubungan dengan tema penelitian yang diambil.

### 1.6.3 Metode Analisis Kebutuhan

Analisis Kebutuhan adalah sebuah proses untuk mendapatkan informasi, model, spesifikasi tentang perangkat lunak yang diinginkan pengguna. Informasi yang diperoleh pengguna yang menjadi acuan untuk melakukan desain perangkat lunak.

### 1.6.4 Metode Analisis Data

Mengacu kepada latar belakang yang telah di sampaikan sebelumnya, penelitian ini mengarah kepada *Naïve Bayes Classifier*, dengan harapan bahwa data yang diunduh dari media sosial twitter yaitu berupa positif dan negatif berbahasa Indonesia dapat terpecahkan oleh konsep penelitian *Text Mining*.

### 1.6.5 Metode Perancangan

Pada tahap ini dilakukan perancangan bagaimana alur sistem ini akan berjalan. Pada penelitian ini menggunakan *Unified Modeling Language (UML)* untuk memvisualisasikan, menspesifikasikan, membangun dan pendokumentasian dari sebuah perancangan dalam pengembangan sistem. Dalam penelitian

ini, menggunakan 4 diagram, yaitu *Use Case Diagram*, *Activity Diagram*, *Class Diagram*, dan *Sequence Diagram*.

### 1.6.6 Metode Pengujian

Pengujian atau *testing* yaitu proses untuk mengetahui keakuratan model yang dibangun pada proses training, umumnya digunakan data yang disebut data teks set untuk memprediksi label.

Dalam proses testing dapat dilakukan dengan 3 cara yaitu :

1. *White Box Testing* merupakan metode pengujian perangkat lunak mengenai pengetahuan internal tentang kemampuan sistem dan pemrograman yang bertujuan untuk mengoptimalkan dari pembuatan sistem.
2. *Black Box Testing* merupakan metode pengujian fungsional dari sebuah sistem yang digunakan untuk menguji perangkat lunak tanpa mengetahui struktur internal kode serta untuk mengetahui dari segi input dan output sistem.
3. *Confusion Matrix* merupakan metode pengujian yang menggunakan tabel matriks, jika data set hanya terdiri dari dua kelas. Kelas yang satu dianggap sebagai positif dan yang lainnya negatif. Performa dari suatu model klasifikasi dapat diukur dengan tingkat akurasi berdasarkan *Confusion matrix* [3].

## **1.7 Sistematika Penulisan**

Secara garis besar sistematika yang dipergunakan dalam penulisan ini sesuai dengan kaidah penulisan laporan Skripsi pada Universitas Amikom Yogyakarta. Adapun penulisannya sebagai berikut :

### **BAB I PENDAHULUAN**

Pada bab pendahuluan berisi tentang latar belakang masalah yang dijadikan dasar bagi penulis untuk merumuskan masalah, batasan masalah, maksud dan tujuan penelitian, manfaat penelitian, metode penelitian dan sistematikan penulisan.

### **BAB II LANDASAN TEORI**

Pada bab landasan teori berisi tentang teori atau konsep yang digunakan dalam penelitian ini. Tinjauan teori diperoleh dari daftar pustaka, jurnal dan penelitian sebelumnya yang terkait dengan masalah penelitian.

### **BAB III METODE PENELITIAN**

Pada bab metodologi penelitian berisi mengenai alur penelitian, penjelasan alur penelitian, analisis kebutuhan, *unified modeling language*, perancangan sistem, dan perhitungan manual.

### **BAB IV HASIL DAN PEMBAHASAN**

Pada bab hasil dan pembahasan ini berisi mengenai implementasi hasil penelitian dan hasil pengujian secara menyeluruh.

**BAB V PENUTUP**

Pada bab penutup berisi mengenai kesimpulan yang didapat dari penelitian pada bab-bab sebelumnya,serta saran yang diharapkan dapat berguna untuk pengembangan penelitian di masa mendatang.

