

# BAB I PENDAHULUAN

## 1.1 Latar Belakang

Dengan penggunaan teknologi digital di Indonesia yang semakin masif, tentunya membuat masyarakat tidak bisa menolak digitalisasi di berbagai sektor. Salah satunya yaitu penggunaan media komunikasi berupa sosial media. Sosial media menyediakan ruang untuk masyarakat melakukan interaksi satu sama lain yang dapat dilakukan kapanpun dan dimanapun. Sosial media juga kerap digunakan masyarakat untuk berbagi informasi terbaru, salah satunya yaitu Twitter.

Twitter merupakan sosial media yang memungkinkan penggunanya dapat menyebarkan informasi dan pendapat tentang suatu topik tertentu melalui pesan publik, yang biasa disebut dengan *tweet*[1]. Twitter memiliki fitur *trending* yang didalamnya berisikan beberapa topik yang sedang ramai diperbincangkan. Salah satunya yaitu kebocoran data registrasi sim yang beberapa waktu belakangan ini menuai berbagai macam pandangan masyarakat yang disampaikan melalui *tweet*.

Dengan adanya *platform* yang mudah diakses dan digunakan masyarakat untuk menyebarkan dan mencari berita terbaru menjadikan twitter sebagai salah satu wadah dalam mendapatkan sumber data *text*. Salah satu data *text* yang dapat ditemukan di twitter yaitu sentimen dan opini masyarakat tentang berita terbaru. Opini-opini tersebut cukup beragam mulai dari yang berisikan komentar positif, negatif, dan juga netral. Terkadang tidak menutup kemungkinan terdapat komentar yang berisikan pendapat yang tidak relevan dengan topik yang sedang dibahas. Dikarenakan banyaknya opini yang ditulis masyarakat tentang kebocoran data registrasi sim maka dibutuhkan analisis sentimen yang dapat membantu mengelompokkan opini berdasarkan sentimen positif, negatif, dan netral.

Penulis menggunakan metode *scraping* data tweet dalam pengambilan data yang akan digunakan dalam penelitian ini. Penelitian ini menggunakan *dataset* dari pencarian tweet dengan kata kunci “kebocoran data sim” yang menggunakan teks bahasa indonesia sebagai topik penelitian. Opini akan diolah melalui metode-

metode yang disesuaikan dalam proses pengambilan text atau *text mining*[2].

Pada penelitian ini metode yang akan digunakan penulis yaitu metode Naive Bayes. Naive Bayes merupakan algoritma yang memiliki kecepatan dalam pemrosesan data dan akurasi yang dihasilkan cukup tinggi bila digunakan pada data yang banyak[3]. Dengan demikian, penggunaan metode Naive Bayes pada penelitian ini diharapkan mendapat hasil yang baik.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang masalah diatas, maka didapatkan rumusan masalah sebagai berikut :

1. Bagaimana penerapan metode Naive Bayes dalam melakukan analisis sentimen pengguna twitter terhadap kebocoran data registrasi sim?
2. Berapa nilai akurasi yang dihasilkan dari metode yang digunakan?

## 1.3 Batasan Masalah

Dalam penelitian ini terdapat beberapa batasan masalah yang tertera dibawah ini :

1. Data *text* yang digunakan pada penelitian ini diambil dari *tweet* berbahasa indonesia.
2. Data yang digunakan diambil dari data *tweet* dengan kata kunci "kebocoran data sim".
3. Pengumpulan data dan pengolahan data menggunakan bahasa pemrograman *Python*.
4. *Tweet* diklasifikasikan menjadi 3 kelompok sentimen yaitu positif, negatif, dan netral.

## 1.4 Tujuan Penelitian

Penelitian ini bertujuan untuk :

1. Mengetahui hasil analisis sentimen pengguna twitter terhadap kebocoran data registasi sim yang dikelompokkan menjadi sentimen positif, negatif,

dan netral.

2. Mendapatkan hasil akurasi dari metode Naïve Bayes dalam melakukan analisis sentimen pengguna twitter terhadap kebocoran data registrasi sim.

### 1.5 Manfaat Penelitian

Penelitian ini memiliki beberapa manfaat sebagai berikut :

1. Dengan adanya penelitian ini diharapkan Kementerian Komunikasi dan Informasi dapat mengetahui sentimen masyarakat tentang kebocoran data registrasi sim yang nantinya dapat digunakan sebagai acuan untuk pengambilan keputusan dalam menyelesaikan masalah kebocoran data ini.
2. Dapat memberikan pengetahuan kepada pembaca tentang penerapan metode Naïve Bayes dalam melakukan analisis sentimen pengguna twitter.

### 1.6 Metode Penelitian

Beberapa metode yang akan digunakan dalam penelitian ini adalah sebagai berikut :

#### 1.6.1 Metode Pengumpulan Data

Penulis melakukan pengumpulan dengan metode *scrapping tweet* pada kolom search di twitter dengan menggunakan kata kunci "kebocoran data sim". *Data text* yang telah di *scrapping* akan melalui tahap *preprocessing* dan *labelling* sebelum nantinya akan digunakan sebagai bahan penelitian.

#### 1.6.2 Metode Analisis Data

Pada tahap ini penulis melakukan *preprocessing* terhadap data yang telah diambil. Setelah melakukan *preprocessing* dan *labelling* data, maka langkah selanjutnya adalah melakukan analisis menggunakan metode Naïve Bayes.

#### 1.6.3 Pengujian

Setelah didapatkan hasil dari analisis menggunakan metode Naïve Bayes selanjutnya adalah menghitung tingkat akurasi dari metode tersebut menggunakan *Confusion Matrix*. Dan dilakukan proses validasi untuk

mengetahui perbandingan nilai akurasi menggunakan *K-fold Cross Validation*.

## 1.7 Sistematika Penulisan

Penelitian ini memiliki sistematika penulisan sebagai berikut:

### **BAB I PENDAHULUAN**

Pada bab ini menjelaskan tentang latar belakang, perumusan masalah, tujuan dan manfaat penelitian, metode penelitian dan sistematika penulisan yang digunakan dalam penelitian.

### **BAB II TINJAUAN PUSTAKA**

Bab ini membahas studi literatur dari penelitian sebelumnya dan dasar-dasar teori yang berhubungan dengan penelitian.

### **BAB III METODE PENELITIAN**

Bab ini membahas tentang tahapan – tahapan yang akan dilakukan dalam perancangan metode penelitian.

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

Bab ini membahas tentang hasil dari perancangan yang telah dibuat beserta metode yang digunakan. Serta membahas evaluasi dari metode yang diterapkan.

### **BAB V PENUTUP**

Pada bab ini menjelaskan tentang kesimpulan dari penelitian yang telah dilakukan dan memberikan saran untuk penelitian kedepannya.