

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Instagram merupakan media sosial yang memungkinkan pengguna untuk berbagi foto dan video. Pada bulan Januari 2018, instagram menduduki peringkat ke tiga dalam hal jumlah pengguna media sosial terbanyak dengan 800 juta pengguna aktif bulanan.

Salah satu negara yang aktif menggunakan instagram adalah Indonesia, dengan 53 juta pengguna aktif bulanan. Instagram semakin banyak digunakan terutama oleh *public figure* di Indonesia. Para pengguna instagram dapat mengunggah foto atau video keseharian mereka, kemudian pengguna dapat menambahkan *caption, hashtag, lokasi, dan tagging* akun instagram lainnya untuk setiap foto atau video yang akan di unggah. Aplikasi instagram dapat digunakan melalui web dan *smartphone* baik platform Android, iOS atau Windows Phone dan bersifat gratis.

Alasan instagram banyak digunakan adalah karena kemudahannya untuk mengunggah foto atau video melalui *smartphone*. Namun instagram juga mempunyai kekurangan, yaitu banyaknya akun instagram yang menanggapi postingan dengan komentar yang dapat di kategorikan sebagai *spam*. Komentar *spam* biasanya banyak ditemui pada setiap postingan akun yang mempunyai *followers* banyak, terutama para *public figure* di Indonesia dan tentunya hal ini sangat mengganggu.

Untuk mengatasi hal ini, sudah ada beberapa solusi antara lain pengunggah dapat menghapus atau melaporkan komentar yang di anggap sebagai *spam* secara manual, tetapi akan membutuhkan waktu yang lama. Solusi lainnya adalah dengan cara membuat akun instagram menjadi *private*, akan tetapi hal ini tentu sulit di lakukan untuk akun yang mempunyai banyak *followers*.

Hal terakhir yang dapat dilakukan adalah dengan mengaktifkan fitur instagram untuk menghapus komentar dengan kata-kata yang telah dimasukan sebelumnya, akan tetapi fitur ini baru mendukung sembilan bahasa yaitu Inggris, Spanyol, Portugis, Arab, Perancis, Jerman, Rusia, Jepang and China. Maka dari itu perlu dibangun suatu sistem untuk mengklasifikasikan komentar spam dalam bahasa lain, salah satunya bahasa 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 Komentar Spam pada Instagram”**.

## 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 komentar *spam* pada instagram?
2. Berapa presentase tingkat akurasi algoritma *Naive Bayes* pada pengklasifikasian komentar *spam* pada instagram?

## 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 instagram tahun 2018.
5. Komentar yang akan di klasifikasikan adalah komentar yang berbahasa Indonesia.
6. Komentar akan dikategorikan ke dalam dua kelas yaitu *spam* dan *not spam*.
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 komentar *spam* atau *not spam*.
2. Mengevaluasi hasil klasifikasi dari model yang di usulkan.

## 1.5 Metode 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 komentar berbahasa Indonesia dari instagram, jurnal-jurnal, skripsi, maupun situs internet mengenai instagram 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 komentar *spam* atau *not spam*.

### 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.