

# BAB I

## PENDAHULUAN

### 1.1 Latar Belakang

Pada Senin 26 Agustus 2019 atau tepatnya setelah 74 tahun Indonesia merdeka, presiden terpilih Joko Widodo melalui kanal Youtube resmi Sekretariat Presiden mengumumkan bahwa pemindahan Ibu Kota Negara Kesatuan Republik Indonesia yang baru yaitu di wilayah administratif Kabupaten Penajam Paser Utara dan Kabupaten Kutai Kartanegara, Kalimantan Timur. Indonesia adalah Negara Kepulauan, terdiri dari 17,504 pulau, 16,056 di antaranya telah dibakukan dan disubmisi ke PBB; luas wilayah darat dan wilayah perairan yaitu 8,3 juta Km<sup>2</sup> (BIG, 2018). Daerah Tertinggal: 122 dari 541 Kabupaten/Kota (22,55%), dan 20 Kab. (16.4%) diantaranya merupakan kabupaten perbatasan (Perpres. 131/2015 tentang Penetapan Daerah Tertinggal tahun 2015-2019).

Data tersebut menjadi satu dari sekian banyak alasan mengapa Ibu Kota Indonesia akan pindah ke Kalimantan Timur. Rencana pemindahan Ibu kota negara dari Jakarta ke Kalimantan Timur didasari karena terjadinya kesenjangan pembangunan, rawannya bencana gempa bumi, karena Jakarta berada di dalam *ring of fire*, lalu lintas di Jakarta macet, penuh sesak dan kumuh lalu masalah rawan banjir (Setkab, BAPPENAS).

Setelah pengumuman itu terdapat banyak pro dan kontra yang terjadi di masyarakat. Dahulu sangat sulit untuk masyarakat mengungkapkan pendapat maupun kritik karena hanya bisa melalui media cetak. Perkembangan teknologi yang sangat pesat terutama teknologi komunikasi membuat masyarakat sangat mudah untuk menyampaikan opini, kritik maupun saran dalam media sosial, dalam hal ini masyarakat menyampaikan pendapatnya terhadap pemindahan Ibu Kota. Salah satu media untuk menyampaikan aspirasi yang sangat populer di kalangan masyarakat adalah Twitter.

Kemunculan twitter menjadikan sarana untuk penggunaanya supaya bisa mengirim pesan maupun berita terkini, gagasan, pendapat (opini) maupun curahan hati. Pesan tersebut biasa disebut dengan *tweet*. *Tweet* merupakan pesan pendek yang hanya berjumlah 140 karakter. Menurut situs alexa pada bulan Februari 2019, twitter menduduki peringkat kesebelas situs global. Sementara dalam akun situs Twitter, Indonesia memiliki 29,4 juta akun Twitter atau menduduki urutan kelima dunia, sehingga sangat mudah untuk menemukan opini, tidak hanya opini yang positif tapi juga negatif.

Penentuan polaritas positif atau negatifnya suatu opini dapat dilakukan secara manual, tetapi seiring bertambahnya sumber opini menjadi semakin banyak tentunya waktu dan usaha yang dibutuhkan untuk mengklasifikasikan polaritas opini tersebut akan semakin banyak terpakai. Oleh karena itu, diajukan penerapan metode pembelajaran mesin untuk mengklasifikasi polaritas opini dari

sumber data yang sangat banyak tersebut dengan menggunakan salah satu fungsi dari text mining, dalam hal ini adalah klasifikasi dokumen.

*Naive Bayes Classifier* memiliki tingkat akurasi yang lebih baik dibanding model *classifier* lainnya (Xhemali, Daniela, Chris J. Hinde, & Roger G. Stone, 2009). Alasan inilah tujuan penulis menggunakan algoritma tersebut supaya mendapatkan akurasi yang tinggi untuk menentukan suatu opini negatif ataupun positif pada data twitter. Pada kasus ini digunakan bahasa pemrograman R, software RStudio dan *R Shiny*.

Sehingga dalam menyelesaikan permasalahan tersebut dilakukan “Analisis Sentimen Pada Media Sosial Twitter Menggunakan Algoritma *Naive Bayes Classifier*”, terhadap pemindahan Ibu Kota Indonesia di Kalimantan Timur.

## **1.2 Rumusan Masalah**

Dari latar belakang yang telah dijelaskan, dapat ditarik beberapa permasalahan antara lain :

1. Bagaimana mengklasifikasikan sebuah data terhadap pemindahan Ibu Kota di Kalimantan Timur dan menentukan opini negatif atau opini positif menggunakan algoritma *Naive Bayes Classifier* ?
2. Bagaimana akurasi yang di peroleh dari penerapan algoritma *Naive Bayes Classifier* ?

### 1.3 Batasan Permasalahan

Agar penelitian ini lebih tepat sarasanya maka diperlukan batasan-batasan sebagai berikut:

- a. Penelitian ini menggunakan teknik text mining berupa Metode Analisis Sentimen menggunakan Algoritma *Naive Bayes Classifier*.
- b. Studi kasus yang digunakan berupa data *tweet* masyarakat terhadap pemindahan Ibu Kota di Kalimantan Timur dengan kata kunci pencarian kunci #ibukotapindah dan #IbuKotaBaru
- c. Data yang digunakan berjumlah 200 data
- d. Data disimpan dalam database menggunakan MySQL
- e. Bahasa Pemrograman yang digunakan pada studi kasus ini adalah Software RStudio dengan Framework R *Shiny*.

### 1.4 Tujuan dan Manfaat Penelitian

Tujuan yang ingin dicapai dari penelitian :

- a. Mengklasifikasikan data menggunakan algoritma *Naive Bayes Classifier* untuk mengetahui tingkat akurasinya.
- b. Menganalisa sebuah data menggunakan metode analisis sentimen untuk menentukan opini dari *tweet* tersebut termasuk opini negatif atau opini positif.
- c. Dapat memudahkan *user* untuk dapat mengakses langsung web interaktif R *Shiny*.

Manfaat yang ingin dicapai dari penelitian ini adalah sebagai berikut :

1. Bagi Penulis

Menjadi media dalam menerapkan teori dan praktek yang sudah didapatkan selama menjalani pendidikan di UNISBANK, dalam hal ini khususnya berupa pembuatan aplikasi text mining dengan metode analisis sentimen.

2. Bagi Pembaca

Memberikan informasi kepada masyarakat tentang penggunaan algoritma *Naïve Bayes Classifier* untuk mengelompokan opini masyarakat terhadap keputusan presiden untuk pemindahan Ibu Kota di Kalimantan Timur yang diambil dari pengguna Twitter.

3. Bagi Universitas

Memperlengkap hasil penelitian yang nanti dikemudian hari digunakan sebagai acuan bagi penelitian lainnya.

## **1.5 Metode Penelitian**

Metode penelitian yang digunakan dalam penulisan ini sebagai berikut :

### **1.5.1 Objek penelitian**

Penelitian yang dilakukan menggunakan opini masyarakat terhadap pemindahan Ibu Kota Indonesia ke Kalimantan Timur melalui media sosial Twitter dengan kata kunci #IbuKotaBaru dan #IbukotaPindah.

### 1.5.2 Metode Pengumpulan Data

Pada penelitian ini, metode yang digunakan untuk mendapatkan data informasi dan untuk menyelesaikan masalah adalah sebagai berikut:

#### 1. Studi Kepustakaan

Dalam pengumpulan data menggunakan metode kepustakaan yang dilakukan dengan pengumpulan jurnal, paper, buku, maupun situs internet sebagai acuan yang ada kaitannya dengan materi penulis khususnya klasifikasi dengan algoritma *Naïve Bayes Classifier*.

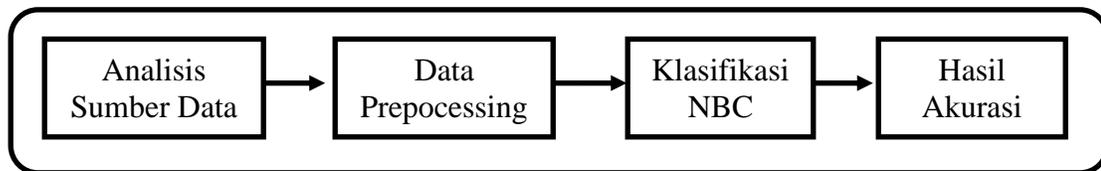
#### 2. Pengumpulan Data *Tweet* dari Twitter

Untuk mendapatkan data dengan cara mengambil data dari twitter dengan menggunakan kata kunci #IbuKotaBaru dan #IbukotaPindah.

Data yang sudah diperoleh disimpan di database MySQL.

### 1.5.3 Metode Analisis Data

Metode analisis data adalah suatu proses yang digunakan dalam menganalisa data *tweet* sehingga diketahui kelas tersebut masuk dalam sentimen positif atau negatif. Berikut adalah gambaran penyelesaian masalah penelitian ini dapat dilihat pada gambar 1.1.



Gambar 1. 1. Alur Sistem Analisis Data.

Berikut adalah gambaran penyelesaian klasifikasi analisis sentimen pada gambar 1.1.

#### 1. Analisis Sumber Data

Dalam penelitian ini menggunakan API Twitter untuk mengambil data dari twitter dan disimpan dalam bentuk file Ms.Excel dalam format CSV lalu diimport ke database, langkah selanjutnya yaitu menentukan klasifikasi secara manual terhadap data yang sudah ada. Data berjumlah 200 data tweet yang terdiri dari 159 data *training* dan 41 data *testing*.

#### 2. Data Preprocessing

Dalam proses ini, klasifikasi *tweet* di twitter yang telah dikumpulkan akan melalui beberapa tahap pemrosesan teks yang terdiri *case folding*, *normalisasi*, *tokenizing*, dan *stopword removal*. Tujuannya adalah agar data yang telah didapatkan lebih tersusun rapi dan akan lebih memudahkan dalam melakukan pengolahan data.

#### 3. Klasifikasi *Naïve Bayes Classifier*

Dalam proses klasifikasi menggunakan algoritma *Naïve Bayes Classifier* dengan cara menghitung probabilitas pada setiap *tweet* yang akan menentukan opini tersebut masuk dalam opini yang negatif atau opini yang positif.

#### 4. Hasil Akurasi

Tahap terakhir adalah menghitung tingkat akurasi, data yang akan dilakukan pengujian merupakan data yang sudah diketahui, sehingga setelah melakukan proses pengklasifikasian maka langkah selanjutnya menghitung tingkat

akurasi atau ketepatan sistem dalam proses pengklasifikasian. Menggunakan metode pengukuran tingkat akurasi *Confusion Matrix*.

## **1.6 Sistematika Penulisan**

Sistematika penulisan penelitian ini adalah sebagai berikut :

### **BAB I PENDAHULUAN**

Meliputi Latar Belakang Penelitian, Perumusan Masalah, Tujuan dan Manfaat Penelitian, Metodologi Penelitian, dan Sistematika Penulisan.

### **BAB II TINJAUAN PUSTAKA**

Memuat uraian sistematis tentang informasi hasil penelitian yang telah dilakukan sebelumnya dan menghubungkannya dengan masalah penelitian yang sedang diteliti.

### **BAB III METODOLOGI PENELITIAN**

Berisi tentang implementasi sistem yang sudah dirancang dalam bab sebelumnya.

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

Menjelaskan mengenai hasil dari penelitian yang telah dikerjakan, serta pembahasan masalah yang ada dalam penelitian tersebut.

### **BAB V PENUTUP**

Meliputi Kesimpulan dan Saran.