

# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar Belakang**

Menurut Haqqizar & Larasyanti, (2019) media sosial merupakan salah satu tempat atau media untuk berekspresi dan berpendapat tentang berbagai macam topik yang ada saat ini. Salah satu sosial media yang saat ini sedang populer digunakan oleh masyarakat adalah Twitter. Twitter merupakan sebuah situs *micro-blogging* yang sangat populer di Indonesia, dan pengguna Twitter juga bervariasi dari pengguna biasa hingga selebritis, perwakilan perusahaan, politikus bahkan presiden.

Namun opini pada twitter banyak sekali makna yang berbeda dari setiap netizen, salah satunya adalah tentang opini *covid-19* (Syarifuddin, 2020). *Covid-19* adalah penyakit yang disebabkan oleh virus *severe acute respiratory syndrome coronavirus 2* (SARS-CoV-2). *Covid-19* dapat menyebabkan gangguan sistem pernapasan, mulai dari gejala yang ringan seperti flu, hingga infeksi paru-paru, seperti pneumonia (Alodokter, 2021). Virus *Covid-19* telah menyebar di Indonesia pertama pada bulan Maret 2020. Penyebaran *Covid-19* semakin meluas, maka pemerintah menerapkan Pembatasan Sosial Berskala Besar (PSBB) demi memutus mata rantai virus *Covid-19*. Keadaan tersebut pastinya akan menjadi sorotan utama dalam pemberitaan di berbagai media massa dan menjadi topik yang hangat diperbincangkan juga di media sosial. Percakapan-percakapan di media sosial ini

mengandung sentimen berupa persepsi para pengguna internet atau *netizen* terhadap diberlakukannya PSBB di DKI Jakarta, baik itu sentimen positif, negatif, maupun netral. Kebisingan di dunia maya ini, membuat berita tentang diberlakukannya PSBB di DKI Jakarta menjadi menarik untuk diamati.

Pembatasan Sosial Berskala Besar (PSBB) adalah pembatasan kegiatan tertentu penduduk dalam suatu wilayah yang diduga terinfeksi *Corona Virus Disease 2019 (Covid-19)* (Kemenko PKM, 2019). Sejak diterapkannya PSBB di DKI Jakarta, masyarakat sering menulis *tweet* tentang adanya PSBB di DKI Jakarta dan dapat menjadi acuan untuk mengetahui sentimen masyarakat terhadap keadaan yang sedang terjadi di Indonesia ini. Kebutuhan dari analisis sentimen terhadap keadaan yang sedang terjadi di Indonesia, biasanya datang apabila ada pihak yang ingin mengetahui tentang sentimen dan tanggapan publik mengenai penerapan PSBB di DKI Jakarta. Oleh karena itu, analisis sentimen terhadap keadaan yang sedang terjadi di DKI Jakarta dari Twitter publik tentang menerapkan PSBB di DKI Jakarta sangat bermanfaat dalam memberikan tambahan wawasan serta gambaran bagi masyarakat tentang keadaan yang sedang terjadi di Indonesia setelah menerapkan PSBB di DKI Jakarta. Twitter sangat cocok untuk dijadikan sumber analisa sentimen dan *opinion mining* karena penggunaan Twitter untuk mengekspresikan opini mereka terhadap berbagai topik.

Tahapan dalam menganalisa sentimen pada topik kali ini dengan metode statistika, salah satu metode statistika yang dapat dilakukan pengkategorian adalah klasifikasi. Klasifikasi adalah suatu pengelompokan data dimana data yang digunakan

tersebut mempunyai kelas label atau target. Sehingga algoritma-algoritma untuk menyelesaikan masalah klasifikasi dikategorisasikan ke dalam supervised learning atau pembelajaran yang diawasi (Wibowo, 2017). Beberapa metode klasifikasi cukup banyak digunakan untuk melakukan klasifikasi berupa teks. Metode klasifikasi tersebut diantaranya adalah *Naïve Bayes Classifier* (NBC), *K-Nearest Neighbour*, dan *Support Vector Machines* (SVM) (Putra Nuansa, 2017). Beberapa kelebihan *Naïve Bayes* diantaranya adalah salah satu algoritma klasifikasi yang sederhana namun memiliki akurasi yang tinggi (Putra Nuansa, 2017). Metode *Naïve Bayes* telah banyak digunakan dalam penelitian mengenai *text mining*.

## 1.2 Rumusan Masalah

Rumusan masalah yang terdapat dari latar belakang adalah sebagai berikut:

1. Bagaimana opini pengguna Twitter pada *tweet* tentang PSBB di DKI Jakarta yang dikelompokkan berdasarkan sentimen positif, negatif dan netral?
2. Seberapa besar ukuran performa model klasifikasi dengan algoritma *Naïve Bayes*?

## 1.3 Batasan Masalah

Terdapat batasan masalah yang akan diteliti dalam penelitian ini yaitu sebagai berikut :

1. Penelitian membahas 4 tagar *tweet* tentang PSBB di DKI Jakarta.
2. Metode yang digunakan klasifikasi dengan algoritma *Naïve Bayes*.
3. *Tweet* yang digunakan berkaitan dengan PSBB di DKI Jakarta.

4. Hasil akhirnya adalah *tweet* dengan kata positif, negatif, maupun netral di antara keempat data *tweet* tersebut.
5. Data yang digunakan bersumber dari Twitter dengan periode waktu 14 September 2020 – 28 September 2020 dan tidak bisa menambahkan data terbaru.
6. *Tools* yang digunakan RapidMiner dan Microsoft Excel.

#### **1.4 Tujuan Penelitian**

Berdasarkan rumusan masalah di atas, tujuan yang ingin dicapai dalam penelitian ini adalah :

1. Mendapatkan opini pengguna *Twitter* pada *tweet* tentang PSBB di DKI Jakarta yang dikelompokkan berdasarkan sentimen positif, negatif dan netral.
2. Mengukur performa model klasifikasi dengan algoritma *Naive Bayes*.

#### **1.5 Manfaat Penelitian**

Penelitian ini mempunyai manfaat sebagai berikut :

1. Hasil Penelitian ini berupa data positif, negatif, maupun netral dari keempat data tagar *tweet*.
2. keakuratan dengan algoritma *Naive Bayes* yang memberikan informasi kepada masyarakat, agar mengetahui seberapa banyak reaksi terhadap penerapan PSBB di DKI Jakarta.

## 1.6 Metodologi Penelitian

Langkah-langkah metodologi penelitian yang akan dilakukan dalam penelitian ini adalah sebagai berikut:

### a. Studi Pustaka

Dilakukan studi pustaka untuk mengumpulkan teori-teori yang digunakan dalam penelitian ini dengan Algoritma *Naive Bayes*. Diperlukan juga teori yang berkaitan dengan analisis sentimen, *Text Mining*, *Data Mining*, dan klasifikasi.

### b. Pengumpulan Data

Dilakukan dengan studi literatur dan praktik untuk mendapatkan dataset. Data yang digunakan dalam penelitian ini diperoleh dengan input data dari media sosial Twitter.

### c. *Input Data*

*Input data* dari Twitter dilakukan pengambilan secara acak dengan menggunakan metode pengumpulan *tweets* manual yaitu memasukkan berbagai macam kata kunci yang berhubungan dengan PSBB di Indonesia pada kolom pencarian Twitter dengan menggunakan *hashtag*.

### d. *Data Preprocessing*

Tahap *preprocessing* merupakan tahap awal untuk mengolah data teks menjadi analisa sentimen. Terdapat beberapa metode yang digunakan pada tahap *preprocessing* seperti :

1. Seleksi Komentar Data yang telah dikumpulkan pada tahap

pengumpulan data ternyata ada data yang berulang dikarenakan adanya *re-tweet* tanpa diberikan komentar.

2. *Cleaning* adalah proses pembersihan data teks yaitu dengan menghilangkan data yang tidak konsisten atau tidak relevan.
3. Normalisasi data digunakan untuk menormalisasikan kalimat. Banyak pengguna dalam menuliskan kata menggunakan singkatan sehingga akan menghambat proses pemahaman teks, seperti contoh “BGT” jika dinormalisasi menjadi “BANGET”, kata perulangan “gara2” menjadi “gara-gara”.

e. *Labeling Data*

Proses *labeling data* dilakukan untuk menentukan sebuah dokumen teks mengandung sentimen positif, netral atau negatif.

f. *Klasifikasi Naive Bayes*

Klasifikasi dengan metode *Naive Bayes* dilakukan dengan menggunakan *tools* RapidMiner untuk mengetahui besar akurasi pembelajaran dengan *Naive Bayes*.

g. *Evaluasi Model Dengan Confusion Matrix*

Untuk mengidentifikasi pola-pola menarik yang ditemukan dalam hal ini adalah tentang reaksi masyarakat tentang adanya PSBB di DKI Jakarta.

## 1.7 Sistematika Penulisan

Sistematika dalam penulisan ini disusun untuk mempermudah pembacaan dan gambaran umum tentang penelitian yang akan dilakukan dengan sistematika

penulisan penelitian ini sebagai berikut :

## BAB I PENDAHULUAN

Dalam BAB I membahas tentang latar belakang penelitian, rumusan masalah, batasan masalah, tujuan penelitian yang ingin dicapai, manfaat penelitian, metodologi penelitian yang merupakan langkah untuk menyelesaikan masalah, dan sistematika penulisan.

## BAB II LANDASAN TEORI

Bab yang membahas berbagai konsep dasar dan teori-teori yang berkaitan dengan penelitian yaitu teori mengenai analisis sentimen, Twitter, *Naive Bayes*, klasifikasi, *Text Mining*, *Data Mining*, dan sebagainya.

## BAB III METODOLOGI

Bab ini menjelaskan tentang pengumpulan data dan metode analisis data, ini adalah tahapan-tahapan pada proses untuk melakukan sebuah penelitian.

## BAB IV HASIL DAN PEMBAHASAN

Merupakan hasil dari proses yang telah dilakukan dari penelitian ini. Lebih tepatnya adalah membahas hasil penelitian yang telah dilakukan dalam mendapatkan data yang di inginkan

## BAB V KESIMPULAN DAN SARAN

Bab yang membahas tentang penarikan kesimpulan dari hasil penelitian yang telah dilakukan, lebih membahas tentang kesimpulan keterkaitan masalah dan tujuan

penelitian yang dilakukan. Juga saran yang bersifat membangun untuk penelitian selanjutnya tentang penelitian ini agar dapat menutupi kekurangan-kekurangan yang ada.