

# BAB I. PENDAHULUAN

## 1.1. Latar Belakang

Didasari oleh perkembangan teknologi informasi begitu cepat bahkan sudah dianggap membumi. Hal itu dapat dilihat dari setiap sendi kehidupan masyarakat yang memanfaatkan teknologi informasi dan komunikasi didalam aktivitasnya sehari-hari[1]. Masyarakat umumnya memanfaatkan media sosial sebagai alat komunikasi ataupun pertukaran sebuah informasi. Pada masyarakat sendiri sudah banyak beredar media sosial yang dimanfaatkan seperti Instagram, facebook, twitter, dsb. Dengan media-media tersebut setiap masyarakat dapat saling berkomunikasi ataupun berbagi informasi mengenai banyak hal yang sedang *trend* seperti kesehatan, gaya hidup maupun opini pribadi dari setiap pengguna media itu sendiri [2].

Dengan banyaknya pengguna yang terus bertambah, media sosialpun menjadi populer tak terkecuali twitter. Sebagai bagian dari sekian banyaknya media sosial, twitter banyak digunakan untuk bertukar opini atau pendapat. Opini yang merupakan buah pikiran dari penggunanyapun diutarakan melalui twitter dengan dengan bentuk teks tetapi ada juga yang dimasukan tambahan video atau gambar, meski demikian secara massif opini yang dituangkan berbentuk teks. Salah satu topik yang banyak menjadi perbincangan di twitter saat ini yaitu mengenai pandemik *virus corona*. Sebuah pandemik yang berdampak besar bagi masyarakat, terutama karena penyebaran *virus corona* yang berjalan sangat cepat membuat kecemasan di berbagai wilayah di Indonesia.

*World Wealth Organization* (WHO) sendiri pada tanggal 11 Maret 2020 mengumumkan kalau *severe acute respiratory syndrome coronavirus 2*(SARSCoV-2) atau *virus corona* merupakan sebuah pandemik. Dalam prosesnya *virus corona* berakibat terinfeksi saluran pernapasan dari ringan sampai sedang dengan gejala seperti flu(Demam diatas 30°C, batuk kering dan sesak napas). Tetapi, *corona* juga dapat mengakibatkan infeksi pernapasan berat yang membuat gejala seperti *pneumonia*, *middle-east respiratory syndrome*

(MERS) dan *severe acute respiratory syndrome* (SARS). Siapa saja dapat terjangkit virus ini, baik dari bayi sampai lansia, namun virus ini rentan menyerang orang yang memiliki imunitas rendah seperti lansia atau anak-anak serta bagi mereka yang mengidap riwayat penyakit tertentu seperti diabetes, infeksi pernapasan akut, asma, penyakit jantung, hipertensi, kanker, serta berbagai kondisi medis lainnya yang menyerang sistem imunitas. Penularan virus ini melalui percikan atau *droplet* yang keluar semisal ketika kita batuk dan tidak dapat menyebar di udara (*airborne disease*). *Droplet* keluar bersamaan ketika batuk dapat menempel disetiap permukaan benda yang apabila tersentuh virus dapat berpindah dan apabila secara tidak sengaja mengusap wajah atau hidung atau makan dengan tangan yang sudah terkena virus, ketika imunitas seseorang sedang tidak optimal akan rentan terinfeksi *corona*. Pemerintah atau dokterpun selalu menghimbau untuk selalu menjaga kebersihan seperti misalnya dengan mencuci tangan lebih intensif, menyemprotkan disinfektan ke benda-benda yang sering kita sentuh, kemudian bagi mereka yang sedang batuk atau flu dianjurkan untuk menggunakan masker agar ketika batuk *droplet* tidak terpapar ke berbagai bidang. *Physical distancing* atau pemberian jarak minimal 4 meter antara seseorang dengan yang lainnya pun menjadi pilihan yang dianjurkan oleh WHO untuk menekan penyebaran *virus corona*. Hingga akhirnya belakangan ini kata-kata *lockdown* sering digaungkan pada sosial media. Banyak orang mendesak Pemerintah untuk melakukan *lockdown*, karena dirasa itu merupakan solusi yang dapat menurunkan penyebaran *virus corona* secara signifikan seperti negara-negara lain yang menerapkan sistem *lockdown*. *Lockdown* itu sendiri secara harafiah artinya dikunci, jadi apabila fakta yang terjadi negara Indonesia benar-benar *lockdown* maka segala aktivitas yang mengharuskan untuk keluar rumah tidak boleh dilaksanakan kecuali bagi mereka yang bertugas seperti tentara, dokter dan pemerintahan yang bergerak menangani wabah ini.

Di Indonesia sendiri, berbagai kalangan masyarakat mulai mengutarakan pertanyaan mengenai program PPKM (Pemberlakuan Pembatasan Kegiatan Masyarakat). Masyarakat mulai memperbincangkan opini mereka mengenai

PPKM di twitter secara masif. Kemudian opini-opini tersebut yang dimanfaatkan agar dapat mengetahui sentimen dari masyarakat. Proses sentimen diawali dengan mempolarisasikan negatif, positif atau netralnya sebuah opini, kemudian diklasifikasikan yang merupakan salah satu bentuk fungsi dari sebuah *text mining* [3].

Terdapat berbagai Teknik mengklasifikasikan sebuah data seperti *Naïve Bayes Classifier*, *Decision Trees*, *Support Vector Machines(SVM)*, *backpropagation neural network* dan masih banyak lagi [4]. Dari sekian banyak teknik yang dapat digunakan pada dokumen ini menggunakan teknik atau metode *naïve bayes classifier*. Di sebuah penelitian berjudul Metode Klasifikasi Data Mining Dan

Teknik Sampling Smote Menangani Class Imbalance Untuk Segmentasi *Customer* Pada Industri Perbankan, menghasilkan sebuah kesimpulan bahwa metode *naïve bayes classifier* merupakan yang tercepat dalam mengklasifikasikan data dibandingkan dengan metode SVM, J48 *decision tree* dan *synthetic minority oversampling technique (SMOTE)* yang kemudian menjadi salah satu alasan mengapa di dokumen ini metode *naïve bayes classifier* dimanfaatkan untuk melakukan analisa sentiment terhadap opini masyarakat mengenai PPKM [5].

Berdasarkan pesatnya perkembangan teknologi informasi seperti media sosial twitter, dimana twitter saat ini menjadi salah wadah bagi masyarakat melakukan *sharing* mengenai opininya terkait dengan isu atau topik panas yang sedang terjadi. Berbagai topik yang muncul di media twitter seperti topik PPKM akibat *virus corona* menjadi salah satu topik yang panas bagi masyarakat untuk menyampaikan opininya. Namun opini yang tersampaikan melalui twitter kita tidak dapat mengetahui apakah bernilai positif atau negative secara garis besar. Oleh karena itu dengan menggunakan metode *naïve bayes classifier* pada penelitian ini akan dapat menghasilkan informasi terkait topik panas yang sedang muncul mengenai PPKM, merupakan sesuatu hal yang positif atau negatif dimata masyarakat.

## 1.2. Rumusan Masalah

Didasari oleh latar belakang yang terjadi, maka rumusan masalah dapat dirangkum seperti berikut :

- 1). Bagaimana proses mengklasifikasi sentimen tentang opini masyarakat mengenai lockdown dengan metode naïve bayes classifier ?
- 2). Berapakah hasil polarisasi dan tingkat akurasi sentimen sebagai implementasi dari metode naïve bayes classifier ?

## 1.3. Batasan Masalah

BerAgar penelitian ini berjalan dengan maksimal, maka diperlukan Batasan sebagai berikut :

- 1). Sumber Dataset hanya berasal dari media twitter
- 2). Pemakaian pencarian data dengan kata kunci yang hanya berhubungan dengan opini PPKM di Indonesia
- 3). Pemanfaatan algoritma pemrograman python sebagai *collecting data*

## 1.4. Tujuan Penelitian

Berdasarkan rumusan masalah yang ada, penulis ingin menghasilkan sebuah nilai sentimen dari hasil analisa opini masyarakat yang beredar di media twitter mengenai PPKM, sebagai dampak dari penyebaran *virus corona*. Dengan metode *naïve bayes classifier* juga penulis ingin mengklasifikasikan polaritas serta mendapatkan nilai tingkat akurasi dari penggunaan metode ini.

## 1.5. Metode Penelitian

### 1.5.1. Deskripsi Problem

Seperti pada sistem filter spam yang terdapat diberbagai layanan email, ketika kita menandai suatu email sebagai spam maka program akan menganatomi email tersebut untukantisipasi email berikutnya yang masuk merupakan kategori spam atau bukan. Namun pada penelitian ini yang menjadi tanda yaitu kata-kata yang terdapat pada data, sehingga program diakhir akan dapat menentukan kata apa saja pada kalimat dan

kemudian mengklasifikasikannya, misalnya dari data terdapat kalimat dan pada kalimat tersebut mengandung kata “beri” maka program dapat melabeli atau masuk kategori positif.

#### **1.5.1.1. Analisis Data**

Pada bagian ini setiap data digabung menjadi satu dataset. Dari dataset yang sudah ada kemudian diidentifikasi setiap fitur yang terdapat pada data, dimana fitur merupakan subset dari atribut yang digunakan pada model.

#### **1.5.1.2. Preprocessing Data**

Merupakan langkah pembersihan data dari berbagai noise seperti simbol-simbol yang tidak terpakai di penelitian ini, Adapun gambaran dari preprocessing data sebagai berikut :

- 1). Cleansing atau pembersihan data dari noise seperti simbol-simbol.
- 2). Case Folding untuk merubah menjadi tipe serupa di tiap datanya, misalnya merubah setiap data menjadi lower case.
- 3). Tokenizing atau memotong teks menjadi per-kata.
- 4). Stemming atau merubah sebuah kata menjadi kata dasar saja.

#### **1.5.2. Dataset**

Seperti istilah *Garbage in – Garbage Out* yang memiliki arti bahwa sebuah input yang tidak optimal akan menghasilkan output yang tidak optimal juga, begitupun dengan data yang akan dipakai pada proses penelitian ini. Data yang digunakan merupakan data kategorik dan berbentuk string. Pada pengujian nya untuk nantinya menghasilkan model yang optimal maka pada penelitian ini data akan diuji dan dibagi menjadi 2 kategori yaitu training set dan test set, dimana pelatihan uji coba model yang pengujiannya terhadap pada test set.

### 1.5.3. Pengembangan Model

Dengan *Library Scikit Learn* data pada penelitian ini dibagi menjadi *train set* sebagai data *training* dan *test set* sebagai data *testing*. Pada prosesnya data *training* dilakukan pengujian model klasifikasi sentimen menggunakan *test set*.

#### 1.5.3.1. Pelatihan dan Evaluasi Model

Pada prosesnya model yang telah dibuat memiliki potensi tingkat *error* tertentu misalnya ketika diuji pada suatu *test set* menghasilkan *error* 5%, kemudian diuji lagi dengan *test set* yang berbeda dan hasil *error* meningkat jadi 22% maka dapat dikatakan model yang dibuat belum optimal. Agar dapat mengatasi model yang belum optimal pada penelitian ini menggunakan *validation set* yang merupakan bagian dari *train set* yang kemudian dipakai pada pengujian model tahap awal. Dengan sederhana menguji beberapa model dengan *hyperparameter* yang berbeda pada data *training* yang sudah dikurangi dengan data validasi. Setelah melakukan pengujian pada *validation set* kemudian kita dapat melatih model pada data *training* utuh untuk mendapatkan model final.

#### 1.5.3.2. Pengujian Model

Pada penggunaan *validation set* terdapat keraguan misalnya ketika ukuran *validation set* terlalu kecil maka kemungkinan kita akan memilih sesuatu model yang tidak optimal, sebaliknya jika *validation set* terlalu besar kemungkinan sisa data pada *train set* lebih kecil dari data *train set* utuh dimana kondisi tidak ideal untuk membandingkan suatu model, oleh karena itu pada penelitian ini menggunakan *K-Fold Cross Validation* atau sering disebut *Cross Validation* sebagai metode pengujian model dimana akan didapatkan persentase dari hasil pengujian yang

lebih optimal.

#### **1.5.4. Penarikan Kesimpulan**

Pada penarikan kesimpulan terakhir setelah data diuji, maka kita dapat mengambil kesimpulan dari hasil uji mengenai setiap data dalam bentuk persentasi yang menggambarkan data tersebut. Hasil kesimpulan akan dapat diketahui data cenderung positif, negatif atau netral.

### **1.6. Sistematika Penulisan**

#### **BAB I PENDAHULUAN**

Pada bab ini berisikan konsep dasar dari pembuatan dan penulisan skripsi yaitu berupa latar belakang, rumusan masalah, batasan masalah, tujuan penelitian, metode penelitian, dan sistematika penulisan.

#### **BAB II TINJAUAN PUSTAKA**

Pada bab ini berisikan beberapa penelitian yang telah terlaksana yang dapat digunakan sebagai acuan untuk penelitian yang dilakukan. Di dalam bab ini juga terdapat tabel perbandingan yang membandingkan penelitian yang telah ada sebelumnya dengan penelitian yang dilakukan.

#### **BAB III LANDASAN TEORI**

Pada bab ini berisikan penjelasan berupa teori-teori yang membantu penulis dalam melaksanakan penelitian.

#### **BAB IV ANALISIS DAN PERANCANGAN EKSPERIMEN**

Pada bab ini berisikan deskripsi problem, analisis data, preprocessing data, datasets, pengembangan model, pelatihan dan evaluasi model, serta pengujian model

#### **BAB V HASIL ANALISA DAN PEMBAHASAN**

Pada bab ini berisikan berisikan deskripsi eksperimen, hasil eksperimen, dan pembahasan eksperimen.

#### **BAB VI PENUTUP**

Pada bab ini berisikan kesimpulan dan saran yang didapatkan penulis dari penelitian yang telah dijalankan.



## **BAB II. TINJAUAN PUSTAKA**