

## BAB I PENDAHULUAN

### 1.1 Latar Belakang

*World Health Organization* (WHO) mendeklarasikan virus COVID-19 sebagai pandemi global pada 11 Maret 2020 WHO melaporkan lebih 52 juta orang terkonfirmasi positif Covid-19 dan 1,2 juta orang meninggal dunia pada minggu kedua bulan November 2020. Sementara Indonesia mencatat 463 ribu orang terkonfirmasi positif dengan korban meninggal telah mencapai 15.148 orang. Penyebaran virus secara cepat melalui kontak fisik memaksa semua negara menerapkan social distancing dan physical distancing untuk mengurangi interaksi. Faktor utama penularan melalui droplets yang dikeluarkan selama berbicara, batuk, atau bersin menjadikan penerapan pembatasan sosial merupakan strategi yang paling banyak diadopsi disaat belum ada vaksin [1].

Pengguna *Twitter* dapat memberikan komentar untuk opininya tentang film yang sedang populer di platform sosial media *Twitter*. Informasi yang didapat berguna untuk pengguna *Twitter* yang lainnya dan juga dapat digunakan untuk mengambil tindakan evaluasi bagi pembuat film dengan film yang akan dibuat. Tweet yang masih acak akan membuat kesulitan untuk mengetahui suatu tweet bernilai positif atau negatif berdasarkan apa yang sudah dijelaskan dalam penelitian *Twitter* sentimen analisis untuk mengklasifikasikan tweet review film berbahasa Indonesia. Data tersebut kemudian di proses dengan melakukan *Text Mining* lalu kemudian dilakukan klasifikasi Tweet berdasarkan pembagian sentimen

yaitu bermuatan Positif dan Negatif. Klasifikasi ini menggunakan algoritme *Support Vector Machine*. Klasifikasi di gunakan untuk memberikan kemudahan dalam melihat opini yang bernilai Positif dan Negatif Tingkat akurasi dari algoritme akan memberikan pengaruh pada hasil klasifikasi. [2].

Media platform *Twitter* adalah salah satu Microblogging yang dimana para pengguna nya dapat menuliskan tentang opini dalam suatu peristiwa atau isu-isu yang mana datanya dapat dilakukan untuk melakukan sentimen analisis yaitu mengidentifikasi pendapat, sikap dan keadaan emosi seseorang terhadap peristiwa tertentu atau objek analisa sentimen dapat dilakukan dengan melihat respon yang dihasilkan dari pengelompokan *tweet* dengan menggunakan *Natural Language Processing* untuk mengekstrak *tweet*, perhitungan sentimen yang di gunakan masyarakat yang berisi sebuah opini tentang isu-isu. Analisa ini mengelompokkan beberapa review yang bermuatan Positif, Negative dan Netral kemudian dibandingkan dan di gambarkan menggunakan grafik histogram [3].

Metode *Support Vector Machine* (SVM) digunakan dalam penelitian ini untuk mengklasifikasi teks sentimen. Dalam komputasi metode ini akan memakan waktu yang lama apabila fitur yang digunakan cukup banyak, dan digunakan fitur fitur untuk menyeleksi yang relevan, menghindari overfitting. Seleksi fitur yang di gunakan adalah seleksi fitur Weight by SVM yang di harapkan bisa meningkatkan nilai akurasi yang signifikan [4].

Algoritme *Support Vector Machine* dapat di gunakan untuk mengklasifikasi opini-opini tersebut. Berdasarkan penelitian yang di lakukan sebelum nya digunakan algoritme *Support Vector Machine* menghasilkan tingkat akurasi yang tinggi dalam melakukan sentimen analisis Penelitian yang di lakukan ditemukan algoritme *Support Vector Machine* disandingkan dengan fitur TF-IDF menghasil tingkat akurasi sebesar 86% lebih unggul dibandingkan dengan metode *Naïve Bayes* yang juga dikolaborasi dengan TF-IDF ketika melakukan analisis opini masyarakat mengenai Gojek pada *Twitter* yang dibagi menjadi dua kelas yaitu positif dan negatif. TF-IDF digunakan agar dapat membantu dalam meningkatkan akurasi apabila disandingkan dengan metode *Support Vector Machine* dibanding ekstrasi fitur seperti *Ratio* dan *N-Gram* [5].

*Python* Merupakan salah satu bahasa pemrograman yang dapat dijalankan pada berbagai sistem operasi baik Linux, MacOS ataupun Windows dengan melakukan konfigurasi terlebih dahulu. *Python* merupakan sebuah bahasa pemrograman tingkat tinggi dikarenakan kode-kode yang ditulis akan di compile menjadi byte code serta dieksekusi sehingga membuat *Python* cocok digunakan untuk bahasa pemrograman scripting, aplikasi web dan lain sebagainya [6].

Pada penelitian yang akan saya lakukan adalah untuk mencari nilai akurasi dari sentimen masyarakat pada platform sosial media *Twitter* yang dimana pada metode yang digunakan terdahulu seperti *Naive Bayes* dan juga penulis membandingkan sesama algoritme metode *Support Vector Machine*. Oleh karena itu penulis mengajukan penelitian terkait

menggunakan metode *Support Vector Machine* menggunakan tiga (3) kernel yang berbeda yaitu *Linear*, *Radial basis Function*, dan *Polynomial* pada analisis sentimen Twitter berbahasa Indonesia pada tweet yang berkaitan dengan Covid-19 Indonesia.

## 1.2 Rumusan Masalah

Berdasarkan latar belakang di atas, dapat dirumuskan permasalahannya yaitu. Berapakah tingkat akurasi terhadap opini tweet berbahasa Indonesia mengenai topik Covid-19 Indonesia menggunakan metode *Support Vector Machine* menggunakan kernel *Linear*, *Radial basis function* dan *Polynomial*?

## 1.3 Batasan Masalah

Beberapa batasan masalah yang digunakan pada penelitian ini adalah sebagai berikut:

1. Dalam penelitian ini menggunakan data yang diambil dari bulan Januari 2020 sampai dengan Agustus 2021
2. Pada penelitian ini penulis menggunakan tiga (3) kernel untuk algoritme *Support vector Machine* yaitu *Linear*, *Radial basis function*, *Polynomial*
3. Pada penelitian ini penulis menggunakan *tweet* yang diambil dari pengguna atau masyarakat Indonesia



#### 1.4 Maksud dan Tujuan Penelitian

Adapun maksud dari tujuan penelitian ini adalah sebagai berikut:

1. Mengetahui tingkat akurasi dari sentimen tweet berbahasa Indonesia dengan tema covid-19 Indonesia dengan menggunakan metode *Support Vector Machine*
2. Menguji metode *Support Vector Machine* untuk mengetahui tingkat akurasi dalam menganalisis sentimen pada twitter

#### 1.5 Manfaat Penelitian

Adapun manfaat dari penelitian ini adalah sebagai berikut:

1. Mendapatkan hasil akurasi pada sentimen dari opini penyebaran virus Covid-19 yang terjadi di Indonesia
2. Memberikan kontribusi pada *Text Mining* dalam masalah sentimen analisis menggunakan data tweet khususnya menggunakan metode *Support Vector Machine*

#### 1.6 Metode Penelitian

Dalam melakukan penelitian ini, Dilakukan melakukan beberapa tahapan-tahapan metode penelitian. Tahapan yang di lakukan sebagai:

##### 1.6.1 Metode Pengumpulan Data

Data yang digunakan dalam penelitian ini merupakan data yang berasal dari WebHarvy, dimana WebHarvy di gunakan untuk melakukan pengumpulan data. Data yang dikumpulkan berjumlah 8576 dataset, data dikumpulkan dilakukan dari bulan Januari 2020 sampai dengan bulan

Agustus 2021. Data yang di ambil merupakan sentimen atau tulisan tweet dari sosial media *Twitter* dan data berupa teks, dari data teks yang tidak terstruktur diolah menjadi data yang siap untuk di proses dalam pengujian penelitian

### **1.6.2 Labeling Dataset**

Setelah melakukan pengumpulan data yang didapat dari WebHarvy maka langkah berikutnya adalah pelabelan data yang berbentuk teks kemudian diolah menjadi data yang siap untuk digunakan melakukan penelitian, label yang di gunakan adalah menggunakan Positif dan Negatif untuk membedakan data dan agar mempermudah dalam melakukan proses latihan dan pengujian data

### **1.6.3 Preprocessing Data**

*Preprocessing* Data didalam penelitian ini dilakukan untuk mengubah data mentah menjadi data siap diolah untuk pengujian dalam tahapan *Preprocessing* terdapat tahapan yaitu *Case Folding*, *stopword removal*, *tokenizing*, *normalisasi* dan *stemming*

### **1.6.4 Pengujian Metode Support Vector Machine**

Setelah melewati proses *preprocessing*, kemudian dilakukan pengujian metode untuk mendapatkan nilai akurasi menggunakan metode *Support Vector Machine (SVM)*, Metode *Support vector machine* yang digunakan menggunakan *Support Vector Machine Kernel Linear*, *Support Vector Machine Kernel Radial basis function* dan *Support Vector Machine Kernel Polynomial*, setelah di lakukan pengujian metode kemudian ditampilkan untuk mengetahui akurasi menggunakan *confusion matrix*

### 1.6.5 Evaluasi

Hasil dari pengujian metode selanjutnya dilakukan proses evaluasi. Pada tahap ini dilihat keberhasilan algoritme untuk klasifikasi sentimen twitter dengan data yang sesuai dengan kelas nya dengan total data uji yang digunakan menggunakan algoritme *confusion matrix*. Dari hal tersebut dapat dilihat tingkat akurasi dan performa algoritme yang digunakan. Dari hal tersebut dapat dilihat akurasi atau performa dari algoritme yang digunakan terhadap data yang ada.

### 1.7 Sistematika Penulisan

Sistematika penulisan yang digunakan dalam penelitian ini terbagi menjadi beberapa bagian, yaitu:

#### BAB I PENDAHULUAN

Bab ini membahas tentang uraian latar belakang, rumusan masalah, batasan masalah, maksud dan tujuan penelitian, manfaat penelitian, metode penelitian serta sistematika penulisan.

#### BAB II LANDASAN TEORI

Bab ini berisi tentang teori-teori yang digunakan serta konsep dasar yang menjadi landasan dari penelitian ini.

### BAB III METODE PENELITIAN

Bab ini berisi tentang pembahasan metode-metode yang digunakan dalam melaksanakan penelitian ini.

### BAB IV HASIL DAN PEMBAHASAN

Bab ini membahas tentang hasil yang didapatkan dari penelitian yang sudah dilakukan oleh penulis.

### BAB V PENUTUP

Bab ini merupakan bab terakhir berisi tentang kesimpulan dari penelitian yang telah dilakukan dan juga berisi saran yang bisa diberikan oleh peneliti kepada peneliti selanjutnya.

