

BAB I

PENDAHULUAN

1.1. LATAR BELAKANG

Media sosial adalah sebuah fitur pada internet yang memungkinkan pengguna internet saling berbagi informasi. Media sosial juga merupakan salah satu alat dimana seseorang dapat saling terhubung dengan segala hal. Dengan media sosial pengguna dapat saling berinteraksi dengan orang lain di seluruh dunia dengan berbagi foto, video pesan singkat dan sebagainya. Media sosial memberikan dampak yang signifikan terhadap keberlangsungan hidup para generasi millennial hal ini terjadi karena masyarakat khususnya anak muda telah tumbuh dan terintegrasi dengan jaringan media sosial. Hasil riset menunjukkan 53 responden yang terdiri dari 35 responden perempuan dan 18 laki-laki menunjukkan mereka telah menggunakan sosial media mulai dari umur 15-20 tahun. Pertumbuhan media sosial juga dibarengi dengan perkembangan smartphone yang semakin mendorong pertumbuhan pengguna media sosial [1].

Sekarang ini terdapat banyak sekali sosial media yang ada di Internet. Facebook, Twitter, Instagram, Youtube dan masih banyak lagi sosial media yang lainnya. Dilansir dari HootSuite Indonesia tahun 2021, Youtube menempati posisi pertama sebagai sosial media dengan pengguna terbanyak yaitu sebesar 107 juta pengguna. Berdasarkan statistik tersebut Youtube dapat menjadi sumber data yang berasal dari komentar para netizen. Kolom komentar menjadi wadah untuk

memberikan opini, respon dan tanggapan terhadap suatu konten video yang diunggah di Youtube. Dengan mengukur seberapa banyak komentar negatif dan positif pada suatu konten di sosial media dapat menjadi sebuah alternatif untuk menilai sentimen publik terhadap konten tersebut, salah satunya vaksin covid-19.

Seperti yang diketahui, vaksin menjadi topik yang menarik untuk dibahas akhir-akhir ini. Banyak stigma yang muncul terhadap vaksin di masyarakat. Maka dari itu untuk mengetahui polaritas masyarakat terhadap vaksin ini perlu dilakukannya analisis sentimen dengan mengolah data komentar. Namun dengan banyaknya jumlah data komentar dan tidak memungkinkannya untuk menghitung satu persatu maka digunakanlah pendekatan *Text Mining* untuk mengolah banyaknya komentar tersebut apakah dikategorikan sebagai komentar positif atau negatif. Dengan menggunakan metode *Support Vector Machine* dataset komentar tersebut akan diolah sehingga menghasilkan pola klasifikasi yang nantinya akan digunakan untuk mengukur tingkat polaritas terhadap topik vaksin covid-19 pada platform Youtube.

1.2. RUMUSAN MASALAH

Rumusan masalah dari latar belakang diatas adalah sebagai berikut:

1. Bagaimana penerapan *Text Mining* dalam penyelesaian masalah berupa pengklasifikasian komentar positif dan negatif pada komentar Youtube menggunakan metode *Support Vector Machine*.

1.3. BATASAN MASALAH

Batasan masalah dalam penulisan skripsi ini adalah sebagai berikut :

1. Pengolahan *Text Mining* ini menggunakan bahasa Python dan menggunakan *tools* berupa Jupyter Notebook.
2. Data yang akan diolah berasal dari komentar pada sebuah konten video di Youtube menggunakan Youtube Data API v3.
3. Penelitian ini akan mengambil dataset komentar dari Youtube Data API v3 dan kemudian akan dilakukan *Text Preprocessing* untuk pembersihan teks. Kemudian akan masuk ke pembobotan *term* untuk mengukur bobot setiap *term* dan terakhir akan dilakukannya pengolahan opini menggunakan *Support Vector Machine* untuk mencari pola kata dan kemudian melakukan prediksi sentimen dari sebuah komentar.
4. Hasil akhir akan ditentukan berdasarkan sentimen terbanyak.

1.4. TUJUAN PENELITIAN

Dari latar belakang dan rumusan masalah diatas, maka tujuan dari penelitian ini adalah sebagai berikut :

1. Melakukan pengolahan opini dari komentar video di Youtube dan menyimpulkan respon dari publik apakah bernilai positif atau negatif.
2. Penelitian ini diharapkan dapat menjadi masukan untuk mengukur tingkat sentimen publik terhadap suatu konten video di Youtube.

1.5 MANFAAT PENELITIAN

Manfaat dari penelitian ini adalah sebagai berikut :

1. Mengetahui respon publik terhadap suatu konten video di Youtube dengan mengolah data komentar.
2. Memberikan informasi berupa kata apa saja yang sering muncul pada kolom komentar beserta sentimen nya.

1.6 METODOLOGI PENELITIAN

Dalam melakukan penelitian ini menggunakan beberapa metode untuk mendapatkan hasil yang diharapkan. Maka penulis menyusun metode yang dilakukan berupa :

1. Studi Literatur

Pada tahap Studi Literatur penulis mencoba memahami konsep dari *text mining*, algoritma, bahasa pemrograman yang digunakan dan *tools* pendukung. Python menjadi bahasa yang populer digunakan untuk *data science*. Hal ini dikarenakan Python menyediakan banyak fungsi analisis data dan fungsi *machine learning*, *data preprocessing tools*, serta visualisasi data. Algoritma *Support Vector Machine* menjadi salah satu pilihan untuk mengklasifikasikan data. Kelebihan dari metode SVM adalah SVM dapat menentukan *hyperplane* atau bidang pemisah dengan memilih bidang dengan optimal *margin* [2].

2. Pengumpulan Data

Data yang akan digunakan dalam pengolahan opini adalah data yang berasal dari komentar video Youtube. Data tersebut dapat diperoleh melalui Youtube Data API v3 yang disediakan oleh Google Console. Youtube Data API v3 memungkinkan untuk mengambil data berupa judul video, *keyword*, id video, komentar dan disediakan secara gratis. Hasil dari pengambilan sampel data komentar tersebut nantinya akan disimpan kedalam file *.csv*.

3. *Text Preprocessing*

Tahapan *text preprocessing* adalah tahapan untuk membersihkan data dari *noise*. *Text preprocessing* perlu dilakukan untuk mendapatkan hasil yang optimal dengan membuang beberapa kata yang tidak diperlukan. Tahapan-tahapan *text preprocessing* antara lain:

1. *Case Folding*

Case Folding adalah tahapan mengubah huruf kapital menjadi huruf kecil.

2. *Text Cleaning*

Proses pembersihan teks dari angka dan karakter yang dianggap tidak memiliki arti.

3. *Tokenizing*

Tokenizing adalah tahapan memecah teks pada dokumen menjadi token secara berurut[1]. Dalam hal ini sebuah kalimat yang terdiri dari beberapa kata akan dipecah menjadi bagian per satu kata.

4. *Normalisasi*

Proses menormalkan kata berupa kata singkatan.

5. *Stopword Removal*

Stopword Removal adalah proses menghilangkan kata hubung atau kata bantu yang kurang bermakna.

6. *Stemming*

Stemming adalah proses mengubah kata menjadi kata dasar atau kata baku dengan menggunakan bantuan library Sastrawi.

4. *Term Frequency Inverse Document Frequency*

Term Frequency Inverse Document Frequency atau TF-IDF adalah suatu metode yang digunakan untuk menilai bobot setiap kata yang umum digunakan. Metode ini akan menghitung seberapa sering suatu kata tersebut muncul dalam sebuah dokumen.

5. *Support Vector Machine*

Setelah data ditimbang dengan TF-IDF maka tahapan selanjutnya adalah melakukan klasifikasi dengan *Support Vector Machine* (SVM). Pada tahapan ini akan dicari bidang pemisah atau disebut *hyperplane* yang membagi antara kata negatif dan positif.

1.7 SISTEMATIKA PENULISAN

Dalam penyusunan skripsi ini ada sistematika penulisan yang digunakan sebagai acuan pokok dalam penyusunan skripsi ini antara lain:

BAB I PENDAHULUAN

Bab ini akan menguraikan tentang latar belakang masalah, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, metodologi penelitian dan sistematika penulisan.

BAB II LANDASAN TEORI

Bab ini akan membahas teori dan metode yang digunakan di dalam penelitian, dapat berupa definisi dan telaah penelitian yang berkaitan dengan penelitian ini.

BAB III METODE PENELITIAN

Bab ini akan membahas mulai dari analisis data yang akan digunakan selama penelitian, sumber data dan perancangan metode secara umum.

BAB IV HASIL DAN PEMBAHASAN

Bab ini akan membahas hasil dari penelitian yang telah dilakukan berupa hasil dari sentimen analisis terhadap video di Youtube.

BAB V PENUTUP

Bab ini berisi kesimpulan dan saran selama penelitian berlangsung. Kesimpulan dapat dikemukakan masalah yang ada pada penelitian serta hasil dari penyelesaian penelitian.

