

BAB I

PENDAHULUAN

4.1 Latar Belakang

Pada awal tahun 2020, dunia dikejutkan dengan munculnya suatu wabah yang kemudian menyebar dengan cepat. Wabah ini bernama coronavirus disease 2019 atau saat ini dikenal dengan COVID-19 yang disebabkan oleh Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-CoV-2). *World Health Organization* (WHO) telah menetapkan secara resmi bahwa wabah virus ini menjadi pandemi pada tanggal 11 Maret 2020 [1]. Sejak munculnya pandemi virus corona terus berkembang pemerintah telah mencoba mengatasi wabah dengan mempertimbangkan serangkaian tindakan. Pesatnya pertumbuhan jumlah kasus secara global telah menimbulkan kepanikan, ketakutan, dan kecemasan di kalangan masyarakat[2].

Kondisi yang demikian memberikan dampak langsung kepada seluruh masyarakat karena pemberlakuan protokol kesehatan yang harus ditetapkan pada seluruh aspek kegiatan, mulai dari pembatasan sosial hingga lockdown total sehingga menghambat seluruh kegiatan masyarakat. Efek dari COVID-19 ini berpotensi membawa tantangan besar bagi sistem kesehatan dunia dan memiliki konsekuensi yang luas pada ekonomi global jika penyebaran virus tidak dikendalikan secara efektif.

Melihat pesatnya penyebaran COVID-19 dan bahaya yang akan muncul apabila tidak segera ditangani, salah satu cara yang sangat mungkin untuk mencegah penyebaran virus ini adalah dengan menerapkan vaksin. Vaksin tidak hanya melindungi mereka yang divaksinasi tetapi dapat mengurangi penyebaran virus COVID-19. Vaksin yang aman dan efektif sangat penting dilakukan karena diharapkan dapat menghentikan penyebaran dan mencegah penyebaran penyakit di masa mendatang. Selain itu virus menyebar dengan sangat cepat maka diperlukan vaksin yang dapat diterapkan dalam waktu singkat sehingga dapat meminimalisir dampaknya [3].

Dalam menanggapi hal tersebut pemerintah menetapkan kebijakan vaksinasi kepada seluruh masyarakat Indonesia [4]. Tetapi kegiatan vaksinasi tersebut haruslah mempertimbangkan segala aspek, mulai dari aspek kelayakan vaksin yang akan digunakan, resiko pasca pemakaian, sampai tahapan dan prosedur dari pemberian vaksin hingga nantinya sampai ke masyarakat. Disamping

itu juga harus mempertimbangkan berbagai masukan, di antaranya adalah dengan melihat bagaimana respon dan opini masyarakat terhadap vaksinasi tersebut.

Sejak Pemerintah Indonesia menerapkan kebijakan vaksin, banyak masyarakat yang memberikan opininya terhadap vaksin di media sosial, baik itu yang bersifat positif ataupun negative. Salah satu media sosial yang paling banyak digunakan oleh masyarakat Indonesia adalah media sosial twitter, yang mencakup lebih dari 52% dari total pengguna media sosial di Indonesia. Hal tersebut menunjukkan adanya peluang sumber data yang sangat besar sehingga dapat dimanfaatkan untuk menghasilkan suatu informasi yang bermanfaat [3].

Atas hal tersebut, maka penelitian ini akan menganalisis bagaimana respon dan opini masyarakat terhadap vaksin COVID-19 dengan memanfaatkan data yang bersumber dari media sosial twitter. Disamping mudah diakses jumlah pengikutnya tidak terbatas, dan jumlah karakter yang digunakan hanya 280 karakter sehingga dapat menyampaikan maksud dengan singkat, padat, dan jelas.

Penelitian analisis sentimen terhadap vaksin sebelumnya pernah dilakukan oleh Muhammad Dani Mulyawan dengan judul penelitian Analisis Sentimen Terkait Vaksin Covid-19 Pada Data Twitter Menggunakan Support Vector Machine [5]. Selain itu penelitian juga dilakukan oleh Putri Amira Sumitro dengan judul Penelitian Analisis Sentimen Terhadap Vaksin Covid-19 di Indonesia Pada Twitter menggunakan metode Lexicon Based [6].

Dalam penelitian ini akan melakukan analisis sentimen dengan dengan mengklasifikasikan respon masyarakat tersebut ke dalam sentimen positif & negatif, dan mengelompokkan opini masyarakat terhadap vaksin COVID-19 menggunakan algoritma Multinomial Naïve Bayes. Algoritma Multinomial Naïve Bayes merupakan pengembangan dari algoritma Naïve Bayes sehingga sangat memungkinkan digunakan pada pembuatan analisis sentiment karena algoritma ini bertujuan untuk mengklasifikasikan respon masyarakat kedalam kategori positif atau negatif.

Kelebihan multinomial naive bayes diantaranya yaitu tingkat akurasi yang lebih baik diantara model classifier lain, bekerja sangat baik dibanding model lainnya, mudah diimplementasikan, Hal ini dibuktikan oleh Xhemali, Hinde Stone dalam jurnalnya "Naïve Bayes Vs. Decision Trees vs. Neural Networks in the classification of Training Web Pages". Selain itu metode ini hanya membutuhkan jumlah data pelatihan (training data) yang kecil untuk menentukan

estimasi parameter yang diperlukan dalam proses pengklasifikasian. Karena yang diasumsikan sebagai variable independent, maka hanya varians dari suatu variable dalam sebuah kelas yang dibutuhkan untuk menentukan klasifikasi, bukan keseluruhan dari matriks kovarians [7].

Berdasarkan uraian diatas penelitian ini berfokus pada “Analisis Sentimen pada Pengguna Media Sosial Twitter Terhadap Vaksin Covid-19 Dengan Menggunakan Algoritma Multinomial Naive Bayes Classifier” untuk penelitian ini algoritma naïve bayes yang digunakan adalah algoritma Multinomial yang merupakan pengembangan dari algoritma Naïve Bayes Classifier dengan tujuan untuk mengklasifikasi opini masyarakat pengguna sosial media twitter kedalam dua sentimen yaitu sentimen positif dan sentimen negative.

4.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dijabarkan diatas, maka permasalahan dapat dirumuskan sebagai berikut :

1. Bagaimana cara mengimplementasikan algoritma Multinomial Naive Bayes kedalam sebuah sistem untuk mengklasifikasi sentimen terhadap vaksin COVID-19?
2. Berapa akurasi yang diperoleh menggunakan algoritma Multinomial Naïve Bayes?

4.3 Batasan Masalah

Berdasarkan dari latar belakang dan rumusan masalah yang telah dijabarkan diatas, maka peneliti membatasi luasnya cakupan ruang lingkup yang akan dibahas pada skripsi ini agar tidak melebar ke topik lain dan memudahkan penulis dalam pengerjaannya. Adapun batasan-batasan masalah sebagai berikut:

1. Bahasa yang diambil dari twitter adalah Bahasa Indonesia.
2. Sentimen Twitter yang dianalisis hanya sentimen yang berhubungan dengan kebijakan vaksin COVID-19.
3. Algoritma klasifikasi yang digunakan hanya Algoritma Multinomial Naïve bayes.
4. Dari sentiment hanya diklasifikasikan kedalam 2 kelas, yaitu sentiment positif dan sentiment negative.

4.4 Maksud dan Tujuan Penelitian

Adapun tujuan penelitian dalam penyusunan skripsi ini yaitu untuk mendapatkan hasil klasifikasi komentar terhadap vaksin COVID-19 dan mencari nilai akurasi yang diperoleh dari

Algoritma Multinomial Naïve Bayes dalam melakukan klasifikasi sentiment terhadap vaksin COVID-19.

4.5 Manfaat Penelitian

Adapun manfaat penelitian dalam penyusunan skripsi ini yaitu diharapkan setelah adanya penelitian ini dapat menjadi referensi untuk penelitian selanjutnya di masa yang akan datang.

4.6 Metode Penelitian

Metode penelitian merupakan langkah yang dilakukan oleh peneliti untuk mengumpulkan informasi atau data. Metode yang digunakan untuk memperoleh data yang digunakan dalam penulisan skripsi ini adalah:

4.6.1 Metode Pengumpulan Data

Metode yang digunakan dalam pengumpulan data untuk merencanakan dan menganalisis adalah sebagai berikut:

4.6.1.1 Metode Kepustakaan

Metode Kepustakaan merupakan metode pengumpulan data dengan mencari dan membaca buku-buku, referensi artikel serta jurnal yang berhubungan dengan analisis sentiment, algoritma Naïve Bayes.

4.6.1.2 Crawling Data

Crawling data merupakan teknik pengumpulan data dari suatu database. Pengumpulan data dari penelitian ini yaitu data yang diunduh dari server twitter berupa user dan tweet beserta atribut-atributnya. Untuk mengakses informasi dari twitter menggunakan twitter API.

4.6.2 Analisis Kebutuhan

Analisa kebutuhan merupakan proses untuk mendapatkan informasi, model, spesifikasi mengenai perangkat lunak yang diinginkan oleh pengguna. Informasi yang diperoleh pengguna akan menjadi acuan untuk melakukan perancangan desain perangkat lunak.

4.6.2.1 Analisis Data

Metode analisis data dalam penelitian ini adalah metode Multinomial Naïve Bayes Classifier yang digunakan untuk melakukan klasifikasi terhadap opini masyarakat pada twitter yang dibagi menjadi 2 kelas yaitu, opini positive dan opini negatif.

4.6.3 Labeling Data

Pengumpulan data dilakukan dengan crawling data tweet menggunakan fasilitas Application Program Interface (API) yang disediakan oleh Twitter. Dalam penelitian ini mengumpulkan data tweet sebanyak 500 tweet yang tersimpan di file.csv. Kemudian tweet yang digunakan sebagai training data diberi label positif dan negatif. Dalam pelabelan setiap tweet dilakukan secara manual.

4.6.4 Preprocessing Data

Tahap Preprocessing bertujuan untuk menyeleksi data yang akan dilakukan pembersihan data tweet yang digunakan pada tahap Analisa sentimen / pengklasifikasian tweet. Data yang diperoleh ditampilkan dalam keadaan apa adanya, sehingga mungkin berupa data yang tidak diinginkan karena merupakan data yang tidak relevan. Data yang tidak relevan tersebut terkadang merupakan data yang sulit ditangani. Oleh karena itu, penghapusan bagian data yang tidak relevan sangat disarankan untuk dilakukan. Berikut struktur yang terdapat pada data tweet yaitu User ID (disimbolkan dengan @), URL, teks, tanggal dan waktu, lokasi, file multimedia (gambar, video, dll), emoticon, hashtag (disimbolkan dengan #). Masing-masing bagian struktur tersebut memiliki signifikansi sendiri selama analisis sentimen tetapi ada beberapa data yang tidak memiliki efek signifikan sehingga disarankan untuk menghilangkan data ini saat melakukan analisis sentiment.

Ada beberapa tahapan dalam preprocessing ini, antara lain sebagai berikut:

1. Stopword Removal

Stopword Removal merupakan proses menghapus kata-kata yang terlalu umum dan kurang penting, ciri-ciri pada kata ini adalah frekuensi kemunculannya yang jumlahnya cukup banyak dibandingkan dengan kata yang lainnya, contoh kata : aku, kamu, dengan, yang,dan, di, dari, dll.

2. Cleansing

Cleaning data merupakan proses menghilangkan tanda baca atau karakter selain teks dengan fungsi punctuation removal.

3. Case Folding

Case Folding yang merupakan proses untuk merubah setiap kata menjadi sama, misal huruf kecil dengan menggunakan fungsi lowercase.

4. Mengubah kata kekinian sesuai PUEBI

Merupakan proses Mengubah kata kekinian sesuai PUEBI merupakan proses untuk merubah setiap kata gaul yang tidak sesuai aturan menjadi sesuai PUEBI (Pedoman Umum Ejaan Bahasa Indonesia).

4.6.5 Pembobotan Kata

TF-IDF (Term Frequency Inverse Document Frequency) merupakan metode yang digunakan untuk menentukan nilai frekuensi sebuah kata di dalam sebuah dokumen atau artikel dan juga frekuensi di dalam banyak dokumen. Perhitungan ini menentukan seberapa relevan sebuah kata di dalam sebuah dokumen [8]. TFIDF adalah sebuah algoritma yang umumnya digunakan untuk pengolahan data besar [8].

TF-IDF pada dasarnya merupakan hasil dari perhitungan antara TF (Term Frequency) dan IDF (Inverse Document Frequency). Banyak cara untuk menentukan nilai yang tepat dari kedua statistik yang ada. Dalam kasus term frequency $tf(t, d)$, cara yang paling sederhana adalah dengan menggunakan raw frequency di dalam dokumen, yaitu berapa kali term t muncul di dokumen d . Jika menyatakan raw frequency t sebagai $f(t, d)$, maka skema tf yang sederhana adalah $tf(t, d) = f(t, d)$. Nilai idf sebuah term (kata) dapat dihitung menggunakan persamaan sebagai berikut:

$$IDF = \text{Log } 10 \left(\frac{D}{dfi} \right) \quad (1)$$

D adalah jumlah dokumen yang berisi term (t) dan dfi adalah jumlah kemunculan (frekuensi) kata terhadap D . Adapun algoritma yang digunakan untuk menghitung bobot (W) masing-masing dokumen terhadap kata kunci (query), yaitu :

$$W_{d,t} = tf_{d,t} \cdot IDF_t$$

Keterangan :

d = dokumen ke- d

t = kata ke- t dari kata kunci

W = bobot dokumen ke- d terhadap kata ke- t

tf = term frekuensi/frekuensi kata

Setelah bobot (W) masing-masing dokumen diketahui, maka dilakukan proses pengurutan (sorting) dimana semakin besar nilai W , semakin besar tingkat kesamaan (similarity) dokumen tersebut terhadap kata yang dicari, demikian pula sebaliknya.

4.6.6 Metode Testing

Metode testing dilakukan dengan cara, yaitu:

1. Confusion matrix merupakan suatu metode yang digunakan untuk mengukur kinerja atau performa suatu metode klasifikasi. Pada dasarnya confusion matrix memberikan informasi perbandingan hasil klasifikasi yang dilakukan oleh sistem dengan hasil klasifikasi yang sebenarnya.

4.7 Sistematika Penulisan

Sistematika dalam penulisan yang digunakan dalam pembuatan skripsi ini dibagi menjadi lima bab, tujuannya untuk memudahkan penulis dalam pembahasa. Penulisanannya antara lain sebagai berikut :

BAB I PENDAHULUAN

Pada bab pendahuluan akan menjelaskan mengenai latar belakang masalah, perumusan masalah, batasan masalah, maksud dan tujuan, manfaat penelitian, metode dan sistematika penulisan penelitian.

BAB II LANDASAN TEORI

Pada bab landasan teori ini akan menjelaskan mengenai teori yang digunakan sebagai dasar dalam analisis penelitian. Tinjauan teori diperoleh dari buku-buku pustaka, jurnal dan penelitian-penelitian sebelumnya terkait dengan masalah penelitian.

BAB III METODE PENELITIAN

Pada bab metodologi penelitian akan menjelaskan mengenai beberapa tahapan yang perlu dilakukan untuk mendapatkan hasil pengujian yang terbaik.

BAB IV IMPLEMENTASI DAN PEMBAHASAN

Bab implementasi dan pembahasan berisi mengenai hasil penelitian secara menyeluruh serta hasil pengujian.

BAB V PENUTUP

Bab ini akan menjelaskan kesimpulan yang didapat dari hasil analisis penelitian ini, serta saran yang diharapkan untuk pengembangan penelitian yang lebih baik lagi.

