

BAB I

PENDAHULUAN

1.1 Latar Belakang

Pada akhir tahun 2019, seluruh dunia dihebohkan dengan munculnya virus covid-19. Virus ini pertama kali muncul di negara China tepatnya di Wuhan, Ibukota Provinsi Hubei Tiongkok. Tercatat virus ini berhasil masuk pertama kali ke indonesia pada tanggal 20 maret 2020 di Depok, Jawa Barat. Virus ini bisa masuk ke indonesia diduga karena salah satu WNI melakukan kontak langsung dengan warga negara jepang yang datang ke indonesia [1].

Pemerintah kemudian menanggapi dengan melakukan beberapa kebijakan yaitu diantaranya Pembatasan Sosial Berskala Besar (PSBB), wajib masker, dan jam malam. Dengan adanya kebijakan tersebut banyak menimbulkan pendapat dari masyarakat salah satunya melalui media sosial twitter [1]. Pemilihan media sosial Twitter ini dikarenakan akses ke twitter itu sendiri tergolong mudah, kemudian jumlah pengikut yang tidak dibatasi dan penyampaian pendapat dari masing masing orang juga singkat, padat, dan jelas karena batas jumlah tweet nya hanya 140 karakter [2]. Faktor paling utama masyarakat menyampaikan opini-opininya adalah karena akses yang begitu mudah dalam menyampaikan pendapat yaitu dengan melalui media sosial salah satunya twitter. Respon dari masyarakat terkait kebijakan penanganan covid-19 di Indonesia pun sangat beragam, mulai dari respon positif seperti memberi dukungan dan semangat hingga respon negatif seperti memberikan kritik bahkan cacian. Respon masyarakat dalam menanggapi terkait kebijakan penanganan covid-19 di Indonesia pun dapat digunakan sebagai pertimbangan pemerintah untuk langkah kedepannya dengan cara melihat lebih banyak lagi respon positif atau negatif yang disampaikan. Selanjutnya dibutuhkan analisis sentimen yang berguna untuk mengkategorikan sebuah tweet di twitter tersebut termasuk ke dalam opini positif atau negatif.

Analisis sentimen merupakan proses untuk mengetahui suatu pendapat atau opini yang disampaikan secara tertulis baik berupa teks maupun dokumen apakah mempunyai sentimen positif atau negatif [3]. Penelitian yang berhubungan dengan analisis sentimen belakangan ini menjadi perhatian karena menghasilkan data seperti opini masyarakat, promosi politik, dan pengambilan keputusan [4]. Tujuan dari analisis sentimen adalah menganalisis suatu komentar dan mengelompokkannya sebagai suatu sentimen positif atau negatif [5]. Analisis sentimen dapat di terapkan dalam pengklasifikasian opini-opini atau tweets masyarakat di twitter terhadap kebijakan penanganan covid-19 di indonesia. Proses analisis sentimen juga membutuhkan suatu model atau suatu metode perhitungan probabilitas yang berfungsi saat proses klasifikasi sebuah tulisan.

Algoritma atau metode yang sering digunakan untuk klasifikasi data, antara lain algoritma *Naïve Bayes*, *Decision Tree C4.5* dan *K-Nearest Neighbors (KNN)* [6].

Penelitian ini menggunakan model algoritma *Naïve Bayes* sebagai algoritma pengklasifikasian sentimen pada *tweet* media sosial twitter. Algoritma *Naïve Bayes* merupakan algoritma sederhana yang menghitung beberapa kemungkinan dengan menghitung jumlah banyaknya data yang muncul serta gabungan dari setiap nilai yang ada dalam sebuah data set [6]. Data dari twitter yang sebelumnya berbentuk teks akan terlebih dahulu diolah, kemudian di preprocessing terlebih dahulu melalui beberapa tahapan preprocessing yang terdiri dari yaitu *Cleansing*, *Case Folding*, *Tokenizing*, *Slangword*, *Stopword*, dan *Stemming*. Preprocessing ini bertujuan agar data tersebut menjadi lebih rapi dan terstruktur. Kemudian selanjutnya di ekstraksi dan dibobotkan sebelum masuk ke proses perhitungan probabilitas menggunakan *naïve bayes*. Salah satu contoh metode ekstraksi dan pembobotan kata adalah *Tf-idf*. *Tf-idf* adalah metode yang digunakan untuk pembobotan kata yang bertujuan untuk mengetahui nilai atau bobot suatu kata dalam dokumen atau kalimat, dimana dokumen atau kalimat tersebut terdapat dalam sekumpulan data. Pemilihan pembobotan *tf-idf* itu sendiri dikarenakan *Tf-idf* merupakan metode pembobotan kata yang dikenal sangat mudah, efisien, dan tentunya memiliki hasil yang akurat. [7]. Fitur yang dihasilkan dari metode *Tf-idf* akan sangat banyak karena data yang digunakan berupa data teks. Oleh karena itu diperlukan proses seleksi fitur untuk memilih fitur-fitur atau atribut-atribut yang dianggap paling penting. Proses seleksi fitur menggunakan metode *information gain*. *Information gain* ini sendiri dapat menemukan fitur-fitur yang paling banyak memiliki informasi berdasarkan kelas tertentu sehingga juga dapat membuang fitur-fitur yang tidak relevan.

Algoritma *Naïve Bayes* digunakan dalam penelitian ini karena banyaknya data yang berupa *tweet* dari twitter yang ingin ditentukan kelas atau labelnya. Kelebihan dari Algoritma *Naïve Bayes* dibandingkan dengan algoritma lain adalah tingkat akurasi *naïve bayes classifier* yang cukup tinggi dan cukup baik digunakan pada perhitungan data dalam jumlah besar [3].

Penelitian sebelumnya pada tahun 2018 yang dilakukan oleh Teguh Ansyor Lorosae, Burhanudin Dwi Prakoso, Saifudin, dan Kusri, tingkat akurasi *naïve bayes* mencapai 95% [8]. Penelitian kedua dilakukan pada tahun 2019 yang dilakukan oleh Nanang Ruhyana, tingkat akurasi *naïve bayes* mencapai 86,67% [9]. Penelitian analisis sentimen ini nantinya menghasilkan suatu informasi yang bermanfaat kepada masyarakat tentang opini pengguna twitter terhadap kebijakan penanganan covid-19 di indonesia.

1.2 Perumusan Masalah

1.2.1 Permasalahan

Kebijakan penanganan covid-19 di Indonesia menuai banyak sekali tanggapan atau sentimen dari masyarakat salah satu contohnya melalui twitter. Tanggapan yang disampaikan pun sangat beragam, dari yang berupa tanggapan positif maupun negatif. Sehingga analisis sentimen ini sangat diperlukan untuk membantu dalam proses perhitungan apakah lebih banyak tanggapan yang positif atau negatif. Algoritma naïve bayes dipilih karena dapat membantu dalam proses pengklasifikasian sentimen yang berupa data teks. Sebelum itu, data yang berupa teks harus diekstraksi terlebih dahulu dan setelah itu dilakukan pembobotan kata menggunakan metode *tf-idf*. Fitur yang dihasilkan dari proses *tf-idf* akan diseleksi untuk memilih fitur-fitur yang relevan menggunakan metode information gain.

1.2.2 Metode Usulan

Penelitian ini mengusulkan *metode naïve* yang berfungsi sebagai algoritma pengklasifikasian sentimen, metode *tf-idf* yang berfungsi sebagai ekstraksi fitur dan pembobotan kata, preprocessing yang berfungsi untuk membuat data menjadi lebih lengkap, rapi, maupun terstruktur, dan metode information gain untuk seleksi fitur.

1.2.3 Pertanyaan Penelitian

1. Bagaimana menentukan klasifikasi opini masyarakat terhadap kebijakan penanganan covid-19 menggunakan metode *Naïve Bayes*?
2. Bagaimana proses penentuan bobot kata dari tweet menggunakan *tf-idf* dan seleksi fitur menggunakan *information gain*?

1.3 Tujuan dan Manfaat

1.3.1 Tujuan

Dibawah ini merupakan tujuan yang ingin dicapai dari penelitian analisis sentimen ini:

1. Untuk mengetahui hasil sentimen publik pada komentar terkait kebijakan penanganan virus covid-19 di indonesia menggunakan Algoritma *Naïve Bayes*.
2. Untuk menentukan bobot setiap kata dalam kalimat data *tweet* dan fitur-fitur yang relevan digunakan untuk klasifikasi

1.3.2 Manfaat

Dibawah ini merupakan manfaat yang didapat dari penelitian analisis sentimen ini:

1. Memperoleh dan mengolah opini dari masyarakat mengenai kebijakan penanganan virus covid-19 di Indonesia melalui media twitter.

2. Sebagai media pendukung untuk mendapatkan timbal balik dari masyarakat untuk proses pengambilan keputusan

1.4 Batasan Masalah

Agar pembuatan skripsi ini lebih terstruktur dan rapi, maka dibutuhkan adanya batasan masalah. Adapun batasan masalah di penelitian analisis sentimen ini adalah:

1. Penelitian ini hanya berfokus pada analisis pengguna media twitter terhadap kebijakan penanganan covid-19 di indonesia dengan kata kunci “Penanganan covid-19 di Indonesia”.
2. Data yang digunakan merupakan *tweet* dari twitter yang berkaitan dengan kebijakan penanganan covid-19 indonesia dari Oktober 2021- Maret 2022 yaitu sebanyak 1500 data, dengan pembagian sebanyak 1200 sebagai data latih dan 300 sebagai data uji.
3. Sentimen dibatasi menjadi 2 kelas, yaitu sentimen positif dan sentimen negatif

1.5 Metodologi Penulisan Skripsi

Berikut merupakan sistematika penulisan yang digunakan dalam penyusunan laporan yang telah disesuaikan dengan format penulisan Skripsi Program Studi Sistem Informasi Fakultas Teknik Universitas Trunojoyo Madura:

BAB I PENDAHULUAN

Pada Bab ini membahas latar belakang, perumusan masalah, tujuan dan manfaat penelitian, batasan masalah, dan sistematika laporan

BAB II KAJIAN PUSTAKA

Pada Bab ini menerangkan tentang tinjauan pustaka, Analisis sentimen, Twitter, Naïve Bayes, Term Frequency – Inverse Document Frequency (*TF-IDF*), Information Gain, dan kerangka teori lain yang berkaitan dengan tema

BAB III METODOLOGI PENELITIAN

Pada bab membahas tentang metode penelitian yang akan dibuat meliputi tahap studi literatur, analisis kebutuhan, membuat model klasifikasi, analisis dan perancangan sistem, dan implementasi sistem.

BAB IV. HASIL DAN PEMBAHASAN

Pada bab ini dijelaskan mengenai implementasi dari analisis sentimen tentang kebijakan penanganan covid-19 di Indonesia pada twitter dengan menggunakan Pembobotan *Tf-idf* dan *Naive Bayes*.

BAB V. PENUTUP

Pada bab ini dijelaskan mengenai kesimpulan hasil akhir dari proses pembelajaran, implementasi, serta uji coba sistem serta berisi saran atas pembuatan skripsi ini.