

BAB 1

PENDAHULUAN

1.1. Latar Belakang

Virus Corona menjadi wabah pertama kali di kota Wuhan, China pada bulan November-Desember 2019. Tingkat penyebaran virus tersebut sangat tinggi sehingga meluas sangat cepat ke seluruh dunia dan bisa dikatakan virus ini adalah virus yang sangat berbahaya. Tahun 2020, WHO mencatat berbagai laporan dari belahan negara lain yang sudah terkena virus corona, termasuk juga negara Indonesia[1]. Adanya warga yang terjangkit virus *Covid-19* pertama kali di Indonesia awal maret 2020, yaitu warga Depok, Jawa Barat, sejumlah dua orang. Penyebaran virus yang semakin luas di Indonesia bukan hanya menimbulkan masalah di bidang kesehatan saja akan tetapi di bidang lainnya seperti pendidikan, ekonomi, dan lainnya. Upaya untuk mengetahui apakah orang terkena dampak dari penyebaran virus corona tersebut, pemerintah Indonesia mengupayakan dengan melakukan tindakan test PCR (*Polymerase Chain Reaction*) terhadap masyarakat yang akan melakukan perjalanan menggunakan transportasi umum antar kota maupun provinsi seperti pesawat terbang, kereta, dll. Dalam hal tersebut Kemenkes menerbitkan Surat Edaran Harga tertinggi untuk melakukan tes PCR, Nomor HK. 02.02/I/3713/2020 terkait batasan tarif tertinggi untuk tes PCR *Real Time Polymerase Chain Reaction*(*RT-PCR*), yakni sebesar 900 ribu. Harga untuk test PCR banyak mendapat tanggapan atau respon dari masyarakat, ada yang baik dan membangun tetapi ada juga yang bertentangan atau menolak bisa dikatakan tanggapan tersebut bersifat positif dan negatif. Salah satu tempat masyarakat untuk menyampaikan pendapat tiap individu secara bebas yaitu melalui social media twitter[2].

Pendapat tiap individu yang dituangkan pada media social twitter, dapat dianalisis untuk bisa mengetahui polaritas pada data tersebut. Terdapat beberapa metode untuk analisis yang bisa digunakan untuk menganalisis respon atau pendapat masyarakat berdasarkan informasi yang ada pada twitter dan metode analisis sentimen merupakan salah satu metode yang tepat. Analisis sentimen adalah salah satu metode dalam menganalisis data yang didapatkan dari internet dan atau media sosial sehingga bisa mengetahui polaritas pada data tersebut. Polaritas dari tanggapan yang ada bisa terkumpul sehingga dapat digunakan untuk memprediksi respon publik atau gambaran perasaan netizen yang bersifat negatif atau positif dengan menggunakan analisis sentimen[3]. Penggunaan analisis sentimen dapat diterapkan dalam mengategorikan / mengklasifikasikan tulisan-tulisan atau tweets di twitter terhadap opini masyarakat terkait vaksinasi. Namun, untuk melakukan analisis sentimen harus ada sebuah model atau metode

perhitungan probabilitas yang dapat membantu dalam proses pengklasifikasian suatu tulisan. Beberapa model algoritma pengklasifikasian data yang banyak digunakan antara lain yaitu *algoritma naïve bayes*, *decision tree C4.5* dan *k-nearest neighbors* (KNN) [4].

Pada penelitian ini menggunakan algoritma *K-Nearest Neighbors* (KNN) sebagai model algoritma pengklasifikasian sentiment sebuah tweet pada media sosial twitter, dan merupakan suatu model algoritma klasifikasi pada objek berdasarkan data *training* yang menggunakan kemiripan terhadap objek atau jarak terdekat pada objek tersebut[5]. Data yang berbentuk teks dari twitter akan diolah terlebih dahulu untuk kemudian dihitung ke *K-Nearest Neighbors* (KNN). Data dari twitter yang berbentuk teks perlu diekstraksi dan dibobotkan terlebih dahulu sebelum dilakukan perhitungan algoritma *K-Nearest Neighbors* (KNN). Tf-Idf dan Tf-Rf merupakan metode ekstraksi fitur dan pembobotan kata. Tf-Idf adalah sebuah ukuran statistik yang digunakan untuk memperbaiki seberapa penting suatu kata di dalam sebuah dokumen atau dalam kelompok sebuah kata. Tf-Rf adalah sebuah metode yang tergolong baru dan menimbang relevansi sebuah dokumen dilihat pada frekuensi terhadap kemunculan term di kategori yang berkaitan untuk dokumen tunggal tiap kalimat dianggap sebagai dokumen[6]. Frekuensi kemunculan kata dalam sebuah dokumen yang diberikan menunjukkan seberapa penting dan seberapa umum kata itu pada dokumen tersebut[7].

Algoritma perhitungan *K-Nearest Neighbors* (KNN) digunakan dalam penelitian ini didasarkan pada banyaknya suatu data berupa tweet dari twitter yang ingin ditentukan kelasnya. Proses terakhir yaitu menentukan kelas dari setiap kalimat dan akurasi model yang dihasilkan oleh algoritma tersebut. Tujuan dari algoritma *K-Nearest Neighbors* (KNN) adalah menentukan objek ke dalam salah satu kelas yang sudah ada pada data sampel yang sebelumnya telah ditentukan[8]. Dalam penelitian ini yang menjadi perbedaan dengan penelitian sebelumnya yaitu membandingkan pembobotan kata Tf-Idf dan Tf-Rf, yang dimana penelitian sebelumnya dengan menggunakan metode *K-Nearest Neighbors* (KNN) dan Tf-Idf memperoleh hasil presentase sebesar 96,61%[9], akurasi 89,4% dan rasio error 10,6%[10]. Pada penelitian diharapkan bisa memberikan suatu informasi yang akurat dan bermanfaat kepada masyarakat mengenai sentiment pengguna twitter terhadap kebijakan pemerintah tentang harga tes PCR.

1.2. Rumusan Masalah

1.2.1. Permasalahan

Permasalahan yang akan diangkat pada penelitian ini adalah:

Kebijakan pemerintah terkait harga tes PCR yang mendapat banyak sekali tanggapan dari masyarakat termasuk melalui media sosial twitter. Tanggapan yang diungkapkan sangat

beragam, baik berbentuk tanggapannya yang bersifat positif ataupun negatif. Sehingga memerlukan analisis sentimen untuk mengetahui apakah lebih banyak sentimen positif atau sentimen negatif dari masyarakat. Pengklasifikasian sentimen yang berupa data teks dapat menggunakan perhitungan algoritma *K-Nearest Neighbors* (KNN). Akan tetapi suatu data yang berbentuk teks harus di ekstraksi terlebih dahulu dengan pembobotan Tf-Idf dan Tf-Rf, dan dilanjutkan menghitung akurasi untuk melihat akurasi perbandingan dari kedua pembobotan kata.

1.2.2. Metode Usulan

Penelitian analisis sentimen ini mengusulkan metode *K-Nearest Neighbors* (KNN) sebagai algoritma pengklasifikasian sentimen, metode Tf-Idf dan Tf-Rf sebagai metode untuk ekstraksi fitur dan pembobotan kata serta metode *k-cross validation* sebagai metode untuk evaluasi model.

1.3. Pertanyaan Penelitian

Berapa akurasi yang akan dihasilkan dari klasifikasi Algoritma *K-Nearest Neighbour* (KNN) dengan pembobotan kata Tf-idf dan Tf-Rf pada analisis sentimen masyarakat terhadap kebijakan pemerintah tentang harga test PCR?

1.4. Tujuan dan manfaat

1.4.1. Tujuan

Menghitung akurasi klasifikasi Algoritma *K-Nearest Neighbour* (KNN) dengan pembobotan kata Tf-Idf dan Tf-Rf untuk mengetahui pembobotan kata yang terbaik.

1.4.2. Manfaat

Memperoleh informasi hasil dari akurasi klasifikasi algoritma *K-Nearest Neighbors* (KNN) yang terbaik dalam membandingkan pembobotan kata Tf-Idf dan Tf-Rf pada analisis sentimen masyarakat terhadap kebijakan pemerintah tentang harga test PCR.

1.5. Batasan Masalah

Penentuan suatu batasan masalah yang dilakukan supaya tidak terjadi penyimpangan ataupun pelebaran dalam pokok masalah pada penelitian sehingga lebih terstruktur dan mempermudah dalam pembahasan, dan tujuan penelitian dapat tercapai. Berikut batasan masalah pada penelitian ini adalah :

1. Luas penelitian hanya pada analisis sentimen pengguna twitter terhadap harga tes PCR dengan kata kunci “harga tes PCR”.
2. Dalam pelabelan manual menggunakan bantuan dari 3 orang guru Bahasa Indonesia (Moh. Didik Suhartono, S.Pd. , Rica Wulandari, S.Pd. , Yuli Indrawati, S.Pd.).

3. Tahapan *preprocessing*, pendekatan *cosine similarity*, perhitungan *K-Nearest Neighbors* (KNN), dan *K-Fold Cross Validation* dilakukan dengan bantuan library. Perhitungan Tf-Idf dan Tf-Rf dilakukan tanpa bantuan library.
4. Platform yang digunakan antara lain *Chrome*, *Jupyter*.

1.6. Sistematika Pembahasan

Sistematika penulisan atau pembahasan yang ada pada penelitian ini terdiri dari lima bab dengan penjelasan secara singkat dari setiap bab adalah seperti berikut:

Bab I : Pendahuluan

Bab ini memuat tentang latar belakang masalah, perumusan masalah, tujuan dan manfaat, batasan masalah, sistematika pembahasan.

Bab II : Landasan Teori

Bab ini berisi tentang penjelasan pada setiap teori yang membantu memuat uraian suatu tinjauan pustaka terdahulu dan kerangka teori relevan yang terkait dengan tema proposal.

Bab III : Metode Penelitian

Bab ini akan membahas tahapan penelitian, pada penelitian menggunakan metode *Cross Industri Standart Process for Data Mining*(CRISP-DM) dimana data akan di proses dengan melakukan beberapa tahapan, diagram system pada penelitian ini, dan skenario pengujian.

Bab IV : Hasil dan Pembahasan

Bab ini yang akan dibahas tentang hasil penelitian yang akan dianalisa sesuai dengan tahapan-tahapan penelitian dari menggunakan metode *Cross Industri Standart Process for Data Mining*(CRISP-DM), source code pada progamnya, dan hasil skenario pengujian.

Bab V : Penutup

Bab ini membahas tentang sebuah kesimpulan dan saran terhadap penelitian yang dilakukan dari hasil pembahasan sebelumnya.