

BAB I

PENDAHULUAN

1.1 LATAR BELAKANG

Pada era globalisasi, media sosial saat ini sudah sangat umum dan banyak digunakan untuk kepentingan masyarakat. Dalam implementasinya, masyarakat cenderung menggunakan media sosial lebih banyak untuk kegiatan jual beli, menyampaikan informasi, bahkan sebagai media untuk mengekspresikan diri. Perkembangan media sosial sangat cepat tidak hanya penggunaannya yang terus meningkat, namun semakin banyaknya media sosial yang ditawarkan melalui aplikasi *mobile* ataupun website.

Salah satu media sosial yang banyak digunakan adalah *Twitter*. Indonesia menduduki peringkat ke lima sebagai negara dengan penduduk yang paling banyak menggunakan *Twitter*. *Twitter* merupakan salah satu media sosial yang memungkinkan penggunanya untuk berbagi informasi dengan sesama secara realtime. Kata yang terkandung dalam *Twitter* adalah bahasa alami manusia yang sering digunakan dalam kehidupan sehari-hari dan berstruktur kompleks. Pengguna *twitter* memposting pesan singkat tentang berbagai topik tertentu dan berlaku secara global. Dengan adanya fitur trending pada *Twitter*, pengguna dapat membagikan tanggapan terhadap suatu hal yang sedang marak dibicarakan. Salah satu topik yang banyak diperbincangkan pada *Twitter* yaitu penyebaran Virus Corona.

Virus Corona adalah virus yang menyerang sistem pernapasan. Penyakit karena infeksi virus ini disebut COVID-19. *Virus Corona* bisa menyebabkan gangguan ringan pada sistem pernapasan, infeksi paru-paru yang berat, hingga kematian. *Virus corona* atau dikenal sebagai pandemi Covid-19 pertama kali dideteksi di Kota Wuhan Desember 2019, sejak penyebaran virus ini masyarakat langsung membicarakannya di media sosial yaitu *Twitter*. Banyaknya jumlah korban dan penyebaran virus yang sangat cepat membuat masyarakat khawatir akan muncul berbagai opini tentang virus *corona* termasuk di Indonesia. Kenaikan kasus Covid-19 di Indonesia pada Juli hingga Agustus 2022 lebih sedikit dibandingkan tahun sebelumnya. Sehingga bebas masker mulai diterapkan oleh masyarakat. Namun

pada awal tahun 2023, kasus covid 19 di Indonesia dikabarkan melonjak karena pasca libur natal dan tahun baru. Hal ini menyebabkan banyaknya jumlah opini yang muncul dan terus meningkat menjadi permasalahan untuk mengetahui sentimen masyarakat, sehingga diperlukan analisis terhadap opini masyarakat secara otomatis. Opini masyarakat inilah yang kemudian di analisis untuk diketahui polaritasnya dengan menggunakan analisis sentimen.

Analisis sentimen adalah bagian dari *teks mining* yang dapat mengelompokkan polaritas dari teks, pengelompokkan tersebut dilakukan untuk melihat bagaimana polaritas dari suatu tweet apakah opini yang diberikan bersifat positif, negatif atau netral (Fanissa et al., 2018) Selain itu untuk mengetahui siapa pemberi opini yang banyak mendapatkan respon dari pengguna *Twitter*. Analisis sentimen juga dapat digunakan untuk mengungkap opini publik terhadap suatu isu, kepuasan pelayanan, kebijakan dan *cyber bullying*.

Penggunaan analisis sentimen dapat diterapkan pada opini masyarakat terhadap virus corona pada dokumen *Twitter* berbahasa Indonesia. Pada penulisan opini masyarakat di *Twitter* terkadang terdapat penulisan *Tweet* yang sulit dibaca. Hal ini disebabkan oleh beberapa faktor seperti penulisan kata yang disingkat, penggunaan bahasa yang modern, salah dalam mengetik huruf dan tidak baku dalam penulisan opini. Beberapa hal tersebut menyebabkan perlu dilakukan analisis terhadap opini netizen di *Twitter* mengenai Corona dengan klasifikasi positif dan negatif representasi ulang kata sebelum dilakukan penentuan sentimen opini. Kata tersebut kemudian akan diklasifikasikan agar dapat diketahui apakah akan masuk ke dalam sentimen positif atau negatif, sehingga dibutuhkan metode klasifikasi untuk menganalisis komentar atau opini.

Pada penelitian sebelumnya yang berkaitan dengan analisis sentimen sudah pernah dilakukan (Fanissa et al., 2018). Pada penelitian ini digunakan metode *Naïve Bayes* untuk analisis sentimen tentang opini film pada dokumen *Twitter* dengan menggunakan kamus kata tidak baku dan normal Normalisasi Levenshtein Distance untuk memperbaiki kata yang tidak baku menjadi kata baku dengan pengklasifikasian *Naïve Bayes* .

Penelitian selanjutnya yang berkaitan dengan analisis sentimen yaitu penelitian yang dilakukan oleh (Kurniawan & Susanto, 2019) dengan menggunakan

metode *K-Means* dan *Naive Bayes Classifier* untuk menganalisis sentimen masyarakat untuk mengetahui kecenderungan opini masyarakat terhadap pemilu. Pada penelitian ini Metode *K-Means* digunakan untuk melakukan klustering pada data latih dan *Naive Bayes classifier* untuk mengklasifikasi pada data testing. Hasil dari pembobotan ini berupa sentimen positif dan negatif. Data diambil dari Twitter mengenai pemilu presiden 2019 sebanyak 500 data tweet. Dari hasil pengujian 100 dan 150 data uji diperoleh akurasi rata-rata 93.35% dan error rate sebesar 6.66%.

Pada penelitian yang dilakukan oleh (Putra, 2013) Klasifikasi yang akan digunakan sebagai perbandingan hasil yaitu *K-Nearest Neighbor (KNN) Classifier* dan *Naive Bayesian Classifier*. Hasil summary dari kedua klasifikasi akan menentukan jenis klasifikasi mana yang lebih cocok diterapkan pada dataset tersebut. Selain itu, akan dilakukan juga perbandingan hasil clustering metode *Simple K-Means* yang menggunakan algoritma *distance function Manhattan Distance* dan *Euclidian Distance*. Perbandingan clustering dilakukan untuk melihat perbedaan pembagian kelas pada kedua function tersebut. Namun pada penelitian ini metode *K-Means cluster* yang diterapkan, menghasilkan perbandingan kelas yang relatif sama antara algoritma *manhattanDistance* dan *EuclidianDistance*

Berdasarkan uraian diatas, peneliti mengusulkan penelitian dengan menggunakan metode *Naive Bayes* serta KNN untuk mengolah hasil opini tersebut. Pemilihan Metode *Naive Bayes* dalam penelitian ini dikarenakan cepat dan efisien terhadap ruang, selain itu hanya memerlukan sedikit data pelatihan untuk melakukan klasifikasi. Sedangkan untuk penggunaan algoritma KNN dikarenakan KNN merupakan algoritma yang efektif untuk data yang besar, tahan terhadap data pelatihan yang *noise*. Kedua metode yang digunakan nantinya akan dibandingkan tingkat akurasi antara *Naive Bayes* dibandingkan dengan *KNN* dalam menganalisis sentimen masyarakat terhadap virus corona berdasarkan opini dari Twitter. Selain itu untuk menangani masalah kata yang tidak baku, akan dilakukan suatu proses perbaikan kata tidak baku dengan menggunakan kamus kata tidak baku. Sehingga hasil penelitian nanti diharapkan memperoleh perbandingan tingkat akurasi yang relevan dengan data yang ada.

1.2 IDENTIFIKASI MASALAH

Identifikasi masalah untuk penelitian yang berjudul “Analisis Sentimen Masyarakat terhadap Virus Corona berdasarkan Opini dari Twitter menggunakan Metode *Naïve Bayes* dan K-Nearest Neighbor” adalah sebagai berikut.

- a. Banyaknya jumlah korban dan penyebaran virus yang sangat cepat membuat masyarakat khawatir akan muncul berbagai opini tentang virus corona termasuk di Indonesia.
- b. Masyarakat sulitnya percaya terhadap berita atau opini yang disebar di media sosial khususnya Twitter tentang informasi terkait Virus Corona.
- c. Masyarakat cenderung tidak percaya terhadap akun-akun berita yang menyebarkan informasi terkait virus corona di media sosial.
- d. Perlunya analisis sentimen untuk menganalisis opini-opini masyarakat di media sosial.

1.3 PEMBATALAN MASALAH

Batasan masalah untuk penelitian yang berjudul “Analisis Sentimen Masyarakat terhadap Virus Corona berdasarkan Opini dari Twitter menggunakan Metode *Naïve Bayes* dan K-Nearest Neighbor” adalah sebagai berikut.

- a. Penelitian ini merupakan penelitian yang mendasar yaitu menggunakan teknik data mining saja tidak sampai pada pengembangan data warehouse.
- b. Penelitian ini hanya mengambil data Tweet dalam bentuk Bahasa Indonesia saja.
- c. Data yang digunakan berasal dari Dokumen Twitter yang mencangkup terkait Virus Corona saja.
- d. Jumlah data tweet yang digunakan adalah 2000 data yang telah di filter yang diambil dari tanggal 01 Maret 2020 sampai dengan Juli 2021
- e. Metode yang digunakan untuk perbandingan analisis sentimen masyarakat terhadap virus corona adalah metode *Naïve Bayes* dan K-Nearest Neighbor.

1.4 RUMUSAN MASALAH

Berdasarkan latar belakang yang dipaparkan sebelumnya, dapat dirumuskan beberapa permasalahan yang akan dibahas dalam penelitian ini, antara lain :

- a. Bagaimana implementasi algoritma *Naïve Bayes* dan *K-Nearest Neighbor* dalam analisis sentimen masyarakat terhadap virus corona berdasarkan opini dari Twitter ?
- b. Bagaimana perbandingan akurasi dari kedua metode dalam menentukan analisis sentimen masyarakat terhadap virus corona berdasarkan opini dari Twitter ?

1.5 TUJUAN PENELITIAN

Adapun tujuan yang diharapkan dapat tercapai dari “Analisis Sentimen Masyarakat terhadap Virus Corona berdasarkan Opini dari Twitter menggunakan Metode *Naïve Bayes* dan *K-Nearest Neighbor*” adalah sebagai berikut.

- a. Untuk mengetahui implementasi algoritma *Naïve Bayes* dan *K-Nearest Neighbor* dalam analisis sentimen masyarakat terhadap virus corona berdasarkan opini dari Twitter.
- b. Untuk mengetahui perbandingan akurasi dari kedua metode dalam menentukan analisis sentimen masyarakat terhadap virus corona berdasarkan opini dari Twitter.

1.6 MANFAAT PENELITIAN

Manfaat yang diharapkan dalam penelitian “Analisis Sentimen Masyarakat terhadap Virus Corona berdasarkan Opini dari Twitter menggunakan Metode *Naïve Bayes* dan *K-Nearest Neighbor*” ini adalah sebagai berikut.

- a. Manfaat Teoritis
 1. Bagi peneliti, penelitian ini diharapkan akan mampu menambah pengetahuan, dapat memahami serta dapat menerapkannya di dalam proses perkuliahan.
 2. Bagi penelitian sejenis, penelitian ini diharapkan dapat menjadi kajian dalam penelitian sejenis yang dilakukan.

b. Manfaat Praktis

1. Bagi Peneliti, dapat mengimplementasikan ilmu yang telah didapat selama perkuliahan dalam Analisis Sentimen Masyarakat terhadap Virus Corona berdasarkan Opini dari Twitter menggunakan Metode *Naïve Bayes* dan K-Nearest Neighbor dan dapat menambah wawasan peneliti tentang perbandingan metode data mining dalam kasus ini.
2. Bagi masyarakat umum, penelitian ini memiliki manfaat sebagai evaluasi untuk penanganan virus corona menurut opini masyarakat di twitter.

