

BAB I

PENDAHULUAN

1.1 Latar Belakang

Di Indonesia, kebiasaan “mudik” merupakan fenomena sosial tahunan. Di sini, mudik diperingati sebagai hari libur umum bagi masyarakat yang bertempat tinggal di kota-kota besar di kampung halamannya (desa atau kota kecil). Dari segi bahasa, hal ini biasa dilakukan menjelang Natal, Tahun Baru, dan Idul Fitri. Biasanya, ini dimulai tujuh hari sebelum Idul Fitri dan berakhir tujuh hari setelah hari raya. Fenomena sosial budaya lain yang juga menjadi rutinitas setiap tahunnya adalah mudik, yang memerlukan pengelolaan baik dari pemerintah agar kelancaran operasionalnya lebaran.

Belum diketahui pasti waktu mulainya acara mudik lebaran. Namun dari sejumlah sumber tercatat bahwa adat atau kejadian ini merupakan wujud hubungan kekerabatan yang pertama kali dilakukan oleh suku-suku migran, khususnya yang tinggal di pulau-pulau padat penduduk seperti Sumatera, Jawa, dan Bali.

Dua tahun kemudian, negara ini dilanda pandemi Covid-19 yang menyebabkan sejumlah tindakan terkait pembatasan dan pengendalian transportasi pada musim mudik 2020 dan 2021. Acara mudik tahun 2023 ini merupakan acara mudik yang kedua kalinya. Hal ini merupakan ungkapan apresiasi kepada semua pihak yang telah membantu merencanakan reuni ini. tidak kewalahan, seperti pada tahun 2020 dan 2021, oleh kekhawatiran terhadap penyebaran dan penularan Covid-19 melalui infrastruktur dan utilitas transportasi.

Salah satu negara dengan jumlah pengguna aktif Twitter harian yang cukup besar adalah Indonesia. Dengan jangkauan 10 juta penonton, Indonesia menempati peringkat kedelapan menurut data Hootsuite. Mengingat penggunaannya yang luas di masyarakat, Twitter menjadi salah satu platform media sosial yang dapat melakukan riset sentimen terhadap berbagai isu.

Analisis sentimen, sering juga disebut analisis opini, adalah studi mengenai opini publik yang diungkapkan dalam tweet. Memprediksi opini yang condong ke

arah negatif, positif, atau netral adalah tujuan analisis sentimen. Analisis sentimen adalah tindakan mengkaji beberapa jenis data, seperti pandangan atau opini, guna menarik kesimpulan dari berbagai opini yang sudah ada sebelumnya. Informasi dalam data ini dibagikan untuk membantu individu atau organisasi dalam mengambil keputusan. Tidak mungkin manusia membaca dan mengevaluasi data dalam jumlah besar secara manual.

Penulis tertarik untuk melakukan penelitian klasifikasi sentimen dengan pendekatan algoritma *K-Nearest Neighbors (KNN)* berdasarkan penelitian yang diselesaikannya dengan melihat tantangan yang dilakukan oleh peneliti lain dalam penelitian klasifikasi. Salah satu manfaat pendekatan *KNN* ini adalah dapat mengkategorikan data analisis sentimen menggunakan data pelatihan dan pengujian terkait opini publik di Twitter. Algoritma *K-Nearest Neighbors (KNN)* adalah metode klasifikasi yang lugas dan mendasar; data yang diberi label memudahkan untuk mengklasifikasikan data ke dalam kelompok yang relevan. Dengan terlebih dahulu memilih nilai *K* tetangga terdekat dengan benar, pendekatan *KNN* dapat mengkategorikan data dengan andal dan menawarkan keuntungan dalam hal kesederhanaan dalam menerjemahkan hasil dan akurasi perkiraan.

Pada penelitian ini, respon masyarakat di twitter dapat menentukan apakah hasil sentiment analisis adalah positif, negatif, dan netral sehingga berdasarkan data tersebut kita dapat mengetahui pendapat masyarakat indonesia secara general terhadap penerapan arus mudik dan arus balik lebaran 2023. Pada penelitian sebelumnya dengan menggunakan metode seleksi fitur *Chi Square* dan *K-Nearest Neighbors* memiliki kelebihan dalam penggunaan data training yang kecil untuk proses pengklasifikasian. Namun kekurangannya yaitu dengan lebih memperhatikan singkatan, gabungan dua kata atau lebih, kata-kata ambigu dan kalimat sarkastik supaya hasil *slangword* lebih sempurna. Kontribusi penelitian ini adalah penambahan *Chi Square* bertujuan untuk mengurangi jumlah fitur sehingga dapat meningkatkan akurasi sehingga tingkat akurasi meningkat dari penelitian sebelumnya.

1.2. Rumusan Masalah

1.2.1 Permasalahan

Mobilisasi mudik antar negara bahkan antar kota-provinsi menyebabkan pembatasan, harga transportasi tiap perjalanan melambung tinggi. Hal ini belum ditambah dengan persyaratan tes antigen sampai *PCR* dan juga vaksinasi. Sehingga keadaan dilapangan terutama masyarakat yang akan melakukan perjalanan menimbulkan pro dan kontra. Terlebih lagi perjalanan ini bukan perjalanan dinas atau liburan melainkan perjalanan mudik-balik dalam rangka hari raya Idul Fitri. Momen Idul Fitri memang sangat berbeda dan merupakan hal yang paling dinantikan bagi perantau.

Analisis sentiment diperlukan untuk mengetahui opini masyarakat. Penelitian menggunakan *KNN* untuk analisis sentiment yang memiliki kekurangan tergantung nilai k , sehingga diperlukan k terbaik yaitu permasalahan memilih nilai k , selain itu klasifikasi *KNN* terpengaruh jumlah fitur yang tidak normal, semakin banyak fitur maka semakin tidak optimal sehingga membutuhkan seleksi fitur. Seleksi fitur yang digunakan adalah *Chi Square* yang membutuhkan nilai k dengan nilai yang terbaik.

1.2.2. Metode Usulan

Peneliti mengusulkan metode *K-Nearest Neighbors (KNN)* dan Seleksi Fitur *Chi Square*, dengan data yang dikumpulkan dari media sosial twitter tentang arus mudik dan arus balik lebaran 2023. Dengan menggunakan metode *K-Nearest Neighbors (KNN)* dan Seleksi Fitur *Chi Square* peneliti dapat mengetahui tentang arus mudik dan arus balik lebaran 2023 apakah berdampak baik atau sebaliknya bagi masyarakat sebagai metode untuk evaluasi model.

1.2.3 Pertanyaan Penelitian

1. Seberapa besar pengaruh Seleksi Fitur *Chi Square* untuk meningkatkan akurasi dan komputasi waktu dari Metode *K-Nearest Neighbors (KNN)*?

1.3. Tujuan dan Manfaat

1.3.1 Tujuan

- a. Untuk mengetahui seberapa besar pengaruh Seleksi Fitur *Chi Square* untuk meningkatkan akurasi dan komputasi dari Metode *K-Nearest Neighbors (KNN)*.

1.3.2 Manfaat

- a. Dapat mengetahui informasi terhadap opini atau persepsi masyarakat terhadap arus mudik dan arus balik lebaran 2023 berupa opini positif, negative, dan netral guna dijadikan bahan evaluasi pemerintah.
- b. Sebagai media pendukung untuk mendapatkan feedback dari masyarakat dalam proses pengambilan keputusan.

1.4 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. Data yang digunakan adalah hasil tweet dari opini masyarakat menggunakan media social twitter terhadap Arus mudik dan arus balik lebaran 2023 di Indonesia selama periode tahun 2023 sebanyak 1000 data.
2. Data yang akan disajikan berupa opini positif, negatif, dan netral di twitter terhadap arus mudik dan arus balik lebaran 2023 di Indonesia.
3. Metode yang dipakai dalam penelitian ini menggunakan *K-Nearest Neighbors (KNN)* dan Seleksi Fitur *Chi Square*.

1.5 Sistematika Peneltitian

Sistematika penelitian proposal tugas akhir ini adalah sebagai berikut :

BAB I PENDAHULUAN

Bab ini membahas tentang latar belakang dari masalah yang dibahas, usulan solusi yang dapat mengatasi permasalahan, tujuan dan manfaat dari penerapan solusi yang diusulkan, serta penyusunan proposal secara sistematis.

BAB II KAJIAN PUSTAKA

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 USULAN

Bab ini akan membahas tahapan penelitian, pada penelitian menggunakan metode *K-Fold Validation* dimana data akan di proses dengan melakukan beberapa tahapan, diagram sistem pada penelitian ini, dan skenario pengujian.

BAB IV HASIL DAN PEMBAHASAN

Dalam bab ini membahas tentang hasil dari sistem dengan acuan pada model sistem yang sudah dibangun.

BAB V KESIMPULAN DAN SARAN

Dalam bab ini membahas tentang kesimpulan skripsi yang dibuat secara keseluruhan serta kekurangan pada suatu sistem yang dibangun.