

BAB I

PENDAHULUAN

1.1 Latar Belakang

Pemilihan umum presiden dan wakil presiden (pilpres) adalah momen krusial dalam sistem demokrasi di Indonesia. Pada tahun 2024, Indonesia akan kembali menggelar pilpres untuk memilih presiden dan wakil presiden yang akan memimpin negara untuk periode berikutnya. Pilpres 2024 diperkirakan akan menjadi periode yang penuh dengan perdebatan dan diskusi di kalangan masyarakat, termasuk di media sosial X. Media sosial seperti *Twitter* memiliki basis pengguna yang sangat besar secara global, memungkinkan kandidat atau partai politik untuk menjangkau audiens yang luas dalam waktu singkat. Namun, media sosial juga sering menjadi arena penyebaran informasi palsu atau hoaks. Informasi yang tidak diverifikasi atau disertai dengan fakta yang salah dapat dengan mudah menyebar dan mempengaruhi opini publik secara negatif.

Analisis sentimen adalah proses komputasional untuk mengekstraksi, mengidentifikasi, atau memahami Opini, perasaan, dan sentimen yang terkandung dalam teks. Analisis sentimen bertujuan utama untuk menentukan apakah teks tersebut memiliki sifat positif, negatif, atau netral, serta untuk memahami sejauh mana opini atau perasaan tersebut dinyatakan dalam teks.

Fokus analisis sentimen ini akan diarahkan pada teks yang terdapat dalam tweet, balasan, dan retweet di *Twitter*. Teks dalam bentuk gambar, video, atau media non-teks akan dikecualikan dari analisis. Untuk menjaga fokus penelitian, skripsi ini akan membatasi ruang lingkup pada salah satu media berita di *Twitter* yang terkait dengan capres dan cawapres 2024, dengan periode pengambilan data sentimen dari bulan 28 November 2023 s/d 10 Februari 2024.

Dalam beberapa tahun terakhir, algoritma pembelajaran mesin seperti *K-Nearest Neighbor* (K-NN) telah menjadi pilihan utama untuk analisis sentimen di media sosial. Algoritma ini mampu memahami pola kompleks dalam data dan mengklasifikasikan sentimen menjadi kategori positif, negatif, atau netral. Oleh karena itu, penerapan K-NN dalam analisis sentimen dapat memberikan wawasan mendalam mengenai dinamika opini masyarakat terhadap calon presiden dan wakil presiden.

Penelitian ini membagi data dengan proporsi 90:10 dan menemukan variasi dalam akurasi berdasarkan nilai k yang digunakan. Pasangan Anies dan Muhaimin menunjukkan akurasi sebesar 67,4% dengan nilai $k=8$, sedangkan pasangan Prabowo dan Gibran memiliki akurasi lebih rendah yaitu 55,7% dengan $k=3$. Sebaliknya, pasangan Ganjar dan Mahfud mencapai akurasi tertinggi sebesar 69,6% dengan nilai $k=10$. Perbandingan hasil penelitian ini dengan studi sebelumnya mengungkapkan bahwa faktor-faktor seperti jumlah data dan fitur pembobotan memainkan peran penting dalam menentukan kinerja algoritma. Temuan ini memberikan wawasan yang berharga tentang persepsi publik terhadap kandidat dan dapat mempengaruhi strategi kampanye serta keputusan politik yang diambil oleh para calon.

Berdasarkan tinjauan penulis terhadap penelitian klasifikasi yang telah dilakukan oleh peneliti lain, Penulis berminat untuk meneliti pengklasifikasian sentimen dengan menggunakan metode algoritma *K-Nearest Neighbor* (K-NN). Metode ini memiliki kelebihan dalam mengklasifikasikan data analisis sentimen terkait opini masyarakat di media sosial seperti *Twitter/X*, dengan memanfaatkan data latih dan data uji. Algoritma *K-Nearest Neighbor* (K-NN) adalah metode klasifikasi yang sederhana dan mudah diterapkan, yang memanfaatkan label data untuk memudahkan pengelompokan ke dalam kategori yang paling sesuai. K-NN unggul dalam hal kemudahan interpretasi hasil dan akurasi prediksi, serta kemampuannya untuk mengklasifikasikan data dengan tepat melalui penentuan nilai K tetangga terdekat yang optimal (Jabal Tursina, 2019).

Berbagai penelitian mengenai analisis sentimen sudah banyak dilakukan. Pada penelitian Fairuz (2021) yang meneliti sentimen masyarakat terhadap *COVID-19* didapatkan akurasi sebesar 85% untuk algoritma *Naïve Bayes* dan 82% untuk algoritma K-NN, ini menunjukkan kinerja kedua algoritma yang digunakan baik.

1.2 Perumusan Masalah

Berdasarkan konteks yang telah dijelaskan, isu-isu yang akan dibahas dalam penelitian ini dapat dirumuskan sebagai berikut:

1. Bagaimana membangun analisis sentimen publik terkait Capres dan Cawapres yang bersaing dalam pemilihan umum presiden 2024 di Indonesia, berdasarkan data dari media sosial *Twitter*?
2. Bagaimana kinerja algoritma K-NN dalam mengklasifikasikan sentimen pengguna terhadap Capres dan Cawapres 2024 di media sosial X?
3. Seberapa akurat dan andalnya algoritma K-NN dalam membedakan antara sentimen positif, negatif, dan netral terhadap Capres dan Cawapres 2024 di media sosial X?

1.3 Batasan Masalah

Berdasarkan permasalahan yang diidentifikasi, peneliti ini akan fokus pada batasan masalah yang akan diselesaikan. adalah:

1. Fokus analisis sentimen akan terpusat pada teks yang terdapat dalam *tweet*, balasan, dan *retweet* di *Twitter*. Teks dalam bentuk gambar, video, atau media non-teks akan dikecualikan dari analisis. Untuk menjaga fokus penelitian dan mencegah pembahasan yang terlalu luas, skripsi ini membatasi ruang lingkup penelitian pada salah satu akun media berita di *Twitter* yang berkaitan dengan calon presiden (capres) dan calon wakil presiden (cawapres) 2024. Pengumpulan data sentimen dilakukan pada

bulan 28 November 2023 s/d 10 Februari 2024. Periode ini merupakan awal kampanye pemilihan presiden dan wakil presiden di mana calon-calon potensial mulai memperkenalkan diri, menyampaikan visi, dan melakukan serangkaian aktivitas kampanye. Oleh karena itu, analisis sentimen pada periode ini dapat memberikan gambaran awal.

2. Penelitian ini akan fokus pada penggunaan algoritma *K-Nearest Neighbor* (K-NN) untuk analisis sentimen, tanpa mengeksplorasi metode atau algoritma lain.
3. Analisis sentimen akan dibatasi pada klasifikasi sentimen menjadi tiga kategori utama: positif, negatif, dan netral. Sentimen yang lebih kompleks atau subkategori sentimen tidak akan dianalisis.

1.4 Tujuan Dan Manfaat

Tujuan dan manfaat yang diharapkan dari penelitian ini adalah:

1. Menganalisis pandangan publik terhadap calon presiden dan calon wakil presiden yang akan berlaga dalam Pemilihan Presiden dan Wakil Presiden 2024.
2. Mengidentifikasi dan mengklasifikasikan tingkat sentimen masyarakat terhadap masing-masing calon menggunakan algoritma *K-Nearest Neighbor* (K-NN). Hal ini dapat memberikan informasi mengenai sejauh mana respons masyarakat terhadap pesan-pesan dan kampanye politik dari Capres dan Cawapres
3. Membangun profil sentimen masyarakat yang terbagi dalam kategori positif, negatif, dan netral terhadap masing-masing calon. Hal ini bertujuan untuk memberikan pandangan holistik tentang bagaimana masyarakat merespons setiap figur politik yang bersaing

Manfaat yang diperoleh dari penelitian ini adalah :

1. Hasil analisis sentimen dapat memberikan panduan bagi tim kampanye politik dalam menyusun strategi komunikasi yang lebih efektif. Mengetahui tren dan pola sentimen masyarakat dapat membantu dalam mengarahkan pesan-pesan yang lebih sesuai dengan kebutuhan dan keinginan pemilih.
2. Penelitian ini akan memberikan pemahaman yang lebih mendalam tentang preferensi dan kecenderungan masyarakat terhadap para calon. Informasi ini dapat digunakan oleh calon presiden dan wakil presiden untuk menyesuaikan platform dan janji-janji kampanye mereka sesuai dengan harapan pemilih.
3. Analisis sentimen dapat menjadi alat evaluasi yang kuat untuk mengevaluasi kinerja kampanye politik. Memahami sejauh mana kampanye mencapai tujuan sentimen positif dapat membantu untuk melakukan penyesuaian strategi jika diperlukan.

1.5 Metodologi Penelitian

Dalam penyusunan laporan ini, penulis menggunakan metode deskriptif untuk menguraikan permasalahan secara rinci berdasarkan data dan fakta yang ada

1. Pengumpulan Data

Data akan dikumpulkan dari *Twitter* menggunakan kata kunci terkait calon presiden dan calon wakil presiden Pilpres 2024. Data yang dikumpulkan akan berupa tweet, retweet, dan balasan, yang akan diunduh untuk periode tertentu sebelum dan setelah pemilihan umum.

2. Pembersihan dan Praproses Data

Data yang diunduh akan menjalani tahap pembersihan dan praproses, termasuk penghapusan tautan, karakter khusus, dan tokenisasi teks. Data juga akan dinormalisasi untuk mengurangi kesalahan dalam analisis sentimen.

3. Klasifikasi Sentimen

Data yang telah diproses akan digunakan untuk melatih model machine learning dengan algoritma *K-Nearest Neighbor* (K-NN). Model ini akan digunakan untuk mengklasifikasikan sentimen tweet menjadi kategori positif, negatif, atau netral.

4. Uji Coba dan Evaluasi Program

Pada tahap ini, penulis akan melakukan uji coba dan evaluasi terhadap program yang telah dibuat untuk memastikan bahwa semua fungsi berjalan sesuai dengan kode sumber yang telah disusun dan dapat memenuhi kebutuhan atau permasalahan yang ada

5. Evaluasi Model

Kualitas model machine learning akan dinilai berdasarkan metrik seperti akurasi, presisi, recall, dan F1-score untuk memastikan bahwa model mampu mengklasifikasikan sentimen dengan tingkat akurasi yang tinggi.

6. Interpretasi Hasil

Hasil analisis sentimen akan diinterpretasikan untuk mengidentifikasi isu-isu utama dan pola dukungan atau kritik terhadap calon presiden dan calon wakil presiden.

1.6 Sistematika Penulisan

Berikut adalah rincian sistematika penelitian yang diterapkan dalam penulisan laporan tugas akhir dengan judul "Analisis Sentimen Capres dan Cawapres 2024 di Media Sosial X Menggunakan Algoritma *K-Nearest Neighbor* (K-NN)":

BAB I PENDAHULUAN

Bab ini membahas secara umum mengenai masalah yang mencakup latar belakang, rumusan masalah, batasan masalah, tujuan dan manfaat penelitian, metodologi penelitian, serta sistematika penelitian.

BAB II TINJAUAN PUSTAKA

Bab ini menyajikan teori-teori yang mendasari penelitian dan membahas berbagai konsep yang berkaitan dengan topik yang sedang dibahas.

BAB III METODOLOGI PENELITIAN

Bab ini menguraikan metode yang digunakan dalam penelitian, termasuk cara pengumpulan data dan teknik-teknik yang diterapkan.

BAB IV HASIL DAN PEMBAHASAN

Bab ini mengulas permasalahan yang telah dirumuskan dan hasil-hasil yang diperoleh dari penelitian yang telah dilaksanakan.

BAB V PENUTUP

Bab ini menyajikan kesimpulan dari penelitian yang telah dilakukan, mencakup proses analisis, pelaksanaan, hasil uji, serta daftar pustaka yang digunakan sebagai referensi.