

1. Pendahuluan

Latar Belakang

Dalam beberapa tahun terakhir, media sosial telah menjadi platform penting untuk mendapatkan informasi baru dan juga untuk bertukar informasi [1]. Platform media sosial yang paling sering digunakan adalah Twitter. Twitter digunakan sebagai tempat untuk berdiskusi, berkomentar, dan memberikan informasi secara online. Salah satu topik yang sering dibicarakan di platform Twitter adalah masalah politik. Oleh karena itu, Twitter telah menjadi sumber berita untuk diskusi politik publik, yang memungkinkan dan memberikan komentar berbasis teks [2]. Selain itu, secara umum mengakui bahwa Twitter memiliki potensi besar untuk menyelidiki perilaku sosial dan politik [3].

Isu politik merupakan topik yang menarik untuk dibahas, terutama dengan Presiden Jokowi yang akan segera mengakhiri masa jabatannya. Pemilihan umum ini dapat menimbulkan opini pro dan kontra terhadap para kandidat yang mencalonkan diri, sehingga akan ada komentar positif, netral, dan negatif. Berdasarkan penelitian, komentar positif memiliki jangkauan yang lebih panjang dibandingkan komentar netral dan negatif, namun komentar netral dan negatif lebih cepat menyebar [4]. Ketika memberikan komentar, pengguna media sosial biasanya mengekspresikan emosi mereka terkait kandidat yang dicalonkan, seperti kesedihan, keterkejutan, ketakutan, kemarahan, kebahagiaan, dan kepercayaan [5]. Selain itu, pemrosesan emosi juga merupakan faktor penting dalam mendeteksi perilaku pengguna, terutama dalam konteks politik.

Penggunaan platform media sosial sangat dipengaruhi oleh jenis konten yang didukung oleh platform tersebut [1]. Dalam konteks politik, teks tweet politik dapat diklasifikasikan ke dalam dua kelas, yaitu opini (subjektif) dan informasi (objektif). Menganalisis perbedaannya dapat memberikan pemahaman tentang pengaruh penggunaan konten terhadap isu politik [6]. Dengan menganalisis peran dan pengaruh kedua jenis konten ini, kita dapat memahami lebih jauh tentang bagaimana opini dan informasi berperan dalam diskusi politik politik di media sosial. Dalam konteks ini, Twitter menjadi platform utama untuk menyuarakan opini dan emosi terkait kandidat yang dipilih.

Pembelajaran tak terawasi dilakukan untuk memproses data tanpa memiliki target informasi atau output yang diinginkan. Tujuan utama dari pembelajaran tak terawasi adalah untuk menemukan pola, struktur, atau informasi yang tidak terlihat dalam data tanpa label atau target [7]. Pengelompokan agglomerative hierarchical adalah metode pembelajaran tanpa pengawasan untuk mengelompokkan data [8]. Metode ini merupakan pendekatan dari bawah ke atas di mana pada tingkat paling awal, titik-titik data diperlakukan sebagai satu individu [9]. Kontribusi utama dari metode ini dalam data twitter adalah untuk membagi dan menggabungkan cluster untuk menghasilkan kategori yang lebih terorganisir [10].

Penelitian ini bertujuan untuk mengklasifikasikan jenis-jenis konten dan emosi yang disampaikan dalam beberapa tagar tentang Anies Baswedan. Data yang digunakan dalam penelitian ini terdiri dari 12.500 data berbahasa Indonesia. Tahap awal pengolahan data akan dilakukan preprocessing data dan kemudian penggunaan TF-IDF Vectorizer untuk mengubah teks menjadi representasi numerik yang dapat digunakan dalam analisis. Selanjutnya, data akan di-cluster menggunakan algoritma Agglomerative Hierarchical Clustering dan menggunakan silhouette score untuk memilih cluster terbaik untuk melakukan analisis jenis konten dan emosi pengguna.

Topik dan Batasannya

Topik dan batasan masalah pada penelitian ini adalah dataset yang digunakan pada penelitian ini sebesar 12.500 data dari tagar seputar kandidat calon presiden Anies Baswedan yang menggunakan Bahasa Indonesia. Data hanya data pada twitter pada rentang waktu 1 tahun dengan menggunakan algoritma Agglomerative Hierarchical Clustering.

Tujuan

Tujuan dari penelitian ini adalah untuk memahami perilaku pengguna, mengelompokkan pengguna berdasarkan jenis konten dan lingkup emosional pada permasalahan politik dan menganalisis kualitas hasil yang dicapai oleh Algoritma Hierarchical clustering dalam menentukan kluster.