

Pengelompokan Jenis Konten dan Peran User Berdasarkan Tweet Text Menggunakan K-Medoids Partitioning Based

^{1,2,3}Fakultas Informatika, Universitas Telkom, Bandung 1raisabenaya@students.telkomuniversity.ac.id,
2yuliant@telkomuniversity.ac.id, 3adityaihsan@telkomuniversity.ac.id

Abstrak

Di era modern ini, penyebaran informasi terjadi dengan cepat melalui media sosial. Salah satu saluran penyebaran informasi adalah melalui platform Twitter. Banyak pengguna twitter yang menyikapi konten yang ada dengan tanggapan positif, negatif maupun netral. Salah satu konten yang hangat untuk ditanggapi adalah konten politik. Konten ini sedang ramai dibicarakan mengingat pemilihan Calon Presiden Republik Indonesia 2024 yang semakin dekat. Salah satu pasangan calon yang dibahas adalah Anies Baswedan. Dengan banyaknya respon dari pengguna twitter, akan sulit untuk melacak apakah para pengguna mendukung Anies Baswedan untuk maju sebagai calon Presiden dikarenakan banyaknya respon yang ada. Penelitian ini bertujuan untuk mengetahui tanggapan pengguna twitter dengan majunya Anies Baswedan sebagai calon Presiden. Metode yang digunakan pada penelitian ini adalah algoritma K-Medoids Partitioning-Based berdasarkan teks pengguna twitter. Algoritma ini dipilih karena mudah untuk diimplementasikan mengingat dasar pembangunan K-Medoids adalah algoritma K-Means tetapi algoritma K-Medoids dapat mengatasi kekurangan dari algoritma K-Means yaitu sensitive terhadap outlier. Pengevaluasian akan dilakukan menggunakan Silhouette Score yang menghasilkan nilai sebesar 0,35 dengan jumlah cluster adalah 2. Kemudian dilakukan analisis terhadap setiap cluster dengan melihat kata-kata dalam cluster. Hasilnya dari kedua cluster yang terbentuk, kedua cluster mengandung konten positif dan menunjukkan bahwa pengguna twitter mendukung Anies baswedan untuk maju sebagai capres 2024.

Kata kunci : Twitter, Clustering, K-Medoids, TF-IDF Vectorizer, Silhouette Score

Abstract

On In this modern era, the spread of information occurs rapidly through social media. One of the channels for disseminating information is through the Twitter platform. Many Twitter users respond to existing content with positive, negative and neutral responses. One of the hot content to respond to is political content. This content is currently being discussed considering the approaching election of the 2024 Presidential Candidate of the Republic of Indonesia. One of the candidate pairs discussed was Anies Baswedan. With so many responses from Twitter users, it will be difficult to track whether users support Anies Baswedan to run as a presidential candidate due to the large number of responses. This study aims to determine the response of twitter users to the advancement of Anies Baswedan as a presidential candidate. The method used in this study is the K-Medoids Partitioning-Based algorithm based on twitter user text. This algorithm was chosen because it is easy to implement considering the basis of K-Medoids development is the K-Means algorithm but the K-Medoids algorithm can overcome the shortcomings of the K-Means algorithm which is sensitive to outliers. The evaluation will be done using Silhouette Score which produces a value of 0.35 with the number of clusters is 2. Then an analysis of each cluster is carried out by looking at the words in the cluster. As a result, from the two clusters formed, both clusters contain positive content and show that Twitter users support Anies Baswedan to run as a 2024 presidential candidate

Keywords: Twitter, Clustering, K-Medoids, TF-IDF Vectorizer, Silhouette Score

1. Pendahuluan

Media sosial adalah platform di mana seseorang dapat bersosialisasi dan menjelajah informasi secara bebas tanpa batasan waktu dan tempat. Ada banyak informasi yang dapat ditemukan di media sosial, baik yang bermanfaat maupun merugikan. Pengguna dapat melakukan berbagai aktivitas seperti mengirim pesan, mengunggah status, berbagi foto, menambahkan teman, dan lain-lain. Setiap pengguna memiliki perilaku yang berbeda-beda, hal ini dapat dilihat dari banyaknya aktivitas yang mereka lakukan di media sosial. Salah satunya merupakan platform media sosial yaitu Twitter. Media sosial ini memungkinkan seseorang untuk mempublikasikan pesan untuk mengekspresikan minat, favorit, pendapat, dan sentimennya terhadap berbagai topik dan isu yang mereka temui sehari-hari [1]. Terkadang, pengguna membuat pesan status (atau "tweet") yang berisi berbagai macam informasi. Tweet yang dibagikan biasanya berisi hal-hal positif (pujian) atau negatif (kritik) mengenai informasi tertentu yang sedang beredar [2] khususnya dibidang politik.

Dataset diperoleh melalui crawling data, lalu diolah melalui tahap preprocessing yang bertujuan untuk mengontrol bentuk data yang sering kali tidak terstruktur dan sulit dikelola sehingga memudahkan analisis data, ini adalah tahap krusial dalam memperbaiki akurasi prediksi dan mengurangi waktu komputasi pada system sebelum dilakukan metode clustering [3].

Pada penelitan sebelumnya [4] dilakukan pengelompokkan pengguna media sosial menggunakan algoritma Hierarchical Clustering dan Non-Hierarchical. Dimana hasil yang diperoleh menunjukkan bahwa penggunaan3ini