

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

<sup>1,2,3</sup>Fakultas Informatika, Universitas Telkom, Bandung [1raisabenaya@students.telkomuniversity.ac.id](mailto:1raisabenaya@students.telkomuniversity.ac.id),  
[2yuliant@telkomuniversity.ac.id](mailto:2yuliant@telkomuniversity.ac.id), [3adityaihsan@telkomuniversity.ac.id](mailto: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

berhasil mengelompokkan pengguna media sosial berdasarkan pola yang serupa sehingga hal ini dapat membantu pengembangan strategi pemasaran yang lebih baik di sosial media. Metode clustering digunakan untuk mengelompokkan objek-objek data menjadi kelompok-kelompok yang homogen berdasarkan kemiripan karakteristik atau atribut yang dimiliki oleh objek-objek data tersebut, metode clustering yang umum digunakan antara lain K-Means, Hierarchical Clustering, dan DBSCAN.

Tweet yang sedang beredar dibidang politik untuk saat ini yaitu mengenai seorang kandidat calon presiden yang mengusulkan diri dalam pemilihan presiden Indonesia 2024, topik ini dipilih karena sebentar lagi akan dilaksanakan pemilihan presiden Indonesia 2024 dan Anies Baswedan telah menjadi salah satu kandidat yang terpilih. Anies dikenal sebagai seorang pemimpin yang cenderung mempraktikkan gaya kepemimpinan yang demokratis. Ia selalu melibatkan warganya dalam setiap upaya perubahan yang dilakukan, dan ia dengan senang hati menerima masukan dan aspirasi dari warga [5]. Saat ia menjabat sebagai gubernur DKI Jakarta, Anies telah berhasil menghadapi banjir di Jakarta yang terjadi sebanyak enam kali sejak awal tahun 2020 hingga akhir februari 2020. Kepemimpinan dan kemampuan ini sangat diapresiasi terutama karena ia secara aktif terlibat dalam penanganan banjir di lapangan dan dianggap sebagai pemimpin yang responsive dalam menangani hal ini [6]. Hal ini menimbulkan simpatik bahkan empati terhadap tindakannya. Sehingga beberapa partai politik tertarik untuk berkolaborasi dengan Anies sebagai calon presiden pada Pilpres 2024 karena dianggap sebagai seorang figure yang layak untuk menjadi Presiden 2024 Berbagai komentar tweet, seperti positif (pujian) atau tweet negatif (hujatan), bermunculan di Twitter hingga menjadi trending topik. Berdasarkan hal tersebut, kami dapat mengumpulkan data perilaku pengguna berdasarkan jenis konten yang mereka tanggap menggunakan algoritma Berbasis Partisi K-Medoids. Dataset akan diambil dari proses crawling data dengan kata kunci “#AniesPr esiden2024”, setelah dataset di peroleh maka akan dilanjutkan dengan proses preprocessing untuk menyamakan format serta struktur kalimat dalam dataset. Tahap preprocessing ini sangat penting untuk meningkatkan akurasi prediksi dan mengurangi waktu komputasi pada sistem sebelum menerapkan metode clustering. Selanjut dataset akan dimasukkan kedalam algoritma K-Medoids untuk dilakukan proses clustering. Silhouette Score digunakan untuk membantu proses pencarian cluster yang optimal. Ketika cluster telah didapatkan maka akan dilakukan proses analisis untuk menentukan jenis konten dan peran pengguna twitter terhadap topik Anies Bawsedan maju sebagai capres 2024.

## 2. Studi Terkait

Banyak penelitian yang telah dilakukan mengenai pengelompokan data pada Twitter, seperti pada penelitian [7] untuk memahami pola dan variasi hashtag pengguna Twitter menggunakan metode Ward's hierarchical, hashtag diklaster menjadi beberapa grup untuk menampilkan struktur hireraki dari seluruh hashtag tersebut. Hasil yang didapat digunakan untuk memverifikasi perbedaan hasil pengelompokan antara setiap periode waktu untuk menganalisis detail variasi tiap kluster, ini dapat digunakan untuk pengembangan pemasaran iklan di Twitter.

Pada penelitian [8] algoritma K-Means digunakan untuk mengelompokkan dan menganalisis opini pengguna Twitter terhadap kasus miras oplosan, dan hasilnya menunjukkan kluster paling optimum terbentuk sebanyak tiga kluster berdasarkan nilai dunn index sebesar 0.8312 dimana kasus miras oplosan masih terpusat pada sosok pengedar miras oplosan, pihak berwenang, dan korban.

Penelitian [9] mengelompokkan opini public tentang bencana alam Indonesia di media social Twitter dengan membandingkan penggunaan algoritma DBSCAN dan K-Medoids. Opini public yang telah dikelompokkan berdasarkan kesamaan topik dan sentimennya menunjukkan bahwa kedua algoritma ini memiliki keunggulan yang berbeda, algoritma K-Medoids menghasilkan kelompok yang lebih jelas dan kohesi sedangkan algoritma DBSCAN lebih baik dalam mengidentifikasi opini public yang berbeda dari mayoritas atau disebut dengan outlier.

Lalu ada penelitian yang ditulis oleh Kharisma dkk [10] yang menggunakan metode K-Medoids clustering dalam menentukan segmentasi opini masyarakat di Twitter. Tweet-tweet yang berisi opini masyarakat akan melewati proses pelabelan, preprocessing, hingga proses clustering. Hasil cluster diuji menggunakan Silhouette Coefficient dan mendapatkan hasil 0.19. Metode clustering ini sukses dalam menentukan opini masyarakat dengan tingkat akurasi 80%.

Syamsul dan Dwi [11] melakukan penelitian terkait pengelompokan mahasiswa yang berpotensi drop out menggunakan metode K-Medoids. Dalam penelitian ini, mahasiswa dikelompokkan menjadi beberapa kelompok dengan karakteristik yang sama untuk mengidentifikasi mahasiswa yang membutuhkan perhatian lebih untuk menghindari mereka dari drop out. Hasil kluster dari 389 data menunjukkan adanya 3 atribut dengan rentang nilai yang berbeda antara kluster, yaitu IPK, IP Semester 1, dan Status Beasiswa. Ketiga atribut ini menjadi fitur yang membedakan antara kluster-kluster tersebut.

## 3. Sistem yang Dibangun

Pada penelitian ini, flowchart pada Figure1 menjelaskan mengenai tahapan proses pengelompokkan jenis konten dan peran user menggunakan K-Medoids. Beberapa tahapan dalam system ini terdiri dari crawling data, preprocessing, TF-IDF, model building, dan evaluasi hasil.