

Klustering Topik Pada Kolom Komentar Instagram Tentang Kabinet Merah Putih Menggunakan Metode K-Means (Clustering Topics in Instagram About Cabinet Merah Putih Using the K-Means Method)

Sefani Cahyo Auliya Rahayu^[1], Ramaditia Dwiyanaputra^[2], Ario Yudo Husodo^[3]

^[1 2 3]Dept Informatics Engineering, Mataram University

Jl. Majapahit 62, Mataram, Lombok NTB, INDONESIA

Email: sefanich.rcrrc49@gmail.com, [rama, ario]@unram.ac.id

Abstract

This research attempts to determine the primary themes that Indonesians talked on President Prabowo Subianto's "Merah Putih" cabinet by using clustering analysis of Instagram comments. Using crawling data from the Instagram platform with the hashtag #KabinetMerahPutih, comments were gathered using the K-Means Clustering approach. Prior to the data being analyzed to create five clusters, the cleaning and pre-processing procedure, which included tokenization with IndoBERT and dimensionality reduction using Principal Component Analysis (PCA), was able to greatly improve the clustering quality, with the Silhouette Coefficient value rising from 0.010 to 0.200. Out of 23.780 initial data, 9.320 clean data were processed for this investigation. The findings demonstrate that the K-Means algorithm can group comments according to pertinent themes and offer profound understanding of support more responsive public policy analysis.

Keywords: Klustering, Instagram, Komentar, Kabinet, K-Means

1 PENDAHULUAN

Media sosial dalam era kemajuan teknologi digital yang cepat telah bertransformasi menjadi saluran komunikasi utama yang sangat memengaruhi persepsi masyarakat. Instagram sebagai salah satu situs web media sosial yang paling terkenal di Indonesia memiliki 91,77 juta pengguna Instagram aktif, dengan 1,07 miliar di antaranya berusia 25 hingga 34 tahun. Instagram menempati urutan ketiga sebagai platform media sosial yang paling sering digunakan, setelah YouTube dan WhatsApp, dengan 36,4% pengguna terdiri dari kelompok usia 18 hingga 24 tahun. Instagram menyediakan konten visual yang menarik, platform ini memiliki forum komentar di mana pengguna dapat berdiskusi, memberikan masukan, dan membagikan pendapat mereka [1]. Oleh karena itu, media sosial berperan sebagai media massa pemerintahan di era sekarang, yang semakin relevan dalam konteks dinamika politik dan sosial.

Salah satu isu yang menyita perhatian masyarakat di postingan Instagram adalah persoalan kabinet baru di musim pergantian presiden dan wakil presiden. Peralihan dari kabinet Jokowi ke kabinet Prabowo, yang dinyatakan sebagai pemenang dalam pemilu yang dilaksanakan pada 14 Februari 2024 dan dilantik secara resmi pada 20 Maret 2024, menjadi perbincangan

hangat di kalangan masyarakat Indonesia [Kompas Nasional.com 2024]. Dalam konteks ini, pemerintahan di era kabinet Prabowo dikenal sebagai kabinet merah putih. Menurut KBBI Kabinet adalah badan atau dewan yang membentuk pemerintah. Nama kabinet merah putih dilatarbelakangi berdasarkan warna bendera nasional bangsa Indonesia yaitu merah putih yang melambangkan semangat persatuan dan identitas nasional [Wikipedia.com 2024]. Deklarasi kabinet merah putih yang dikeluarkan oleh Presiden Prabowo Subianto pada tanggal 20 Oktober 2024 menandai awal dari perubahan signifikansi dalam struktur pemerintahan. Selanjutnya, pelantikan pada tanggal 22 Oktober 2024, kabinet merah putih yang dipimpin oleh Presiden Prabowo Subianto memulai pekerjaannya. Perubahan struktur dalam kabinet ini mencakup penambahan jumlah kementerian menjadi 48, berbeda dengan kabinet sebelumnya yang hanya 34 kementerian. Pada era kabinet ini, terdapat 17 menteri dari kabinet sebelumnya yang kembali diangkat untuk periode 2024-2029 [Sekretariat Kabinet Republik Indonesia.com 2024]. Dengan demikian, perubahan ini tidak hanya mencerminkan dinamika politik, tetapi juga menciptakan ruang bagi masyarakat untuk berpartisipasi dalam diskusi mengenai kebijakan baru yang diusung oleh kabinet.

Berkaitan dengan situasi ini, pembahasan mengenai kabinet era Prabowo Subianto Joyohadikusumo bersama Gibran Rakabuming Raka telah menjadi isu yang populer dan menarik di kalangan masyarakat, banyak individu yang menyampaikan pandangan terkait kabinet merah putih melalui postingan Instagram dengan tagar #KabinetMerahPutih. Hal ini menunjukkan bahwa masyarakat yang semakin aktif dalam menyuarakan pendapat mereka di platform media sosial. Selanjutnya, informasi yang tersedia dari interaksi ini dapat dimanfaatkan untuk melakukan klustering topik yang muncul dalam diskusi di Instagram. Klustering adalah teknik dalam analisis data, metode di mana sekumpulan objek atau data dikelompokkan menjadi kelompok atau kluster berdasarkan kesamaan atau ciri-ciri tertentu [2]. Klustering topik dapat membantu dalam mengidentifikasi tema-tema utama yang dibahas oleh masyarakat, serta memberikan wawasan tentang bagaimana isu-isu tersebut saling berkaitan.

Dalam konteks ini, klustering menjadi metode yang relevan untuk mengelompokkan data komentar yang diperoleh. K-Means adalah algoritma pengelompokan yang membagi data menjadi beberapa kluster dengan menghitung jarak minimum antara titik data dan titik kluster [3]. Persamaan dan perbedaan ditentukan dari nilai atribut yang mewakili objek dan sering kali melibatkan perhitungan jarak atau *measures distance*. Jenis pembelajaran ini termasuk pembelajaran tanpa pengawasan yang tidak menggunakan data pelatihan, sehingga data yang ada dikelompokkan menjadi dua bagian, tiga bagian, atau lebih [4]. Dengan demikian, K-Means dapat memberikan wawasan yang lebih mendalam mengenai tema-tema yang muncul dalam diskusi masyarakat terkait kabinet merah putih.

K-Means juga merupakan algoritma pengelompokan yang dapat digunakan untuk mengelompokkan sejumlah besar data dengan metode partisi berbasis titik dengan waktu komputasi yang cepat. Metode pengelompokan non-hierarki ini mempertisi data menjadi dua kelompok atau lebih, kemudian memilih data sesuai dengan beberapa kriteria, dan kemudian digabungkan ke dalam satu *cluster* [5].

Oleh karena itu, menggunakan algoritma ini untuk klustering topik di Instagram dapat meningkatkan pemahaman pandangan masyarakat terhadap kabinet merah putih. Dengan menggunakan klustering, kita dapat mengidentifikasi tema-tema utama yang dibahas oleh masyarakat, sehingga analisis ini tidak hanya memberikan gambaran tentang opini publik, tetapi juga membantu dalam mengembangkan

kebijakan yang lebih sesuai dengan tuntutan masyarakat.

2 TINJAUAN PUSTAKA

Berikut adalah hasil dari tinjauan literatur yang telah dilakukan pada berbagai jurnal penelitian terdahulu yang berhubungan dengan topik *Clustering Metode K-Means*, yang dapat diuraikan sebagai berikut:

Penelitian [6] membahas tentang analisis menggunakan metode K-Means *Clustering* untuk mendorong pengguna Twitter untuk permainan lato-lato tradisional. Kemudian, data *crawling* dikategorikan menjadi tiga kategori sentimen: positif, netral, dan negatif. Hasil analisis sentimen yang dominan yaitu negatif sebanyak 962 *tweet* negatif, 112 *tweet* positif, dan 95 *tweet* netral.

Penelitian [7] membahas tentang penerapan algoritma K-Means untuk mengelompokkan respon pengguna layanan maskapai di Skytax. Terdapat 1.060 ulasan dari 3 maskapai dari tahun 2015 hingga 2019, Garuda Indonesia, Air Asia, dan Lion Air menerima hasil yang beragam, baik positif maupun negatif. Garuda Indonesia mencapai 71.8%, Lion Air 60.5%, dan Air Asia 52.8%. sehingga menunjukkan bahwa algoritma K-Means sangat efektif dalam menyediakan informasi bagi konsumen berdasarkan mutu layanan maskapai.

Penelitian [8] menganalisis sentimen pengguna Twitter terhadap Tokopedia dengan menggunakan K-Means; mereka menemukan 92.80% akurasi untuk analisis kata dan 89.80% akurasi untuk analisis tweet harian. Penelitian lain [9] mempelajari sentimen keluhan pengguna Indosat di Twitter (300 *tweet*, dari Desember 2018 hingga April 2019) dengan menggunakan clustering K-Means melalui tokenisasi, pemfilteran, dan stemming, yang menghasilkan tiga *cluster*: keluhan jaringan (*cluster 0*), permintaan perbaikan (*cluster 1*), dan masalah sinyal (*cluster 2*).

Penelitian lainnya yang membandingkan metode K-Means pada penelitian [10] dan [11]. Penelitian [10] membahas klustering puskesmas berdasarkan gizi balita, penelitian membandingkan K-Means dan Fuzzy C-Means. K-Means memiliki nilai *Silhouette Coefficient* 0.5183, sedangkan Fuzzy C-Means hanya memiliki nilai 0.4966. Penelitian [11] memeriksa pengelompokan rumah sakit di Jakarta menggunakan DBSCAN, GMM, dan *Agglomerative Hierarchical*. GMM memiliki nilai *Davies-Bouldin* terendah dan proporsi *cluster* tertinggi, masing-masing 62.83%, 36.65%, dan 0.52%. Sebaliknya, *Agglomerative Hierarchical* dan DBSCAN memiliki nilai *Davies-Bouldin* yang lebih tinggi, masing-masing 89.84% dan 3.45% di atas GMM.

Penelitian lain yang membandingkan pada penelitian [12] dan [13]. Penelitian [12] menggabungkan K-Means dan SVM untuk meningkatkan klasifikasi tema tugas akhir siswa, meningkatkan akurasi dari 85.38% menjadi 86.21%. Penelitian [13] mengklasifikasikan *blogger* musiman dan profesional, menggunakan dataset 100 data UCI. Akurasi 76.27%—naik 1.27% dari penelitian sebelumnya—dihasilkan oleh analisis *Rapidminer*, dengan kesalahan klasifikasi 23.73%.

Algoritma K-Means efektif untuk klustering, terbukti dalam analisis sentimen dan ulasan maskapai di Twitter. Kelebihannya mencakup kemampuan menangani dataset besar, menghasilkan nilai *Silhouette Coefficient* tinggi, dan meningkatkan akurasi klasifikasi saat dikombinasikan dengan algoritma lain. Metode ini cocok untuk mengelompokkan komentar Instagram tentang kabinet "Merah Putih" berdasarkan tema dan sentimen.

2.1 Klustering

Klustering adalah metode untuk mengklasifikasikan kumpulan data menjadi berbagai kelompok berdasarkan tingkat kecocokan yang diinginkan. Klustering melibatkan kumpulan data atau objek dalam beberapa grup atau *cluster*. Data di setiap *cluster* hampir sama dengan aslinya dan objek di dalamnya dapat dibedakan dari objek di *cluster* lainnya [5].

2.2 Instagram

Instagram memungkinkan pengguna berinteraksi melalui tanda suka, komentar, dan pesan yang terdiri dari teks, foto, atau video. Komentar dapat dibatasi untuk menghindari konten yang tidak relevan atau melampaui batas. Komentar yang memiliki jumlah tertinggi di bagian atas daftar ditampilkan dalam fitur komentar terbaik [14].

2.3 Komentar

Komentar yaitu bentuk ungkapan atau respon seseorang mengenai suatu topik, karya, tulisan, dan lain-lain. Komentar ini bisa dalam bentuk opini, kritik, saran, atau evaluasi untuk memperluas wawasan, memberikan klarifikasi, atau menyampaikan masukan terhadap suatu subjek yang dibahas [15].

2.4 Kabinet

Menurut [Wikipedia.com 2024] kabinet merupakan sekumpulan menteri yang ditunjuk oleh kepala negara sebagai pelaksana peran eksekutif dalam pemerintahan. Kabinet mendukung kepala negara dalam penyusunan dan penetapan kebijakan publik, mengatur urusan pemerintahan, serta membuat Keputusan yang berhubungan dengan administrasi negara.

2.5 K-Means

K-Means sebuah metode pengklasteran data berbasis partisi yang dibagi menjadi beberapa kelompok berbeda dengan pendekatan yang berulang, dan melalui iterasi K-Means dapat mengurangi rata-rata jarak antar setiap data dan klusternya [3]. Metode ini adalah pendekatan data tanpa label (*unsupervised learning*) yang mengelompokkan data berdasarkan fitur tertentu. Algoritma diawali dengan penetapan nilai *k* sebagai *cluster* lalu memilih secara acak *K* titik awal data sebagai *cluster* pusat (*centroid*). Kemudian menghitung jarak antar data dan centroid dengan *Euclidean Distance* sebagai berikut: [7].

$$d(x_i, \mu_j) = \sqrt{\sum_{k=1}^n (x_{ik} - \mu_{jk})^2}$$

Keterangan:

$d(x_i, \mu_j)$ = Jarak antara data ke-*i* dan centroid ke-*j*

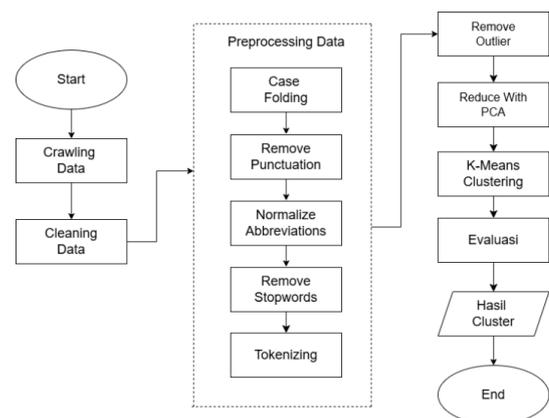
x_i = Data ke-*i*, dimana $i = 1, 2, \dots, m$

μ_j = Centroid ke-*j*, dimana $j = 1, 2, \dots, k$

n = Jumlah fitur dari data

3 METODE PENELITIAN

Berikut beberapa proses yang dilakukan antara lain pengumpulan atau *crawling* data dari platform media sosial Instagram, preprocessing data, *remove outlier*, *reduce with PCA*, model *K-Means Clustering*, evaluasi, dan terakhir hasil *cluster*.



Gambar 1. Flowchart Metode Penelitian

3.1 Crawling Data

Teknik *crawling* menggunakan *extension Instant Data Scrapper* untuk mengumpulkan data mentah dalam format csv. Pada Tabel 1 menunjukkan bahwa data awal mengandung banyak tulisan yang tidak diperlukan, seperti *hashtag*, emoji, simbol-simbol, URL, data duplikat, dan *@mention*, dll.

TABEL I. CRAWLING DATA

| No | Teks |
|----|--|
| 1 | Menyala terus bapak presiden, saya yakin bapak pasti bisa 💎 bangga sekali punya pemimpin negara seperti bapak, sehat sehat terus bapak dan sekeluarga semoga selalu dalam lindungan Allah SWT dan dipermudah segala urusannya aamiin,, |
| 2 | Pak Prabowo gratisan dunia pendidikan dan kesehatan. Itu akan meringankan rakyatnya dari kemiskinan dan perang pengangguran . |
| .. | .. |

3.2 Cleaning Data

Data mentah yang dikumpulkan melalui crawling masih mengandung elemen yang perlu dibersihkan, seperti data duplikat, *@mention*, *hashtag*, emoji, URL, dan elemen lainnya yang tidak relevan dengan tujuan penelitian ini. Oleh karena itu, dari 23.780 data komentar yang berhasil dibersihkan, 9.320 data masih dilanjutkan ke tahap berikutnya. Tabel 2 berikut menunjukkan hasil data setelah proses perbaikan.

TABEL II. CLEANING DATA

| No | Teks |
|-----|--|
| 1 | Menyala terus bapak presiden, saya yakin bapak pasti bisa bangga sekali punya pemimpin negara seperti bapak, sehat sehat terus bapak dan sekeluarga semoga selalu dalam lindungan Allah SWT dan dipermudah segala urusannya aamiin,, |
| 2 | Pak Prabowo gratisan dunia pendidikan dan kesehatan. Itu akan meringankan rakyatnya dari kemiskinan dan perang pengangguran |
| ... | ... |

3.3 Preprocessing Data

3.3.1 Case Folding

Untuk menyeragamkan karakter pada data, proses *case folding* digunakan setelah data diproses pada tahap sebelumnya. Untuk menyeragamkan kata, setiap huruf dalam penelitian ini diubah menjadi huruf kecil.

TABEL III. DATA CASE FOLDING

| No | Teks |
|----|---|
| 1 | menyala presiden bangga pemimpin negara sehat sehat sekeluarga semoga |

| | |
|----|--|
| | lindungan allah subhanahu wata'ala dipermudah urusannya amin |
| 2 | prabowo gratisan dunia pendidikan kesehatan meringankan rakyatnya kemiskinan perang pengangguran |
| .. | .. |

3.3.2 Remove Punctuation

Pada tahap ini, tanda baca seperti titik, koma, dan tanda tanya diekstrak dari teks. Tanda baca biasanya tidak memberikan informasi yang berguna untuk analisis teks, dan mereka juga dapat mengganggu proses tokenisasi dan *stemming*.

TABEL IV. DATA REMOVE PUNCTUATION

| No | Teks |
|----|---|
| 1 | menyala presiden bangga pemimpin negara sehat sehat sekeluarga semoga lindungan allah subhanahu wataala dipermudah urusannya amin |
| 2 | prabowo gratisan dunia pendidikan kesehatan meringankan rakyatnya kemiskinan perang pengangguran |
| .. | .. |

3.3.3 Normalize Abbreviations

Pada tahap normalisasi, kamus KBBI digunakan untuk mengubah kata-kata yang tidak baku menjadi kata baku. Tujuan dari proses ini adalah untuk mengubah kata-kata tersebut menjadi bentuk yang lebih umum dan sesuai dengan entri dalam Kamus Besar Bahasa Indonesia (KBBI).

TABEL V. DATA NORMALIZE ABBREVIATIONS

| No | Teks |
|----|---|
| 1 | [nyala, presiden, bangga, pemimpin negara, sehat, keluarga, semoga,.....] |
| 2 | [prabowo, gratis, dunia, pendidikan, sehat, ringan, rakyat, miskin.....] |
| .. | .. |

3.3.4 Remove Stopwords

Proses untuk meningkatkan efisiensi analisis teks dengan mengurangi jumlah kata yang diproses dan meningkatkan kualitas analisis dengan menghilangkan kata-kata yang tidak relevan.

TABEL VI. DATA REMOVE STOPWORDS

| No | Teks |
|----|--|
| 1 | [presiden, bangga, pemimpin, negara sehat, keluarga, semoga, lindungan.....] |

| | |
|----|--|
| 2 | [probowo, gratis, pendidikan, sehat ringan, rakyat, miskin.....] |
| .. | .. |

3.3.5 Tokenizing

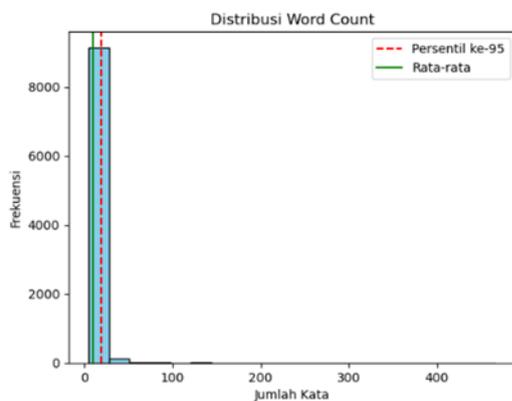
Untuk mempersiapkan teks untuk analisis lebih lanjut, tokenisasi dilakukan dengan menggunakan model IndoBERT untuk memecah teks menjadi bagian yang lebih kecil (token).

TABEL VII. DATA TOKENIZING

| No | Teks |
|----|---|
| 1 | [menyala, presiden, bangga, pemimpin, negara, sehat, sekeluarga, semoga, lindungan.....] |
| 2 | [probowo, gratiskan, dunia, pendidikan, kesehatan, meringankan, rakyatnya, kemiskinan.....] |
| .. | .. |

3.4 Remove Outlier

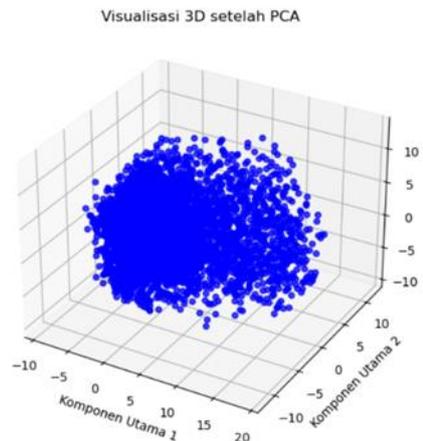
Data *outlier* harus dihapus untuk menjaga analisis representatif. Gambar 2 menunjukkan histogram distribusi jumlah kata. Sebagian besar teks memiliki antara 0 dan 50 kata, tetapi beberapa memiliki banyak kata. Garis hijau menunjukkan jumlah kata rata-rata, dan garis putus-putus merah menunjukkan persentil 95 sebagai titik tertinggi. Ketidakeimbangan data ditunjukkan oleh pola distribusi yang diskew ke kiri.



Gambar 2. Visualisasi *Remove Outlier*

3.5 Reduce with PCA

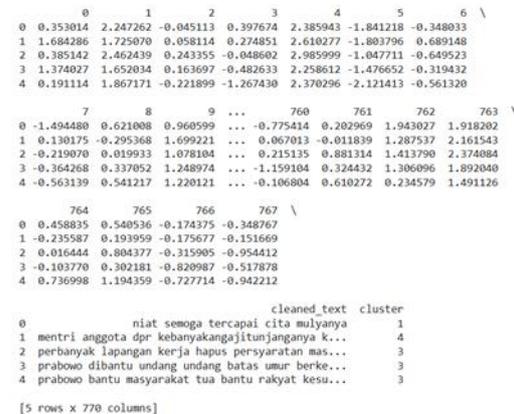
Teknik PCA mengurangi dimensi data sambil mempertahankan sebagian besar variansi. Ini ditunjukkan dalam visualisasi tiga dimensi oleh hasil PCA, di mana data tersebar membentuk pola elips simetris tanpa *cluster* yang jelas, seperti yang ditunjukkan dalam Gambar 3. Metode ini membantu menemukan pola global dalam data yang sangat besar.



Gambar 3. Visualisasi 3D PCA

3.5 K-Means Clustering

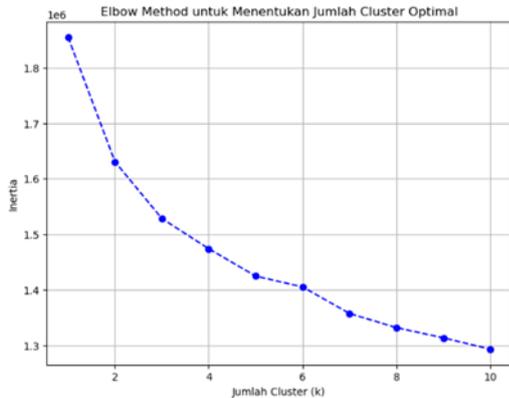
Pusat *cluster* awal digambarkan dalam ruang vektor 768 dimensi pada Gambar 4. Dokumen ditambahkan ke *cluster* dalam iterasi K-Means berdasarkan jarak geometris antara vektor dokumen dan pusat *cluster*. Perwakilan numerik untuk perhitungan jarak tersedia di kolom 0-767. Pusat *cluster* diperbarui sebagai vektor dokumen *cluster* rata-rata. Proses ini dilakukan berulang kali sampai konvergensi, yang berarti dokumen tidak berpindah atau pusat *cluster* tidak berubah.



Gambar 4. Model K-Means Clustering

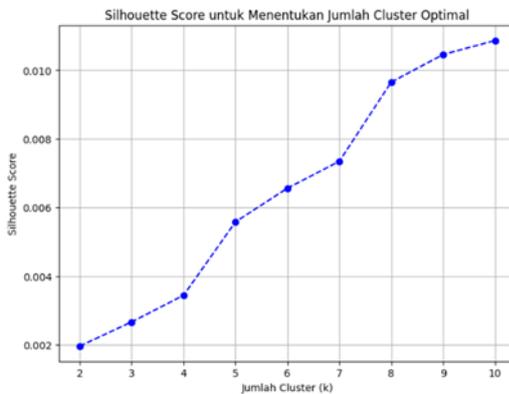
3.6 Evaluasi

Metode *Elbow* dan *Silhouette Coefficient* digunakan untuk menilai model *clustering*. Sebelum menggunakan Metode *Elbow* untuk menentukan jumlah *cluster* terbaik berdasarkan nilai inerti, PCA digunakan untuk mereduksi data. Gambar 5 menunjukkan visualisasi Metode *Elbow*, dengan jumlah *cluster* (*k*) diwakili oleh sumbu X dan nilai inerti, yaitu jumlah kuadrat jarak titik data ke pusat *cluster* terdekat. Jumlah *cluster* optimal adalah 4, karena titik "elbow", di mana penurunan inerti melambat, terlihat di sekitar *k* = 4.

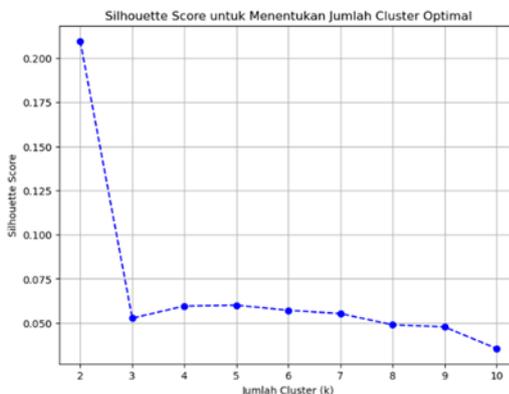


Gambar 5. Grafik Elbow Method

Silhouette Coefficient digunakan untuk mengukur kualitas *clustering*, dengan nilai antara -1 dan 1 yang menunjukkan kedekatan objek dalam *cluster*. Nilai *Silhouette Coefficient* tanpa PCA adalah 0.010 yang menunjukkan *clustering* yang kurang optimal. Namun, setelah menggunakan PCA untuk mengurangi dimensi kata nilainya meningkat menjadi 0.200 menunjukkan bahwa PCA membantu memperjelas struktur data dan memisahkan *cluster* dengan lebih baik, seperti yang ditunjukkan pada Gambar 6 dan 7.



Gambar 6. Grafik Tanpa PCA



Gambar 7. Grafik Dengan PCA

3.7 Hasil Cluster

Menghasilkan kelompok-kelompok berdasarkan topik atau tema dari 5 *cluster* yang berbeda untuk menampilkan hasil *clustering* setelah konvergensi, sehingga dapat menganalisis pola sentimen dari masing-masing *cluster* dengan representasi vektor numerik untuk setiap komentar yang dihasilkan melalui penerapan metode *word embeddings* (Word2Vec) untuk hasil visualisasi *cluster* dan interpretasi nilai setiap *cluster*.

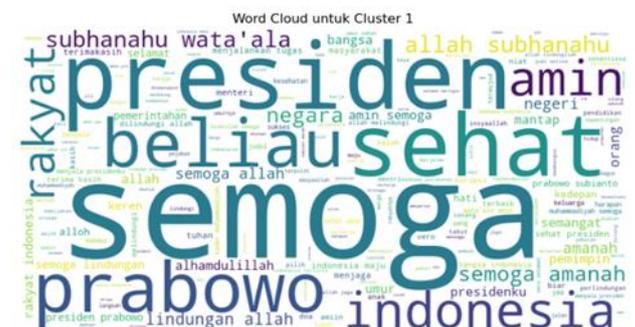
4 HASIL DAN PEMBAHASAN

Penelitian ini mengumpulkan 23.780 komentar Instagram dengan kata kunci "#KabinetMerahPutih" sejak November 2024. *Wordcloud* digunakan untuk visualisasi untuk menemukan kata kunci yang paling banyak dibicarakan. Hasilnya dibagi menjadi lima *cluster* dengan 20 kata yang paling sering muncul dari *cluster* 0 hingga *cluster* 4.



Gambar 8. WordCloud Cluster 0

Cluster ini berkonsentrasi pada masalah sosial dan ekonomi yang terkait dengan kehidupan sehari-hari masyarakat. Kata-kata seperti "kaya", "duit", "uang", dan "kerja" berfokus pada kesejahteraan ekonomi, sedangkan "rakyat" dan "negara" berfokus pada hubungan antara pemerintah dan masyarakat. Secara umum, kelompok ini dapat mencakup pembicaraan tentang situasi sosial-ekonomi dan harapan masyarakat terhadap pemerintah.



Gambar 9. WordCloud Cluster 1

Cluster ini berfokus pada doa dan harapan untuk kemakmuran bangsa dan pemimpinnya. Kata-kata seperti "semoga", "amin", dan "sehat" membawa semangat dan optimisme ke masa depan. Ini bisa mencakup percakapan tentang kebijakan pemerintah yang diharapkan akan menguntungkan masyarakat.



Gambar 10. WordCloud Cluster 2

Cluster ini menunjukkan bahwa masalah politik dan ekonomi yang berkaitan dengan pemimpin dan kebijakan. Ada perdebatan tentang kepemimpinan dan kebijakan pemerintah dengan menggunakan istilah seperti "presiden", "prabowo", dan "menteri". Ini bisa termasuk melihat bagaimana kebijakan ekonomi memengaruhi masyarakat.



Gambar 11. WordCloud Cluster 3

Fokus cluster ini adalah korupsi dan masalah politik Indonesia. Ada diskusi tentang tokoh-tokoh politik dan kebijakan melalui penggunaan kata-kata seperti "prabowo", "presiden", dan "menteri". Kata-kata seperti "korupsi", "anggaran", dan "pejabat" menunjukkan bahwa integritas dan transparansi pemerintah menjadi masalah.



Gambar 12. WordCloud Cluster 4

Cluster ini menunjukkan perhatian pada masalah sosial dan ekonomi yang lebih luas, seperti kebijakan pemerintah dan utang. Ada perdebatan tentang hubungan antara masyarakat dan pemerintah ketika digunakan istilah "rakyat", "negara", dan "presiden". Kata-kata seperti "hutang", "uang", dan "anggaran" menunjukkan fokus pada pengelolaan sumber daya dan masalah keuangan.

5 KESIMPULAN DAN SARAN

Hasil penelitian menunjukkan bahwa kualitas clustering secara signifikan ditingkatkan dengan preprocessing data. Dengan menggunakan PCA, nilai koefisien *Silhouette* meningkat dari 0.010 menjadi 0.200, menunjukkan efektivitasnya dalam pengurangan dimensi. Untuk dataset ini, algoritma K-Means terbukti terbaik karena mampu mengelompokkan komentar menjadi lima cluster relevan secara cepat, efisien, dan menghasilkan hasil yang jelas. Algoritma ini dipilih karena kemudahan penggunaan, kecepatan komputasi yang tinggi, dan kemampuan untuk menangani dataset yang lebih besar. Oleh karena itu, algoritma ini cocok untuk menganalisis komentar media sosial yang kompleks. Masalah sosial, harapan pemimpin, dan politik adalah topik utama hasil clustering.

Penelitian mendatang harus menggabungkan metode clustering aglomeratif atau DBSCAN untuk perbandingan. Analisis sentimen dapat melengkapi kelompokan dan memberikan wawasan lebih mendalam tentang persepsi masyarakat terhadap kabinet "Merah Putih". Analisis temporal untuk mengamati perubahan pendapat juga dapat dimasukkan dalam penelitian lanjutan.

UCAPAN TERIMA KASIH

Kami mengucapkan terima kasih kepada semua yang telah membantu kami menyelesaikan jurnal ini dengan baik. Kami berharap pembaca dan bidang terkait mendapatkan manfaat darinya.

DAFTAR PUSTAKA

- [1] A. A. Suwu, N. Widyaputri, and J. Tumiwa, "Analisis Pemanfaatan Aplikasi Instagram Dalam Pemasaran Bisnis Online Shop Di Kota Manado," *Jurnal Ilmiah Society*, vol. 2, no. 2, pp. 1-10, 2022. ISSN: 2337-4004.
- [2] A. K. Rahmansyah, A. T. Shidqul Aziz, N. Novianto, and D. Rolliawati, "Perbandingan algoritma K-Means dan Fuzzy C-Means untuk clustering Puskesmas berdasarkan gizi balita di Surabaya," *Processor: Jurnal Ilmiah Sistem Informasi, Teknologi Informasi dan Sistem Komputer*, vol. 18, no. 1, pp. 83-88, Apr. 2023. doi: 10.33998/processor.2023.18.1.696.
- [3] E. A. Saputra and Y. Nataliani, "Analisis pengelompokan data nilai siswa untuk menentukan siswa berprestasi menggunakan metode clustering K-Means," *Journal of Information Systems and Informatics*, vol. 3, no. 3, pp. 1-10, Sep. 2021.
- [4] F. I. Priatna, W. H. Fajri, R. R. Novariyanti, and M. Z. Rizquloh, "Analisis sentimen masyarakat terhadap konten akun Instagram PSSI: Studi kasus pada reaksi terhadap postingan terkini," *JUMASH (Jurnal Mahasiswa Sosial Humaniora)*, vol. 1, no. 1, pp. 56-74, Juli–Desember 2024.
- [5] K. E. Setiawan dan A. Kurniawan, "Pengelompokan Rumah Sakit di Jakarta Menggunakan Model DBSCAN, Gaussian Mixture, dan Hierarchical Clustering," *J. Informatika Terpadu*, vol. 9, no. 2, pp. 149-156, Sep. 2023.
- [6] K. L. Berutu, H. Rumapea, and H. G. Manullang, "Analisa sentimen penggunaan permainan tradisional Lato – Lato pada sosial media Twitter menggunakan metode K-Means," *Methotika: Jurnal Ilmiah Teknik Informatika*, vol. 4, no. 1, pp. 42-51, Apr. 2024.
- [7] M. A. Z. Larasati, N. A. S. Winarsih, M. S. Rohman, and G. W. Saraswati, "Penerapan metode K-Means clustering dalam menganalisis sentimen masyarakat terhadap K-Poppers pada Twitter," *Progresif: Jurnal Ilmiah Komputer*, vol. 1, no. 1, pp. 1 -10, 2024.
- [8] N. Larasati1 and R. N. Fahmi, "Implementasi metode K-Means clustering dalam analisis persebaran UMKM di Jawa Barat," *Journal of Information System*, vol. 6, no. 2, pp. 211-220, Nov. 2021. doi: 10.33633/joins.v6i2.5310.
- [9] N. Widiastuti, A. Hermawan, dan D. Avianto, "Implementasi Metode Naïve Bayes untuk Klasifikasi Data Blogger," *J. Ilmiah Penelitian Informatika*, vol. 8, no. 3, pp. 985-999, Sep. 2023. doi: 10.29100/jipi.v8i3.3713.
- [10] O. Somantri, S. Wiyono, and Dairoh, "Metode K-Means untuk optimasi klasifikasi tema tugas akhir mahasiswa menggunakan Support Vector Machine (SVM)," *Scientific Journal of Informatics*, vol. 3, no. 1, pp. 35-44, Mei 2016.
- [11] S. I. Safitri, C. Suhery, and S. Bahri, "Implementasi algoritma K – Means untuk clustering sentimen pada opini kualitas pelayanan jasa penerbangan," *Jurnal Komputer dan Aplikasi*, vol. 9, no. 2, pp. 186-197, 2021.
- [12] T. I. Saputra and R. Arianty, "Implementasi algoritma K-Means clustering pada analisis sentimen keluhan pengguna Indosat," *Jurnal Ilmiah Informatika Komputer*, vol. 24, no. 3, pp. 191-197, Des. 2019.
- [13] Wicaksono, B., and Sari, N., "Jurnal Media dan Komunikasi Digital," vol. 4, no. 4, pp. 221-235, 2018.
- [14] W. S. Gemilang, P. Purwantoro, and C. Carudin, "Analisis sentimen pengguna Instagram pada calon presiden 2024 menggunakan algoritma Support Vector Machine," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 7, no. 4, pp. 1-10, Agustus 2023.
- [15] Y. Andrianus, W. Wasino, and T. Sutrisno, "Implementasi algoritma K-Means terhadap opini masyarakat mengenai perkiraan pemilu 2024 pada Twitter," *J. Sistem Informasi dan Teknik Komput.*, vol. 8, no. 2, pp. 305-316, 2023.
- [16] F. N. Ulya., and D. Prabowo. "Kompas Nasional.com" Tersedia: <https://nasional.kompas.com/read/2024/12/25/12244691/kaleidoskop-2024-bergantinya-kepemimpinan-jokowi-ke-prabowo?page=all> pada tanggal 14 November 2024.
- [17] Wikipedia, "Bendera Indonesia," *Wikipedia.com* Tersedia:https://id.wikipedia.org/wiki/Bendera_Indonesia. Diakses: 14 November 2024.
- [18] Sekretariat Kabinet Republik Indonesia, "Inilah Kementerian Negara Kabinet Merah Putih," Tersedia: <https://setkab.go.id/inilah-kementerian-negara-kabinet-merah-putih/>. Diakses: 14 November 2024.