

Analisis Persepsi Publik Terhadap Pilkada Jakarta 2024 dengan Clustering dan Sentimen pada Artikel Berita

Anggi Nur Zavira¹, Ilma Fathiyarahmani², Khansa Nadhifa³, Kinanti Anindia Putri⁴, Widya Amanda Mulyadi⁵, Zalfa Syawaliana⁶

^{1,2,3,4,5,6} Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Jakarta, Indonesia

Email: ¹anggi_1314622009@mhs.unj.ac.id, ²ilma_1314622023@mhs.unj.ac.id, ³khansa_1314622032@mhs.unj.ac.id, ⁴kinanti_1314622024@mhs.unj.ac.id, ⁵widya_1314622025@mhs.unj.ac.id, ⁶zalfa_1314622027@mhs.unj.ac.id

ABSTRAK

Penelitian ini menganalisis pola pemberitaan media massa terkait Pilkada Jakarta 2024 menggunakan pendekatan *text mining* dengan algoritma K-Means Clustering dan analisis sentimen berbasis Lexicon. Data diperoleh melalui *web scraping* dari RSS Feed Google News, menghasilkan 100 artikel yang dianalisis setelah melalui proses *preprocessing* berupa *tokenizing*, *filtering*, dan *stemming*. Algoritma K-Means digunakan untuk mengelompokkan artikel ke dalam delapan *cluster* berdasarkan tema dominan, seperti dukungan politik, kegagalan pencalonan, dan isu strategis terkait Pilkada. Algoritma ini bekerja dengan menghitung jarak antara data dan centroid, yang terus diperbarui hingga mencapai *cluster* optimal berdasarkan metode *Silhouette Score*. Analisis sentimen menunjukkan bahwa sebagian besar artikel memiliki sentimen netral, mencerminkan objektivitas media, meskipun beberapa *cluster* memperlihatkan adanya sentimen positif dan negatif, yang mengindikasikan potensi bias dalam pemberitaan. Temuan ini memberikan wawasan mengenai peran media dalam membentuk opini publik serta pengaruh pola pemberitaan terhadap persepsi masyarakat dalam proses demokrasi. Penelitian ini diharapkan dapat meningkatkan literasi politik masyarakat dan mendorong partisipasi yang lebih kritis, sekaligus membuka peluang pengembangan metode analisis lebih lanjut untuk memahami bias media secara lebih mendalam.

Kata Kunci: *Text Mining, K-Means Clustering, Sentiment Analysis*

ABSTRACT

This study analyzes the pattern of mass media coverage related to the 2024 Jakarta Gubernatorial Election using a text mining approach with the K-Means Clustering algorithm and Lexicon-based sentiment analysis. Data were obtained through web scraping from Google News RSS Feeds, resulting in 100 articles that were analyzed after undergoing preprocessing processes such as tokenizing, filtering, and stemming. The K-Means algorithm was used to cluster the articles into eight clusters based on dominant themes, such as political support, candidacy failures, and strategic issues related to the election. This algorithm works by calculating the distance between data points and centroids, which are continuously updated until an optimal cluster is achieved based on the Silhouette Score method. Sentiment analysis revealed that most articles had a neutral sentiment, reflecting media objectivity, although some clusters showed positive and negative sentiments, indicating potential bias in the coverage. These findings provide insights into the role of media in shaping public opinion and the influence of news coverage on public perceptions in the democratic process. This study is expected to enhance political literacy among the public and encourage more critical participation, while also opening opportunities for further development of analysis methods to better understand media bias.

Keywords: *Text Mining, K-Means Clustering, Sentiment Analysis*

Penulis Korespondensi:

Widya Amanda Mulyadi

Email: widya_1314622025@mhs.unj.ac.id

Article Info

Diterima: 26 Desember 2024

Direvisi: 3 Januari 2025

Disetujui: 12 Februari 2025

This is an open access article under the [CC BY](https://creativecommons.org/licenses/by/4.0/) license.



1. PENDAHULUAN

Pilkada merupakan momen yang sangat penting dalam sistem demokrasi di Indonesia terutama pada Pilkada Jakarta dimana memiliki dampak signifikan terhadap dinamika politik nasional. Salah satu sumber informasi bagi masyarakat untuk mengetahui informasi terbaru terkait perkembangan isu politik, kompetisi para calon, serta strategi kampanye yang telah diterapkan oleh partai politik ataupun individu adalah media massa terutama media massa *online* yang sangat *online* sangat memudahkan masyarakat untuk mendapatkan informasi terkait Pilkada Jakarta 2024.

Media massa memiliki peran penting dalam pesta demokrasi di Indonesia dengan menyalurkan berbagai informasi politik, media massa dapat memengaruhi masyarakat untuk pro aktif dalam mendukung terciptanya sistem demokrasi melalui pemilihan umum yang sesuai dengan amanat undang-undang yakni langsung, umum, bebas, rahasia, jujur dan adil (LUBER JURDIL). Namun, netralitas media massa sering kali menjadi isu yang diperdebatkan, terutama dalam konteks pemberitaan politik. Menurut penelitian pada Pilkada Manado 2020 yang dilakukan oleh Claudia Rahmim, dkk [1] terdapat dua media massa yang tidak netral dalam penyajian data yaitu, media lokal Sulut Online dan Berita Manado, ketidaknetralan ini terlihat dari porsi pemberitaan yang tidak seimbang dan isi berita yang sering memihak pada kandidat tertentu.

Indonesia Corruption Watch (ICW) membuka diskusi publik mengenai kritisi kampanye Pilkada Jakarta 2024 yang dianggap minim gagasan konkret juga kurang solutif terhadap isu-isu masyarakat. Para paslon dinilai terlalu kota-sentris sehingga mengabaikan permasalahan mendesak seperti perampasan ruang hidup, krisis iklim, kerentanan ekonomi kelas menengah, dan maraknya praktik korupsi pada pengelolaan anggaran daerah. Perubahan status Jakarta menjadi pusat ekonomi nasional juga mempengaruhi jalannya Pilkada Jakarta. Ternormalisasinya politik uang, lemahnya visi antikorupsi, serta kegagalan memahami kebutuhan masyarakat semakin memperburuk legitimasi proses demokrasi.

Pilkada Jakarta tidak hanya berdampak pada dinamika politik, tetapi juga pada aspek-aspek sosial dalam masyarakat. Pada proses pemilihan sering kali memicu diskusi publik yang cukup intens sehingga memengaruhi kohesi sosial dan membentuk kesadaran politik di kalangan masyarakat. Dalam hal ini, peran media massa sangatlah penting untuk membentuk opini-opini publik maka dari itu media massa haruslah menyuguhkan pemberitaan yang bersifat netral agar tidak memicu bias atau konflik lebih lanjut, dengan memahami pola pemberitaan dan sentimen media massa sangat penting untuk menangkap fenomena sosial yang berkembang selama Pilkada Jakarta. Tingginya volume informasi yang tersedia menyebabkan tantangan tersendiri, baik bagi masyarakat umum maupun bagi peneliti yang ingin menganalisis tren pemberitaan. Adanya data berita yang sangat beragam sering kali sulit dipahami tanpa adanya sistem pengelompokan atau analisis yang terstruktur.

Untuk mengatasi tantangan tersebut, teknologi seperti *text mining* dan algoritma *K-Means Clustering* menjadi solusi yang efektif untuk mengelompokkan artikel-artikel berita berdasarkan kesamaan tema, sehingga dapat membantu dalam memahami pola pemberitaan secara lebih jelas. Selain pengelompokan berita, penelitian ini akan menambahkan analisis sentimen pada setiap kluster. Analisis sentimen akan dilakukan menggunakan *text mining* untuk mengklasifikasikan sentimen menjadi tiga kategori utama, yaitu sentimen positif, di mana artikel memberikan pandangan atau opini yang mendukung calon tertentu; sentimen negatif, di mana artikel berisi kritik, isu kontroversial, atau sentimen buruk terhadap calon atau kebijakan tertentu; dan sentimen netral, di mana artikel bersifat informatif dan tidak memihak.

Penelitian yang dilakukan oleh Yepin Andrianus, dkk [2] membahas implementasi algoritma *K-Means Clustering* untuk menganalisis opini masyarakat terkait pemilu 2024 berdasarkan data dari Twitter. Hasil yang didapat menunjukkan bahwa mayoritas opini masyarakat bersifat netral dengan persentase 92,50%, sementara sentimen positif hanya 4,30% dan sentimen negatif sebesar 3,20%. Tingkat akurasi pengelompokan diuji menggunakan metode *Silhouette Coefficient*, yang menunjukkan hasil kuat dengan nilai 0,981. Penelitian ini membuktikan bahwa *K-Means* efektif dalam mengelompokkan opini publik dan memberikan gambaran tentang sentimen masyarakat terhadap pemilu.

Selain itu, penelitian terdahulu oleh Atika Rahmawati, dkk [3] membandingkan algoritma *Support Vector Machine (SVM)* dan *K-Means* untuk analisis sentimen publik terhadap pelaksanaan Pilkada serentak. Hasil yang didapatkan melalui metode SVM mendapatkan akurasi tertinggi sebesar 91% sedangkan *K-Means* hanya mencatat 83%. Penelitian ini menunjukkan efektivitas algoritma SVM lebih unggul dalam klasifikasi sentimen, namun *K-Means* tetap relevan untuk pengelompokan tema dalam *dataset* yang besar dan heterogen.

Tujuan penelitian ini tidak hanya memberikan kontribusi bagi para peneliti dalam memahami pola pemberitaan dan sentimen media terkait Pilkada Jakarta 2024, tetapi juga memberikan manfaat nyata bagi masyarakat umum. Melalui pengelompokan artikel berita menggunakan algoritma *K-Means Clustering* dan analisis sentimen, dengan memanfaatkan data berita yang diambil melalui teknik *web scraping* dari RSS Flead Google News, penelitian ini diharapkan masyarakat dapat memperoleh informasi terkait isu-isu yang paling banyak dibicarakan dari Pilkada Jakarta 2024 pada media massa *online* yaitu pada portal Google News. Dengan demikian masyarakat dapat memahami lebih jelas bagaimana media massa *online* membentuk opini publik terhadap Pilkada Jakarta 2024. Hasil penelitian ini diharapkan dapat meningkatkan literasi politik masyarakat, membantu mereka dalam mengidentifikasi pemberitaan yang relevan, dan mendorong partisipasi aktif serta kritis dalam proses demokrasi yang lebih berkualitas.

2. METODE PENELITIAN

2.1 Teknik Pengambilan Data

Data dalam penelitian ini diperoleh dengan menggunakan metode *web scraping* untuk mengumpulkan artikel berita dari Google News RSS Feed terkait topik “Pilkada Jakarta 2024”. Asfandyer et. al. (2023) dalam penelitian *Automatic Scientific Literature Gathering and Analysis Using Web Scraping and Locality Sensitive Hashing* menyebutkan bahwa metode web metode *scraping* terbukti efektif dalam pengumpulan data teks terstruktur untuk kepentingan analisis ilmiah. Sejalan dengan Papadopoulus et. al. (2022) menilai bahwa *web scraping* mendukung analisis berbasis data secara komprehensif dan akurat dalam penelitian *Big Data Analysis and Web Mining Techniques for News Articles*. Proses pengambilan data menggunakan metode *web scraping* dilakukan dengan memanfaatkan fungsi *feedparser* dalam Python yang secara otomatis mengambil, menguraikan, hingga memecah data menjadi bagian-bagian yang lebih kecil dalam XML atau JSON dari sumber RSS Feed.

Data yang dikumpulkan meliputi judul artikel, tautan berita, tanggal publikasi, serta ringkasan isi artikel dalam format terstruktur seperti CSV atau DataFrame untuk dianalisis lebih lanjut. Tahapan pengambilan data dimulai dengan mengakses RSS Feed, lalu menentukan kata kunci pencarian yang relevan, yaitu “Pilkada Jakarta 2024” melalui tautan <https://news.google.com/rss/search?q=pilkada%20jakarta%202024&hl=en-ID&gl=ID&ceid=ID%3Aen>. Selanjutnya, *feedparser* akan membaca dan mengekstrak konten dari setiap entri RSS Feed untuk memastikan konsistensi data sehingga dinyatakan layak untuk proses analisis lebih lanjut. Teknik ini memungkinkan peneliti untuk mengumpulkan data skala besar secara otomatis, sistematis, dan meminimalisir waktu dan upaya manual pengumpulan data berita dari berbagai sumber berbeda.

2.2 Analisis Clustering dengan K-Means

Analisis *clustering* dilakukan dengan menggunakan algoritma *K-Means* untuk mengelompokkan artikel berita menjadi beberapa kluster berdasarkan kemiripan tema agar memudahkan pemahaman mengenai berbagai topik yang dibahas dalam artikel berita itu sendiri. Tahapan analisis *clustering* dengan *K-Means* didahului dengan *pre-processing* data mengikuti prosedur berikut.

1. Proses *Tokenizing*.

Proses *Tokenizing* adalah proses pemotongan sebuah kalimat menjadi unit-unit kata yang lebih kecil sehingga setiap kata memiliki makna yang berbeda. Anggap kalimat “Pilkada Jakarta 2024” akan ditokenisasi menjadi kata “Pilkada”, “Jakarta”, dan “2024”.

2. Proses *Filtering*.

Proses *Filtering* merupakan proses penghapusan kata yang tidak memiliki nilai informasi seperti karakter-karakter non-alfabet lain meliputi angka, tanda baca, atau simbol. Daftar kata umum atau *stopwords* seperti kata “atau”, “dan”, atau “yang” akan turut dihapus pada proses ini.

3. Proses *Stemming*

Proses *Stemming* merupakan proses perubahan kata menjadi bentuk dasarnya mengikuti kaidah KBBI. Sebagai contoh, kata “Pemilihan” diubah menjadi kata dasar “Pilih”. Oleh karena itu, proses dapat dikatakan juga sebagai lanjutan dari proses *Filtering*.

4. Proses Ekstraksi Fitur

Metode TF-IDF atau *Term Frequency-Inverse Document Frequency* akan digunakan dalam proses ekstraksi fitur untuk menampilkan teks dalam bentuk numerik berdasarkan bobot kepentingan kata dalam korpus. Pada akhirnya, nilai tersebut digunakan dalam menentukan relevansi suatu kata terdapat prediksi sentimen.

Setelah *pre-processing* dilakukan, metode *Elbow* akan diterapkan guna menentukan jumlah optimal kluster dengan menghitung varians intra-kluster atau inersia pada berbagai nilai k. Inersia mengukur seberapa dekat titik data dalam suatu kluster. Nilai inersia menunjukkan kepadatan kluster. Apabila nilai inersia semakin kecil, maka anggota-anggota kluster memiliki kesamaan yang lebih tinggi dengan jarak antar titik-titik data lebih kecil. Nilai k yang optimal ditemukan pada titik *elbow*, yaitu titik di mana penurunan inersia mulai melambat sekaligus menunjukkan jumlah kluster terbaik untuk menggambarkan data secara efektif. *Silhouette Score* juga dilakukan sebagai evaluasi tambahan untuk mengukur seberapa baik artikel-artikel dikelompokkan dalam kluster dengan rentang nilai dari -1 sampai 1. Semakin tinggi *Silhouette Score* menunjukkan pengelompokan yang lebih baik.

Mengikuti perolehan nilai k optimal, algoritma *K-Means* dijalankan untuk pengelompokan data. Hasil pengelompokan dianalisis lebih lanjut dengan mengidentifikasi kata kunci atau *topwords* dalam setiap kluster berdasarkan nilai TF-IDF tertinggi. Penggunaan alat visualisasi *Word Cloud* turut digunakan untuk menunjang visualisasi kluster yang lebih intuitif. Kata-kata yang memiliki frekuensi muncul lebih banyak akan ditampilkan dalam ukuran yang lebih besar.

2.3 Analisis Sentimen dengan Lexicon

Analisis sentimen diterapkan untuk memahami distribusi nada pemberitaan dari setiap kluster dengan pendekatan Lexicon dan didukung oleh *TextBlob* untuk memudahkan analisis teks dalam berbagai aplikasi, seperti klasifikasi sentimen

atau ekstraksi informasi. Nilai sentimen atau *sentiment score* dihitung berdasarkan nilai polaritas atau *polarity score* dengan cara menjumlahkan keseluruhan bobot kata yang terdeteksi sistem sehingga dapat diklasifikasikan ke dalam jenis sentimen berikut ini.

Jika nilai polaritas > 0	termasuk dalam sentimen positif.
Jika nilai polaritas = 0	termasuk dalam sentimen netral.
Jika nilai polaritas < 0	termasuk dalam sentimen negatif.

Klasifikasi sentimen ini memberikan pemahaman persebaran sentimen untuk masing-masing tema sekaligus perkiraan nada pemberitaan dalam berbagai topik yang muncul dari hasil *clustering*. Melalui penelitian *Sentiment Analysis Using Lexicon-Based Methods for News Data*, Medhat et. al. (2022) beranggapan pendekatan berbasis Lexicon sangat efektif untuk menganalisis sentimen teks dalam skala besar karena sederhana, kredibel, dan tidak memerlukan pelatihan data. Tidak berbeda dengan Thelwall (2023) dalam penelitian *Lexicon-Based Sentiment Analysis for Big Data Applications* menekankan keunggulan metode Lexicon dalam analisis sentimen teks yang bersifat informal, salah satunya adalah artikel berita.

3. HASIL DAN PEMBAHASAN

3.1 Web Scraping

Proses *web scraping* pada penelitian ini mengumpulkan 100 artikel berita terkait Pilkada Jakarta 2024. Tahap pengumpulan tautan menggunakan RSS *feed* Google News yang menghasilkan daftar tautan berita yang relevan. Kumpulan tautan tersebut disimpan dalam file "links.txt". Selanjutnya, pada tahap pengunduhan konten, digunakan *library* Python seperti "BeautifulSoup4" untuk mengekstrak isi artikel dari masing-masing tautan. Data tersebut kemudian digunakan untuk menganalisis lebih lanjut, termasuk *clustering* dan analisis sentimen.

3.2 Preprocessing Data

Tahapan *preprocessing* teks melibatkan tiga langkah utama: *tokenizing*, *filtering*, dan *stemming*. *Tokenizing* adalah proses memecah kalimat menjadi unit kata-kata kecil agar setiap kata dapat memiliki makna yang berbeda, misalnya kalimat "PAN membuka peluang untuk mendukung Anies Baswedan di Pilgub Jakarta 2024" menjadi ['PAN', 'membuka', 'peluang', 'untuk', 'mendukung', 'Anies', 'Baswedan', 'di', 'Pilgub', 'Jakarta', '2024']. *Filtering* dilakukan untuk menghapus elemen yang tidak relevan seperti angka, tanda baca, dan kata-kata umum (*stopwords*) sehingga hasilnya menjadi ['PAN', 'membuka', 'peluang', 'mendukung', 'Anies', 'Baswedan', 'Pilgub', 'Jakarta']. *Stemming* adalah tahap akhir di mana kata-kata diubah ke bentuk dasar, contohnya "membuka" menjadi "buka" dan "mendukung" menjadi "dukung", menghasilkan ['pan', 'buka', 'peluang', 'dukung', 'anies', 'baswedan', 'pilgub', 'jakarta']. Proses ini menyederhanakan teks untuk memudahkan analisis lebih lanjut.

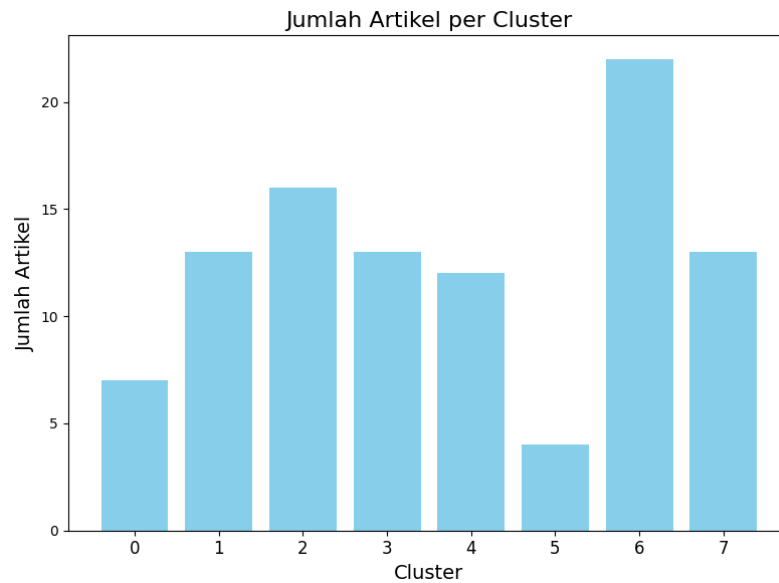
3.3 Hasil Analisis Clustering

Berikan pernyataan bahwa apa yang diharapkan, seperti yang dinyatakan pada bab "Pendahuluan" pada akhirnya dapat menghasilkan bab "Hasil dan Analisis", sehingga ada kesesuaian. Selain itu juga dapat ditambahkan prospek pengembangan hasil penelitian dan penerapan prospek studi lanjut ke selanjutnya (berdasarkan Penentuan jumlah kluster dilakukan dengan menggunakan *Silhouette Score* yang dapat mengukur seberapa baik setiap titik data dikelompokkan, dengan nilai lebih tinggi menunjukkan pemisahan kluster yang lebih baik. Berdasarkan tabel 4 jumlah kluster dari 1 hingga 10 yang menunjukkan optimal adalah 8 kluster dengan nilai *Silhouette Score* tertinggi yaitu 0.01612.

Tabel 1. Nilai *Silhouette Score* Jumlah Kluster

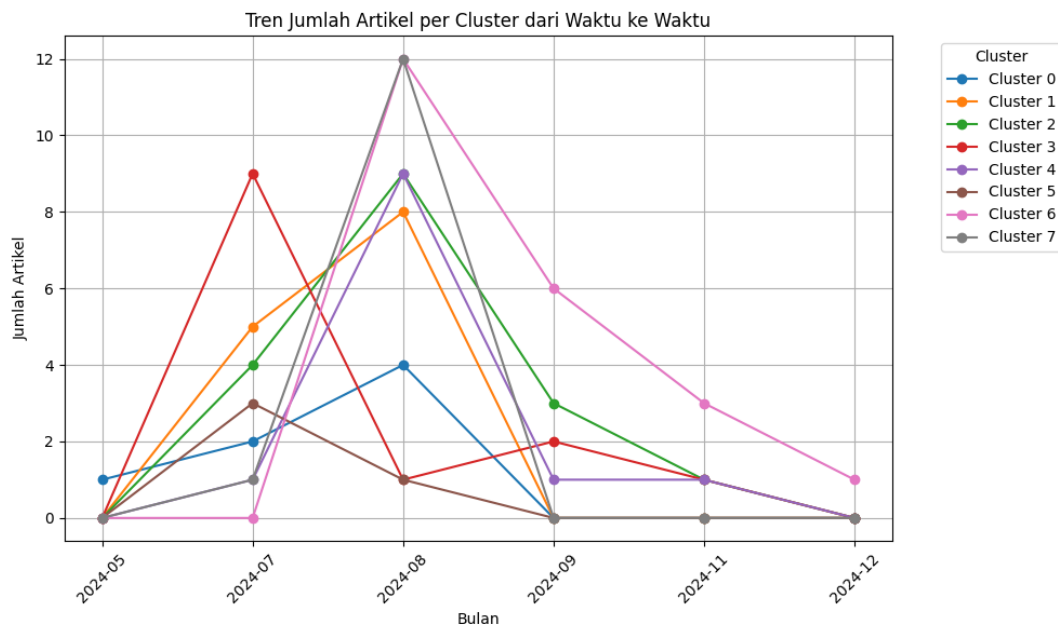
Jumlah Cluster	Silhouette Score
2 Cluster	0.00602
3 Cluster	0.00134
4 Cluster	0.01026
5 Cluster	0.01234
6 Cluster	0.01206
7 Cluster	0.01305
8 Cluster	0.01612
9 Cluster	0.01611
10 Cluster	0.01348

Dari jumlah 8 kluster tersebut, kluster dengan jumlah artikel terbanyak adalah kluster 6 dengan 22 artikel diikuti dengan kluster ke 2 yaitu 16 artikel. Sementara itu jumlah artikel sedikit ada pada kluster 0 sebanyak 7 artikel dan kluster ke 5 sebanyak 4 artikel. Kemudian untuk kluster 1, 3 dan 4 memiliki jumlah artikel yang sama yaitu sebanyak 13 artikel.



Gambar 1. Histogram Jumlah Artikel per Kluster

Pada gambar 2 menunjukkan tren jumlah artikel yang dikelompokkan berdasarkan kluster (Kluster 0 hingga 7) dari waktu ke waktu, dengan sumbu horizontal mewakili bulan dari Mei hingga Desember 2024, dan sumbu vertikal menunjukkan jumlah artikel.



Gambar 2. Grafik Tren Jumlah Artikel per Kluster dari Waktu ke Waktu

Terlihat bahwa jumlah artikel untuk hampir semua kluster mencapai puncaknya pada bulan Agustus 2024, sebelum mengalami penurunan tajam pada bulan-bulan berikutnya. Karena pada bulan Agustus menjadi bulan dimana isu-isu dalam Pilkada Jakarta 2024 mulai ramai dibicarakan mulai terkait pencalonan, permasalahan antar pencalonan dan partai politik. Hasil analisis menunjukkan bahwa setiap kluster memiliki topik pembahasan spesifik terkait Pilkada Jakarta 2024, yang ditentukan dengan mengidentifikasi kata kunci sebagai kata yang paling sering muncul dalam artikel di setiap kluster.

Kluster 0, misalnya, berfokus pada dukungan dan kandidat Pilkada dengan kata kunci seperti "dukungan kandidat" dan "pilkada," yang mencerminkan pembahasan tentang kandidat dari partai politik seperti PDIP dan PKB, serta kandidat independen.



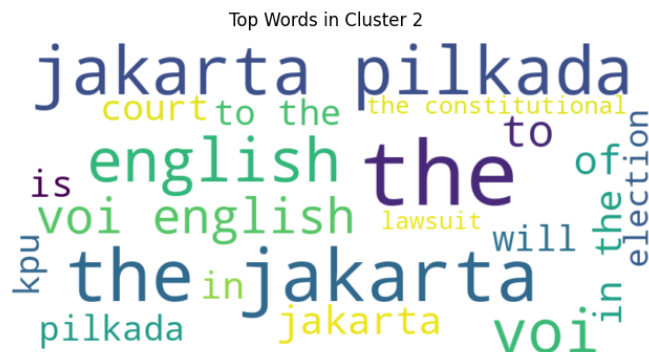
Gambar 3. Word Clouds dalam Kluster 0

Kluster 1 menonjol dengan kata-kata seperti "nasdem" dan "anies," yang menunjukkan fokus pada pencalonan Anies Baswedan yang secara resmi didukung Partai Nasdem.



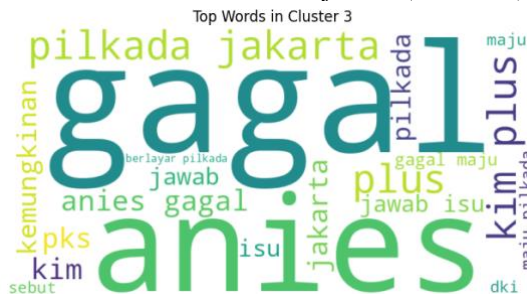
Gambar 4. Word Clouds dalam Kluster 1

Kluster 2, dengan kata-kata seperti "English" dan "voi," mengelompokkan artikel berbahasa Inggris dari media VOI yang terkait dengan Pilkada Jakarta.



Gambar 5. Word Clouds dalam Kluster 2

Sementara itu, kluster 3 membahas kegagalan Anies Baswedan dalam pencalonan yang tercermin dari kata "gagal" dan "kim" yang merujuk pada dinamika politik terkait Koalisi Indonesia Maju Plus (KIM Plus).



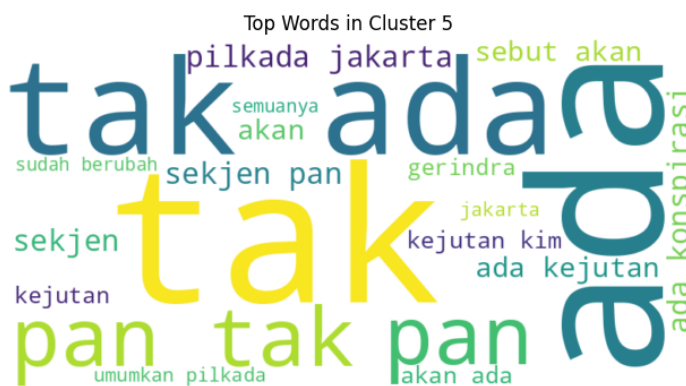
Gambar 6. Word Clouds dalam Kluster 3

Kluster 4 dengan kata kunci "anies" dan "jokowi" mengangkat isu dukungan politik Anies, ambang batas pencalonan, dan peran Jokowi dalam Pilgub DKI Jakarta dan Jawa Tengah.



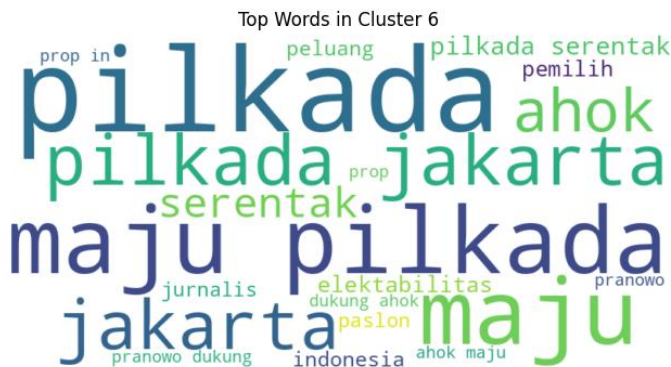
Gambar 7. Word Clouds dalam Kluster 4

Kluster 5 menyoroti isu ketidakpastian dan teori konspirasi dalam pilgub yang diwakili oleh kata "tak ada" dan "konspirasi" yang menegaskan klarifikasi PAN bahwa tidak ada upaya untuk menggagalkan Anies maupun Koalisi Indonesia Maju (KIM).



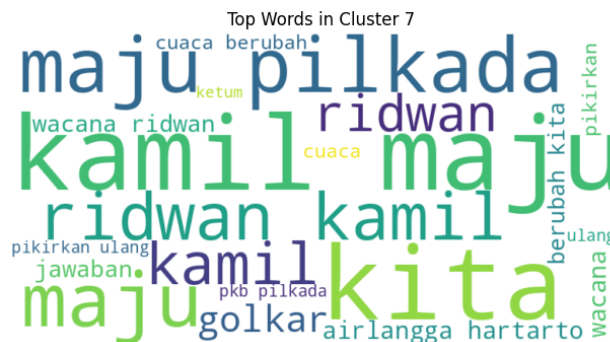
Gambar 8. Word Clouds dalam Kluster 5

Selanjutnya, Kluster 6 membahas Pilkada Serentak dan dinamika politik di Jakarta, dengan kata-kata seperti "pilkada" dan "serentak," yang mencakup isu persiapan kandidat, peluang bagi kandidat seperti Anies, serta dampak Pilkada terhadap ekonomi Jakarta.



Gambar 9. Word Clouds dalam Kluster 6

Terakhir, kluster 7 menyoroti Ridwan Kamil sebagai kandidat potensial dengan dukungan kuat dari Golkar dan Koalisi Indonesia Maju (KIM), tercermin dari kata-kata seperti "ridwan kamil" dan "golkar," serta rencana pengumuman pasangannya.



Gambar 10. Word Clouds dalam Kluster 7

Tren pembahasan artikel mencapai puncaknya pada Agustus 2024, dengan topik yang hangat diperbincangkan meliputi isu pencalonan, konflik politik, dan dukungan partai. Hal ini mencerminkan tingginya tingkat perhatian publik terhadap pemilihan gubernur Jakarta, sebagaimana terlihat dari pola kata yang dominan di setiap kluster.

3.4 Hasil Analisis Sentimen tiap Kluster

Hasil *clustering* yang terdiri dari delapan *cluster* berdasarkan topik utama kemudian dinilai apakah kumpulan berita tersebut bernilai positif, negatif, atau netral menggunakan pendekatan Lexicon. Mayoritas artikel dalam tiap *cluster* memiliki sentimen netral yang menunjukkan bahwa media massa umumnya menjaga objektivitas. Akan tetapi, pada *cluster* 2 (Media berita VOI dengan bahasa Inggris) terdapat indikasi sentimen positif sebanyak 5 artikel, negatif sebanyak 1 artikel, dan netral sebanyak 10 artikel. Hal ini bisa disebabkan oleh penggunaan kata atau bahasa tertentu dalam artikel.

Tabel 2. Hasil Analisis Sentimen tiap Cluster

Kluster	Topik Kluster	Netral	Positif	Negatif
0	Dukungan dan Kandidat Pilkada	7	0	0
1	Partai Nasdem mengusung Anies untuk Pilkada Jakarta	13	0	0
2	Media berita VOI dengan bahasa Inggris mengenai Pilkada Jakarta	10	5	1
3	Kegagalan Anies dalam pencalonan Pilkada Jakarta	13	0	0
4	Anies Baswedan dan Dukungan Politik	12	0	0
5	Tak Ada Dukungan dan Konspirasi dalam Pilkada	4	0	0
6	Pemilihan Serentak dan Dukungan Pilkada Jakarta 2024	22	0	0
7	Ridwan Kamil dan Koalisi Politik dalam Pilkada Jakarta 2024	13	0	0

Hasil tersebut menunjukkan bahwa sebagian besar artikel bersifat netral sesuai dengan peran media sebagai penyampai informasi yang objektif. Akan tetapi, indikasi bias sentimen pada salah satu *cluster* dapat digarisbawahi untuk analisis lebih lanjut agar mengetahui faktor-faktor yang memengaruhi bias tersebut.

4. KESIMPULAN

Hasil penelitian menunjukkan bahwa melalui algoritma *K-Means*, pola pemberitaan media dari 100 artikel mengenai Pilkada Jakarta 2024 dapat dikelompokkan menjadi delapan tema utama dengan fokus pada isu dukungan politik, calon kandidat, hingga dinamika politik yang terjadi selama Pilkada Jakarta 2024 berlangsung. Sebagian besar artikel-artikel di setiap kluster menunjukkan distribusi sentimen yang bersifat netral, meskipun ditemukan adanya sentimen positif dan negatif yang mengindikasikan adanya bias dalam artikel-artikel pada kluster 2 mengenai topik Media Berita VOI dengan bahasa Inggris mengenai Pilkada Jakarta.

Pendekatan *text mining* yang melibatkan analisis *clustering* dengan algoritma *K-Means* dan analisis sentimen berbasis Lexicon terbukti efektif dalam mengidentifikasi pola pemberitaan dan sentimen media massa terkait Pilkada Jakarta 2024 sekaligus memberikan pemahaman yang lebih mendalam tentang perkembangan dinamika politik. Hasil ini diharapkan dapat menjadi acuan bagi pemerintah, media, dan masyarakat dalam meningkatkan pemahaman tentang persepsi publik, menjaga netralitas pemberitaan, dan mendukung proses demokrasi yang berkualitas.

REFERENSI

- [1] F. V. Sari dan A. Wibowo, "ANALISIS SENTIMEN PELANGGAN TOKO ONLINE JD.ID MENGGUNAKAN METODE NAÏVE BAYES CLASSIFIER BERBASIS KONVERSI IKON EMOSI," *Simetris: Jurnal Teknik Mesin, Elektro dan Ilmu Komputer*, vol. 10, no. 2, hlm. 681–686, Nov 2019, Diakses: 26 Desember 2024. [Daring]. Tersedia pada: <https://jurnal.umk.ac.id/index.php/simet/article/view/3487>
- [2] A. Goyal, V. Gupta, dan M. Kumar, "Recent Named Entity Recognition and Classification techniques: A systematic review," *Comput Sci Rev*, vol. 29, hlm. 21–43, Agu 2018, doi: 10.1016/J.COSREV.2018.06.001.

- [3] T. ALASALI dan Y. ORTAKCI, "Clustering Techniques in Data Mining: A Survey of Methods, Challenges, and Applications," *Computer Science*, Mar 2024, doi: 10.53070/BBD.1421527.
- [4] D. M. Thomas dan S. Mathur, "Data Analysis by Web Scraping using Python," *Proceedings of the 3rd International Conference on Electronics and Communication and Aerospace Technology, ICECA 2019*, hlm. 450–454, Jun 2019, doi: 10.1109/ICECA.2019.8822022.
- [5] N. Syamsu Wais Al Qorni, "PENERAPAN ALGORITMA K-MEANS CLUSTERING PADA ULASAN PENGGUNA MERDEKA MENGAJAR DI PLAY STORE PROGRAM STUDI TEKNIK INFORMATIKA SEKOLAH TINGGI MANAJEMEN INFORMATIKA dan KOMPUTER WIDYA CIPTA DHARMA 2024".
- [6] M. Taboada, J. Brooke, M. Tofiloski, K. Voll, dan M. Stede, "Lexicon-based methods for sentiment analysis," *Computational Linguistics*, vol. 37, no. 2, hlm. 267–307, 2011, doi: 10.1162/COLI_A_00049.
- [7] "View of IMPLEMENTASI ALGORITMA K-MEANS TERHADAP OPINI MASYARAKAT MENGENAI PERKIRAAN PEMILU 2024 PADA TWITTER." Diakses: 26 Desember 2024. [Daring]. Tersedia pada: <https://ejournal.caturisakti.ac.id/index.php/simtek/article/view/271/233>
- [8] A. Rahmawati, A. Marjuni, J. Zeniarja, J. Informatika, U. Dian, dan N. Semarang, "ANALISIS SENTIMEN PUBLIK PADA MEDIA SOSIAL TWITTER TERHADAP PELAKSANAAN PILKADA SERENTAK MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE PUBLIC SENTIMENT ANALYSIS ON TWITTER SOCIAL MEDIA TO PILKADA SERENTAK EVENT USING SUPPORT VECTOR MACHINE ALGORITHM," vol. 10, no. 2, 2017.
- [9] "Kampanye Tiga Pasangan Calon Pilkada Jakarta 2024: Masalah Krusial Warga Tak Terjamah, Minim Solusi Nyata | ICW." Diakses: 26 Desember 2024. [Daring]. Tersedia pada: <https://antikorupsi.org/id/kampanye-tiga-pasangan-calon-pilkada-jakarta-2024-masalah-krusial-warga-tak-terjamah-minim-solusi>
- [10] O. R. Romli dan E. Nurlia, "Menimbang Isu Politik Identitas Pada Pilkada Dki Jakarta Tahun 2024," *Jurnal Ilmu Politik dan Pemerintahan*, vol. 9, no. 2, Des 2023, doi: 10.37058/jipp.v9i2.8944.
- [11] Oleh, "PROGRAM STUDI SARJANA ILMU HUKUM FAKULTAS HUKUM UNIVERSITAS HASANUDDIN MAKASSAR 2024 SKRIPSI TINJAUAN YURIDIS TERHADAP PENGATURAN HUKUM PILKADA SERENTAK TAHUN 2024 DAN IMPLIKASINYA JUDICIAL REVIEW OF LEGAL REGULATIONS OF SIMULTANEOUS REGIONAL ELECTIONS IN 2024 AND THE IMPLICATIONS".
- [12] C. Rahim, B. Niode, A. B. Pati, D. M. Liando, dan M. Mamentu, "NETRALITAS MEDIA MASSA BERBASIS ONLINE PADA PEMILIHAN WALI KOTA DAN WAKIL WALI KOTA MANADO TAHUN 2020," *Jurnal Review Pendidikan dan Pengajaran (JRPP)*, vol. 7, no. 1, hlm. 776–792, Jan 2024, doi: 10.31004/JRPP.V7I1.24692.
- [13] "MEDIA MONITORING ISU PILKADA PROVINSI DKI JAKARTA TAHUN 2024".
- [14] J. Homepage *dkk.*, "Analisis Sentimen terhadap Perkuliahan Daring di Indonesia dari Twitter Dataset Menggunakan InSet Lexicon," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 1, no. 1, hlm. 24–33, Mar 2021, doi: 10.57152/MALCOM.V1I1.20.
- [15] A. P. Giovanni, A. Ardiansyah, T. Haryanti, L. Kurniawati, dan W. Gata, "ANALISIS SENTIMEN APLIKASI RUANG GURU DI TWITTER MENGGUNAKAN ALGORITMA KLASIFIKASI," *Jurnal Teknoinfo*, vol. 14, no. 2, hlm. 115, Jul 2020, doi: 10.33365/JTI.V14I2.679.
- [16] "(PDF) McQuail's Mass Communication Theory | Khalsa Alrahbi - Academia.edu." Diakses: 26 Desember 2024. [Daring]. Tersedia pada: https://www.academia.edu/36441695/McQuails_Mass_Communication_Theory
- [17] R. P. Bastoni, "Analisis Sentimen Opini Publik Terhadap Calon Presiden Indonesia 2024 Menggunakan Metode Naive Bayes Dan Lexicon Based Pada Platform YouTube Skripsi Oleh".
- [18] E. Susanto, V. C. Mawardi, dan M. D. Lauro, "APLIKASI CLUSTERING BERITA DENGAN METODE K MEANS DAN PERINGKAS BERITA DENGAN METODE MAXIMUM MARGINAL RELEVANCE".
- [19] "View of PENCARIAN DATA QUICK COUNT PILPRES DENGAN TEKNIK WEB SCRAPING." Diakses: 26 Desember 2024. [Daring]. Tersedia pada: <https://bajangjournal.com/index.php/JIRK/article/view/6695/5147>
- [20] "ANALISIS SENTIMEN UNTUK MEMPREDIKSI HASIL CALON PEMILU PRESIDEN MENGGUNAKAN LEXICON BASED DAN RANDOM FOREST." Diakses: 26 Desember 2024. [Daring]. Tersedia pada: <https://ejournal.upbatam.ac.id/index.php/jif/article/view/7987/3319>