

# Analisis Sentimen dan Clustering Komentar Twitter Menggunakan Metode Lexicon-Based dan Algoritma K-Means

Anugrah Resky Samudra\*<sup>1</sup> | Chyquitha Danuputri<sup>2</sup> | Titin Wahyuni<sup>2</sup>

<sup>1</sup>Mahasiswa Program Studi Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Makassar, Indonesia.

[105841105421@student.unismuh.ac.id](mailto:105841105421@student.unismuh.ac.id)

<sup>2</sup>Program Studi Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Makassar, Indonesia.

[chyquithadanuputri@unismuh.ac.id](mailto:chyquithadanuputri@unismuh.ac.id)

[titinwahyuni@unismuh.ac.id](mailto:titinwahyuni@unismuh.ac.id)

## Korespondensi

\*Anugrah Resky Samudra R

[105841105421@student.unismuh.ac.id](mailto:105841105421@student.unismuh.ac.id)

## ABSTRAK:

Media sosial Twitter telah menjadi salah satu platform utama bagi masyarakat dalam menyampaikan opini dan persepsi terhadap berbagai isu publik, termasuk isu ekonomi nasional. Besarnya volume data teks yang dihasilkan menuntut adanya metode analisis yang mampu mengekstraksi informasi secara efektif. Penelitian ini bertujuan untuk memetakan opini publik di Twitter berdasarkan polaritas sentimen serta kemiripan topik pembahasan. Metode yang digunakan meliputi analisis sentimen berbasis leksikon untuk menentukan kecenderungan sentimen positif, negatif, dan netral, serta teknik clustering teks untuk mengelompokkan komentar berdasarkan kesamaan karakteristik kontennya. Data penelitian diperoleh dari hasil pengumpulan komentar Twitter yang relevan dengan isu ekonomi nasional, kemudian melalui tahapan preprocessing teks, pembobotan kata menggunakan TF-IDF, dan proses pengelompokan data. Hasil penelitian menunjukkan bahwa integrasi analisis sentimen dan clustering mampu memberikan gambaran yang lebih komprehensif mengenai pola opini publik, baik dari sisi kecenderungan sentimen maupun topik diskusi yang dominan. Temuan ini diharapkan dapat menjadi referensi dalam pemanfaatan text mining untuk analisis opini publik berbasis media sosial.

## KATA KUNCI

Analisis Sentimen, Text Mining, Media Sosial, Clustering Teks, Twitter.

## ABSTRACT:

Twitter has become one of the main social media platforms for the public to express opinions and perceptions on various public issues, including national economic issues. The large volume of textual data generated requires analytical methods capable of extracting information effectively. This study aims to map public opinion on Twitter based on sentiment polarity and topic similarity. The methods employed include lexicon-based sentiment analysis to identify positive, negative, and neutral sentiments, as well as text clustering techniques to group comments according to content similarity. The research data were obtained from the collection of Twitter comments related to national economic issues, followed by text preprocessing, term weighting using TF-IDF, and clustering processes. The results indicate that the integration of sentiment analysis and text clustering provides a more comprehensive overview of public opinion patterns, both in terms of sentiment tendencies and dominant discussion topics. These findings are expected to serve as a reference for the application of text mining in social media-based public opinion analysis.

## Keywords:

Sentiment Analysis, Text Mining, Social Media, Text Clustering, Twitter.

## 1 | PENDAHULUAN

Perkembangan teknologi informasi dan komunikasi telah mendorong meningkatnya penggunaan media sosial sebagai sarana utama masyarakat dalam menyampaikan opini dan berinteraksi secara digital. Media sosial memungkinkan pengguna untuk mengekspresikan pandangan serta respons terhadap berbagai isu publik secara cepat dan luas, termasuk melalui teks pendek seperti di Twitter. Fenomena ini telah memicu banyak penelitian yang menggunakan data Twitter untuk memahami opini publik secara komputasional karena sifatnya yang real-time dan volume data teks yang sangat besar (A. Amelia and R. Yusuf, 2025).

Isu ekonomi nasional merupakan salah satu topik yang banyak menarik perhatian masyarakat karena berkaitan langsung dengan kesejahteraan dan dinamika sosial yang kompleks. Opini publik terhadap isu ekonomi yang berkembang di media sosial dapat mencerminkan persepsi dan sikap masyarakat secara umum. Namun, besarnya volume data komentar yang dihasilkan dari Twitter menjadikan analisis manual tidak efektif, sehingga diperlukan metode berbasis komputasi untuk mengolah dan menganalisis data teks tersebut secara otomatis. Pendekatan text mining menjadi pilihan karena mampu mengekstraksi pola dan insight dari kumpulan data teks yang tidak terstruktur (Fuad Amirullah, Syariful Alam, M.Imam Sulistyono, 2023).

Analisis sentimen merupakan salah satu teknik dalam text mining yang digunakan untuk mengidentifikasi polaritas opini atau sikap pengguna terhadap suatu topik tertentu. Teknik ini sering kali diimplementasikan pada data media sosial untuk memahami kecenderungan emosional publik terhadap isu tertentu. Studi-studi terbaru menunjukkan bahwa analisis sentimen pada data media sosial, seperti yang dilakukan pada Twitter, terus berkembang dengan penerapan model statistik dan pembelajaran mesin untuk meningkatkan akurasi dan interpretasi data. (A. Amelia and R. Yusuf, 2025).

Analisis sentimen, atau sering disebut juga opinion mining, adalah bidang studi dalam Natural Language Processing (NLP) yang berfokus pada identifikasi, ekstraksi, dan kuantifikasi keadaan afektif atau informasi subjektif dalam data teks. Tujuan utamanya adalah menentukan polaritas opini penulis apakah positif, negatif, atau netral terhadap suatu entitas, seperti produk, layanan, atau isu tertentu ( Firmansyah & Herawan, 2022).

Meskipun analisis sentimen dapat mengungkap kecenderungan emosional publik, pendekatan ini memiliki keterbatasan dalam menangkap struktur topik utama yang terdapat dalam data teks. Misalnya, komentar yang memiliki sentimen sama belum tentu memiliki tema atau konteks diskusi yang sama. Hal ini menuntut penggunaan metode tambahan seperti clustering teks, yang digunakan untuk mengelompokkan data berdasarkan kemiripan konten agar dapat mengidentifikasi pola topik diskusi yang dominan dalam kumpulan teks. (Dini K.Astutik, Artanti, Fenny F, 2022).

Clustering teks telah menjadi teknik penting dalam text mining untuk memetakan kelompok dokumen atau komentar berdasarkan kesamaan karakteristik kata, sehingga pola diskusi dapat dianalisis secara lebih mendalam. Teknik ini membantu mengidentifikasi kelompok tema dalam opini publik, sekaligus memberikan konteks topik yang tidak selalu terdeteksi melalui analisis sentimen saja. (Anggina, S., Setiawan, N. Y., & Bachtiar, F. A, 2022).

Penelitian ini mengintegrasikan analisis sentimen berbasis lexicon dengan clustering teks untuk memetakan opini publik terhadap isu ekonomi nasional di Twitter. Pendekatan berbasis leksikon dipilih karena tidak memerlukan data latih dan relatif mudah diimplementasikan, terutama pada dataset berukuran besar, sedangkan clustering teks diaplikasikan untuk memahami struktur topik utama dalam kumpulan komentar. Dengan menggabungkan kedua pendekatan ini, penelitian diharapkan memberikan gambaran yang lebih komprehensif tentang opini publik, tidak hanya dari sisi sentimen, tetapi juga struktur topik diskusi yang berkembang di media sosial. (Amanda, A., & Nurmawati, E, 2023).

Dengan menggabungkan analisis sentimen dan clustering teks, penelitian ini bertujuan untuk memberikan gambaran yang lebih komprehensif mengenai opini publik, tidak hanya dari sisi polaritas sentimen, tetapi juga dari sisi struktur topik pembahasannya. Hasil penelitian diharapkan dapat memberikan kontribusi akademik dalam pengembangan metode analisis opini publik berbasis text mining serta menjadi referensi bagi penelitian selanjutnya di bidang analisis media sosial.

Penelitian ini mengadopsi Algoritma K-Means. K-Means merupakan salah satu metode data mining unsupervised yang populer untuk mengelompokkan data berdasarkan kemiripan karakteristiknya. Keandalan K-Means telah terbukti efektif dalam melakukan text clustering pada data media sosial di Indonesia untuk mengidentifikasi topik-topik utama dari beragam komentar publik, seperti pada respons masyarakat terhadap isu sosial (Alfian, 2023). Keberhasilan ini menegaskan bahwa K-Means adalah alat yang relevan untuk penelitian studi kasus ini

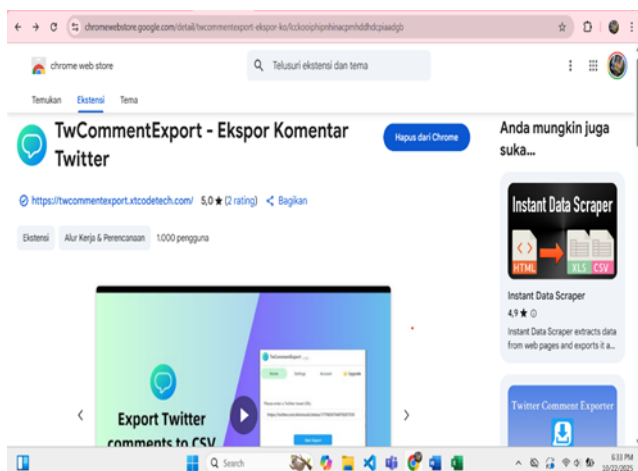
## 2 | METODE

Penelitian ini menggunakan pendekatan kuantitatif dengan metode text mining untuk menganalisis opini publik terhadap isu ekonomi nasional yang disampaikan melalui media sosial Twitter. Tahapan penelitian dirancang secara sistematis mulai dari pengumpulan data, preprocessing teks, analisis sentimen, hingga proses clustering teks untuk memperoleh pola opini publik yang lebih komprehensif.

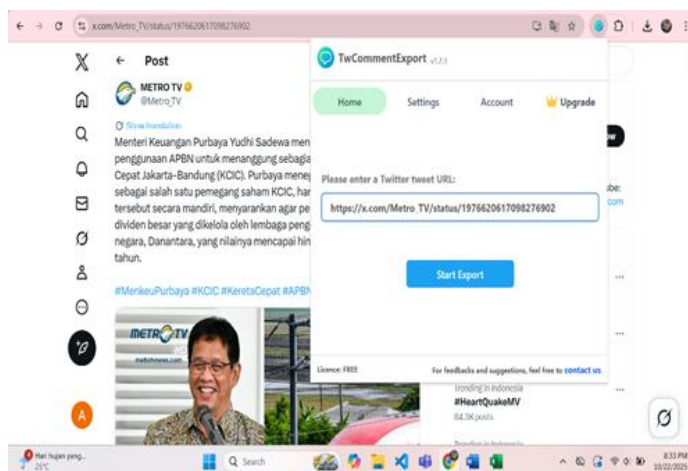
### 2.1 | Sumber dan Pengumpulan Data

Data yang digunakan dalam penelitian ini berupa komentar atau cuitan (tweet) dari media sosial Twitter yang berkaitan dengan isu ekonomi nasional. Pengumpulan data dilakukan dengan memanfaatkan Application Programming Interface (API) Twitter menggunakan kata kunci tertentu yang relevan dengan topik penelitian. Data yang diperoleh berupa teks komentar pengguna yang kemudian disimpan dalam bentuk dataset untuk diproses lebih lanjut.

Jumlah data yang digunakan dalam penelitian ini disesuaikan dengan kebutuhan analisis dan ketersediaan data pada periode pengambilan tertentu. Dataset yang telah terkumpul selanjutnya melalui proses seleksi untuk memastikan bahwa data yang digunakan relevan dengan topik penelitian dan tidak mengandung duplikasi.



GAMBAR 1. Proses Install Ekstensi TwCommentExport - Ekspor Komentar Twitter



GAMBAR 2. Proses Scrapping Data

### 2.2 | Teknik Pengumpulan Data

Tahap preprocessing bertujuan untuk membersihkan dan menyiapkan data teks agar dapat dianalisis secara optimal. Proses preprocessing dilakukan melalui beberapa tahapan, yaitu:

- **Case Folding**, tahap ini dilakukan dengan mengubah seluruh huruf dalam teks menjadi huruf kecil (lowercase) untuk menghindari perbedaan makna akibat perbedaan penulisan huruf kapital.
- **Cleaning**, proses pembersihan teks dilakukan dengan menghapus karakter yang tidak relevan seperti URL, mention (@username), hashtag (#), angka, tanda baca, dan simbol-simbol lainnya yang tidak berkontribusi terhadap analisis sentimen.
- **Tokenize**, pada tahap ini, teks dipecah menjadi unit-unit kata (token) berdasarkan spasi atau pemisah tertentu untuk memudahkan proses analisis selanjutnya.
- **Stopword Removal**, stopwords removal dilakukan untuk menghapus kata-kata umum yang tidak memiliki makna signifikan dalam analisis sentimen, seperti kata hubung dan kata depan.
- **Stemming**, tahap stemming bertujuan untuk mengubah kata berimbuhan menjadi kata dasar sehingga dapat mengurangi variasi kata dan meningkatkan akurasi analisis.

Hasil dari tahap preprocessing ini adalah data teks yang lebih bersih dan terstruktur, sehingga siap digunakan dalam proses analisis sentimen dan clustering.

## 2.3 | Analisis Sentimen Berbasis Lexicon

Analisis sentimen dalam penelitian ini dilakukan menggunakan pendekatan berbasis leksikon (lexicon-based). Metode ini mengklasifikasikan sentimen suatu teks berdasarkan kumpulan kata yang telah memiliki bobot sentimen tertentu, seperti positif, negatif, atau netral.

Setiap kata dalam teks hasil preprocessing akan dicocokkan dengan kamus leksikon sentimen. Skor sentimen suatu komentar dihitung berdasarkan akumulasi bobot sentimen dari kata-kata yang terkandung di dalamnya. Jika skor sentimen bernilai positif, maka komentar diklasifikasikan sebagai sentimen positif. Sebaliknya, jika skor bernilai negatif maka diklasifikasikan sebagai sentimen negatif, sedangkan skor mendekati nol dikategorikan sebagai sentimen netral.

Pendekatan berbasis leksikon dipilih karena tidak memerlukan data latih (training data) dan relatif mudah diimplementasikan, serta cocok digunakan untuk analisis opini publik dengan skala data yang besar.

## 2.4 | Analisis Sentimen Berbasis Lexicon

Salah satu skema pembobotan yang paling efektif dan populer untuk menciptakan vektor ini adalah Term Frequency-Inverse Document Frequency (TF-IDF). TF-IDF mengukur tingkat pentingnya sebuah kata (term) dalam sebuah dokumen yang merupakan bagian dari sebuah koleksi dokumen atau korpus. Bobot ini merupakan hasil perkalian dari dua metrik (Junior et al., 2022).

Secara matematis, TF-IDF dihitung menggunakan persamaan berikut:

- Term Frequency (TF) menunjukkan seberapa sering suatu kata muncul dalam satu dokumen.
- Inverse Document Frequency (IDF) menunjukkan tingkat keunikan suatu kata dalam seluruh dokumen.

Bobot TF-IDF yang dihasilkan digunakan sebagai vektor fitur yang merepresentasikan setiap komentar Twitter, sehingga memungkinkan proses pengelompokan data dilakukan secara komputasional.

## 2.5 | Clustering Teks

Proses clustering dilakukan untuk mengelompokkan komentar Twitter berdasarkan kemiripan konten teks. Dengan menggunakan representasi TF-IDF, setiap komentar dipetakan ke dalam ruang vektor berdimensi tinggi. Teknik clustering digunakan untuk mengelompokkan komentar ke dalam beberapa cluster yang memiliki karakteristik serupa.

Clustering bertujuan untuk mengidentifikasi topik diskusi yang dominan dalam opini publik. Dengan demikian, meskipun komentar memiliki sentimen yang sama, clustering memungkinkan analisis lebih lanjut terhadap perbedaan tema atau fokus pembahasan yang muncul dalam setiap kelompok. (Hendrastuty, 2024).

## 2.6 | Evaluasi Hasil Clustering

Untuk mengevaluasi kualitas hasil clustering, penelitian ini menggunakan metode evaluasi internal berupa Silhouette Score. Metode ini mengukur tingkat kesamaan suatu data dengan cluster-nya sendiri dibandingkan dengan cluster lainnya.

Nilai Silhouette berada pada rentang -1 hingga 1. Nilai yang mendekati 1 menunjukkan bahwa data berada pada cluster yang tepat, sedangkan nilai mendekati 0 menunjukkan adanya tumpang tindih antar cluster. Evaluasi ini digunakan untuk menentukan jumlah cluster yang optimal serta menilai efektivitas proses pengelompokan data.

## 2.7 | Alur Penelitian

Secara umum, alur penelitian dimulai dari pengumpulan data Twitter, preprocessing teks, analisis sentimen berbasis leksikon, pembobotan kata menggunakan TF-IDF, proses clustering teks, hingga evaluasi hasil clustering. Tahapan-tahapan tersebut dirancang untuk menghasilkan pemetaan opini publik yang komprehensif berdasarkan sentimen dan topik pembahasan.

### 3 | HASIL

Bagian ini menyajikan hasil dari proses analisis sentimen dan clustering teks terhadap komentar Twitter yang berkaitan dengan isu ekonomi nasional. Hasil penelitian diperoleh setelah seluruh data melalui tahapan preprocessing, pembobotan kata menggunakan TF-IDF, serta proses analisis sentimen dan pengelompokan data.

#### 3.1 | Hasil Pre-Processing Data

Tahap preprocessing menghasilkan data teks yang telah dibersihkan dari unsur-unsur yang tidak relevan, seperti tanda baca, simbol, URL, dan kata-kata umum yang tidak memiliki makna signifikan. Proses ini bertujuan untuk meningkatkan kualitas data sebelum dilakukan analisis lebih lanjut.

Setelah preprocessing, setiap komentar Twitter direpresentasikan dalam bentuk teks yang lebih terstruktur, sehingga dapat digunakan secara efektif dalam proses analisis sentimen dan clustering. Hasil preprocessing ini menjadi dasar bagi tahap pembobotan kata dan analisis selanjutnya.

#### 3.2 | Distribusi Sentimen Opini Publik

Berdasarkan hasil analisis sentimen berbasis leksikon, komentar Twitter yang dianalisis diklasifikasikan ke dalam tiga kategori sentimen, yaitu positif, negatif, dan netral. Distribusi sentimen menunjukkan adanya variasi respon publik terhadap isu ekonomi nasional.

Hasil pengolahan data menunjukkan bahwa sebagian besar komentar berada pada kategori sentimen netral, diikuti oleh sentimen negatif dan sentimen positif. Hal ini mengindikasikan bahwa banyak pengguna Twitter yang menyampaikan informasi atau opini tanpa kecenderungan emosional yang kuat, sementara sebagian lainnya menunjukkan ketidakpuasan atau dukungan terhadap isu ekonomi yang dibahas.

Distribusi sentimen ini memberikan gambaran awal mengenai kecenderungan opini publik, namun belum cukup untuk memahami topik diskusi yang berkembang. Oleh karena itu, diperlukan analisis lanjutan melalui proses clustering teks.

#### 3.3 | Hasil Clustering Teks

Proses clustering dilakukan terhadap data teks yang telah direpresentasikan dalam bentuk vektor TF-IDF. Pengelompokan komentar Twitter menghasilkan beberapa cluster yang merepresentasikan topik diskusi yang berbeda dalam isu ekonomi nasional.

Setiap cluster memiliki karakteristik kata dominan yang mencerminkan tema pembahasan tertentu. Analisis terhadap kata-kata yang sering muncul pada masing-masing cluster menunjukkan adanya perbedaan fokus diskusi, seperti kebijakan ekonomi, harga kebutuhan pokok, kondisi lapangan kerja, dan kesejahteraan masyarakat.

Hasil clustering ini menunjukkan bahwa meskipun komentar berada dalam kategori sentimen yang sama, topik pembahasannya dapat berbeda. Dengan demikian, clustering teks memberikan informasi tambahan yang tidak dapat diperoleh hanya dari analisis sentimen saja.

#### 3.4 | Evaluasi Kualitas Clustering

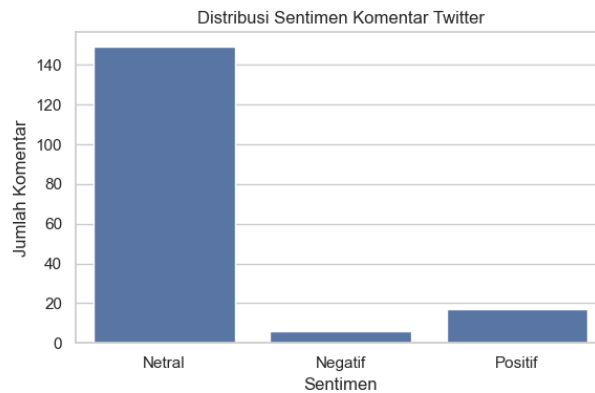
Evaluasi kualitas hasil clustering dilakukan menggunakan metode Silhouette Score. Nilai Silhouette yang diperoleh menunjukkan bahwa proses pengelompokan data telah menghasilkan cluster yang cukup baik, dengan tingkat kemiripan data yang tinggi di dalam cluster dan perbedaan yang jelas antar cluster.

Nilai evaluasi ini digunakan sebagai dasar dalam menentukan jumlah cluster yang optimal serta menilai efektivitas metode clustering yang digunakan. Hasil evaluasi menunjukkan bahwa pendekatan clustering teks mampu mengelompokkan komentar Twitter secara representatif berdasarkan kemiripan konten.

#### 3.5 | Visualisasi Hasil

Untuk memperjelas hasil analisis, visualisasi data digunakan dalam bentuk tabel dan grafik. Visualisasi distribusi sentimen menunjukkan perbandingan jumlah komentar pada masing-masing kategori sentimen, sedangkan visualisasi hasil clustering menggambarkan sebaran data berdasarkan kelompok topik.

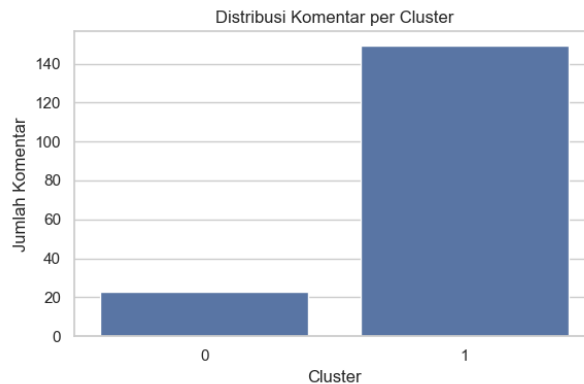
Penggunaan visualisasi ini membantu dalam memahami pola opini publik secara lebih intuitif dan mendukung interpretasi hasil penelitian secara keseluruhan.



**GAMBAR 3.** Hasil Grafik Sentimen

Pada Gambar 3 hasil ini menunjukkan bahwa sebagian besar pengguna Twitter memberikan komentar yang bersifat informatif, deskriptif, atau netral secara emosional, tanpa ekspresi dukungan atau penolakan yang kuat. Komentar positif muncul dalam proporsi kecil, mencerminkan adanya apresiasi atau pandangan optimistis terhadap isu ekonomi nasional atau kebijakan yang dibahas. Sedangkan komentar negatif yang jumlahnya sangat sedikit mencerminkan sebagian kecil pengguna yang bersikap kritis atau skeptis terhadap topik tersebut.

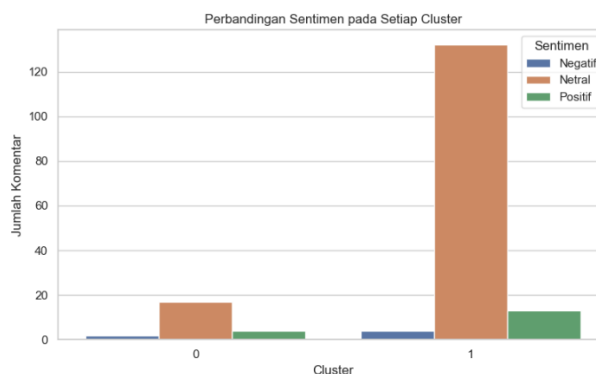
Hasil distribusi ini juga menjadi input awal bagi tahap Clustering (K-Means) untuk mengelompokkan komentar berdasarkan topik pembahasan yang kemudian membentuk dua tema utama, yaitu Kebijakan publik dan efektivitas anggaran dan isu utang dan kebijakan fiskal.



**GAMBAR 4.** Hasil Grafik Komentar per Cluster

Sementara itu, pada Gambar 4, Komentar dalam cluster 0 (Kebijakan & Anggaran), cenderung membahas efektivitas penggunaan anggaran, serapan dana, dan program kerja pemerintah. Sentimen dominan di cluster ini adalah netral hingga positif, menandakan pembahasan lebih bersifat informatif.

Komentar dalam cluster 1 (Utang & Fiskal), banyak menyinggung isu utang negara, kebijakan APBN, dan kritik terhadap pemerintah. Meskipun sebagian besar bersifat netral, terdapat komentar bernada negatif yang menggambarkan persepsi publik yang kritis terhadap kebijakan ekonomi.



**GAMBAR 5.** Hasil Grafik Sentimen per Cluster

Hasil ini menunjukkan bahwa kedua cluster memiliki dominasi sentimen netral, yang memperkuat temuan sebelumnya bahwa mayoritas komentar Twitter bersifat informatif dan tidak emosional. Namun, perbedaan kecil antara cluster menandakan adanya kecenderungan emosional yang spesifik pada tiap topik.

Cluster 0 (Kebijakan & Anggaran) lebih banyak sentimen positif, mencerminkan optimisme atau dukungan terhadap kinerja pemerintah. Cluster 1 (Utang & Fiskal) muncul lebih banyak sentimen negatif, mencerminkan pandangan kritis terhadap pengelolaan keuangan negara.

Secara keseluruhan, hasil pengujian menunjukkan bahwa, distribusi sentimen memperlihatkan kecenderungan netral, namun topik utang dan proyek nasional memunculkan opini negatif yang lebih kuat. Top terms mampu merepresentasikan tema dominan dari setiap cluster secara semantik. Konsistensi model menegaskan bahwa pola pengelompokan yang dihasilkan K-Means bukan kebetulan, tetapi hasil dari pola kata yang stabil di dalam kata.

## 4 | PEMBAHASAN

Pembahasan ini menginterpretasikan hasil analisis sentimen dan clustering teks terhadap komentar Twitter yang berkaitan dengan isu ekonomi nasional. Hasil penelitian menunjukkan bahwa pendekatan integratif antara analisis sentimen berbasis lexicon dan clustering teks mampu memberikan pemahaman yang lebih mendalam mengenai pola opini publik di media sosial.

Dominasi sentimen netral dalam hasil analisis menunjukkan bahwa sebagian besar pengguna Twitter cenderung menyampaikan informasi, opini faktual, atau tanggapan tanpa ekspresi emosional yang kuat. Temuan ini sejalan dengan karakteristik media sosial Twitter yang sering digunakan sebagai sarana berbagi informasi dan diskusi publik secara terbuka. Sementara itu, keberadaan sentimen negatif yang cukup signifikan mengindikasikan adanya ketidakpuasan atau kekhawatiran masyarakat terhadap kondisi ekonomi nasional, seperti kenaikan harga kebutuhan pokok, lapangan pekerjaan, dan kebijakan ekonomi pemerintah.

Sentimen positif yang muncul dalam jumlah lebih kecil mencerminkan adanya dukungan atau optimisme sebagian masyarakat terhadap isu ekonomi tertentu. Hal ini menunjukkan bahwa opini publik terhadap isu ekonomi nasional bersifat heterogen dan dipengaruhi oleh berbagai faktor, baik kondisi ekonomi aktual maupun persepsi individu pengguna media sosial.

Hasil clustering teks memberikan kontribusi penting dalam memperkaya analisis sentimen. Dengan mengelompokkan komentar berdasarkan kemiripan konten, penelitian ini mampu mengidentifikasi topik-topik diskusi yang dominan dalam opini publik. Setiap cluster merepresentasikan fokus pembahasan yang berbeda, meskipun berada dalam kategori sentimen yang sama. Temuan ini menegaskan bahwa analisis sentimen saja belum cukup untuk memahami kompleksitas opini publik, sehingga diperlukan pendekatan tambahan seperti clustering teks.

Integrasi analisis sentimen dan clustering teks memungkinkan pemetaan opini publik secara lebih komprehensif. Pendekatan ini tidak hanya mengungkap kecenderungan emosional masyarakat, tetapi juga struktur topik diskusi yang berkembang. Hal ini sejalan dengan penelitian-penelitian sebelumnya yang menyatakan bahwa penggabungan metode text mining dapat meningkatkan kualitas analisis opini publik dibandingkan dengan penggunaan satu metode tunggal.

Dari sisi metodologis, penggunaan pendekatan berbasis leksikon dalam analisis sentimen memberikan kemudahan dalam implementasi serta tidak memerlukan data latih. Namun demikian, pendekatan ini memiliki keterbatasan dalam menangkap konteks

kalimat, ironi, atau sarkasme yang sering muncul dalam bahasa media sosial. Oleh karena itu, hasil klasifikasi sentimen perlu dipahami sebagai representasi umum dari kecenderungan opini publik, bukan sebagai interpretasi absolut.

Hasil evaluasi clustering menggunakan Silhouette Score menunjukkan bahwa proses pengelompokan data telah menghasilkan cluster yang cukup representatif. Nilai evaluasi yang diperoleh mengindikasikan bahwa komentar Twitter yang dianalisis memiliki struktur topik yang dapat dipisahkan dengan baik. Hal ini memperkuat validitas penggunaan teknik clustering teks dalam penelitian ini.

Secara keseluruhan, hasil pembahasan menunjukkan bahwa metode yang digunakan dalam penelitian ini mampu memberikan gambaran yang lebih kaya mengenai opini publik terhadap isu ekonomi nasional di Twitter. Temuan ini memiliki implikasi penting bagi akademisi, peneliti, maupun pengambil kebijakan dalam memahami respons masyarakat berbasis data media sosial. Selain itu, pendekatan yang digunakan dalam penelitian ini dapat dikembangkan lebih lanjut dengan mengintegrasikan metode pembelajaran mesin atau analisis semantik untuk meningkatkan akurasi dan kedalaman analisis.

## 5 | KESIMPULAN

Berdasarkan hasil analisis dan pembahasan yang telah dilakukan, dapat disimpulkan bahwa pemanfaatan analisis sentimen berbasis leksikon dan clustering teks mampu memberikan gambaran yang komprehensif mengenai opini publik terhadap isu ekonomi nasional di media sosial Twitter. Analisis sentimen berhasil mengklasifikasikan opini masyarakat ke dalam kategori positif, negatif, dan netral, sehingga memberikan informasi mengenai kecenderungan emosional publik terhadap isu yang dibahas.

Hasil clustering teks menunjukkan bahwa komentar Twitter dapat dikelompokkan ke dalam beberapa cluster yang merepresentasikan topik diskusi yang berbeda. Temuan ini membuktikan bahwa opini publik tidak hanya beragam dari sisi sentimen, tetapi juga dari sisi fokus pembahasan. Integrasi analisis sentimen dan clustering teks memungkinkan pemetaan opini publik yang lebih mendalam dibandingkan dengan penggunaan satu metode analisis saja.

Secara keseluruhan, penelitian ini memberikan kontribusi dalam penerapan metode text mining untuk analisis opini publik berbasis media sosial, khususnya dalam konteks isu ekonomi nasional. Pendekatan yang digunakan dapat dijadikan sebagai referensi dalam penelitian sejenis serta dapat dikembangkan lebih lanjut dengan mengintegrasikan metode pembelajaran mesin atau analisis semantik untuk meningkatkan akurasi hasil analisis di masa mendatang.

## Daftar Pustaka

- Fuad Amirullah, Syariful Alam, M.Imam Sulistyono S (2023). Analisis Sentimen Terhadap Kinerja KPU Menjelang Pemilu 2024 Berdasarkan Opini Twitter Menggunakan Naïve Bayes. *Jurnal Ilmiah Teknik dan Ilmu Komputer*, Vol. 2No. 3, Agustus 2023, 69-76e-ISSN: 2828-5344. <https://doi.org/10.55123/storage.v2i3.2293>
- A. Amelia and R. Yusuf, "ANALISIS SENTIMEN MASYARAKAT INDONESIA PADA PLATFORM X TERHADAP ISU FUFUFAFA MENGGUNAKAN BIDIRECTIONAL ENCODER REPRESENTATIONS FROM TRANSFORMERS", *JINTEKS*, vol. 7, no. 1, pp. 72–80, Feb. 2025. <https://doi.org/10.51401/jinteks.v7i1.5160>
- Dini K.Astutik, Artanti, Fenny F. Penerapan Text Mining pada Analisis Sentimen Pengguna Twitter Layanan Transportasi Online Menggunakan Metode Density Based Spatial Clustering of Applications With Noise (DBSCAN) dan K-Means. (2022). *J Statistika: Jurnal Ilmiah Teori Dan Aplikasi Statistika*, 15(1). <https://doi.org/10.36456/jstat.vol15.no1.a5983>
- Amanda, A., & Nurmawati, E. (2023). Analisis Sentimen dan Pemodelan Topik Pada Tweet Terkait Data Badan Pusat Statistik. *Jurnal Sistem Informasi Dan Informatika (Simika)* P-ISSN, 6(2), 2622–6901. <https://doi.org/10.47080/simika.v6i2.2789>
- Anggina, S., Setiawan, N. Y., & Bachtiar, F. A. (2022). Analisis Ulasan Pelanggan Menggunakan Multinomial Naïve Bayes Classifier dengan Lexicon-Based dan TF-IDF Pada Formaggio Coffee and Resto. *@ is The Best: Accounting Information Systems and Information Technology Business Enterprise*, 7(1), 76-90. <https://doi.org/10.34010/aisthebest.v7i1.7072>
- Alamsyah, P., Hakim, L. N., Wijaya, G., & Wicaksono, A. (2024). Debunking disinformation on YouTube: a fact check on the 2024 Indonesian election. *Jurnal Studi Komunikasi (Indonesian Journal of Communications Studies)*, 8(3), 547–560.

<https://doi.org/10.25139/jsk.v8i3.8348>

- Alfian, I. (2023). Penerapan Metode K-Means Dalam Melakukan Pengelompokan Bencana Alam di Indonesia Dilakukan dengan Memanfaatkan Teknik Text Mining. <https://jurnal.itg.ac.id/>
- Firmansyah, I., & Herawan, B. H. (2022). Analisis Sentimen Citayam Fashion Week menggunakan Support Vector Machine Analysis of Citayam Fashion Week Sentiment Using Support Vector Machine. <https://doi.org/10.26418/justin.v10i4.56664>
- Fitriyah, N., Warsito, B., Asih, D., & Maruddani, I. (2021). Analisis Sentimen Gojek Pada Media Sosial Twitter Dengan Klasifikasi Support Vector Machine (SVM). *JURNAL GAUSSIAN*, 9(3), 376–390. <https://ejournal3.undip.ac.id/index.php/gaussian/>
- Hendrastuty, N. (2024). Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Dalam Evaluasi Hasil Pembelajaran Siswa. *Jurnal Ilmiah Informatika Dan Ilmu Komputer (JIMA-ILKOM)*, 3(1), 46–56. <https://doi.org/10.58602/jima-ilkom.v3i1.26>
- Junior, A. R., Handayani, H. H., Fitri, A., & Masruriyah, N. (2022). Analisis Sentimen Menggunakan Algoritma K-Means untuk Mengetahui Kalimat Positif maupun Negatif pada Buletin APTIKOM. III(1), 113.
- Hoiriyah, H., Mardiana, H., Walid, M., & Darmawan, A. K. (2023). Lexicon-based and naive bayes sentiment analysis for recommending the best marketplace selection as a marketing strategy for msme. *Jurnal Pilar Nusa Mandiri*, 19(2), 65–76. <https://doi.org/10.33480/pilar.v19i1.4176>
- ISMAL, Amalia Rachmadana; HAKIM, Raden Bagus Fajriya. Implementasi Lexicon Based Untuk Analisis Sentimen Dalam Menentukan Rekomendasi Pantai Di DI Yogyakarta Berdasarkan Data Twitter: Implementasi Lexicon Based. *Emerging Statistics and Data Science Journal*, 2023, 1.1: 37–46. <https://doi.org/10.20885/esds.vol1.iss.1.art5>
- Prasetyo, K., & Mahendra, R. (2025). Analisis Kinerja Convolutional Neural Networks Baseline untuk Identifikasi Jenis Penyakit Kentang. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 5(2), 609–615. <https://doi.org/10.57152/malcom.v5i2.1722>
- Tri Elsa, M., Berutu, S. S., & Maedjaja, F. (2024). Analisis Perbincangan Pada Grup Whatsapp Dengan K-Means Clustering. *JUPI (Jurnal Ilmiah Penelitian Dan Pembelajaran Informatika)*, 9(4), 1960–1971. <https://doi.org/10.29100/jipi.v9i4.5554>
- AMALIAH, Fitrah; NURYANA, I. Kadek Dwi. Perbandingan Akurasi Metode Lexicon Based Dan Naive Bayes Classifier Pada Analisis Sentimen Pendapat Masyarakat Terhadap Aplikasi Investasi Pada Media Twitter. *Journal of Informatics and Computer Science (JINACS)*, 2022, 3.03: 384–393. <https://doi.org/10.26740/jinacs.v3n03.p384-393>
- Sumitro, P. A., Mulyana, D. I., & Saputro, W. (2021). Analisis Sentimen Terhadap Vaksin Covid-19 di Indonesia pada Twitter Menggunakan Metode Lexicon Based. *Jurnal Informatika dan Teknologi Komputer (J-ICOM)*, 2(2), 50–56. <https://doi.org/10.55377/j-icom.v2i2.4009>