p-ISSN : 2443-2210
e-ISSN : 2443-2229

Jurnal Teknik Informatika dan Sistem Informasi
Volume 11 Nomor 3 Desember 2025

Analisis Ulasan Daring Menggunakan Metode *Density-Based Spatial Clustering of Applications with Noise*

http://dx.doi.org/10.28932/jutisi.v11i3.12363

Riwayat Artikel
Received: 30 Juni 2025 | *Final Revision*: 10 November 2025 | *Accepted*: 22 November 2025

Creative Commons License 4.0 (CC BY – NC)

Edwin Hartono[#1], Charitas Fibriani[✉*2]

[#] *Teknik Informatika, Fakultas Teknologi Informasi, Universitas Kristen Satya Wacana, Jl. Diponegoro No. 52-60, Salatiga, Kec. Sidorejo, Kota Salatiga, Jawa Tengah, Indonesia*
[1]672022016@student.uksw.edu

[*]*Sistem Informasi, Fakultas Teknologi Informasi, Universitas Kristen Satya Wacana, Jl. Diponegoro No. 52-60, Salatiga, Kec. Sidorejo, Kota Salatiga, Jawa Tengah, Indonesia*
[2]charitas.fibriani@uksw.edu

[✉]Corresponding author: charitas.fibriani@uksw.edu

Abstrak* — Penelitian ini menerapkan metode klaster berbasis kepadatan untuk menganalisis persepsi pengguna berdasarkan ulasan di Google Maps. Fokus penelitian ini terletak pada pengolahan data ulasan yang bersifat dinamis dan tidak memiliki label untuk menjawab kebutuhan pengelola dalam memahami sentimen publik. Sebanyak 399 data telah dikumpulkan melalui Apify, kemudian data diproses melalui tahap pembersihan, normalisasi, dan *stemming*. Representasi teks dilakukan dengan pembobotan frekuensi kata terhadap dokumen, sementara visualisasi WordCloud dimanfaatkan untuk mengidentifikasi kata-kata dominan yang mencerminkan persepsi positif sehingga dapat membantu memahami konteks sebelum proses pengelompokan. Metode *Density-Based Spatial Clustering of Applications with Noise* diterapkan untuk membentuk klaster ulasan. Hasil analisis menunjukkan bahwa metode ini mampu mengelompokkan ulasan dalam klaster berdasarkan kemiripan konten, meskipun sebagian data teridentifikasi sebagai gangguan. Temuan tersebut memberikan wawasan yang bermanfaat dalam memahami persepsi masyarakat, sehingga dapat membantu dalam pengambilan keputusan strategis. Dengan pemilihan parameter yang tepat, metode ini mampu menjadi pendekatan efektif untuk analisis sentimen ulasan publik lebih lanjut.

***Kata kunci*— Analisis Sentimen; DBSCAN; Google Maps; Klasterisasi.**

Online Review Analysis Using Density-Based Spatial Clustering of Applications with Noise

Abstract* — *This study applies a density-based clustering method to analyze user perceptions based on reviews on Google Maps. The focus of this research lies in processing dynamic, unlabeled review data to address managers' needs in understanding public sentiment. A total of 399 data sets were collected through Apify, then the data were processed through cleaning, normalization, and stemming stages. Text representation was performed by weighting word frequencies across documents, while WordCloud visualization was utilized to identify dominant words reflecting positive perceptions to help understand the context before the clustering process. The Density-Based Spatial Clustering of Applications with Noise method was applied to form review clusters. The analysis results show that this method is able to group reviews into clusters based on content similarity, although some data were identified as noise. These findings provide useful insights in understanding public perception, thus aiding in strategic decision-making. With the right parameter selection, this method can be an effective approach for further public review sentiment analysis.

***Keywords*— *Clustering*; *DBSCAN*; *Google Maps*; *Sentiment Analysis*.**

I. PENDAHULUAN

Perkembangan teknologi informasi mendorong *platform digital* seperti Google Maps menjadi salah satu sumber informasi mengenai berbagai tempat. Google Maps mendukung pengambilan keputusan dengan memberikan pengetahuan mendalam mengenai kualitas suatu tempat [1]. Analisis ulasan pengguna pada *platform* Google Maps telah menjadi salah satu pendekatan dalam memahami persepsi publik terhadap institusi pendidikan, termasuk Universitas Kristen Satya Wacana (UKSW). Analisis ulasan penting bagi pengelola tempat dikarenakan mereka dapat mengetahui sentimen positif dan negatif yang diberikan oleh pengguna sehingga mereka dapat memperbaiki atau meningkatkan kualitas pelayanan mereka [2]. Hal tersebut membuat pengorganisasian dan pengelolaan data menjadi sangat penting untuk mendapatkan sebuah makna yang berguna di dalam penelitian [3].

Sebagian besar penelitian sebelumnya lebih banyak memfokuskan pada metode *supervised learning* atau *clustering* tradisional seperti K-Means, yang mengharuskan penentuan jumlah klaster sejak awal [4]. Pendekatan ini kurang ideal untuk data ulasan yang bersifat dinamis dan tidak berlabel. Metode *clustering* seperti DBSCAN (*Density-Based Spatial Clustering of Applications with Noise*) dapat menjadi alternatif lain dalam membentuk klaster secara otomatis tanpa memerlukan penentuan jumlah klaster terlebih dahulu. DBSCAN mampu mengabaikan data yang dianggap sebagai *noise* atau *outliner*, menjadikan algoritma ini lebih adaptif terhadap data ulasan yang beragam [5].

Algoritma DBSCAN menggunakan dua parameter utama, yaitu *epsilon* (ε) dan MinPts. Parameter *epsilon* (ε) digunakan untuk menentukan jarak maksimum untuk memasukkan data tetangga, sedangkan MinPts digunakan untuk menentukan jumlah minimum titik dalam radius ε yang mengklasifikasikan sebuah titik sebagai *core point*. Proses algoritma DBSCAN meliputi indentifikasi *core points*, ekspansi klaster dengan penambahan titik tetangga, serta penandaan *noise* untuk titik yang tidak memenuhi syarat *density*. Kinerja algoritma ini sangat dipengaruhi oleh pemilihan parameter ε dan MinPts, serta dapat mengalami penurunan efektivitas ketika diterapkan pada data dengan dimensi yang sangat tinggi tanpa adanya proses reduksi dimensi seperti t-SNE.

Beberapa penelitian terdahulu telah membahas teknik analisis teks untuk memahami sentimen publik melalui media sosial. Sebagian besar studi berfokus pada *platform* seperti Twitter dengan menggunakan metode *clustering* tradisional seperti K-*Means* dan *Hierarchical*. Meskipun kedua metode ini efektif dalam mengelompokkan data, metode ini masih memiliki keterbatasan dalam pengolahan daya berdimensi tinggi dikarenakan perlunya penentuan jumlah klaster sejak awal serta sensitivitas terhadap *noise* [6]. Keterbatasan ini menunjukkan bahwa pendekatan tersebut kurang ideal untuk mengolah data ulasan yang bersifat dinamis dan tidak berlabel, seperti ulasan pengguna terhadap layanan dan fasilitas kampus pada melalui Google Maps.

Studi penelitian lainnya membandingkan kinerja performa algoritma K-*Means* dan DBSCAN dalam menganalisis sentimen teks. Hasilnya menunjukkan bahwa DBSCAN memiliki kemampuan yang lebih baik dalam mendeteksi *noise* dan mampu membentuk klaster secara otomatis tanpa perlu menentukan jumlah klaster sebelumnya. Performa DBSCAN sangat sensitif terhadap pemilihan parameter nilai ε (*epsilon*) dan MinPts, serta kualitas tahap pra-pemrosesan yang dilakukan sebelumnya [7]. Temuan ini memungkinkan DBSCAN untuk menganalisis sentimen teks dengan lebih baik, akan tetapi perlu dilengkapi *preprocessing* yang matang dan pemilihan parameter yang tepat untuk memaksimalkan efektivitas algoritma tersebut.

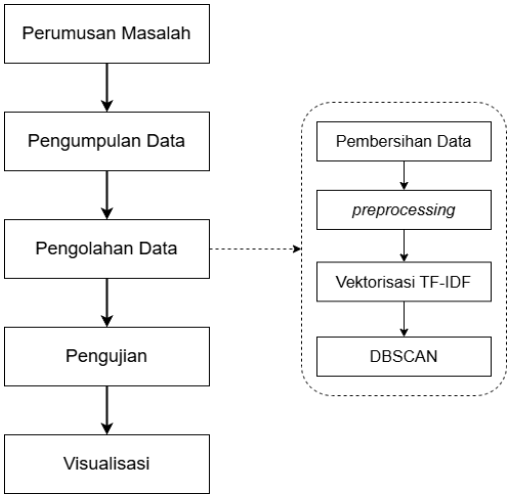
Penerapan DBSCAN juga ditemukan pada studi kasus analisis data medis untuk mengelompokkan data pasien penyakit ginjal kronis. Hasil studi ini memperlihatkan bahwa DBSCAN mampu membentuk klaster yang bergantung pada pemilihan parameter serta strategi pra-pemrosesan data dengan baik, termasuk penerapan representasi data TF-IDF [8]. Penerapan DBSCAN dalam konteks analisis persepsi publik terhadap institusi pendidikan melalui Google Maps masih terbatas, sehingga memberikan peluang untuk melakukan eksplorasi lebih lanjut terhadap potensi algoritma ini.

Berdasarkan hasil penelitian terdahulu, DBSCAN mampu menunjukkan potensinya dalam analisis data teks, dan penerapan terhadap konteks ulasan kampus masih jarang dilakukan. Penelitian ini difokuskan untuk menjawab dua pertanyaan utama. Pertama, bagaimana pola persepsi pengguna terhadap kampus UKSW dapat dikenali melalui ulasan yang diberikan pengguna di Google Maps. Kedua, seberapa efektif algoritma DBSCAN dalam melakukan klasterisasi terhadap data ulasan yang tidak memiliki label. Menjawab pertanyaan pertama akan membantu pihak universitas memahami kekuatan dan kelemahan berdasarkan pengalaman langsung pengguna. Pertanyaan kedua bertujuan untuk menguji kemampuan DBSCAN dalam mengelompokkan data teks berdimensi tinggi tanpa intervensi manual untuk menentukan jumlah klaster.

Penelitian ini berhipotesis bahwa DBSCAN mampu mengelompokkan ulasan secara efektif tanpa menentukan jumlah klaster sebelumnya serta memberikan gambaran yang bermakna mengenai persepsi publik. Kebaharuan penelitian ini terletak pada penerapan DBSCAN untuk menganalisis ulasan Google Maps pada konteks institusi pendidikan, serta mengintegrasikan TF-IDF dan WordCloud untuk memperkuat interpretasi hasil analisis.

II. METODE PENELITIAN

Metode penelitian merupakan serangkaian tahapan yang disusun secara terstruktur. Berikut ini merupakan penjelasan dari metode penelitian, penerapan dan pengujian algoritma pada hasil percobaan yang diterapkan pada ulasan Google Maps UKSW. Tahapan penelitian ditunjukkan pada Gambar 1, menjelaskan 5 tahapan utama: perumusan masalah, pengumpulan data, pengolahan data, pengujian klaster, dan visualisasi hasil.



Gambar 1. Alur Penelitian

A. Perumusan Masalah

Rumusan masalah didapat dengan mengamati fenomena ulasan pengguna pada Google Maps terkait kampus UKSW. Ulasan pengguna ini mencakup variasi opini publik mengenai fasilitas, layanan, dan pengalaman di lingkungan kampus. Melalui pengamatan ini, ditemukan dua faktor permasalahan yang menjadikan fokus utama penelitian. Pertama, memahami pola persepsi publik terhadap UKSW untuk mengetahui apakah persepsi publik ini bersifat positif atau negatif. Kedua, menguji efektivitas algoritma DBSCAN dalam mengelompokkan ulasan kampus UKSW tanpa menggunakan label.

B. Pengumpulan Data

Penelitian ini mengumpulkan data menggunakan aplikasi Apify yang memungkinkan *scraping* data ulasan Google Maps dilakukan secara otomatis dan efisien. Pemilihan aplikasi Apify digunakan untuk mempercepat proses ekstraksi informasi sehingga tidak memakan waktu yang lama. Data yang dikumpulkan mencakup ulasan dari periode Januari 2012 hingga Maret 2025, menghasilkan total 399 ulasan berbahasa Indonesia dan Inggris yang berisikan pengalaman pengguna terhadap fasilitas, layanan, serta lingkungan kampus UKSW. Data ulasan ini mampu memberikan wawasan yang kaya untuk analisis lebih lanjut. Data mentah kemudian di filter untuk menghapus duplikasi dan ulasan tidak relevan sehingga siap dilanjutkan ke tahap berikutnya.

C. Pengolahan Data

Pengolahan data dilakukan melalui proses *preprocessing* yang mengubah data mentah menjadi format yang siap untuk vektorisasi dan *clustering* [9]. Proses pembersihan teks meliputi penghapusan tag HTML, emotikon, URL, dan karakter non-alfanumerik. Tidak hanya itu, terdapat proses normalisasi (*lowercasing*), tokenisasi dengan NLTK, penghilangan *stopword* Bahasa Indonesia dan Bahasa Inggris, serta *stemming* menggunakan Sastrawi. Setelah proses pembersihan, data tersebut diubah menjadi vektor numerik melalui TF-IDF yang melakukan ekstraksi fitur untuk menghitung bobot dan mengukur pentingnya suatu kata dalam sebuah dokumen [10]. Tahap pengolahan data menggunakan beberapa rumus matematis untuk merepresentasikan teks, mengelompokkan data, dan mengevaluasi hasil. Representasi teks dilakukan menggunakan teknik TF-IDF (*Term Frequency-Inverse Document Frequency*) yang menggabungkan dua komponen utama: *Term Frequency* (TF) dan *Inverse Document Frequency* (IDF). TF menghitung frekuensi kemunculan kata dalam suatu dokumen menggunakan Rumus (1). IDF berperan dalam mengurangi bobot kata yang terlalu umum di seluruh dokumen dengan cara menghitung nilai logaritma dari perbandingan antara total jumlah dokumen N dan jumlah dokumen yang memuat kata t menggunakan Rumus (2). Nilai akhir TF-IDF untuk suatu kata t dalam dokumen d diperoleh dengan mengalikan TF dan IDF menggunakan Rumus (3). Istilah yang sering muncul dalam satu dokumen namun jarang ditemukan di dokumen lain akan memperoleh bobot yang tinggi.

Setelah tahap pembersihan, proses selanjutnya adalah klasifikasi data menggunakan algoritma DBSCAN (*Density-Based Spatial Clustering of Applications with Noise*). Algoritma ini mendefinisikan ε-*neighborhood* dari suatu titik p sebagai himpunan titik-titik dalam q, dalam *dataset* D yang memenuhi Rumus (4). Suatu titik p disebut sebagai *core point* apabila ukuran *ε-neighborhood* mencapai atau melebihi MinPts, dilihat pada Rumus (5). Titik-titik yang tidak memenuhi syarat tetapi berada dalam N dari *core point* disebut sebagai *border point*, sedangkan titik yang bukan *core* maupun *border* dianggap sebagai *noise* atau *outlier*. Suatu data harus memiliki sifat *density-reachable*, yang artinya titik p secara langsung dapat dijangkau dari *core point* q jika p berada dalam *ε-neighborhood*, menggunakan Rumus (6). Nilai parameter ditentukan melalui eksplorasi awal terhadap distribusi jarak data dan merujuk terhadap penelitian terdahulu, dengan kisaran nilai eps di antara 0.2 – 0.6 dan MinPts 3 – 6 yang memberikan keseimbangan antara jumlah klaster dan *noise* [11], [12], [13]. Setelah proses klasterisasi, dilakukan analisis tematik klaster untuk mengidentifikasi kata kunci dominan dari setiap kelompok yang terbentuk. Kata kunci dominan ditentukan dengan menghitung rata-rata bobot TF-IDF tertinggi di dalam klaster. Metode ini digunakan untuk penamaan klaster secara tematik dan penarikan kesimpulan strategis bagi UKSW.

$$tf(t,d) = \frac{f_{t,d}}{\sum_{t \epsilon d} f_{t,d}} \quad (1)$$

$$idf(t,D) = log\ log\ \frac{N}{|[d \epsilon D; t \epsilon d]|} \quad (2)$$

$$TF - IDF(t,d) = tf(t,d)\ x\ df(t,D) \quad (3)$$

$$N_\varepsilon(p) = \{q \epsilon D \mid dist(p,q) < \varepsilon\} \quad (4)$$

$$|N_\varepsilon(p)| \geq MinPts \quad (5)$$

$$p \in N_\varepsilon(q)\ dan\ |N_\varepsilon(q)| \geq MinPts \quad (6)$$

Keterangan:
$f_{t,d}$ = jumlah kemunculan kata t dalam dokumen d.
$\sum_{t \epsilon d} f_{t,d}$ = total jumlah kata dalam dokumen d.
N = total jumlah dokumen dalam korpus D.
$|[d \in D; t \in d]|$ = jumlah dokumen yang memuat kata t.
$dist(p,q)$ = jarak antara titik p dan q.
$MinPts$ = jumlah minimum untuk membentuk *core point.*

D. Pengujian Klaster

Tahap pengujian ini menggunakan metrik *silhouette score* yang menggabungkan metode *cohesion* dan *separation*. Metode *cohesion* diukur dengan menghitung jumlah keseluruhan objek dalam sebuah klaster, sedangkan *separation* ditentukan dengan menghitung jarak rata-rata antar objek dari klaster ke klaster terdekatnya. Jarak antar data dihitung menggunakan rumus *euclidean distance*. Untuk memberikan gambaran seberapa baik hasil *clustering*, nilai *silhouette* dihitung untuk setiap *cluster* menggunakan persamaan Rumus (7). Perhitungan tersebut akan dihitung sebagai rata-rata dari semua nilai *silhouette* menjadi *overall silhouette score* menggunakan Rumus (8) [14]. Alternatif metrik lain seperti *Davies-Bouldin Index* juga dipertimbangkan dalam analisis awal. *Silhouette score* dipilih karena mampu memberikan interpretasi lebih intuitif dan cocok untuk mengukur kinerja algoritma pada data berdimensi tinggi.

$$s(i) = \frac{b(i) - a(i)}{max\ (a(i), b(i))} \quad (7)$$

$$Sil(K) = \frac{1}{|K|} \sum_{i=1}^{K} s(c_i) \quad (8)$$

Keterangan:
$a(i)$ = rata-rata jarak titik data i ke titik lain dalam klaster yang sama.
$b(i)$ = rata-rata jarak titik data i ke titik dalam klaster terdekat lainnya.
K = jumlah total klaster dalam hasil *clustering.*
$s(c_i)$ = rata – rata *silhouette score* dari klaster i.

E. Visualisasi

Tahapan visualisasi menggunakan reduksi t-SNE dan WorlCloud. Reduksi dimensi t-SNE diterapkan untuk memetakan vektor TF-IDF berdimensi tinggi ke dalam ruang 2D. Hasilnya nanti akan ditampilkan dalam *scatterplot* di mana titik yang memiliki pola kemiripan tinggi akan terkumpul berdekatan, sedangkan titik *noise* atau *outlier* akan tampak terpisah. Struktur klaster DBSCAN dapat langsung diamati dari segi kepadatan maupun penyebarannya [15]. Selain visualisasi klaster yang dihasilkan oleh DBSCAN melalui reduksi dimensi t-SNE, penelitian ini juga memanfaatkan WordCloud untuk memberikan gambaran awal dan kontekstual mengenai isi ulasan yang menjadi objek klasterisasi. Visualisasi WordCloud tidak secara langsung terlibat dalam pembentukan klaster algoritma DBSCAN, tetapi visualisasi ini membantu untuk memahami tema-tema dominan dan sentimen umum yang terkandung dalam data ulasan Google Maps UKSW sebelum klasterisasi dilakukan. WordCloud memvisualisasikan kata kunci dominan, membantu peneliti memahami persepsi publik secara menyeluruh. Pemahaman ini menjadi dasar interpretasi yang lebih kaya terhadap klaster yang terbentuk oleh DBSCAN, memungkinkan peneliti untuk memahami makna di balik pengelompokan ulasan yang telah diidentifikasi secara global melalui WordCloud.

III. HASIL DAN PEMBAHASAN

Penelitian membentuk kerangka kerja implementasi clustering ulasan menggunakan algoritma DBSCAN, mulai dari persiapan data hingga visualisasi hasil. Untuk pemahaman alur pemrosesan lebih baik, dibentuklah *pseudocode* agar mampu menggambarkan bagaimana data mentah diolah, diubah ke representasi numerik, diolah algoritma DBSCAN, dan akhirnya divisualisasikan.

A. Format Teks

Tahap pertama pada Gambar 2 memastikan semua pustaka yang diperlukan tersedia. Penelitian ini melakukan *import library* untuk manipulasi data menggunakan pandas, pemrosesan teks (re, tokenisasi, *stopword removal*, emoji, dan Sastrawi), representasi vektor dengan TF-IDF, algoritma *clustering* menggunakan DBSCAN, evaluasi kualitas dengan *silhouette score*, reduksi dimensi dengan t-SNE, serta alat visualisasi yaitu matplotlib untuk *scatter plot* dan WordCloud untuk representasi TF-IDF. Persiapan semua modul ini di awal agar dapat dipanggil langsung tanpa adanya gangguan.

START
1. *Import library* yang dibutuhkan:
   - *Library* untuk membaca *file* Excel
   - *Library* untuk pemrosesan teks (tokenisasi, *stopword removal*, re, emoji, Sastrawi, numpy)
   - *Library* untuk representasi vektor TF-IDF
   - *Library* untuk algoritma DBSCAN
   - *Library* untuk reduksi data TSNE
   - *Library* untuk evaluasi *clustering Silhouette Score*
   - *Library* untuk visualisasi (WordCloud, plt)

Gambar 2. Persiapan *Library*

B. Pemuatan Dataset

Tahap kedua pada Gambar 3 merupakan proses pemuatan dataset Excel yang bertujuan untuk mengakses data ulasan secara sistematis. Data dibaca dari *file* Excel bernama datasets_UKSW.xlsx untuk dimuat ke dalam program. Pengambilan dataset hanya mengambil kolom *review_text* yang berisikan teks ulasan yang akan dianalisis lebih lanjut. Pemilihan ini memastikan bahwa hanya informasi yang diperlukan untuk digunakan dalam tahap analisis berikutnya.

2. Baca data dari *file* Excel:
   - Buka *file* "datasets_UKSW.xlsx"
   - Ambil kolom "review_text" sebagai sumber data teks

Gambar 3. Pemuatan Dataset Excel ke Program

C. Preprocessing Data

Tahap ketiga pada Gambar 4 merupakan proses *preprocessing*. Setiap ulasan akan diproses melalui beberapa sub-tahap. Pertama, seluruh karakter akan diubah menjadi huruf kecil untuk menghindari duplikasi fitur berbasis kapitalisasi. Kemudian, tanda baca dan angka dihapus, sehingga hanya tersisa huruf dan spasi. Proses berikutnya akan menggunakan

stopword Bahasa Indonesia dan Inggris untuk memperkaya makna vektor. Setelah itu, dilakukan normalisasi dan *stemming* untuk mengkonsolidasikan bentuk kata dasar. Keseluruhan proses ini akan menghasilkan dokumen bersih yang mampu mengurangi *noise* sebelum vektorisasi.

3. Lakukan *preprocessing* terhadap setiap teks:
   FOR setiap *review* IN kolom "review_text":
      - Ubah ke huruf kecil
      - Hapus karakter tanda baca
      - Hapus angka
      - Hapus *stopword* Bahasa Indonesia dan Inggris
      - *Stemming* dan Normalisasi kata
      - Gabungkan kembali kata-kata menjadi *string* bersih
   END FOR

Gambar 4. Proses *Preprocessing*

D. Vektorisasi TF-IDF

Tahap keempat pada Gambar 5 merupakan proses transformasi ke TF-IDF. Dokumen yang telah melewati tahapan *preprocessing* akan diubah menjadi representasi numerik menggunakan TF-IDF *Vectorizer*. Teknik IDF menggabungkan dua komponen utama: *Term Frequency* (TF) yang dihitung menggunakan Rumus (1), dan *Inverse Document Frequency* (IDF) yang dihitung menggunakan Rumus (2). Nilai akhir TF-IDF diperoleh dengan mengalikan TF dan IDF menggunakan Rumus (3). Parameter init yang dipilih adalah max_features = 2000, min_df = 3, max_df = 0,7, ngram_range = 1,2 untuk menyeleksi fitur agar tidak memasukkan kata yang terlalu jarang muncul maupun terlalu sering digunakan. Proses fit_transform di akhir akan menghasilkan matriks *sparse* berukuran n_samples x n_features yang siap dipakai oleh DBSCAN.

4. Transformasi teks menjadi representasi numerik:
   - Inisialisasi TF-IDF *Vectorizer* dengan parameter:
      - max_features = 2000
      - min_df = 3
      - max_df = 0.7
      - ngram_range = (1, 2)
      - stop_words = kombinasi *stopword* Inggris dan Indonesia
   - Fit dan *transform teks* menjadi matrix TF-IDF

Gambar 5. Vektorisasi TF-IDF

E. Visualisasi WordCloud

Tahap kelima pada Gambar 6 adalah pembuatan visualisasi WordCloud yang dihasilkan dari skor TF-IDF gabungan seluruh dokumen. Kata-kata dengan bobot tertinggi akan muncul lebih besar, menunjukkan istilah yang dominan dalam korpus. Representasi ini memudahkan peneliti mengenali tema utama dalam data tersebut.

5. Visualisasi konten teks menggunakan WordCloud:
   - Gabungkan semua teks bersih
   - Hitung frekuensi kata dari hasil TF-IDF
   - Tampilkan WordCloud berdasarkan bobot TF-IDF tertinggi

Gambar 6. Pembuatan Visualisasi WordCloud

Visualisasi yang dihasilkan pada Gambar 7, dapat ditemukan tema-tema utama yang mendominasi persepsi pengguna. Kata-kata seperti “kampus”, “uksw”, “tempat, dan “nyaman”, muncul dengan ukuran paling menonjol, mengindikasikan frekuensi dan bobot yang sangat tinggi dalam korpus data. Kata-kata lain yang sering muncul dan memiliki konotasi positif antara lain “universitas”, “indonesia”, “sejuk”, “baik”, dan “bagus”. Pengelompokkan kata-kata dominan ini secara kolektif mencerminkan persepsi publik yang positif terhadap kampus Universitas Kristen Satya Wacana (UKSW). Secara khusus,

dominasi kata “nyaman”, dan “sejuk” menunjukkan bahwa pengguna menghargai lingkungan dan fasilitas kampus sebagai tempat yang menyenangkan dan kondusif. Representasi visual ini membantu dalam mengidentifikasi kata kunci penting dan tema utama dalam data ulasan, sehingga memudahkan peneliti untuk menarik kesimpulan mengenai sentimen keseluruhan.

Gambar 7. Visualisasi WordCloud

F. Algoritma DBSCAN

Tahap keenam pada Gambar 8 menunjukkan proses penerapan algoritma DBSCAN dalam membentuk klaster dari matriks TF-IDF yang telah dihasilkan sebelumnya. Proses klasterisasi DBSCAN dilakukan dengan penyesuaian parameter ε (epsilon) dan MinPts (min_samples) menggunakan *metric cosine*, menggunakan Rumus (4), (5), dan (6). Evaluasi klaster menggunakan metrik *silhouette score* untuk mengukur seberapa baik data dalam satu klaster saling berdekatan dan terpisah dari klaster lainnya menggunakan persamaan Rumus (7), dan perhitungan rata-rata dari nilai *silhouette* menjadi *overral silhouette score* menggunakan Rumus (8). Nilai *silhouette score* menjadi indikator utama dalam menilai keberhasilan klasterisasi. Penilaian yang nantinya akan menjadi penentu efektivitas model dan dalam mengelompokkan data dengan kepadatan bervariasi. Hasil uji coba parameter dapat dilihat melalui Tabel 1.

6. Terapkan algoritma DBSCAN:
   - Atur parameter: Epsilon, MinPts, metric = cosine
   - Jalankan DBSCAN pada matrix TF-IDF
   - Setiap *review* akan mendapatkan label *cluster*
   - Nilai label -1 menunjukkan data yang tidak termasuk dalam *cluster* mana pun (*outlier*/*noise*)
   - Hitung kembali *silhouette score* menggunakan label hasil DBSCAN (*noise* dan tanpa *noise*)

Gambar 8. Proses Algoritma DBSCAN

TABEL 1
HASIL UJI COBA PARAMETER

Eps	MinPts	Jumlah Klaster	*Silhouette Score* (Tanpa *Noise*)	*Silhouette Score* (*Noise*)
0.2	3	13	0.92	0,04
0.2	4	5	0,97	0,04
0.2	5	3	0,95	0,02
0.2	6	0	-1,00	-1,00
0.3	3	19	0,70	0,07
0.3	4	11	0,72	0,04
0,3	5	5	0,77	0,03
0,3	6	1	-1,00	0,04
0,4	3	26	0,42	0,09

Eps	MinPts	Jumlah Klaster	*Silhouette Score* (Tanpa *Noise*)	*Silhouette Score* (*Noise*)
0,4	4	14	0,48	0,06
0,4	5	10	0,49	0,04
0,4	6	4	0,42	0,03
0,5	3	25	0,14	0,07
0,5	4	17	0,20	0,07
0,5	5	14	0,21	0,06
0,5	6	11	0,27	0,06
0,6	3	10	0,02	-0,01
0,6	4	3	0,03	-0,0003
0,6	5	2	0,02	0,01
0,6	6	2	0,02	0,01

Tabel 1 menunjukkan bahwa pemilihan nilai epsilon dan MinPts memiliki dampak terhadap metrik evaluasi. Nilai epsilon yang terlalu kecil menghasilkan klaster padata dengan *silhouette score* tinggi (tanpa *noise*), namun sering mengorbankan banyak data sebagai *noise*. Sebaliknya, nilai epsilon yang terlalu besar cenderung menggabungkan titik secara berlebihan, menurunkan kepadatan klaster sehingga nilai *silhouette score* menjadi kecil. Nilai MinPts yang lebih tinggi dapat menghasilkan klaster yang lebih bermakna, tetapi juga berisiko meningkatkan jumlah *noise* pada data yang jarang. Proporsi *noise* yang tinggi dalam sebagian besar konfigurasi menjadi tantangan utama dalam analisis ini. Tingginya *noise* menunjukkan bahwa banyak ulasan bersifat unik atau tidak memiliki kesamaan untuk tergabung dalam sebuah klaster. Untuk mengatasi hal ini, terdapat beberapa pendekatan yang dapat dipertimbangkan, seperti melakukan *preprocessing* lanjutan atau *filtering* kata yang kurang relevan, serta menerapkan teknik reduksi dimensi tambahan seperti PCA sebelum proses klasterisasi.

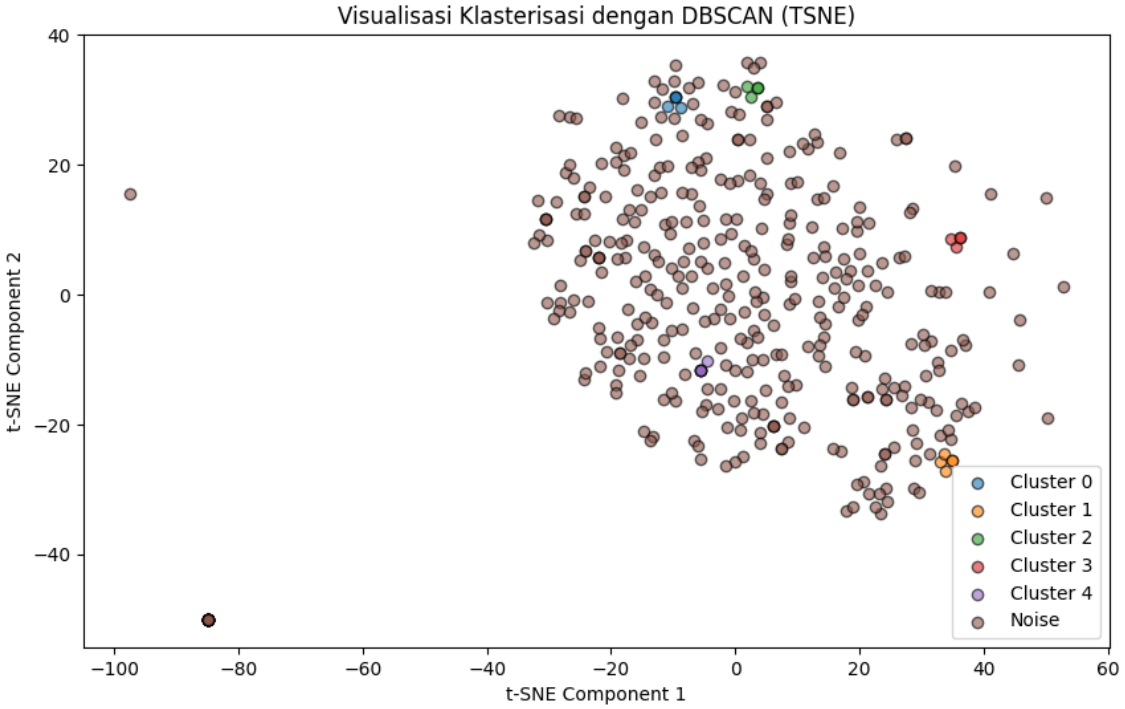
G. Reduksi Dimensi dan Visualisasi Scatter Plot

Tahap ketujuh pada Gambar 9 adalah proses reduksi data dan visualisasi hasil *clustering*. Untuk memvisualisasikan struktur klaster dalam ruang dua dimensi, digunakan t-SNE dengan parameter yang telah disesuaikan. Matriks TF-IDF direduksi, kemudian *scatterplot* dihasilkan dengan pewarnaan sesuai klaster label. Visualisasi ini memudahkan pengamatan pola kelompok dan keberadaan *outlier*, sekaligus memperlihatkan sejauh mana DBSCAN berhasil menangkap kepadatan data asli. Untuk memberikan pemahaman yang lebih mendalam, tiga konfigurasi parameter yang dipilih akan dianalisis lebih lanjut dengan gambaran visualisasi yang diberikan. Visualisasi ini membantu dalam memahami bagaimana ulasan-ulasan dikelompokkan dan sejauh mana *noise* mempengaruhi struktur klaster. Selain visualisasi, tampilkan juga kata kunci dominan dengan mengambil rata-rata bobot TF-IDF paling tinggi di dalam sebuah klaster.

7. Visualisasi hasil clustering dengan t-SNE:
   - Reduksi dimensi matrix TF-IDF menjadi 2D menggunakan t-SNE
   - Buat *scatter plot* dengan warna berdasarkan label klaster
   - Menampilkan kata kunci dominan tiap klaster

Gambar 9. Reduksi t-SNE dan Visualisasi Scatter Plot

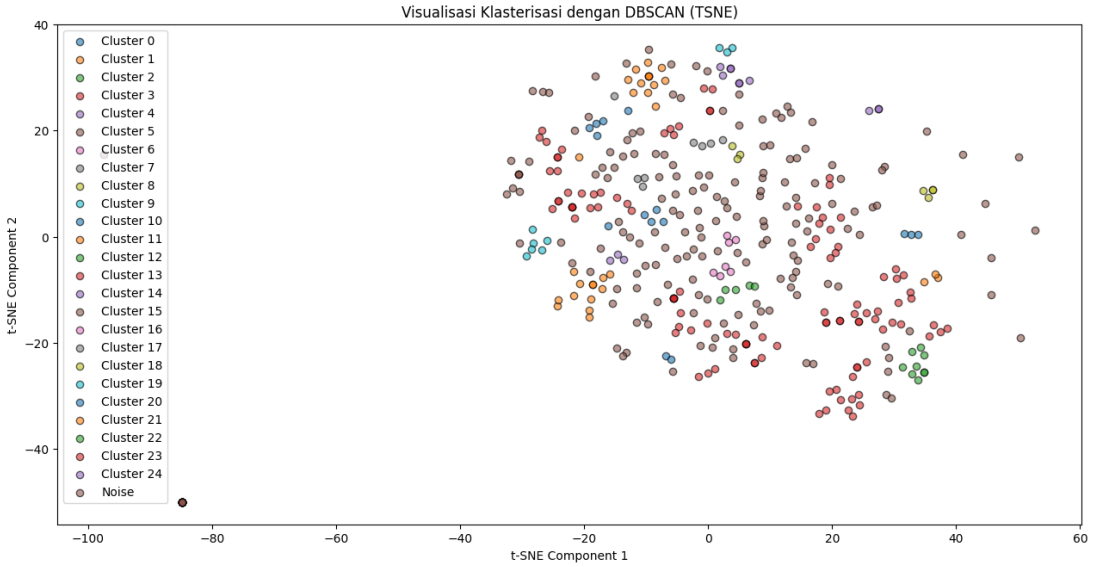
Konfigurasi pertama pada Gambar 10, menggunakan nilai eps = 0,2 dan MinPts = 4 yang memperoleh 5 klaster jika *noise* tidak dihitung. Nilai *silhouette score* tanpa *noise* mencapai 0,97, yang berarti bahwa klaster-klaster inti yang terbentuk memiliki tingkat kepadatan dan pemisahan yang sangat tinggi. Data dalam masing-masing klaster inti sangat mirip satu sama lain dan berbeda signifikan dari klaster inti lainnya. Kelemahan konfigurasi ini menghasilkan jumlah *noise* yang sangat tinggi. Hal ini menunjukkan bahwa meskipun beberapa kelompok ulasan dapat diidentifikasi dengan jelas, sebagian besar ulasan lainnya sangat beragam sehingga tidak dapat dimasukkan ke dalam klaster dengan radius eps yang kecil. Nilai *silhouette score* dengan *noise* yang sangat rendah (0,04) membuat struktur klaster menjadi kurang jelas akibat banyaknya *noise*. Kelima klaster yang terbentuk didominasi kata kunci yang generik dan positif seperti “bagus”, “baik”, dan “good”. Klaster ini merepresentasikan kepuasan dasar terhadap kampus UKSW.



Gambar 10. Visualisasi Klaster dengan eps = 0,2 dan MinPts = 4

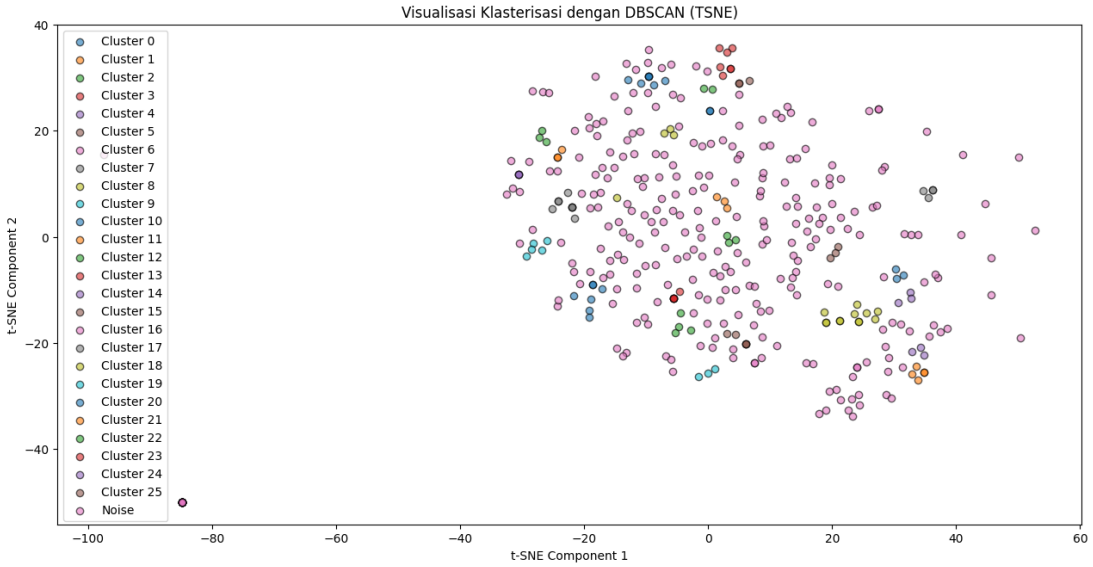
Konfigurasi kedua pada Gambar 11 menggunakan nilai eps = 0,5 dan MinPts = 3. Jumlah klaster yang terbentuk meningkat menjadi 25, sementara jumlah *noise* berkurang. Peningkatan eps memungkinkan lebih banyak titik data untuk bergabung ke dalam klaster. Namun, kualitas internal klaster akan menurun, dapat dilihat pada nilai *silhouette score* yang mengabaikan *noise* sebesar 0,14. Nilai ini menunjukkan klaster yang terbentuk cenderung kurang padat dan mungkin menjadi tumpang tindih. Nilai *silhouette score* dengan *noise* menjadi sedikit lebih baik (0,07) dibandingkan dengan konfigurasi pertama. Peningkatan tersebut menunjukkan bahwa sedikit data yang terisolasi sebagai *noise*, meskipun kualitas pemisahan antar klaster masih rendah. Visualisasi yang dihasilkan Gambar 11 menampilkan sebaran titik yang lebih merata ke dalam berbagai klaster, namun dengan batas antar klaster yang kurang baik. Klaster terbesar yang teridentifikasi sebanyak n = 98, didominasi ulasan positif seperti “sejuk”, “asri”, dan “nyaman”. Hasil ini menggarisbawahi bahwa ulasan tersebut menggambarkan kepuasan dan estetika kampus UKSW.

p-ISSN : 2443-2210
e-ISSN : 2443-2229



Gambar 11. Visualisasi Klaster dengan eps = 0,5 dan MinPts = 3

Konfigurasi ketiga pada Gambar 12 dengan eps 0,4 dan MinPts =3, membentuk 26 klaster di dalamnya. Nilai *silhouette score* tanpa *noise* adalah 0,42, menunjukkan kualitas klaster internal yang moderat (lebih baik dari konfigurasi kedua tetapi tidak setinggi konfigurasi pertama). Hal menariknya adalah pada nilai *silhouette score* dengan *noise* yaitu sebesar 0,09, menjadi yang tertinggi di antara ketiga skenario ini. Hal ini mengindikasikan bahwa konfigurasi ini memberikan keseimbangan yang baik dalam menangkap struktur data secara keseluruhan. Visualisasi pada Gambar 12 menunjukkan sejumlah klaster yang relatif jelas, meskipun tidak sepadat pada Gambar 10. Klaster terbesar yang teridentifikasi sebanyak n = 15, didominasi kata kunci seperti “place”, “great”, “university”, “nice”, dan “lot”. Klaster ini merepresentasikan dukungan positif terhadap institusi pendidikan UKSW. Ini menegaskan bahwa citra UKSW menjadi universitas baik yang telah diterima secara umum oleh pengguna.



Gambar 12. Visualisasi Klaster dengan eps = 0,4 dan MinPts = 3

DBSCAN mampu mengolah ulasan Google Maps Kampus UKSW dalam menemukan klaster ulasan yang seragam, terutama saat parameter eps diatur rendah (eps = 0,2 dengan MinPts = 4) menghasilkan nilai *silhouette score* tanpa *noise*

yang baik (0,97). Terdapat tantangan yang muncul akibat tingginya proporsi *noise* yang terjadi pada hampir semua skenario, terlihat dari nilai *silhouette score* dengan *noise* yang secara konsisten rendah. Hal tersebut mengindikasikan bahwa ulasan yang ada bisa jadi sangat beragam, atau mungkin karakteristik teks berbahasa Indonesia serta sensitivitas DBSCAN terhadap parameter eps dan MinPts memerlukan penyesuaian yang optimal. Secara keseluruhan, DBSCAN menawarkan potensi untuk memberikan gambaran awal mengenai persepsi publik melalui identifikasi klaster inti, meskipun memiliki keterbatasan dalam menangani variabilitas data ulasan secara menyeluruh tanpa membuat banyak *noise*, sehingga penentuan parameter yang seimbang menjadi sangat penting untuk pengembangan analisis lanjutan.

IV. SIMPULAN

Penelitian ini telah berhasil mengimplementasikan metode clustering DBSCAN untuk menganalisis pola persepsi pengguna terhadap Kampus UKSW melalui ulasan yang diberikan di Google Maps. Algoritma DBSCAN menunjukkan kemampuannya dalam mengidentifikasi klaster ulasan yang sangat padat dan seragam, terutama ketika parameter epsilon (ε) diatur pada nilai rendah (eps = 0,2 dengan MinPts = 4), menghasilkan *silhouette score* tanpa *noise* yang sangat baik (0,97). Hal tersebut mengindikasikan bahwa ulasan dalam klaster inti tersebut memiliki kemiripan yang tinggi dan terpisah dengan jelas dari klaster inti lainnya. Selain itu, visualisasi WordCloud pada tahap pra-analisis berhasil mengungkap tema-tema utama seperti "kampus", "uksw", "tempat", dan "nyaman", yang mencerminkan sentimen positif secara keseluruhan.

Penelitian ini tentunya memiliki sejumlah keterbatasan metodologis yang perlu diperhatikan. Pertama, ukuran dataset yang relatif kecil serta kemungkinan adanya rentang bias waktu dapat mempengaruhi hasil penelitian. Kedua, data ulasan yang bervariasi dan beragam membuat banyak data terklasifikasi sebagai *noise*. Ketiga, sensitivitas DBSCAN dalam pemilihan parameter dapat mempengaruhi kualitas hasil klasterisasi.

Beberapa rekomendasi teknis dapat dipertimbangkan. Perbaikan dapat dilakukan dengan meningkatkan ukuran dataset. Sistem pra-pemrosesan teks yang lebih dalam, misalnya *lemmatization*. Penambahan teknik reduksi untuk mengurangi data *noise*. Perbandingan performa DBSCAN dengan algoritma lain seperti K-*Means*, BERT, atau metode lainnya dapat memberikan wawasan yang lebih mendalam.

Temuan penelitian ini dapat memberikan implikasi praktis bagi pengelola kampus. Hasil klasterisasi yang diperoleh dapat dijadikan evaluasi terhadap kualitas layanan dan fasilitas kampus. Klaster yang menunjukkan ulasan negatif dapat menjadi acuan untuk perbaikan. Sebaliknya, klaster dengan sentimen positif dapat dimanfaatkan untuk meningkatkan citra institusi. Pendekatan tersebut membuat kampus dapat merancang intervensi berbasis data dan responsif terhadap kebutuhan masyarakat.

DAFTAR PUSTAKA

[1] P. M. Aryanto and R. Mardhiyyah, "Analisis Sentimen Terhadap Review Google Maps Jogja City Mall Menggunakan Algoritma Support Vector Machine," *Journal of Computer System and Informatics (JOSYC)*, vol. 6, pp. 25–35, 2024.
[2] S. N. Budiman, S. Lesanti, and Erwan, "Analisis Sentimen Berdasarkan Hasil Review Lokasi Google Map Menggunakan Natural Language Toolkit TextBlob dan Naïve Bayes," *JAMI: Jurnal Ahli Muda Indonesia*, vol. 5, no. 2, pp. 114–126, Nov. 2024.
[3] I. A. Siregar, "Analisis Dan Interpretasi Data Kuantitatif," *ALACRITY : Journal Of Education*, vol. 1, no. 2, pp. 39–48, 2021.
[4] 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-Popers pada Twitter," *Progresif: Jurnal Ilmiah Komputer*, vol. 18, no. 2, pp. 201–210, 2022.
[5] I. N. Simbolon and P. D. Friskila, "Analisis dan Evaluasi Algoritma DBSCAN (Density-based Spatial Clustering of Applications with Noise) pada Tuberkulosis," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 12, no. 3S1, Nov. 2024.
[6] V. D. Păvăloaia, "Clustering Algorithms in Sentiment Analysis Techniques in Social Media-A Rapid Literature Review," *Int J Adv Comput Sci Appl*, vol. 15, no. 3, 2024.
[7] F. Andriyani and Y. Puspitarani, "Performance Comparison of K-Means and DBScan Algorithms for Text Clustering Product Reviews," *SinkrOn: Jurnal Dan Penelitian Teknik Informatika*, vol. 6, no. 3, pp. 944–949, 2022.
[8] R. Anggara and A. Rahman, "Implementasi Algoritma DBSCAN Dalam Mengelompokan Data Pasien Terdiagnosa Penyakit Ginjal Kronis (PGK)," *Jurnal Algoritme*, vol. 3, no. 1, pp. 114–123, 2022.
[9] A. C. T. Angel, V. H. Pranatawijaya, and W. Widiatry, "Analisis Sentimen dan Emosi dari Ulasan Google Maps untuk Layanan Rumah Sakit di Palangka Raya Menggunakan Machine Learning," *KONSTELASI: Konvergensi Teknologi dan Sistem Informasi*, vol. 4, no. 1, pp. 35–49, 2024.
[10] N. Purnomo and W. Gata, "Pengelompokan Analisis Sentimen Komentar Youtube Terhadap Pengambilalihan Jalan Rusak di Lampung Menggunakan Algoritma Clustering," *Progresif: Jurnal Ilmiah Komputer*, vol. 20, no. 2, pp. 701–713, 2024.
[11] M. S. Amrullah, A. G. Putrada, M. N. Fauzan, and N. Alamsyah, "ETLE Sentiment Analysis Performance Increasement with TF-IDF, MDI Feature Selection, and SVM," *Sistemasi: Jurnal Sistem Informasi*, vol. 13, no. 4, pp. 1308–1318, 2024.
[12] D. H. Bangkalang, "Opinion Mining of Regional Heads in Indonesia Using The Support Vector Machine (SVM) Method," *JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 9, no. 3, pp. 1622–1627, 2024.
[13] D. Armiady, "Analisis Metode DBSCAN (Density-Based Spatial Clustering of Application with Noise) dalam Mendeteksi Data Outlier," *JURIKOM (Jurnal Riset Komputer)*, vol. 9, no. 6, p. 2158, 2022.
[14] S. Paembonan and H. Abduh, "Penerapan Metode Silhouette Coeficient Untuk Evaluasi Clutering Obat," *PENA TEKNIK: Jurnal Ilmiah Ilmu-Ilmu Teknik*, vol. 6, no. 2, pp. 48–54, 2021.
[15] S. P. Putra and D. A. Anggoro, "Analisis Clustering Global Living Cost Berdasarkan Socioeconomic Status Menggunakan Algoritma DBSCAN," *Jurnal Media Informatika Budidarma*, vol. 8, no. 2, p. 820, Nov. 2024.