

Analisis Sentimen Diseminasi Produk Iklim Menggunakan Metode Recurrent Neural Network (RNN) dalam Klasifikasi dan Density-Based Spatial Clustering of Applications with Noise (DBSCAN) untuk Klasterisasi

Noris Mestika, Aji Supriyanto*

Fakultas Teknologi Informasi dan Industri, Program Studi Teknologi Informasi, Universitas Stikubank, Semarang, Indonesia

Email: ¹norismestika0023@mhs.unisbank.ac.id, ^{2*}ajisup@edu.unisbank.ac.id

Email Penulis Korespondensi: ajisup@edu.unisbank.ac.id

Submitted: 24/12/2025; Accepted: 05/03/2026; Published: 06/03/2026

Abstrak—Perubahan iklim dan cuaca ekstrem berdampak signifikan pada berbagai sektor kehidupan, sehingga diseminasi informasi iklim yang akurat dan responsif menjadi sangat penting. Sentimen masyarakat bisa menjadi salah satu indikator penilaian publik terhadap diseminasi iklim. Implikasi dari analisis sentimen dapat dijadikan sebagai strategi komunikasi dari penyedia informasi terhadap publik. Studi ini bertujuan untuk menganalisis sentimen masyarakat terhadap diseminasi produk iklim yang dilakukan oleh Stasiun Klimatologi Jawa Tengah melalui media sosial Instagram dengan akun @bmkngjateng dan X dengan akun @bmkng_semarang. Analisis dilakukan dengan kerangka kerja hybrid yang mengintegrasikan metode *Recurrent Neural Network* (RNN) untuk klasifikasi sentimen dan *Density-Based Spatial Clustering of Applications with Noise* (DBSCAN) untuk pengelompokan topik diskusi serta identifikasi outlier. Data sebanyak 12.847 komentar dikumpulkan melalui web scraping pada periode 2020–2024. Hasil klasifikasi RNN menunjukkan dominasi respons netral (76,41%), diikuti sentimen negatif (13,15%) dan positif (10,44%). Model mencapai performa sangat baik dengan akurasi 96% dan *F1-Score* rata-rata 0,96. DBSCAN berhasil mengidentifikasi 82 kluster topik dan mengklasifikasikan 74,5% data sebagai *noise*, yang sebagian besar berupa interaksi non-topik atau spam. Validitas struktur kluster dikonfirmasi dengan nilai *Silhouette Coefficient* 0,3675, *Davies-Bouldin Index* 0,504, dan *Calinski-Harabasz Index* 191,395 yang menandakan bahwa kluster topik yang terbentuk memiliki struktur yang kokoh dan terpisah secara tegas satu sama lain. Analisis integratif mengungkap bahwa sentimen negatif secara konsisten terfokus pada kluster isu spesifik seperti banjir dan panas ekstrem, berbeda dengan sentimen positif yang lebih menyebar pada apresiasi layanan. Temuan ini merekomendasikan perlunya penerapan sistem filtrasi otomatis untuk memprioritaskan penanganan keluhan pada kluster topik kritis guna efisiensi strategi komunikasi publik.

Kata Kunci: Analisis Sentimen; Diseminasi Iklim; RNN; DBSCAN; Media Sosial

Abstract—Climate change and extreme weather events have a significant impact on various sectors of life, making the accurate and timely dissemination of climate information crucial. Public sentiment can be an indicator of public assessment of climate dissemination. The implications of the sentiment analysis itself can be used as a communication strategy from information providers to the public. This study aims to analyze public sentiment toward the dissemination of climate products by the Central Java Climatology Station through social media platforms Instagram (@bmkngjateng) and X (@bmkng_semarang). The analysis was conducted using a hybrid framework integrating the *Recurrent Neural Network* (RNN) method for sentiment classification and the *Density-Based Spatial Clustering of Applications with Noise* (DBSCAN) for topic clustering and outlier identification. A total of 12,847 comments were collected via web scraping from 2020 to 2024. The RNN classification results revealed a dominance of neutral responses (76.41%), followed by negative (13.15%) and positive (10.44%) sentiments. The model achieved high performance with 96% accuracy and a weighted average *F1-Score* of 0.96. DBSCAN successfully identified 82 topic clusters and classified 74.5% of the data as noise, largely consisting of non-topical interactions or spam. The validity of the cluster structure was confirmed by a *Silhouette Coefficient* of 0.3675, a *Davies-Bouldin Index* of 0.504, and a *Calinski-Harabasz Index* of 191.395, indicating that the formed topic clusters possess a robust structure and are distinctly separated from one another. Integrative analysis revealed that negative sentiments were consistently focused on specific issue clusters such as floods and extreme heat, whereas positive sentiments were dispersed across service appreciation. These findings suggest the necessity of implementing an automatic filtration system to prioritize handling complaints within critical topic clusters for efficient public communication strategies.

Keyword: Sentiment Analysis; Climate Dissemination; RNN; DBSCAN; Social Media

1. PENDAHULUAN

Perubahan iklim dan kejadian cuaca ekstrem telah memberikan dampak signifikan terhadap berbagai sektor vital masyarakat, seperti pertanian, transportasi, dan kesehatan. Dalam konteks ini, ketersediaan informasi iklim yang akurat dan tepat waktu menjadi krusial bagi pengambilan keputusan oleh masyarakat, pelaku usaha, dan pembuat kebijakan [1]. Stasiun Klimatologi Jawa Tengah, sebagai unit pelaksana teknis Badan Meteorologi, Klimatologi, dan Geofisika (BMKG), memegang peran strategis dalam menyediakan informasi tersebut melalui produk-produk seperti prakiraan musim dan peringatan dini bencana hidrometeorologi [2]. Di era digital, media sosial, telah menjadi saluran utama diseminasi informasi sekaligus ruang interaksi dua arah antara lembaga dan masyarakat. Kolom komentar pada salah satu platform media sosial, contohnya Instagram menyimpan kekayaan data berupa tanggapan, pertanyaan, dan kritik publik yang dapat dianalisis untuk memahami persepsi dan efektivitas komunikasi iklim [3] [4]. Analisis sentimen terhadap komentar-komentar ini merupakan pendekatan penting untuk mengekstraksi opini publik secara sistematis, meskipun menghadapi tantangan kompleksitas linguistik teks media sosial yang singkat, informal, dan seringkali ambigu [5].

Berdasarkan latar belakang tersebut, penelitian ini mengidentifikasi empat rumusan masalah utama: (1) Bagaimana persepsi atau sentimen masyarakat terhadap diseminasi produk iklim oleh Stasiun Klimatologi Jawa Tengah di media sosial, khususnya Instagram dan X? (2) Seberapa efektif metode Recurrent Neural Network (RNN) dalam mengklasifikasikan sentimen komentar tersebut? (3) Bagaimana DBSCAN dapat mengelompokkan pola komentar berdasarkan kemiripan konten? (4) Apa implikasi hasil analisis bagi strategi komunikasi iklim lembaga?.

Gap penelitian yang diisi oleh studi ini terletak pada integrasi metode klasifikasi mendalam (*deep learning*) dan *clustering* berbasis kerapatan dalam konteks analisis respons publik terhadap informasi iklim dari lembaga pemerintah. Penelitian terdahulu menunjukkan bahwa analisis sentimen di media sosial untuk layanan publik telah dilakukan, namun masih terbatas dalam beberapa aspek. Berdasarkan penelitian Mustakim dkk. pada tahun 2019, algoritma DBSCAN telah terbukti efektif dalam melakukan clustering teks pada topik-topik tren di media sosial Twitter [6]. Sementara itu, berdasarkan penelitian Kurniasari dan Setyanto pada tahun 2020, penggunaan RNN sangat krusial untuk menangkap konteks sekuensial dalam analisis sentimen [7].

Penelitian oleh Firmansyah, dkk pada tahun 2020 berhasil mengklasifikasi berbagai jenis kalimat ilmiah dengan membuktikan efektivitas optimasi SGD pada model RNN berbasis Word2Vec yang mampu mencapai akurasi signifikan sebesar 77,48% [8]. Studi Bodapati dkk. (2021) menunjukkan bahwa meski CNN lebih efisien secara komputasi, metode RNN-LSTM lebih unggul dalam meningkatkan hasil analisis sentimen karena kemampuannya menangani konteks dan dependensi jangka panjang (*long-term dependencies*) pada teks [9].

Lebih spesifik, dalam domain klimatologi dan informasi iklim, penelitian oleh Rasmila, dkk di tahun 2025 cenderung fokus pada prediksi cuaca atau analisis data sensor [10], sedangkan analisis terhadap interaksi dan persepsi masyarakat di media sosial tentang cuaca dan iklim masih jarang dieksplorasi. Penelitian sebelumnya banyak berfokus pada platform seperti Twitter (X) atau ulasan e-commerce dengan pendekatan tunggal, baik hanya klasifikasi atau hanya klustering saja. Penelitian yang dilakukan Ramdhani & Al-Fadillah (2021) membuktikan keunggulan metode *Deep Learning* yang mencapai akurasi sempurna sebesar 100% dalam mengklasifikasikan sentimen masyarakat di Twitter terhadap kebijakan pembelajaran daring, di mana mayoritas publik memberikan respons positif sebesar 73% [11]. Untuk teks media sosial, Laksito dkk. (2024) menunjukkan keberhasilan RNN dalam mengklasifikasikan tweet terkait COVID-19 [12]. Kajian yang secara simultan mengintegrasikan klasifikasi sentimen berbasis RNN (untuk menangkap konteks dan dependensi sequential dalam teks) dengan pengelompokan topik otomatis menggunakan DBSCAN (untuk mengidentifikasi pola diskusi yang *robust* terhadap *noise*) pada komentar Instagram khusus untuk konten layanan publik klimatologi masih sangat terbatas.

Kebaruan dari penelitian ini adalah kerangka kerja *hybrid* yang tidak hanya mengukur valensi emosional (positif/negatif/netral), tetapi juga memetakan tema spesifik yang memicu sentimen tersebut, serta secara eksplisit mengidentifikasi dan mengkarakterisasi *noise* dalam diskusi publik. Pendekatan ini diharapkan dapat memberikan pemahaman yang lebih berlapis dan operasional bagi lembaga dalam merancang strategi komunikasi yang lebih efektif dan responsif. Penelitian terdahulu seperti yang dilakukan oleh Dang, dkk. (2020) mengusulkan pendekatan hibrida yang mengombinasikan arsitektur *deep learning* (DNN, CNN, dan RNN) dengan teknik *word embedding* untuk meningkatkan akurasi klasifikasi sentimen sekaligus mengoptimalkan efisiensi biaya komputasi. [13]. Pada tahun 2021, Dang, dkk. menunjukkan bahwa pendekatan hibrida yang mengintegrasikan model konvensional (CNN, LSTM, atau SVM) dengan arsitektur BERT untuk ekstraksi fitur mampu meminimalisir kesalahan pada data kompleks dan mencapai akurasi hingga 90% [14].

Secara spesifik, penelitian ini bertujuan untuk: (1) Menganalisis sentimen masyarakat terhadap diseminasi produk iklim di Instagram dan X; (2) Mengembangkan dan menguji model RNN untuk klasifikasi sentimen komentar; (3) Menerapkan algoritma DBSCAN untuk mengelompokkan komentar berdasarkan kedekatan konten guna mengidentifikasi topik dominan; serta (4) Memberikan rekomendasi strategis berbasis data untuk meningkatkan efektivitas komunikasi publik Stasiun Klimatologi Jawa Tengah. Penelitian klasifikasi data Twitter BMKG nasional pernah dikaji oleh Darwis, dkk pada tahun 2021 dengan menggunakan Naïve Bayes namun belum menggunakan pendekatan *deep learning* yang kompleks [15].

Bidang analisis sentimen telah banyak dikembangkan dengan berbagai pendekatan *machine learning* dan *deep learning*. Pada konteks klasifikasi, *Recurrent Neural Network* (RNN) dan variannya seperti LSTM telah terbukti efektif menangani data sekuensial seperti teks. Penanganan data sekuensial menggunakan metode analisis yang tepat sangat menentukan akurasi peramalan. Penelitian oleh Kurniasari dan Setyanto pada tahun 2020 mencapai akurasi 91.9% untuk analisis sentimen bahasa Indonesia pada ulasan Traveloka menggunakan RNN dan Word2Vec [7]. Namun, penelitian tersebut umumnya berfokus pada domain komersial. Sebagai pengembangan RNN, Long Short-Term Memory (LSTM) unggul dalam menangani masalah *long-term dependency* pada data sekuensial. Yarbo dkk. Pada tahun 2024 menerapkan metode ini untuk menganalisis sentimen ulasan Traveloka di Google Play Store. Temuan menunjukkan bahwa kata “disappointed” menjadi istilah paling dominan, diikuti oleh “easy”, “buy”, “fast”, dan “good”. Dengan pembagian data 80:20, model LSTM tersebut menghasilkan akurasi sebesar 83%, yang mengindikasikan bahwa performa model tetap dipengaruhi oleh prapemrosesan dan karakteristik dataset ulasan pengguna [16].

Di sisi lain, untuk pengelompokan data teks tanpa label, algoritma DBSCAN dikenal keunggulannya dalam membentuk kluster berbentuk tidak teratur dan mendeteksi *noise*. Penelitian Mustakim dkk. pada tahun 2021, algoritma DBSCAN unggul dari K-Medoids dalam mengidentifikasi opini publik terkait bencana alam dan sangat toleran terhadap outlier [17]. Beberapa kajian literatur menganalisis sentimen dengan DBSCAN menggunakan

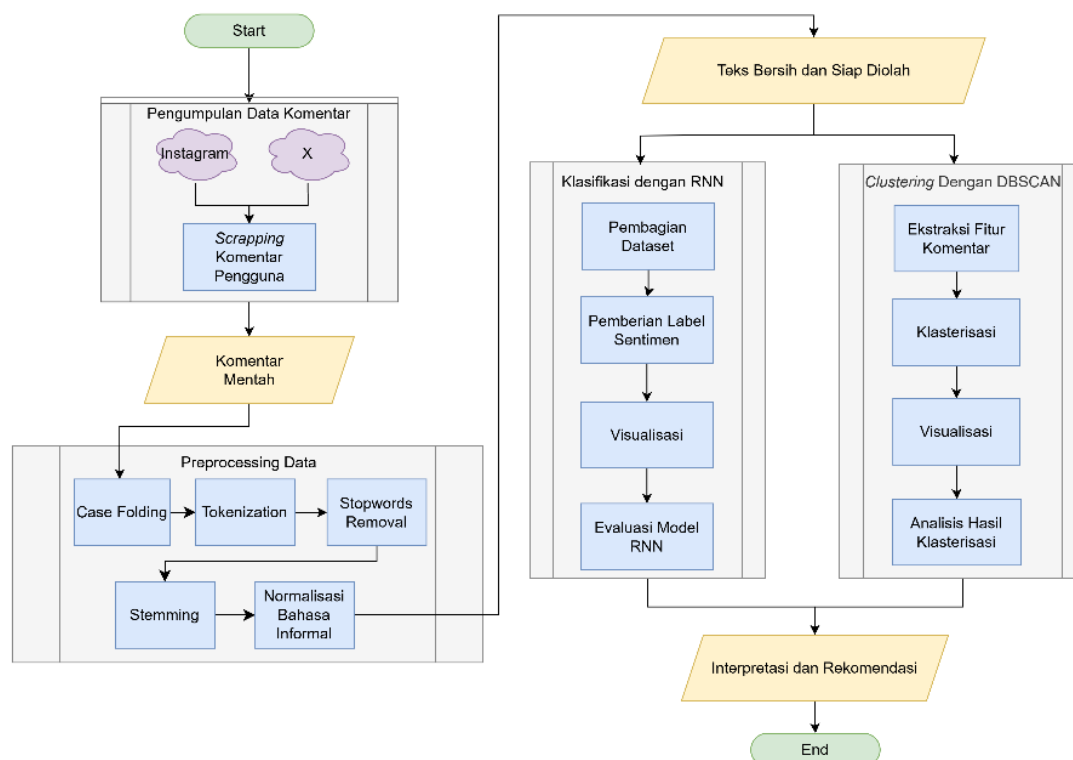
pembobotan TF-IDF. Penelitian oleh Andriyani dan Puspitarani pada tahun 2022 membandingkan algoritma K-Means dan DBSCAN dalam mengelompokkan ulasan produk Cetaphil menggunakan pembobotan TF-IDF untuk mengonversi frekuensi kata menjadi nilai bobot term yang lebih representatif. Hasilnya menunjukkan bahwa DBSCAN lebih unggul dengan akurasi 99,80% dibandingkan K-Means (99,50%), membuktikan efektivitas metode berbasis kepadatan dalam memproses representasi teks tersebut [18]. Penelitian Hartono dan Fibriani pada tahun 2025 menerapkan reduksi t-SNE untuk memetakan vektor TF-IDF berdimensi tinggi ke dalam ruang 2D [19]. Hasilnya yaitu titik yang memiliki pola kemiripan tinggi akan terkumpul berdekatan, sedangkan titik noise atau outlier akan tampak terpisah.

Beberapa studi mulai menggabungkan pendekatan klasifikasi dan klustering. Penelitian oleh Tsamratul 'ain dan Supatmi pada tahun 2023 menggunakan DBSCAN-SVM untuk analisis keluhan pengguna aplikasi PeduliLindungi [20]. Terakhir, merujuk pada tinjauan bibliometrik oleh Ferdiansyah dkk. (2024), kombinasi K-Means dan DBSCAN terus berkembang sebagai standar industri dalam pengelompokan data yang kompleks [21]. Namun, integrasi RNN (untuk klasifikasi mendalam yang kontekstual) dan DBSCAN (untuk pengelompokan topik yang *robust* terhadap *noise*) khusus pada domain informasi iklim dan media sosial Instagram masih terbatas. Penelitian ini berupaya menjembatani gap tersebut dengan memadukan kedua metode untuk ekstraksi wawasan yang lebih komprehensif dari interaksi digital masyarakat terhadap informasi publik yang kritis.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Penelitian ini mengadopsi paradigma penelitian kuantitatif dengan pendekatan eksploratif-analitik. Tahapan utama penelitian dirancang secara sistematis yang diilustrasikan pada alur diagram penelitian pada Gambar 1. Penelitian ini mengintegrasikan dua metode inti, yaitu *Recurrent Neural Network* (RNN) untuk klasifikasi sentimen dan *Density-Based Spatial Clustering of Applications with Noise* (DBSCAN) untuk pengelompokan topik guna mengekstraksi wawasan yang berlipis dari umpan balik publik digital.



Gambar 1. Alur Diagram penelitian

2.2 Pengumpulan Data

Data primer penelitian ini berupa komentar publik yang diambil akun media sosial Stasiun Klimatologi Jawa Tengah yaitu Instagram dengan akun @bmkjg Jateng dan X dengan akun @bmkjg Semarang. Periode pengambilan data mencakup lima tahun, dari Januari 2020 hingga Desember 2024. Teknik pengumpulan data dilakukan melalui web scraping menggunakan ekstensi IGCommentsExport pada Google Chrome pada platform Instagram, yang memungkinkan ekstraksi komentar beserta metadata terkait. Data mentah yang terkumpul kemudian disimpan dalam format CSV untuk diproses lebih lanjut.

2.3 Pengolahan Data (*Preprocessing*)

Sebelum dianalisis, data teks mentah menjalani serangkaian tahap preprocessing untuk membersihkan dan menstandarisasi format. Tahapan ini penting untuk meningkatkan kualitas data masukan (input) bagi model komputasional. Proses diawali dengan Case Folding yang mengubah semua karakter menjadi huruf kecil untuk keseragaman, kemudian dilanjutkan dengan Cleansing untuk menghapus noise seperti tanda baca, simbol, URL, mention (@), hashtag (#), dan karakter non-alfanumerik. Setelah itu, dilakukan Tokenization untuk memecah kalimat menjadi unit-unit kata individual (tokens) dan Stopword Removal untuk menghilangkan kata umum yang tidak informatif seperti “yang”, “di”, dan “dan”. Tahap berikutnya adalah Stemming yang mengembalikan kata ke bentuk dasarnya menggunakan algoritma Nazief & Adriani yang diimplementasikan melalui pustaka Python seperti NLTK atau cistem, di mana algoritma ini dikenal akurat untuk stemming bahasa Indonesia seperti contohnya kata “kebersamaan” menjadi “sama” dan “bergerak” menjadi “gerak”. Terakhir, dilakukan Normalisasi Bahasa Informal untuk mengganti kata slang, singkatan, dan typo dengan bentuk bakunya seperti contoh “bgt” menjadi “banget” dan “gak” menjadi “tidak” menggunakan kamus buatan sendiri.

2.4 Analisis Data

2.4.1 Klasifikasi Sentimen Dengan RNN-LSTM

Model klasifikasi sentimen dibangun menggunakan arsitektur RNN dengan lapisan LSTM (*Long Short-Term Memory*). LSTM dipilih untuk kemampuannya menangani *long-range dependencies* dalam data sekuensial seperti teks. Alur pemodelan diawali dengan Pelabelan Data di mana data yang telah diproses diberi label sentimen (Positif, Netral, Negatif) secara semi-otomatis menggunakan kamus leksikon sederhana dan penanganan negasi yang menghasilkan distribusi 38,1% positif, 32,1% netral, dan 29,8% negatif, serta dilakukan validasi manual terhadap 500 komentar (sekitar 4% dari total data) untuk memastikan kualitas label dalam menangkap konteks khusus iklim, nuansa bahasa media sosial, dan potensi sarkasme sebelum menghasilkan ground truth yang akurat melalui konsensus. Selanjutnya dilakukan Vektorisasi dan Pembagian Data dengan mengubah teks menjadi vektor numerik menggunakan Tokenizer dari Keras dan membagi dataset menggunakan rasio 80:20 untuk data latih dan data uji. Arsitektur model yang digunakan terdiri dari lapisan *Embedding* dengan dimensi 100, lapisan *SpatialDropout1D* dengan rate 0.2, lapisan LSTM sebanyak 64 unit, serta lapisan Dense akhir dengan aktivasi softmax untuk klasifikasi tiga kelas. Proses ini diakhiri dengan Pelatihan dan Evaluasi di mana model dikompilasi menggunakan *loss function categorical_crossentropy* dan *optimizer Adam*, serta diukur performanya menggunakan metrik akurasi, *precision*, *recall*, F1-score, dan *confusion matrix*.

2.4.2 Klusterisasi Dengan DBSCAN

Untuk mengelompokkan komentar berdasarkan kemiripan tematik digunakan algoritma DBSCAN yang mampu menemukan kluster dengan bentuk arbitrer dan secara otomatis mengidentifikasi outlier, yang diawali dengan Ekstraksi Fitur melalui pengubahan teks hasil preprocessing menjadi representasi numerik menggunakan TF-IDF (*Term Frequency-Inverse Document Frequency*) dengan 5000 fitur terbanyak. Langkah selanjutnya adalah Reduksi Dimensi di mana matriks TF-IDF yang berdimensi tinggi direduksi menggunakan PCA (*Principal Component Analysis*) menjadi 100 komponen utama untuk efisiensi dan kinerja DBSCAN. Kemudian dilakukan Penentuan Parameter utama DBSCAN, yaitu *eps* (jarak tetangga) dan *min_samples* (titik minimum dalam satu kluster), yang ditentukan melalui analisis *K-distance plot* dan eksperimen iteratif. Proses ini diakhiri dengan Klusterisasi dan Analisis di mana model DBSCAN diterapkan untuk menghasilkan label kluster, lalu kluster yang terbentuk dianalisis dengan melihat kata kunci dominan pada setiap kelompok menggunakan word cloud dan analisis manual terhadap sampel komentar.

2.5 Integrasi Dan Interpretasi Hasil

Hasil dari klasifikasi RNN (distribusi sentimen) dan klusterisasi DBSCAN (kelompok topik) diintegrasikan secara kualitatif. Analisis dilakukan dengan memetakan sentimen dominan pada setiap kluster topik utama. Integrasi ini menghasilkan wawasan yang lebih mendalam, misalnya “topik banjir didominasi oleh sentimen negatif yang berisi keluhan” atau “topik informasi rutin cuaca didominasi sentimen netral”. Interpretasi akhir dari pola-pola ini menjadi dasar perumusan rekomendasi strategis bagi institusi. Kesimpulan Metodologi: Kerangka metodologi hybrid RNN-DBSCAN yang dijelaskan di atas dirancang untuk memberikan analisis dua lapis: lapisan pertama memahami emosi publik (melalui RNN), dan lapisan kedua memahami pokok pembicaraan publik (melalui DBSCAN). Kombinasi ini diharapkan mampu memberi landasan data yang kuat untuk pengambilan keputusan komunikasi yang lebih efektif dan terukur.

2.6 Algoritma Metode RNN dan DBSCAN

Penelitian ini menggunakan pendekatan kuantitatif dengan desain analitik eksploratif, yang bertujuan untuk mengidentifikasi pola sentimen dan kluster topik dari komentar publik di media sosial. Metode utama yang diterapkan adalah kombinasi dari klasifikasi *supervised learning* menggunakan RNN dan *clustering unsupervised learning*

menggunakan DBSCAN. Pendekatan campuran (*hybrid*) ini dipilih untuk tidak hanya mengkategorikan sentimen (positif, negatif, netral) tetapi juga mengungkap struktur tematik alami dari diskusi publik tanpa prasangka label awal.

2.6.1 Recurrent Neural Network (RNN)

a. Persamaan Umum *Recurrent Neural Network* (RNN)

Untuk setiap langkah waktu t , *hidden state* h_t dihitung dengan:

$$h_t = \sigma(W_{hx}x_t + W_{hh}h_{t-1} + b_h) \quad (1)$$

$$\hat{y}_t = \text{softmax}(W_{hy}h_t + b_y) \quad (2)$$

Dalam representasi matematisnya, variabel-variabel tersebut didefinisikan sebagai h_t yang merupakan hidden state pada waktu t , x_t sebagai vektor input pada waktu t , serta W_{hx} yang merupakan matriks bobot dari input ke hidden layer. Selain itu, terdapat W_{hh} sebagai matriks bobot dari hidden state sebelumnya ke hidden state saat ini, b sebagai bias, dan σ sebagai fungsi aktivasi yang biasanya menggunakan tanh atau ReLU. Seluruh komponen ini bekerja secara bersamaan untuk menghasilkan \hat{y}_t , yaitu output prediksi pada waktu t .

b. Arsitektur LSTM (Varian RNN yang Digunakan)

LSTM memiliki tiga *gate* dan *cell state*:

Forget Gate:

$$f_t = \sigma(W_f \cdot [h_{t-1}, x_t] + b_f) \quad (3)$$

Input Gate:

$$i_t = \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \quad (4)$$

$$\tilde{C}_t = \tanh(W_c \cdot [h_{t-1}, x_t] + b_c) \quad (5)$$

Cell State Update:

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \quad (6)$$

Output Gate:

$$o_t = \sigma(W_o \cdot [h_{t-1}, x_t] + b_o) \quad (7)$$

$$h_t = o_t * \tanh(C_t) \quad (8)$$

Dalam arsitektur LSTM yang digunakan, terdapat beberapa komponen utama yang mengatur aliran informasi, yaitu C_t yang berfungsi sebagai cell state atau memori jangka panjang untuk menyimpan informasi relevan dalam waktu lama. Komponen pengatur aliran tersebut terdiri dari f_t, i_t, o_t yang masing-masing merupakan forget gate, input gate, dan output gate dengan nilai aktivasi berada di rentang antara 0 dan 1. Selain itu, terdapat \tilde{C}_t yang berperan sebagai kandidat nilai baru untuk memperbarui cell state, serta simbol asterisk (*) yang melambangkan operasi perkalian elemen atau Hadamard product dalam perhitungan matematisnya.

c. Algoritma Evaluasi Model Klasifikasi (RNN)

Tabel 1. Algoritma Confusion Matrix

	Prediksi Positif	Prediksi Negatif
Aktual Positif	True Positive (TP)	False Negative (FN)
Aktual Negatif	False Positive (FP)	True Negative (TN)

d. Metrik Evaluasi Klasifikasi

1. Akurasi (Accuracy):

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (9)$$

2. Presisi (Precision):

$$\text{Precision} = \frac{TP}{TP+FP} \quad (10)$$

3. Recall (Sensitivity):

$$\text{Recall} = \frac{TP}{TP+FN} \quad (11)$$

4. F1-Score:

$$F1 = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \quad (12)$$

2.6.2 Density-Based Spatial Clustering of Applications with Noise (DBSCAN)

a. Konsep Dasar DBSCAN (Clustering)

Definisi ε -neighborhood:

$$N_\varepsilon(p) = \{q \in D \mid \text{dist}(p, q) \leq \varepsilon\} \quad (13)$$

Pada algoritma DBSCAN, penentuan kluster didasarkan pada hubungan antara titik-titik data, di mana p didefinisikan sebagai titik pusat dan q sebagai titik lain yang berada di dalam dataset. Jangkauan di sekitar titik pusat ditentukan oleh ε yang merupakan radius maksimum atau dikenal sebagai parameter eps. Untuk menentukan apakah suatu titik berada dalam jangkauan tersebut, digunakan fungsi $\text{dist}(p, q)$: yang menghitung jarak Euclidean atau metrik jarak lainnya untuk mengukur kedekatan antar titik tersebut.

b. Klasifikasi Titik Dalam DBSCAN

1. Core Point:

$$|N_\varepsilon(p)| \geq \text{MinPts} \quad (14)$$

(Memiliki setidaknya MinPts tetangga dalam radius ε)

2. Border Point:

$$|N_\varepsilon(p)| < \text{MinPts} \text{ dan } p \in N_\varepsilon(q) \text{ untuk suatu core point } q \quad (15)$$

3. Noise/Outlier:

$$|N_\varepsilon(p)| < \text{MinPts} \text{ dan } p \notin N_\varepsilon(q) \text{ untuk semua core points } q \quad (16)$$

c. Algoritma Evaluasi Klustering DBSCAN

1. Silhouette Coefficient

Untuk setiap titik data i :

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \quad (17)$$

Kualitas hasil klusterisasi dapat diukur menggunakan nilai Silhouette, di mana $a(i)$ merepresentasikan rata-rata jarak titik i ke semua titik lain dalam kluster yang sama, sedangkan $b(i)$ adalah rata-rata jarak minimum dari titik i ke semua titik di kluster lain yang paling dekat. Rentang nilai untuk metrik ini berada di antara $-1 \leq s(i) \leq 1$. Jika nilai $s(i)$ semakin mendekati 1, hal tersebut menunjukkan bahwa kualitas kluster yang dihasilkan sudah semakin baik dan terpisah secara optimal.

2. Davies-Bouldin Index (Dbi)

$$DB = \frac{1}{k} \sum_{i=1}^k \max_{j \neq i} \left(\frac{\sigma_i + \sigma_j}{d(c_i, c_j)} \right) \quad (18)$$

Metrik evaluasi lainnya melibatkan k sebagai jumlah total kluster yang terbentuk, dengan σ_i yang menyatakan rata-rata jarak semua titik di kluster i menuju ke titik pusatnya atau centroid c_i . Selain itu, digunakan juga variabel $d(c_i, c_j)$ untuk mengukur jarak antara centroid kluster i dan kluster j . Dalam pengujian ini, nilai hasil perhitungan yang lebih rendah menunjukkan bahwa kluster yang terbentuk memiliki karakteristik yang lebih kompak dan terpisah dengan baik satu sama lain.

3. Calinski-Harabasz Index (*Variance Ratio Criterion*)

$$CH = \frac{\text{Tr}(B_k)}{\text{Tr}(W_k)} \times \frac{N-k}{k-1} \quad (19)$$

Evaluasi kepadatan kluster juga menggunakan metrik yang melibatkan $\text{Tr}(B_k)$ sebagai *trace* matriks scatter antar-kluster dan $\text{Tr}(W_k)$ sebagai *trace* matriks scatter dalam-kluster. Perhitungan ini juga memperhitungkan N yang merupakan jumlah total sampel dan k sebagai jumlah kluster yang tersedia. Berbeda dengan metrik sebelumnya, pada indeks ini nilai yang lebih tinggi menunjukkan bahwa kluster yang dihasilkan cenderung lebih padat dan memiliki pemisahan yang sangat baik antar kelompoknya.

d. Rumus Pendukung (*Preprocessing & Feature Extraction*)

1. Tf-Idf (*Term Frequency-Inverse Document Frequency*)

$$\text{TF-IDF}(t, d) = \text{TF}(t, d) \times \text{IDF}(t) \quad (20)$$

$$\text{TF}(t, d) = \frac{\text{freq}(t, d)}{\text{total terms in } d} \quad (21)$$

$$\text{IDF}(t) = \log \left(\frac{N}{1 + \text{df}(t)} \right) + 1 \quad (\text{smooth IDF}) \quad (22)$$

Dalam proses ekstraksi fitur teks, digunakan beberapa variabel utama yaitu t yang mewakili term atau kata tertentu dan d yang mewakili dokumen atau komentar tempat kata tersebut muncul. Untuk menghitung bobot kepentingannya, digunakan N sebagai jumlah total seluruh dokumen dalam koleksi data serta $\text{df}(t)$ yang menunjukkan jumlah dokumen spesifik yang mengandung term t tersebut.

2. Principal Component Analysis (PCA)

$$Cv = \lambda v \tag{23}$$

Untuk mengatasi masalah curse of dimensionality pada vektor TF-IDF, sebelum proses klusterisasi diterapkan *Principal Component Analysis* (PCA). PCA bekerja dengan mereduksi dimensi data melalui dekomposisi nilai eigen dari matriks kovarians, di mana C adalah matriks kovarians dari fitur input, λ adalah nilai eigen yang merepresentasikan besaran varians, dan v adalah vektor eigen yang menentukan arah ortogonal baru. Transformasi ini menghasilkan set fitur baru yang lebih ringkas namun tetap mempertahankan informasi varians terpenting dari data asli, sehingga optimal untuk dihitung menggunakan metrik jarak Euclidean pada tahap klusterisasi selanjutnya.

3. Euclidean Distance

$$d(p, q) = \sqrt{\sum_{i=1}^n (q_i - p_i)^2} \tag{24}$$

Dalam penentuan densitas lokal dan pembentukan neighborhood pada algoritma DBSCAN, penelitian ini menerapkan metrik Euclidean Distance untuk mengukur kedekatan geometris antar titik data pada ruang fitur yang telah direduksi. Jarak antara dua vektor data, p dan q dihitung dimana $d(p, q)$ merepresentasikan nilai jarak Euclidean yang dihasilkan, n mengacu pada jumlah dimensi fitur (komponen utama hasil PCA), sedangkan p_i dan q_i adalah nilai skalar fitur ke- i pada masing-masing titik data yang dibandingkan. Penggunaan metrik ini didasarkan pada asumsi bahwa setelah proses reduksi dimensi PCA, fitur-fitur data menjadi ortogonal dan padat, sehingga jarak linear menjadi representasi kemiripan yang valid dan efisien secara komputasi

3. HASIL DAN PEMBAHASAN

Data yang digunakan dalam penelitian ini dikumpulkan dari komentar pada akun Instagram (@bmkgtateng) dan X (@bmkg_semarang) Stasiun Klimatologi Jawa Tengah periode 2020-2024. Sebanyak 12.847 komentar berhasil diambil dan digabungkan. Setelah melalui tahap *preprocessing* yang terdiri dari *case folding*, *cleansing*, *tokenization*, *stopword removal*, *stemming*, dan normalisasi bahasa informal, diperoleh total data bersih yang digunakan sebanyak 9.164 komentar. Hasil pelabelan awal menunjukkan distribusi sentimen yang sangat tidak seimbang (*imbalanced dataset*), di mana mayoritas data merupakan kelas Netral dengan jumlah 7.022 komentar. Sementara itu, sentimen Negatif berjumlah 1.202 komentar dan sentimen Positif dengan 904 komentar merupakan kelas minoritas. Ketimpangan ini menunjukkan bahwa sebagian besar respons publik cenderung bersifat informatif atau objektif, bukan emosional.

3.1 Hasil Klasifikasi Sentimen Dengan RNN – LSTM

Sebelum dimasukkan ke dalam arsitektur model, data teks bersih dikonversi menjadi representasi numerik melalui proses integer encoding. Selanjutnya, teknik sequence padding diterapkan untuk menyeragamkan panjang input setiap komentar agar sesuai dengan dimensi masukan yang dibutuhkan model. Model RNN dengan arsitektur LSTM berhasil dikembangkan dengan konfigurasi yaitu, lapisan embedding (dimensi=100), *SpatialDropout1D* (rate=0.2), lapisan LSTM (64 unit), dan lapisan *dense* akhir dengan aktivasi *softmax*. Pelatihan dilakukan selama 5 epoch dengan batch size 64 menggunakan skema pembagian data 80:20, di mana 20% data (sebanyak 2.001 komentar) dialokasikan khusus untuk pengujian.

Tabel 2. Laporan Hasil Pelatihan Model RNN

Epoch	Training Accuracy	Training Loss	Validation Accuracy	Validation Loss
1	0.7552	0.7449	0.8363	0.4487
2	0.8862	0.3277	0.9345	0.2149
3	0.9637	0.1191	0.9553	0.1516
4	0.9838	0.0624	0.9580	0.1538
5	0.9867	0.0485	0.9634	0.1414

Hasil pelatihan model terlihat pada Tabel 2, dimana model mencapai akurasi validasi tertinggi sebesar 96,34% pada epoch ke-5. Selisih antara Training Accuracy (98,67%) dan Validation Accuracy (96,34%) yang relatif kecil (sekitar 2,3%) mengindikasikan bahwa model memiliki kemampuan generalisasi yang baik (*Good Fit*). Hal ini menunjukkan bahwa strategi regularisasi (*Dropout*) yang diterapkan berhasil meminimalkan terjadinya overfitting, meskipun data yang digunakan tidak seimbang (*imbalanced dataset*).

Tabel 3. Laporan Hasil Klasifikasi Model RNN Data Uji

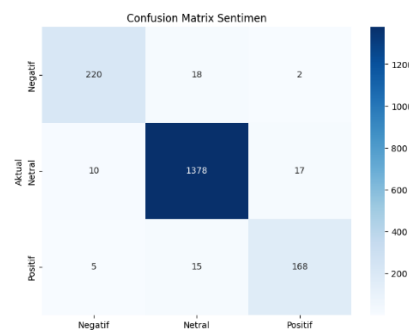
Kelas	Precision	Recall	F1 - Score	Support
Negatif	0,94	0,92	0,93	240
Netral	0,98	0,98	0,98	1405

Kelas	Precision	Recall	F1 - Score	Support
Positif	0,90	0,89	0,90	188
Akurasi	-	-	0,96	1833
Macro Avg	0,94	0,93	0,94	1833
Weighted Avg	0,96	0,96	0,98	1833

Performa model memperoleh hasil yang sangat memuaskan meskipun dihadapkan pada ketimpangan distribusi data (*imbalanced dataset*). Berdasarkan evaluasi pada data uji, model mencapai akurasi keseluruhan sebesar 96%. Mengacu pada laporan hasil klarifikasi model pada Tabel 3 dan *confusion matrix* pada Gambar 2, model menunjukkan performa dominan pada kelas Netral (F1-Score 0.98) yang merupakan kelas mayoritas. Namun, hal yang paling menonjol adalah kemampuan model dalam menangani kelas minoritas yaitu sentimen Positif dan Negatif. Model berhasil mengenali sentimen Negatif dengan sangat baik dengan Recall 0.92 dan Precision 0.94, serta mengenali sentimen Positif dengan cukup baik dengan Recall 0.89 dan Precision 0.90.

Berdasarkan Confusion Matrix yang dihasilkan, model menunjukkan performa klasifikasi yang solid dengan pola kesalahan yang sangat minim dan terpola. Visualisasi Confusion Matrix memperlihatkan diagonal utama yang berwarna biru sangat pekat, terutama pada label Netral (1.378 data terprediksi benar). Interpretasi visual ini mengonfirmasi stabilitas model.

Kesalahan yang terjadi (kotak biru muda di luar diagonal) jumlahnya sangat minim dan bersifat minor. Contoh, ada beberapa komentar Negatif yang terprediksi sebagai Netral, atau sebaliknya. Hal ini dikarenakan adanya ambiguitas Bahasa dimana terkadang batas antara “mengeluh panas” (Negatif) dan “melaporkan suhu tinggi” (Netral) menjadi sangat tipis. Model menunjukkan keandalan tinggi dengan hampir tidak melakukan kesalahan fatal. Sangat sedikit kasus di mana komentar Positif diprediksi sebagai Negatif, begitu pun sebaliknya. Ini berarti model sukses menangkang polaritas sentimen yang bertolak belakang.

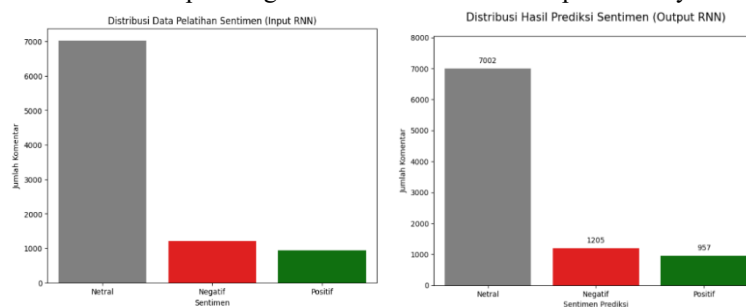


Gambar 2. Confusion Matrix hasil klasifikasi sentimen

3.2 Analisis Distribusi Sentimen dan Hasil RNN

Visualisasi distribusi sentimen pada tahap pelabelan awal (ground truth) memperlihatkan dominasi kelas Netral sebesar 76,63%. Dominasi ini merefleksikan karakteristik interaksi pada akun Stasiun Klimatologi Jawa Tengah yang cenderung bersifat informatif-saintifik dan objektif. Kelas minoritas terdiri dari sentimen Negatif (13,12%) dan Positif (10,26%), yang umumnya muncul sebagai respons terhadap kondisi cuaca ekstrem atau kualitas pelayanan.

Setelah dilakukan pelatihan model dan prediksi ulang terhadap seluruh dataset, distribusi sentimen mengalami penyesuaian minor. Hasil prediksi final RNN menunjukkan proporsi sentimen Netral terkoreksi menjadi 76,41%, sementara sentimen Negatif dan Positif mengalami sedikit peningkatan masing-masing menjadi 13,15% dan 10,44%. Fluktuasi angka yang sangat kecil antara label awal dan hasil prediksi ini mengindikasikan bahwa model RNN memiliki konsistensi tinggi (*robustness*) dan kemampuan generalisasi yang baik, di mana model mampu mempertahankan proporsi kelas asli tanpa mengalami bias ekstrem terhadap kelas mayoritas.



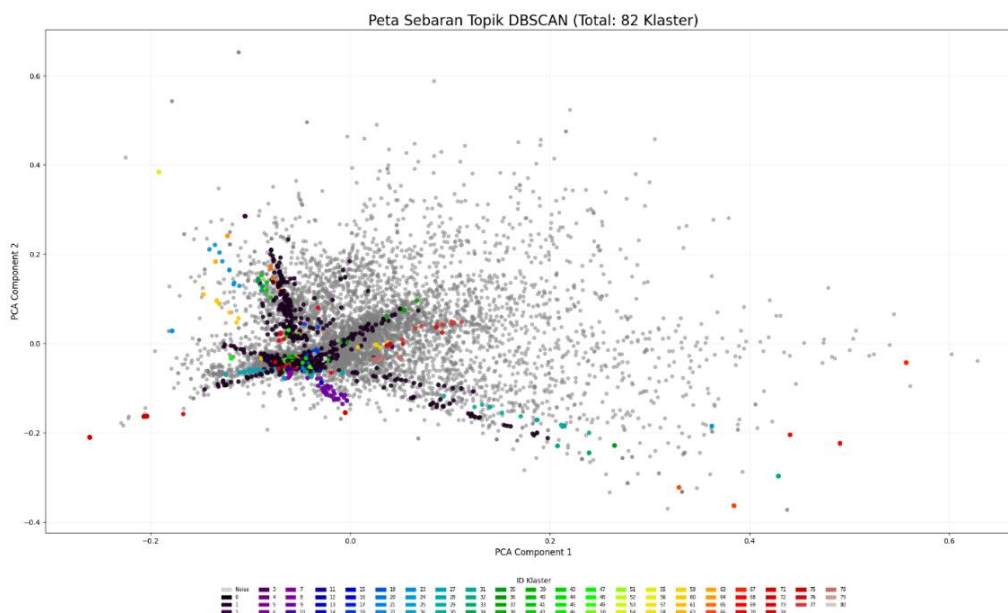
Gambar 3. Diagram batang distribusi sentimen Input (kiri) dan Output (kanan) RNN

Visualisasi Word Cloud pada Gambar 4 menunjukkan perbandingan distribusi kata kunci antara tiga kategori sentimen utama yang dihasilkan dari model klasifikasi RNN, yaitu Netral, Positif, dan Negatif. Pada kelas Netral

ID Klaster	Jumlah Komentar	Persentase	Kata Kunci Dominan	Interpretasi Topik
12	28	0,30%	<i>mas, ra, nek, si, ndi</i>	Percakapan Santai (Interaksi menggunakan bahasa Jawa)
-1 (Noise)	6842	74,50%	-	Outlier/Data tanpa pola gerombolan yang kuat
Lainnya (76 klaster)	673	7,33%	(Bervariasi)	Topik-topik spesifik lainnya

3.4 Visualisasi Hasil Klaster

Visualisasi hasil klasterisasi menggunakan proyeksi dua dimensi *Principal Component Analysis* (PCA) menampilkan sebaran data komentar yang telah dikelompokkan oleh algoritma DBSCAN. Pada scatter plot yang dihasilkan, setiap titik mewakili satu komentar, diwarnai berdasarkan label klaster yang ditetapkan, sementara titik-titik berwarna abu-abu menunjukkan komentar yang dikategorikan sebagai noise atau outlier. Pola pengelompokan yang terbentuk mengindikasikan bahwa komentar dengan topik serupa cenderung berkumpul dalam ruang fitur yang sama, meskipun beberapa klaster menunjukkan tumpang tindih yang merefleksikan kesamaan semantik antar topik pembahasan. Visualisasi ini tidak hanya memvalidasi efektivitas DBSCAN dalam mengidentifikasi pola diskusi, tetapi juga memberikan wawasan intuitif mengenai struktur tematik yang berkembang dalam respons masyarakat terhadap informasi iklim. Berikut adalah hasil *scatter plot* ditunjukkan pada Gambar 5.



Gambar 5. Visualisasi *Scatter Plot* DBSCAN

3.5 Evaluasi DBSCAN

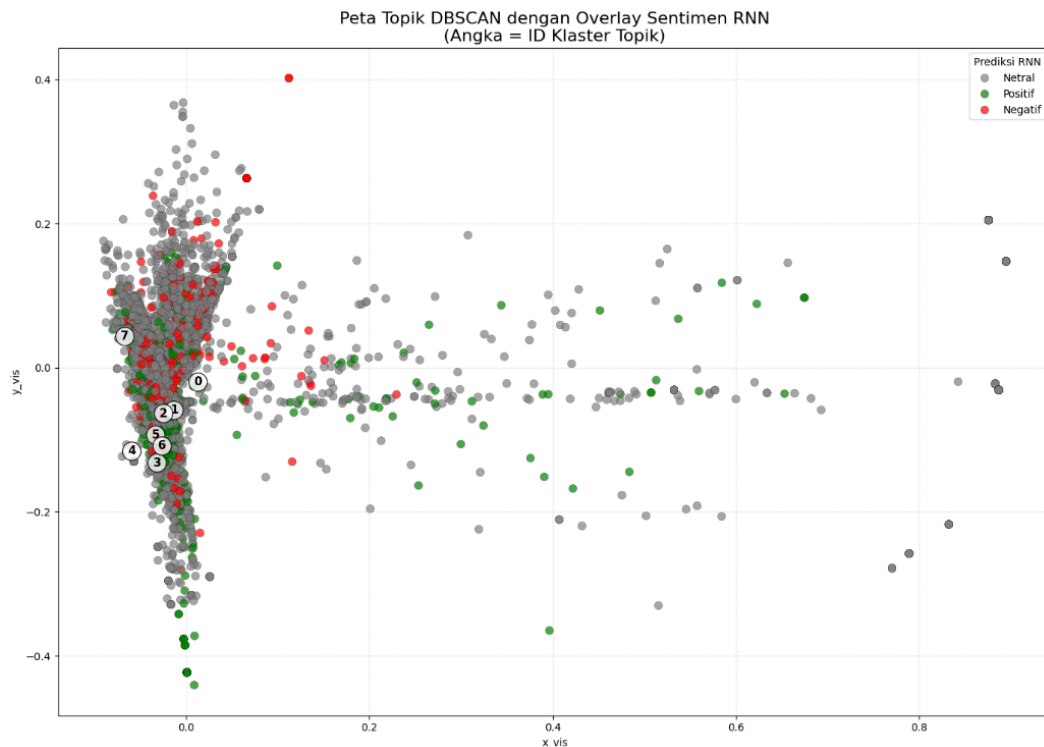
Kualitas pengelompokan 82 klaster yang dihasilkan oleh algoritma DBSCAN dievaluasi menggunakan tiga metrik validasi internal. Hasil pengukuran menunjukkan nilai Silhouette Coefficient sebesar 0,3675. Meskipun nilai ini berada pada kategori moderat, hal tersebut dapat dianggap wajar mengingat karakteristik data teks media sosial yang memiliki irisan semantik alami antar-topik akibat penggunaan kosakata umum yang sering tumpang tindih. Namun demikian, validitas struktur klaster dikonfirmasi lebih kuat oleh nilai Davies-Bouldin Index (DBI) yang rendah, yaitu 0,504 (bernilai di bawah satu). Rendahnya DBI ini menandakan bahwa klaster-klaster yang terbentuk memiliki tingkat kekompakan intra-klaster yang baik serta pemisahan antar-klaster yang cukup tegas. Temuan ini diperkuat oleh nilai Calinski-Harabasz Index (CHI) sebesar 191,3953, yang merefleksikan bahwa varians antar-klaster jauh lebih besar dibandingkan varians di dalam klaster. Dengan demikian, struktur topik yang dihasilkan DBSCAN dapat dikatakan distingtif dan bermakna. Secara keseluruhan, integrasi ketiga metrik tersebut memvalidasi bahwa hasil klasterisasi bukan merupakan artifak acak, melainkan representasi nyata dari pola diskusi publik yang terbentuk secara sistematis dalam data komentar media sosial.

3.6 Integrasi Hasil RNN dan DBSCAN

3.6.1 Pemetaan dan Visualisasi Sentimen pada Struktur Topik

Untuk memvalidasi hubungan antara polaritas emosi publik dan konteks diskusi yang mendasarinya, penelitian ini mengintegrasikan hasil klasifikasi sentimen RNN (positif, netral, negatif) dengan ruang fitur topik yang dibentuk oleh DBSCAN. Dengan parameter $\epsilon = 0,4$ dan $\text{min_samples} = 10$, DBSCAN membentuk delapan klaster utama yang

merepresentasikan topik diskusi publik yang berbeda. Kluster terbesar didominasi topik informatif seputar hujan, cuaca, dan interaksi dengan admin BMKG dengan kecenderungan sentimen netral, sementara kluster lain menunjukkan isu yang lebih spesifik, seperti keluhan cuaca panas ekstrem yang didominasi sentimen negatif serta kluster kecil berisi ungkapan harapan cuaca cerah dengan dominasi sentimen positif. Visualisasi dalam bentuk scatter plot dua dimensi yang ditunjukkan pada Gambar 6 menunjukkan bahwa kluster-kluster besar cenderung netral dan informatif, sedangkan kluster keluhan terkonsentrasi pada area tertentu. Secara keseluruhan, integrasi DBSCAN dan RNN memberikan gambaran yang lebih kaya terhadap opini publik, karena tidak hanya memetakan topik pembahasan, tetapi juga memperlihatkan kecenderungan emosional masyarakat terhadap setiap topik secara terstruktur.



Gambar 6. Sebaran Sentimen pada Grafik Topik (Scatter Plot)

3.6.2 Wawasan Intergrasi (Insight) RNN-DBSCAN

Integrasi RNN dan DBSCAN menghasilkan pemahaman berlapis tentang sentimen dan topik diskusi publik. Analisis menunjukkan sentimen negatif terpusat pada isu spesifik seperti banjir, sementara sentimen positif tersebar lebih luas. Hasil ini memberikan peta strategis untuk penyesuaian komunikasi dan respons institusi. Berikut adalah tabel integrasi hasil RNN dan DBSCAN yang ditunjukkan pada Tabel 5.

Tabel 5. Matriks Integrasi Temuan RNN dan DBSCAN

Aspek Analisis	Temuan Model RNN (Klasifikasi)	Temuan Model DBSCAN (Klusterisasi)	Wawasan Terintegrasi (Insight)
Pola Data	Distribusi sentimen didominasi oleh kelas Netral (76,41%), sedangkan sentimen Negatif dan Positif berada pada proporsi yang lebih kecil.	Hanya sekitar 25,5% data membentuk kluster topik yang padat, sementara 74,5% lainnya merupakan noise (komentar acak/tidak bertema kuat).	Sebagian besar interaksi publik bersifat informatif dan rutin, sedangkan isu penting tersembunyi pada segmen data kecil namun bermakna.
Fokus Isu	Sentimen Negatif terdeteksi dengan presisi tinggi (94%), menunjukkan keandalan model dalam menangkap emosi keluhan publik.	DBSCAN membentuk kluster-kluster spesifik, misalnya kluster isu banjir dan cuaca panas ekstrem pada wilayah tertentu.	Keluhan publik tidak bersifat umum, tetapi terfokus pada peristiwa nyata dan lokasi spesifik yang membutuhkan respons cepat.
Sebaran Emosi	Sentimen Positif dan Netral tersebar pada banyak topik umum seperti apresiasi	Kluster-kluster kecil memuat topik spesifik dengan kepadatan tinggi,	Respon emosional masyarakat berbeda antar topik: isu kritis sebagai emosi kuat, isu umum sebagai emosi datar/positif.



Aspek Analisis	Temuan Model RNN (Klasifikasi)	Temuan Model DBSCAN (Klasterisasi)	Wawasan Terintegrasi (Insight)
Implikasi Sistem	layanan dan permintaan informasi. Model RNN dapat dimanfaatkan sebagai early warning system berbasis deteksi sentimen negatif.	sedangkan topik umum cenderung menyebar. DBSCAN efektif untuk pemetaan topik dan lokasi isu secara tematik.	Integrasi keduanya memungkinkan sistem memprioritaskan kluster topik kritis dengan sentimen negatif untuk efisiensi operasional.
Rekomendasi Strategis	Fokus pada pemantauan sentimen negatif secara real-time.	Fokus pada kluster padat dan abaikan noise untuk analisis inti.	Strategi komunikasi BMKG sebaiknya diarahkan pada irisan antara sentimen negatif dan kluster topik padat agar lebih tepat sasaran.

Perbandingan kontribusi analitik antara metode RNN dan DBSCAN, serta wawasan strategis yang dihasilkan. Kolom “Temuan Model RNN” menggambarkan distribusi sentimen secara umum, sementara “Temuan Model DBSCAN” mengungkap struktur topik dan identifikasi data yang tidak relevan. Kolom “Wawasan Terintegrasi (Insight)” menyimpulkan bahwa sentimen negatif publik terfokus pada isu-isu konkret seperti banjir, dan dapat disimpulkan bahwa 74,5% data merupakan *noise* mengindikasikan perlunya sistem filtrasi otomatis agar strategi komunikasi BMKG dapat berfokus pada informasi inti yang krusial. Dengan demikian, integrasi RNN dan DBSCAN yang didukung oleh bukti statistik kualitas kluster ini menghasilkan peta persepsi publik yang tidak hanya deskriptif tetapi juga secara metodologis kokoh.

4. KESIMPULAN

Penelitian mengenai analisis sentimen terhadap diseminasi informasi iklim oleh Stasiun Klimatologi Jawa Tengah di media sosial (Instagram dan X) mengungkap bahwa mayoritas audiens memberikan respons yang positif. Melalui implementasi algoritma *Recurrent Neural Network* (RNN) dan *Density-Based Spatial Clustering of Applications with Noise* (DBSCAN), terlihat adanya pola interaksi yang dinamis antara penyedia informasi dan masyarakat. Berdasarkan klasifikasi sentimen, Penerapan metode *Deep Learning* RNN-LSTM terbukti sangat efektif dalam analisis sentimen produk iklim, mencapai akurasi pengujian sebesar 96%. Model ini memiliki sensitivitas tinggi dalam mendeteksi sentimen negatif masyarakat yang krusial bagi evaluasi layanan. Di sisi lain, penggunaan DBSCAN memberikan nilai tambah melalui kemampuannya mengelompokkan diskusi ke dalam puluhan kluster tematik secara otomatis. Algoritma DBSCAN berhasil memetakan struktur diskusi publik menjadi 82 kluster topik spesifik. Tingginya rasio *noise* (74,5%) menunjukkan karakteristik data media sosial yang heterogen, sekaligus menegaskan pentingnya metode klasterisasi berbasis densitas untuk memisahkan informasi inti dari komentar acak. Secara keseluruhan, Integrasi kedua metode menghasilkan wawasan strategis bahwa sentimen negatif publik terfokus pada dampak fisik bencana (seperti banjir di lokasi spesifik), sedangkan sentimen positif tersebar pada apresiasi layanan informasi. Temuan ini merekomendasikan Stasiun Klimatologi Jawa Tengah untuk beralih dari pemantauan manual ke sistem filtrasi otomatis berbasis prioritas, mengingat dominasi data *noise* yang signifikan. Secara teknis, penelitian selanjutnya sangat disarankan untuk mengadopsi arsitektur *Transformer* (seperti *IndoBERT*) dan teknik reduksi dimensi non-linear (t-SNE atau UMAP) guna mengatasi kompleksitas bahasa informal serta mempertajam separasi topik yang terbentuk. Selain itu, integrasi model ke dalam dashboard pemantauan real-time dan penerapan metode *Active Learning* untuk pelabelan data menjadi langkah krusial berikutnya, memastikan bahwa mitigasi isu publik dan evaluasi layanan informasi iklim dapat dilakukan secara instan, akurat, dan komprehensif di berbagai kanal digital.

REFERENCES

- [1] A. Meilya and R. Arivadany, “Dampak Perubahan Iklim Terhadap Penyebaran Demam Berdarah : Tinjauan Literatur,” *J. Kesehat. Tambusai*, vol. 5, no. 3, pp. 7107–7119, 2024, doi: 10.31004/jkt.v5i3.30576.
- [2] R. Setiawan, W. I. Palanganga, and A. Akhriana, “Perbandingan Aplikasi Prakiraan Cuaca Weather , BMKG ,” *J. Dipakomti*, vol. 16, no. 2, 2025, doi: 10.36774/dipakomti.v16i2.1780.
- [3] G. S. Lasatira, K. D. Hartomo, and I. Sembiring, “Analisis Sentimen Terhadap Pengaruh Minat Belanja Berdasarkan Komentar di Marketplace Menggunakan Metode Recurrent Neural Network (RNN),” *J. Sist. Inf. Bisnis*, vol. 13, no. 2, pp. 112–119, 2023, doi: 10.21456/vol13iss2pp112-119.
- [4] Yusril, W. Fuadi, and Y. Afrillia, “Analisis sentimen review aplikasi stockbit di google play store dan x(twitter) menggunakan support vector machine,” *J. Teknol. dan Sist. Inf. Univrab*, vol. 10, no. 2, pp. 1050–1062, 2025, doi: 10.36341/rabit.v10i2.6446.
- [5] A. Aljabar and A. A. Karim, Abd, “Analisis Sentimen Menggunakan Algoritma Lstm Pada Media Sosial,” *J. Publ. Ilmu Komput. dan Multimed.*, vol. 1, no. 3, pp. 181–187, 2022, doi: 10.55606/jupikom.v1i3.517
- [6] Mustakim et al., “DBSCAN algorithm: Twitter text clustering of trend topic pilkada pekanbaru,” *J. Phys. Conf. Ser.*, vol. 1363, no. 1, 2019, doi: 10.1088/1742-6596/1363/1/012001.
- [7] L. Kurniasari and A. Setyanto, “Sentiment Analysis using Recurrent Neural Network,” *J. Phys. Conf. Ser.*, vol. 1471, no. 1, 2020, doi: 10.1088/1742-6596/1471/1/012018.



- [8] M. R. Firmansyah, R. Ilyas, and F. Kasyidi, “Klasifikasi Kalimat Ilmiah Menggunakan Recurrent Neural Network,” *Pros. 11th Ind. Res. Work. Natl. Semin.*, vol. 11, no. 1, pp. 488–495, 2020, doi: 10.35313/irwns.v11i1.2055.
- [9] S. Bodapati, H. Bandarupally, R. N. Shaw, and A. Ghosh, “Comparison and Analysis of RNN-LSTMs and CNNs for Social Reviews Classification,” *Adv. Appl. Data-Driven Comput. (Advances Intell. Syst. Comput. vol. 1319)*, J. C. Bansal, L. C. C. Fung, M. Simic, A. Ghosh, Eds. Singapore Springer Singapore, vol. 1319, pp. 49–59, 2021, doi: 10.1007/978-981-33-6919-1_4.
- [10] Rasmila, Y. Saputri, F. Syaki, and N. Hadinata, “Sentiment Analysis of Trending Topics on Social Media X Using Natural Language Processing and LSTM,” *J. Appl. Informatics Comput.*, vol. 9, no. 6, pp. 3034–3041, 2025, doi: 10.30871/jaic.v9i6.10931.
- [11] N. Ramdhani and R. H. Al-Fadillah, “Analisis Sentimen Pengguna Twitter Terhadap Belajar Daring Selama Pandemi Covid-19 Dengan Deep Learning,” *J. Siliwangi*, vol. 7, no. 2, p. 2021, 2021, doi: 0.37058/jssainstek.v7i2.4281.
- [12] A. Laksito, N. Wiranda, S. N. Karimah, and M. Hayaty, “The COVID-19 Tweets Classification Based on Recurrent Neural Network,” *Int. J. Adv. Sci. Eng. Inf. Technol.*, vol. 14, no. 1, pp. 358–364, 2024, doi: 10.18517/ijaseit.14.1.18832.
- [13] N. C. Dang, M. N. Moreno-Garcia, and F. D. la Prieta, “Sentiment Analysis Based on Deep Learning in E-Commerce,” *Electronics*, vol. 9, no. 483, 2020, doi: 10.1007/978-3-031-10986-7_40.
- [14] C. N. Dang, M. N. Moreno-Garcia, and F. De La Prieta, “Hybrid Deep Learning Models for Sentiment Analysis,” *Complexity*, vol. 2021, 2021, doi: 10.1155/2021/9986920.
- [15] D. Darwis, N. Siskawati, and Z. Abidin, “Penerapan Algoritma Naive Bayes Untuk Analisis Sentimen Review Data Twitter Bmkg Nasional,” *J. Tekno Kompak*, vol. 15, no. 1, pp. 131–145, 2021, doi: 10.33365/jtk.v15i1.744.
- [16] H. Yarbo, R. Rais, and I. Setiawan, “Sentiment Analysis Using Recurrent Neural Network (Rnn) Method With Long Short Term Memory (Lstm) On Traveloka Application Comment Review,” *Proc. 5th Int. Semin. Sci. Technol. (ISST 2023)*, *Adv. Phys. Res.*, vol. 10, pp. 69–77, 2024, doi: 10.2991/978-94-6463-520-1_12.
- [17] Mustakim, M. Z. Fauzi, Mustafa, A. Abdullah, and Rohayati, “Clustering of Public Opinion on Natural Disasters in Indonesia Using DBSCAN and K-Medoids Algorithms,” *J. Phys. Conf. Ser.*, vol. 1783, no. 1, 2021, doi: 10.1088/1742-6596/1783/1/012016.
- [18] F. Andriyani and Y. Puspitarani, “Performance Comparison of K-Means and DBScan Algorithms for Text Clustering Product Reviews,” *Sinkron*, vol. 6, no. 3, pp. 944–949, 2022, doi: 10.33395/sinkron.v7i3.11569.
- [19] E. Hartono and C. Fibriani, “Analisis Ulasan Daring Menggunakan Metode Density-Based Spatial Clustering of Applications with Noise Online Review Analysis Using Density-Based Spatial Clustering of Applications with Noise,” *J. Tek. Inform. dan Sist. Inf.*, vol. 11, no. 3, pp. 475–485, 2025, doi: 10.28932/jutisi.v11i3.12363.
- [20] A. Tsamratul ‘Ain and S. Supatmi, “Implementasi Algoritma DBSCAN dan Support Vector Machine Dalam Penentuan Sentimen Pengguna Aplikasi PeduliLindungi,” *J. Tata Kelola dan Kerangka Kerja Teknol. Inf.*, vol. 9, no. 1, pp. 8–18, 2023, doi: 10.34010/jtk3ti.v9i1.8950
- [21] F. R. Ferdiansyah, R. W. Nugraha, R. Sofian, H. Purwanto, D. Saepudin, and E. Andriansyah, “Implementation of K-Means and DBSCAN algorithms : A Bibliometric Review,” *Proc. Widyatama Int. Conf. Eng. 2024 (WICOENG 2024)*, *Adv. Eng. Res.*, vol. 252, 2024, doi: 10.2991/978-94-6463-618-5.