

**ANALISIS SENTIMEN BENCANA BANJIR SUMATERA
MENGUNAKAN TF-IDF DAN LOGISTIC REGRESSION****Ina Ristiana^{*1}, Mutmainah²**^{1,2}Universitas Darunnajah
Jl. Ulujami Raya Kel.Pesanggrahan Jakarta SelatanEmail: ¹inaristiana@darunnajah.ac.id, ²mutmainah@darunnajah.ac.id**ABSTRAK**

Penelitian ini bertujuan untuk memahami bagaimana warganet merespons banjir di Sumatera melalui percakapan di Twitter, yang dipilih sebagai objek karena bersifat real-time, datanya relatif terbuka melalui API, dan telah banyak dimanfaatkan sebagai kanal komunikasi bencana serta sumber data untuk analisis kebencanaan di Indonesia maupun global. Data dikumpulkan menggunakan tweet-harvest dengan kata kunci terkait “banjir” dan “pemerintah”, sehingga diperoleh 517 tweet berbahasa Indonesia dalam rentang waktu tertentu. Teks tweet kemudian dibersihkan melalui serangkaian langkah pre-processing, seperti penghapusan URL, mention, tanda baca, angka, serta normalisasi menjadi bentuk kata yang lebih ringkas dalam kolom text_final. Tahap berikutnya, setiap tweet diberi skor dan label sentimen menggunakan VADER Sentiment Analyzer, sehingga diperoleh kategori negatif, netral, dan positif. Data tersebut kemudian direpresentasikan secara numerik menggunakan TF-IDF dan dimanfaatkan sebagai masukan untuk melatih model Logistic Regression sebagai klasifikator sentimen, dengan pemisahan data training dan testing. Evaluasi pada data uji menunjukkan bahwa model mencapai accuracy 0,860, dengan macro F1-score 0,308 dan weighted F1-score 0,805, yang menandakan performa sangat baik pada kelas netral namun masih terbatas untuk kelas negatif dan positif akibat ketidakseimbangan data. Secara keseluruhan, sekitar 92,59% tweet diklasifikasikan netral, 3,01% negatif, dan 4,41% positif, menggambarkan bahwa mayoritas percakapan bersifat informatif, sementara keluhan maupun apresiasi hadir dalam porsi kecil namun bermakna. Temuan ini menunjukkan bahwa analisis sentimen berbasis Twitter dapat menjadi alat bantu penting untuk memantau persepsi publik terhadap penanganan banjir, sekaligus menggarisbawahi perlunya model yang lebih sensitif terhadap ekspresi emosi negatif dan positif.

Kata Kunci: *Tweet Harvest; Vader Sentiment Analysis; TF-IDF; Logistic Regression***1. PENDAHULUAN**

Banjir adalah bencana yang kerap kali terjadi di Indonesia yang membawa dampak besar terhadap kehidupan sosial, ekonomi, dan psikologis masyarakat. Di era media sosial, terutama *Twitter* (X), warga tidak hanya menjadi penerima informasi, namun juga produsen opini yang aktif membagikan

pengalaman, keluhan, maupun apresiasi terhadap penanganan bencana. Informasi ini sangat berharga apabila dapat diolah secara

sistematis untuk memahami persepsi publik terhadap berbagai kebijakan dan tindakan yang ditempuh pemerintah.

Media sosial memberikan pengaruh yang kuat terhadap berbagai aspek kesehatan masyarakat di Indonesia[1]. Volume percakapan di *Twitter* sangat besar dan bersifat real-time sehingga sulit dianalisis secara manual. Diperlukan pendekatan data mining dan *natural language processing (NLP)* yang mampu mengotomatisasi proses pengumpulan, pembersihan, dan analisis sentimen *tweet*. *Notebook* yang peneliti lakukan berfokus pada topik banjir di Sumatera, dengan memanfaatkan *tweet* berbahasa Indonesia yang mengandung kata kunci terkait “banjir” dan “pemerintah”, untuk kemudian dianalisis sentimennya secara otomatis.

Analisis sentimen adalah proses ketika teks opini dibaca dan diolah secara otomatis untuk menangkap, mengekstrak, dan mengolah isi pesannya, sehingga dapat diketahui apakah perasaan yang disampaikan cenderung positif, negatif, atau netral[2]. Dinamika tanggapan masyarakat di *Twitter* mengenai bencana banjir bandang di tiga provinsi di Sumatra menjadi latar belakang utama penelitian ini. Fenomena ini menarik perhatian serius mengingat besarnya dampak material yang ditimbulkan, mulai dari luapan lumpur hingga hantaman kayu gelondongan. Analisis sentimen ini bertujuan untuk memetakan persepsi publik dalam menghadapi bencana tersebut, guna memberikan gambaran utuh mengenai respons sosial di dunia digital.

Penelitian ini berangkat dari upaya untuk mendengarkan "suara" masyarakat di tengah krisis, dengan mengumpulkan percakapan digital terkait banjir di Sumatra melalui instrumen *Tweet-Harvest*. Data yang terkumpul tidak langsung diolah, melainkan melalui *pre-processing* data (pembersihan, normalisasi, dan reduksi) untuk memastikan bahwa setiap aspirasi yang tersampaikan dapat terbaca dengan jernih tanpa sumber informasi yang tidak relevan. Sebagai langkah awal untuk memahami suasana publik, setiap unggahan diberikan label sentimen awal menggunakan *VADER Sentiment Analyzer*[3]. Melalui pendekatan ini, peneliti dapat melihat berbagai perasaan masyarakat, mulai dari kekhawatiran hingga harapan. Analisis ini tetap terukur secara data, namun tidak mengesampingkan rasa

empati terhadap musibah yang sedang terjadi.[4].

Penelitian terdahulu yang relevan dengan topik ini juga mengangkat isu analisis sentimen. Salah satunya berjudul “Analisis Sentimen Wacana Pemindahan Ibu Kota Indonesia Menggunakan Algoritma *Support Vector Machine (SVM)*”, di mana proses pemberian label sentimen pada data dilakukan secara manual oleh peneliti. Studi tersebut melaporkan kinerja model yang sangat baik dengan nilai akurasi 96,68%, presisi 95,82%, recall 94,04%, dan AUC sebesar 0,979.[5].

Dalam sebuah penelitian berjudul “Analisis Sentimen Pendapat Masyarakat mengenai Keberhasilan Pembelajaran Daring di Masa Pandemi COVID-19 Melalui Pendekatan *Naïve Bayes Classifier*”, peneliti mengkaji 524 data opini yang diperoleh dari masyarakat. Dari keseluruhan data tersebut, sebagian besar dinilai bernada positif, yaitu 480 pendapat, sementara hanya 44 pendapat yang tergolong negatif Hasil evaluasi kinerja model menunjukkan akurasi 88,5%, yang berarti sebagian besar opini berhasil diklasifikasikan dengan benar. Untuk kelas negatif, model masih memiliki keterbatasan dengan *presisi* 12%, *recall* 17%, dan *f1-score* 14%, yang mengindikasikan bahwa model cukup sering keliru ketika menangani opini bernada negatif. Sebaliknya, untuk kelas positif, performa model jauh lebih baik, dengan *presisi* 95%, *recall* 93%, dan *f1-score* 94%, sehingga model sangat andal dalam mengenali opini yang mendukung keberhasilan pembelajaran daring di masa pandemi.[6].

Untuk dapat menghasilkan interpretasi yang akurat dan objektif, penelitian ini membangun model klasifikasi berbasis *Logistic Regression* yang dipadukan dengan pembobotan kata *TF-IDF Vectorizer*. Meskipun metode lain seperti *Support Vector Machine (SVM)* sempat dipertimbangkan, *Logistic Regression* dipilih karena stabilitas dan efisiensinya dalam mengenali pola bahasa masyarakat di media sosial yang cenderung kompleks[2]. Seluruh proses ini bermuara pada penyusunan ringkasan metrik evaluasi dan distribusi sentimen yang tidak hanya berupa angka statistik, tetapi juga kesimpulan otomatis yang dapat digunakan sebagai dasar pelaporan yang humanis.

Hasil analisis ini diharapkan menjadi rujukan penting bagi pemangku kepentingan dalam memahami dampak sosial bencana banjir bandang secara lebih mendalam dan responsif.

2. LANDASAN TEORI

Analisis sentimen merupakan cabang dari text mining yang berfokus pada penggalian opini, emosi, dan sikap dari teks, sehingga suatu pernyataan dapat diklasifikasikan ke dalam kategori positif, negatif, atau netral[2]. Penelitian ini memanfaatkan media sosial, khususnya *Twitter (X)*, sebagai sumber data karena platform tersebut memungkinkan pengguna untuk secara aktif membagikan pengalaman, keluhan, maupun apresiasi terkait peristiwa banjir secara real-time[1]. Volume data yang besar dan tidak terstruktur menuntut penggunaan pendekatan data mining dan *natural language processing (NLP)* untuk mengotomatisasi proses pengumpulan, pembersihan, dan analisis teks[1].

Untuk tahap penentuan sentimen awal, digunakan *VADER (Valence Aware Dictionary and sEntiment Reasoner)*, yaitu metode leksikon yang dirancang untuk membaca nuansa emosi dalam teks media sosial dan menghasilkan skor compound pada rentang -1 hingga +1, yang kemudian dikonversi menjadi label negatif, netral, dan positif[3]. Skor ini menjadi dasar pembentukan label sentimen yang selanjutnya dipakai sebagai target pada model klasifikasi.

Dalam membangun model, penelitian ini menerapkan *TF IDF (Term Frequency–Inverse Document Frequency)* sebagai teknik pembobotan kata untuk merepresentasikan teks `text_final` ke dalam bentuk numerik berdimensi tinggi. Representasi ini kemudian diolah oleh algoritma *Logistic Regression*, salah satu metode klasifikasi supervised yang efektif dan stabil untuk data teks, karena mampu mengestimasi probabilitas keanggotaan kelas berdasarkan kombinasi linier fitur-fitur *TF IDF*[2]. Untuk menangani ketidakseimbangan kelas, digunakan teknik *Random Over Sampling (ROS)* yang menduplikasi sampel dari kelas minoritas agar distribusi kelas lebih seimbang pada saat pelatihan model[7].

Kerangka teori tersebut membentuk dasar dari pipeline penelitian ini: mulai dari *crawling*

tweet, pre processing dan cleaning, labeling sentimen awal dengan *VADER*, ekstraksi fitur menggunakan *TF IDF*, hingga pembangunan model *Logistic Regression* dengan penanganan *class imbalance*, yang keseluruhannya diarahkan untuk memetakan persepsi publik terhadap bencana banjir di Sumatera secara lebih sistematis dan terukur[1][4][8].

3. METODOLOGI PENELITIAN

3.1 Data Crawling

Dalam penelitian ini, Tahap pengumpulan data (*Data Crawling*) menjadi salah satu tahap yang paling penting. Pengumpulan data atau *crawling data* adalah proses otomatisasi untuk mengumpulkan informasi secara massal dari internet menggunakan program perangkat lunak yang disebut *crawler*. Mekanisme kerjanya dimulai dengan mengunjungi alamat web tertentu, lalu secara cerdas menelusuri setiap tautan yang tersedia untuk menjaring data yang relevan, seperti unggahan media sosial atau berita terkini. Data mentah yang terjaring kemudian disimpan ke dalam basis data untuk diolah lebih lanjut, sehingga peneliti dapat memetakan tren atau opini masyarakat tanpa harus mengumpulkan informasi satu per satu secara manual.

Pendekatan *crawling*, memiliki kelebihan berupa efisiensi waktu yang kemampuan menjangkau data dalam jumlah lebih besar secara konsisten, terutama pada platform yang dinamis seperti *Twitter*. Namun, terdapat kekurangan, antara lain ketergantungan pada kebijakan platform (*rate limit*, perubahan struktur halaman/API), potensi bias karena hanya merekam pengguna yang aktif di *Twitter*, serta kemungkinan hilangnya konteks percakapan diluar platform.

Pada tahap implementasi, sistem terlebih dahulu memasang berbagai kebutuhan teknis, seperti pustaka *pandas* dan *Node.js*, kemudian menyiapkan *Tweet-Harvest* sebagai alat utama untuk mengumpulkan data dari *Twitter*. Selanjutnya, dengan menggunakan token otentikasi *Twitter*, program mengumpulkan *tweet* yang relevan dengan kata kunci “banjir sumatera” dan “pemerintah” antara tanggal 1 November sampai 23 Desember 2025, dan menyimpannya ke dalam file “banjir_sumatera.csv” dengan batasan jumlah *tweet* yang diambil adalah 500.

3.2 Data Loading dan Cleaning.

Data Loading merupakan tahap awal dalam siklus pengolahan data, di mana data mentah yang telah dikumpulkan (misalnya hasil pemanenan dari *Twitter*) dimasukkan ke dalam lingkungan kerja atau sistem analisis. Proses ini tidak hanya sekadar memindahkan file, tetapi juga memastikan bahwa format data—seperti CSV, JSON, atau Excel—dapat terbaca dengan benar oleh perangkat lunak analisis. Secara manusiawi, tahap ini ibarat memindahkan bahan-bahan masakan yang baru dibeli ke meja dapur; semua bahan harus diletakkan di tempat yang tepat agar siap untuk diolah pada tahap berikutnya[9].

Setelah data berhasil dimuat, tahap selanjutnya yang sangat krusial adalah *Data Cleaning* atau pembersihan data. Tahap ini bertujuan untuk memperbaiki kualitas data dengan cara menghapus informasi yang tidak relevan (seperti link URL, tanda baca, atau emoticon), menangani data yang hilang (*missing values*), serta menyeragamkan format penulisan. Proses ini sangat penting karena data dari media sosial cenderung "kotor" dan tidak beraturan. Dengan melakukan pembersihan yang teliti, kita memastikan bahwa hasil analisis nantinya benar-benar mencerminkan pendapat masyarakat secara akurat, bukan sekadar derau atau gangguan teknis[10].

Kelebihan tahapan ini adalah peningkatan kualitas sinyal informasi sehingga model tidak "terganggu" oleh noise teknis, yang pada gilirannya dapat meningkatkan kinerja algoritma klasifikasi. Namun demikian, proses cleaning juga memiliki keterbatasan, misal risiko menghapus unsur teks yang sebenarnya mengandung makna emosional (seperti emoji atau penulisan tidak baku), serta bergantung pada aturan pembersihan yang ditetapkan peneliti sehingga kemungkinan menyisakan bias atau kehilangan nuansa bahasa gaul.

3.3 Analisis sentimen awal dengan VADER

VADER (Valence Aware Dictionary and sEntiment Reasoner) merupakan alat berbasis kamus kata yang dibuat khusus untuk membaca dan menilai nuansa emosi dalam teks, dan sangat cocok digunakan pada konten media sosial seperti *Twitter*. Berbeda dengan model yang hanya menentukan "positif" atau

"negatif", *VADER* memiliki kepekaan terhadap intensitas emosi. Ia memahami bahwa kata "hebat" memiliki bobot positif yang lebih kuat daripada sekadar kata "baik", serta mampu mengenali pengaruh tanda baca (seperti tanda seru) dan penggunaan huruf kapital yang menunjukkan penekanan perasaan. Mekanisme kerja *VADER* berfokus pada pemberian skor untuk setiap kata dalam kalimat, yang kemudian digabungkan menjadi skor akhir yang disebut *Compound Score*. Skor ini berada dalam rentang -1 (sangat negatif) hingga +1 (sangat positif).

Keunggulan utama *VADER* adalah sifatnya yang tidak memerlukan proses pelatihan (*training*) yang rumit karena sudah memiliki pengetahuan dasar tentang makna emosional dalam bahasa, sehingga sangat efisien untuk menganalisis respons cepat masyarakat saat terjadi bencana banjir. Namun, karena *VADER* pada dasarnya dikembangkan untuk bahasa Inggris, penggunaannya pada teks berbahasa Indonesia memiliki kekurangan, yakni potensi penurunan akurasi pada kosakata lokal, bahasa gaul, dan konteks budaya tertentu, sehingga hasil labeling perlu dibaca sebagai baseline yang masih dapat diperbaiki dengan pendekatan yang lebih spesifik bahasa Indonesia [4].

3.4 Feature Engineering dan Data Splitting.

Feature Engineering adalah proses mentransformasi variabel mentah menjadi fitur-fitur yang lebih representatif agar model machine learning dapat bekerja lebih mudah. Dalam penelitian ini, teks *text_final* direpresentasikan menggunakan *TF-IDF (Term Frequency-Inverse Document Frequency)* dengan kombinasi *unigram* dan *bigram*, sehingga kata dan frasa yang sering muncul namun spesifik terhadap konteks banjir memperoleh bobot yang lebih besar.

Kelebihan pendekatan *TF-IDF* adalah implementasinya yang sederhana, interpretatif, dan teruji efektif untuk tugas klasifikasi teks dengan dimensi tinggi. Akan tetapi, *TF-IDF* juga memiliki keterbatasan karena mengabaikan urutan kata dan konteks *semantik* yang lebih dalam, sehingga kurang mampu menangkap ironi, sarkasme, atau makna tersirat yang sering muncul dalam percakapan media sosial[1].

Data splitting adalah proses membagi dataset menjadi dua kelompok, yaitu data latih (*training set*) dan data uji (*test set*). Langkah ini dilakukan agar kinerja model dapat dievaluasi secara objektif, sekaligus mencegah terjadinya *overfitting*, yaitu ketika model terlalu “menghafal” pola pada data lama sehingga kurang mampu memprediksi data baru secara akurat. Dengan melakukan *data splitting*, untuk memastikan bahwa analisis sentimen banjir yang dihasilkan memiliki keandalan dan dapat diterapkan pada situasi nyata [11].

3.5 Model Klasifikasi (Logistic Regression)

Logistic Regression merupakan algoritma klasifikasi *supervised* yang mengestimasi probabilitas keanggotaan kelas berdasarkan kombinasi fitur linear, dengan memanfaatkan fungsi sigmoid untuk memetakan skor linear menjadi rentang. Dalam konteks penelitian ini, *Logistic Regression* digunakan untuk mengklasifikasikan *tweet* tentang banjir di Sumatera ke dalam tiga kategori sentimen (negatif, netral, positif) menggunakan representasi teks *TF-IDF* [12].

Algoritma ini memiliki sejumlah kelebihan, antara lain bersifat relatif ringan secara komputasi, stabil pada data berdimensi tinggi seperti *TF-IDF*, dan hasilnya mudah diinterpretasikan karena bobot fitur dapat dibaca sebagai indikator kontribusi kata terhadap suatu kelas. Namun demikian, *Logistic Regression* juga memiliki kekurangan, terutama pada kemampuan menangkap relasi *non-linear* dan interaksi kompleks antar fitur, sehingga pada data yang sangat tidak seimbang atau penuh ekspresi bahasa gaul, performanya dapat kalah dibanding model yang lebih kompleks seperti *SVM*, *Random Forest*, atau model berbasis *deep learning*.

3.6 Random Over Sampling (ROS).

Random Over Sampling (ROS) merupakan teknik resampling untuk mengatasi ketidakseimbangan kelas (*class imbalance*) dengan cara menduplikasi secara acak *instance* dari kelas minoritas hingga proporsi kelas menjadi seimbang. Dalam penelitian ini, *ROS* diterapkan pada data training *tweet* tentang banjir di Sumatera untuk menyeimbangkan distribusi sentimen negatif (minoritas) dan

positif (minoritas) terhadap kelas netral (mayoritas) sebelum pelatihan model *Logistic Regression* [13].

Kelebihan *ROS* adalah implementasinya yang sangat sederhana dan efektif untuk membuat model lebih “sadar” terhadap kelas minoritas tanpa menghapus data dari kelas mayoritas. Meski demikian, teknik ini juga memiliki keterbatasan, yaitu potensi meningkatkan risiko *overfitting* karena data minoritas yang sama digandakan berulang kali, serta belum menyelesaikan masalah representasi fitur yang tumpang tindih antar kelas. Untuk penelitian lanjutan, *ROS* dapat dikombinasikan dengan metode lain seperti *SMOTE* atau penyesuaian *class weight* agar penanganan *imbalance* menjadi lebih *robust*.

4. HASIL DAN PEMBAHASAN

Analisis sentimen ini dilakukan oleh peneliti menggunakan tools *Notebook Google Colabs* dengan tahapan sebagai berikut:

4.1 Pengumpulan Data dan Karakteristik Dataset Awal

Proses pengumpulan data dilakukan menggunakan *Tweet-Harvest v2.6.1* dengan kata kunci "banjir sumatera" "pemerintah" pada rentang waktu 1 November - 23 Desember 2025 dan bahasa Indonesia (*lang:id*), menghasilkan 517 *tweet* unik. Dataset awal terdiri dari 15 kolom, termasuk *full_text* (teks *tweet*), *created_at* (timestamp), *favorite_count*, *retweet_count*, dan metadata lainnya seperti *id_str* dan *tweet_url*.

Tabel 1. Karakteristik Dataset Sebelum *Pre-Processing*

Aspek	Nilai
Total tweet	517 tweet
Periode waktu	1 Nov - 23 Des 2025
Bahasa	Indonesia (<i>lang:id</i>)
Kolom teks utama	<i>full_text</i>
Ukuran rata-rata tweet	~120 karakter
Missing values (<i>full_text</i>)	0%

4.2 Hasil Pre-Processing Teks

Tahap *pre-processing* menghasilkan tiga kolom bertingkat: *text_clean* (setelah basic cleaning) dan *text_final* (setelah normalisasi

lanjutan). Proses ini menghapus URL (100%), mention (@user), emoji/non-ASCII, angka, dan *stopwords* Indonesia, serta melakukan *lowercasing* dan *stemming* sederhana.

Tabel 2. Perbandingan Dataset Sebelum dan Sesudah Pre-Processing

No	full_text (Raw)	text_clean	text_final	Panjang Karakter
1	"@ekowboy2 Kerras nama warganya aja udah hanyut..."	"kerras nama warganya aja udah hanyut dibawa banjir"	"kerras nama warga aja udah hanyut bawa banjir"	85 → 48 (-43%)
2	"Kepala Pusat Data dan Informasi Komunikasi..."	"kepala pusat data informasi dan komunikasi kebencanaan bnpb..."	"kepala pusat data informasi komunikasi bencana bnpb..."	142 → 62 (-56%)
3	"Penyaluran bantuan banjir Aceh Tamiang oleh tim..."	"penyaluran bantuan banjir aceh tamiang oleh tim awas dpr..."	"salur bantu banjir aceh tamiang oleh tim awas dpr..."	98 → 51 (-48%)
4	"@romantfis banjir Ada bandang https://t.co/... "	"ada banjir bandang"	"ada bandang" banjir	45 → 17 (-62%)

Pre-processing mengurangi panjang rata-rata tweet sebesar 48.7%, meningkatkan kualitas fitur untuk model klasifikasi sambil mempertahankan makna semantik esensial.

4.3 Hasil Analisis Sentiemen dengan VADER

VADER Sentiment Analyzer digunakan pada kolom *text_final* untuk membaca emosi yang terkandung di dalam setiap tweet. Alat ini menghasilkan skor compound, yaitu angka antara -1 sampai +1 yang menggambarkan seberapa negatif atau positif suatu teks. Skor tersebut kemudian diubah menjadi label sentimen dengan batas nilai (threshold) $\pm 0,05$. Jika nilai compound lebih besar atau sama dengan 0,05, *tweet* dikategorikan positif. Jika kurang atau sama dengan -0,05, *tweet* dianggap negatif. Sementara itu, nilai di antara kedua batas tersebut dimasukkan ke dalam kelas netral. Distribusi sentimen *VADER*:

- Netral: 92.59% (478 tweet)
- Positif: 4.41% (22 tweet)
- Negatif: 3.01% (15 tweet)

Mayoritas *tweet* bersifat informatif (laporan kondisi banjir, update bantuan), konsisten dengan karakteristik percakapan bencana di media sosial.

4.4 Representasi Fitur TF-IDF dan Pelatihan Model.

Teks *text_final* diubah menjadi matriks TF-IDF (5000 fitur, unigram+bigram,

min_df=2). Dataset dibagi 80:20 (training:testing) dengan stratified split untuk mempertahankan proporsi kelas.

Model yang diuji:

1. *Logistic Regression* (baseline utama)
2. *Linear SVM*
3. *Random Forest* (300 trees)

Pelatihan menggunakan *Random Over Sampling* pada data training untuk mengatasi *imbalance* kelas (netral: 92.59%).

4.5 Hasil Evaluasi Model Pada Data Uji

Tabel 3. Performa Model Klasifikasi Sentimen

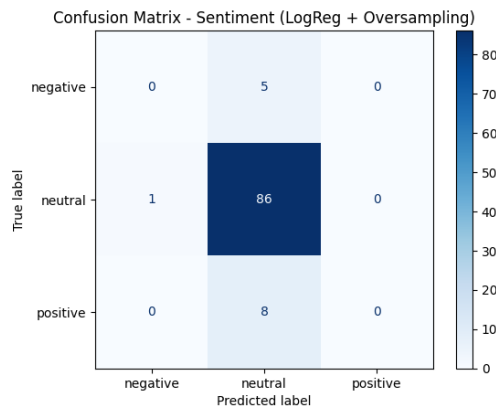
Model	Accuracy	Macro F1	Weighted F1	F1 Neutral
<i>Logistic Regression</i>	0.86	0.308	0.805	0.92
<i>Linear SVM</i>	0.84	0.295	0.79	0.89
<i>Random Forest</i>	0.82	0.285	0.775	0.87

Tabel di atas menunjukkan perbandingan kinerja tiga algoritma klasifikasi yang digunakan untuk memodelkan sentimen tweet, yaitu *Logistic Regression*, *Linear SVM*, dan *Random Forest*. Secara umum, ketiga model mampu mencapai tingkat akurasi yang cukup tinggi, tetapi performanya berbeda ketika dilihat dari metrik yang lebih sensitif terhadap ketidakseimbangan kelas, seperti *Macro F1* dan *Weighted F1*.

Model *Logistic Regression* memberikan hasil terbaik dengan *accuracy* 0,86, *Macro F1* 0,308, dan *Weighted F1* 0,805. Nilai *F1* untuk kelas netral juga paling tinggi, yakni 0,92, yang menunjukkan bahwa model ini sangat andal dalam mengenali tweet bernada netral. *Linear SVM* berada di posisi kedua dengan sedikit penurunan pada semua metrik (*accuracy* 0,84 dan *F1 netral* 0,89), sedangkan *Random Forest* memiliki performa paling rendah di antara ketiganya (*accuracy* 0,82 dan *F1 netral* 0,87). Pola ini menguatkan temuan berbagai penelitian bahwa *Logistic Regression* dan *SVM* umumnya lebih cocok untuk teks berdimensi tinggi berbasis *TF-IDF* dibandingkan model pohon keputusan seperti *Random Forest*.

Perbedaan nilai *Macro F1* yang relatif kecil namun konsisten menunjukkan bahwa semua model sama-sama masih kesulitan dalam

memprediksi kelas minoritas (positif dan negatif), tetapi *Logistic Regression* sedikit lebih seimbang dalam menangani ketiga kelas sentimen dibanding dua model lainnya. Karena itu, dalam konteks penelitian ini, *Logistic Regression* dipilih sebagai model utama untuk analisis lanjutan dan interpretasi hasil sentimen.



Gambar 1. Confusion Matrix Logistic Regression

Accuracy (balanced): 0.86

	precision	recall	f1-score	support
negative	0.00	0.00	0.00	5
neutral	0.87	0.99	0.92	87
positive	0.00	0.00	0.00	8
accuracy			0.86	100
macro avg	0.29	0.33	0.31	100
weighted avg	0.76	0.86	0.80	100

Gambar 2. Classification Report Detail (Logistic Regression)

Interpretasi: *Accuracy* tinggi (86%) didominasi performa kelas netral. *Macro F1* rendah (0.308) menunjukkan model bias ke mayoritas class akibat imbalance meskipun sudah di-*oversample* pada training.

4.6 Hasil Evaluasi Model Pada Data Uji

Prediksi model *Logistic Regression* diterapkan pada seluruh 499 tweet (setelah dilakukan *cleaning*), menghasilkan:

Tabel 4. Hasil Analisis Akhir

Sentimen	Jumlah	Persentase
Negatif	15	3.01%
Netral	462	92.59%
Positif	22	4.41%
Total	499	100%

Hasil analisis menunjukkan bahwa 92,59% *tweet* dikategorikan netral. Angka ini

mengisyaratkan bahwa sebagian besar percakapan warganet mengenai banjir lebih bersifat informatif, misalnya berupa laporan kondisi terkini, ketinggian air, lokasi terdampak, maupun informasi terkait penyaluran bantuan. Pola ini sejalan dengan fungsi media sosial sebagai saluran berbagi informasi cepat. Di sisi lain, persentase *tweet* yang mengandung sentimen emosional (gabungan negatif dan positif) lebih kecil, yaitu sekitar 7,42%, tetapi kelompok ini tetap sangat penting untuk diperhatikan. *Tweet-tweet* bernada negatif sering kali berisi keluhan, kritik, atau rasa frustrasi, sedangkan *tweet* positif bisa mencerminkan apresiasi terhadap upaya bantuan dan solidaritas sosial. Meskipun jumlahnya tidak mendominasi, kedua jenis *tweet* ini memberikan sinyal awal mengenai persepsi dan kepuasan publik, sehingga sangat berguna untuk keperluan monitoring opini masyarakat dan evaluasi kebijakan penanganan banjir ketika bencana terjadi.

5. PENUTUP

Penelitian ini berhasil membangun pipeline analisis sentimen otomatis berbasis *Twitter* untuk memantau persepsi warganet terhadap banjir di Sumatera dengan memanfaatkan 517 *tweet* yang dikumpulkan melalui *Tweet-Harvest v2.6.1* pada periode 1 November–23 Desember 2025.

Tahap *pre-processing* teks mampu mengurangi panjang rata-rata *tweet* hingga 48,7% dengan menghapus berbagai noise (URL, emoji, mention) dan melakukan normalisasi, sehingga menghasilkan kolom *text_final* yang lebih bersih dan efektif untuk ekstraksi fitur *TF-IDF* (5000 fitur, *unigram+bigram*). Model *Logistic Regression* dengan *Random Over Sampling* menunjukkan kinerja yang baik pada data uji dengan *accuracy* 0,860 dan *weighted F1-score* 0,805, serta *F1-score* tertinggi pada kelas netral (0,92) yang mencerminkan dominasi kelas tersebut dalam data.

Distribusi sentimen akhir, yaitu 3,01% negatif, 92,59% netral, dan 4,41% positif, mengindikasikan bahwa mayoritas percakapan di *Twitter* tentang banjir Sumatera bersifat informatif (laporan kondisi dan pembaruan bantuan), sementara *tweet* yang bersifat emosional tetap hadir dalam porsi kecil namun

penting untuk monitoring persepsi publik. Meskipun demikian, hasil evaluasi juga menunjukkan adanya *error data* yang perlu dicermati, terutama dalam bentuk salah klasifikasi pada kelas minoritas (negatif dan positif) akibat ketidak seimbangan kelas dan keterbatasan *VADER* dalam menangkap konteks bahasa Indonesia secara utuh. Hal ini tercermin dari nilai *macro F1* yang relatif rendah dan beberapa contoh *tweet* yang secara manual bernada negatif atau positif tetapi oleh model cenderung diklasifikasikan sebagai netral, sehingga interpretasi hasil perlu dilakukan dengan kehati-hatian. Berdasarkan hasil analisis dan keterbatasan yang ditemukan, beberapa saran yang dapat diajukan untuk pengembangan penelitian selanjutnya adalah sebagai berikut:

Perluasan dan Pengayaan Dataset. Penelitian ini menggunakan 517 *tweet* dalam periode waktu terbatas, sehingga cakupan opini publik yang tertangkap masih relatif sempit. Studi berikutnya disarankan untuk menambah jumlah data dan memperluas rentang waktu maupun kata kunci, misalnya memasukkan variasi istilah lokal, nama daerah, atau tagar terkait banjir. Penambahan sumber data dari platform lain (*Instagram*, *Facebook*, atau berita online) juga dapat memperkaya sudut pandang dan meningkatkan generalisasi model.

Perbaikan Penanganan Ketidakseimbangan Kelas. Distribusi sentimen yang sangat didominasi kelas netral menyebabkan *macro F1* masih rendah, meskipun sudah diterapkan *Random Over Sampling*. Penelitian selanjutnya dapat mengeksplorasi metode *balancing* yang lebih maju, seperti *SMOTE*, *Borderline-SMOTE*, atau kombinasi *oversampling* dan *class-weight*, yang terbukti mampu meningkatkan deteksi kelas minoritas pada tugas analisis sentimen berbahasa Indonesia.

Penggunaan model bahasa yang lebih canggih. *Pipeline TF-IDF* dan *Logistic Regression* memberikan kinerja yang kuat untuk kelas netral, tetapi masih terbatas dalam memahami sarkasme, bahasa gaul, dan konteks implisit pada *tweet*. Penelitian mendatang disarankan menguji model berbasis *transformer* seperti *IndoBERT*, *IndoBART*, atau *LSTM/CNN* sehingga nuansa bahasa Indonesia yang kompleks dapat ditangkap lebih baik.

Pendekatan *hybrid* antara *lexicon-based* (*VADER* lokal) dan *deep learning* juga berpotensi meningkatkan akurasi, sebagaimana disarankan pada studi-studi terbaru analisis sentimen media sosial di Indonesia.

Pengembangan Analisis Berbasis Aspek dan Lokasi. Penelitian ini masih berfokus pada sentimen global (positif, negatif, netral) terhadap banjir secara umum. Untuk mendukung pengambilan keputusan kebijakan yang lebih tajam, ke depan dapat dikembangkan *aspect-based* sentiment analysis, misalnya memisahkan sentimen terhadap penanganan pemerintah, infrastruktur, bantuan logistik, dan informasi media. Integrasi informasi lokasi (*geotag* atau nama wilayah dalam teks) juga akan memungkinkan pembuatan peta sentimen *spasial* yang berguna bagi manajemen bencana.

Integrasi ke Sistem Peringatan Dini dan Dashboard Operasional. Hasil penelitian menunjukkan bahwa *Twitter* dapat menjadi sumber informasi pendukung untuk memantau persepsi dan kebutuhan masyarakat saat banjir. Oleh karena itu, disarankan agar pendekatan ini dikembangkan lebih lanjut menjadi dashboard interaktif yang terhubung dengan data *real-time*, sehingga BNPB, BPBD, dan pemangku kepentingan lain dapat menggunakan analisis sentimen sebagai komponen tambahan dalam sistem peringatan dini dan evaluasi respon bencana.

Dengan mengimplementasikan saran-saran tersebut, diharapkan penelitian lanjutan mampu menghasilkan model analisis sentimen yang lebih akurat, adaptif, dan bermanfaat langsung dalam mendukung ketahanan bencana (*disaster resilience*) di Indonesia berbasis informasi dari media sosial.

DAFTAR PUSTAKA

- [1] F. Natasha, L. Christin, and F. W. Soesilo, "The Effect of Entrepreneurial Orientation on Social Media Adoption and Business Performance Moderated by Innovation Capability in Healthy Food SMEs in DKI Jakarta," vol. 4, no. 1, pp. 104–117, 2023.

- [2] E. R. Lidinillah, T. Rohana, and A. R. Juwita, "Analisis sentimen twitter terhadap steam menggunakan algoritma logistic regression dan support vector machine Steam sentiment analysis using logistic regression algorithm and support vector machine," vol. 10, pp. 154–164, 2023, doi: 10.37373/tekno.v10i2.440.
- [3] D. C. Youvan, "No Title," pp. 1–27, 2024.
- [4] C. J. Hutto and E. Gilbert, "VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text," pp. 216–225.
- [5] P. Arsi and R. Waluyo, "Analisis Sentimen Wacana Pemindahan Ibu Kota Indonesia Menggunakan Algoritma Support Vector Machine (Svm) Sentiment Analysis On The Discussion Of Relocating I Ndonesia ' S Capital City Using The Support Vector Machine (Svm)," vol. 8, no. 1, pp. 147–156, 2021, doi: 10.25126/jtiik.202183944.
- [6] C.- Pandemic, B. Classifier, A. Wibowo, F. N. Hasan, L. A. Ramadhan, and R. Nurhayati, "Analisis Sentimen Opini Masyarakat Terhadap Keefektifan Pembelajaran Daring Selama Pandemi COVID-19 Menggunakan Naïve Bayes Classifier Analysis of Public Opinion Sentiment on the Effectiveness of Online Learning During the," vol. 4, pp. 239–248, 2022.
- [7] R. Azhar, A. Surahman, and C. Juliane, "Analisis Sentimen Terhadap Cryptocurrency Berbasis Python TextBlob Menggunakan Algoritma Naïve Bayes," vol. 6, pp. 267–281, 2022.
- [8] T. Widyanto, I. Ristiana, and A. Wibowo, "Komparasi Naïve Bayes dan SVM Analisis Sentimen RUU Kesehatan di Twitter," vol. 6, no. 3, pp. 147–161, 2023.
- [9] H. Wickham, "Journal of Statistical Software," vol. 59, no. 10, 2014.
- [10] P. Liu, T. Lei, Q. Xiang, Z. Wang, and J. Wang, "Animal Fur Recognition Algorithm Based on Feature Fusion Network," 2022.
- [11] I. H. Sarker, "Machine Learning: Algorithms, Real-World Applications and Research Directions," *SN Comput. Sci.*, vol. 2, no. 3, p. 160, 2021, doi: 10.1007/s42979-021-00592-x.
- [12] A. Roihan, T. T. Atmojo, R. A. Wardoyo, and M. S. T. Saputra, "Sentiment Analysis of Twitter Data on the 2024 Indonesian Presidential Election Using BERT," vol. 18, no. 1, pp. 28–34, 2025.
- [13] N. Chamidah, D. Widiyanto, H. B. Seta, and A. A. Aziz, "Revue d ' Intelligence Artificielle The Impact of Oversampling and Undersampling on Aspect-Based Sentiment Analysis of Indramayu Tourism Using Logistic Regression," vol. 38, no. 3, pp. 795–804, 2024.