
PERBANDINGAN NLP DAN SVM DALAM ANALISIS SENTIMEN KOMENTAR INSTAGRAM TERKAIT STIGMA MASYARAKAT TERHADAP BANJIR SUMATERA 2025

Muhammad Akbar Firdaus¹, Maisya Fitri Anugrah², Sri Hidayati³, Dedy Rahman Harahap⁴, Rendy Rabensi Sembiring⁵, Muhammad Syahputra Novelan⁶
Universitas Panca Budi, Medan
e-mail: afir832@gmail.com

Abstract: *This study aims to analyze public sentiment regarding the 2025 Sumatra flood based on Instagram comments using a Natural Language Processing (NLP) approach. The methods applied include IndoBERT and Support Vector Machine (SVM) with TF-IDF features. A total of 711 comments were collected through a crawling process and processed using preprocessing techniques. The results show that negative sentiment dominates at 44.7%, followed by positive (30.4%) and neutral (24.9%) sentiments. Model evaluation indicates that IndoBERT outperforms SVM with an accuracy of 74.8% compared to 66.4%. WordCloud visualization reveals dominant terms such as flood, government, forest, and palm oil, reflecting public concerns about environmental issues and government policies.*

Keyword: *Sentiment Analysis, Natural Language Processing (NLP), IndoBERT, Support Vector Machine (SVM), Sumatra Flood, Instagram*

Abstrak: Penelitian ini bertujuan menganalisis sentimen masyarakat terhadap banjir Sumatera 2025 berdasarkan komentar Instagram menggunakan pendekatan *Natural Language Processing* (NLP). Metode yang digunakan adalah IndoBERT dan *Support Vector Machine* (SVM) dengan fitur TF-IDF. Data sebanyak 711 komentar dikumpulkan melalui proses crawling dan diproses melalui tahap preprocessing. Hasil analisis menunjukkan bahwa sentimen negatif mendominasi sebesar 44,7%, diikuti positif 30,4% dan netral 24,9%. Evaluasi model menunjukkan IndoBERT memiliki akurasi lebih tinggi sebesar 74,8% dibandingkan SVM sebesar 66,4%. Visualisasi WordCloud menunjukkan kata dominan seperti banjir, pemerintah, hutan, dan sawit, yang mencerminkan fokus perhatian masyarakat terhadap isu lingkungan dan kebijakan pemerintah.

Kata kunci: Analisis Sentimen, *Natural Language Processing*(NLP), IndoBERT, *Support Vector Machine* (SVM), Banjir Sumatera, Instagram

PENDAHULUAN

Perkembangan teknologi informasi yang begitu pesat telah mengubah cara manusia berkomunikasi, berinteraksi, dan memperoleh informasi. Seiring bertambahnya waktu, komunikasi tidak lagi terbatas pada pesan singkat (SMS) atau panggilan telepon, tetapi telah beralih ke platform digital berbasis internet yang menawarkan kecepatan, efisiensi, dan jangkauan yang lebih luas, satu diantaranya adalah akun jejaring sosial. Jejaring sosial merupakan sebuah situs

berbasis pelayanan yang memungkinkan untuk membuat profil, melihat daftar pengguna yang tersedia, serta mengundang atau menerima teman untuk bergabung dalam situs tersebut (Firmansyah & Aditya, 2010). Jejaring sosial juga hadir sebagai ruang baru bagi masyarakat untuk berbagi informasi, berpendapat, serta membangun jejaring sosial secara real-time tanpa batasan geografis. Tampilan dasar situs jejaring sosial ini menampilkan halaman profil pengguna, yang di dalamnya terdiri dari identitas diri dan foto pengguna. Jejaring

sosial tersebut seperti facebook, twitter dan instagram. Perkembangan penggunaan jejaring sosial di Indonesia bisa dikatakan sangat pesat. Menurut studi *Nielsen*, tingkat pertumbuhan penggunaan Internet di Indonesia telah mencapai 26%, dan masyarakat Indonesia menghabiskan 1,5 jam sehari di Internet. Menurut ICT Watch, saat ini terdapat 180 juta pengguna ponsel di Indonesia dari 220 juta penduduk Indonesia. Menurut CNN Indonesia, Instagram mencapai 22 juta pengguna (Lesmana, 2012).

Jejaring Sosial atau Media sosial seperti Facebook, Twitter, dan Instagram menjadi sarana penting dalam aktivitas sosial masyarakat modern. Media sosial memfasilitasi pengguna untuk membuat konten, berinteraksi, dan membangun jejaring secara digital (Ellison & B, 2008). Instagram, khususnya, berkembang menjadi platform visual yang memungkinkan pengguna berbagi informasi berupa berita, foto, video, dan cerita secara kreatif serta berinteraksi melalui komentar dan berbagai fitur lainnya (Hu et al., 2014). Selain menjadi media berbagi informasi, Instagram juga berfungsi sebagai ruang diskusi publik di mana berbagai isu sosial, termasuk peristiwa bencana, sering kali dibahas secara intens. Karakteristik inilah yang membuat Instagram menjadi sumber data yang menarik untuk dianalisis, terutama terkait ekspresi sentimen penggunanya dikolom komentar instagram.

Salah satu kejadian yang menarik banyak perhatian di media sosial adalah Banjir Sumatera 2025, di mana berbagai akun berita baik lokal maupun nasional membagikan informasi tentang dampak, area yang terkena, dan respons dari pemerintah. Kejadian ini memicu banyaknya komentar dari masyarakat, terutama mengenai penyebab banjir tersebut. Komentar-komentar tersebut mencerminkan pandangan publik terhadap berbagai faktor penyebab bencana, seperti kerusakan lingkungan, pembangunan yang tidak terencana, perubahan iklim, serta kritik terhadap kinerja pemerintah. Ini menunjukkan

bahwa media sosial tidak hanya berfungsi sebagai sarana penyebaran informasi, tetapi juga refleksi opini masyarakat digital terhadap isu sosial dan kebencanaan (Ellison & B, 2008).

Tingginya jumlah komentar dari masyarakat memberikan berbagai macam reaksi, mulai dari dukungan, rasa empati, kritik yang membangun, hingga komentar yang penuh emosi yang menunjukkan pandangan positif dan negatif. Untuk mengetahui bagaimana masyarakat memahami faktor-faktor yang menyebabkan banjir, diperlukan analisis yang terorganisir tentang sentimen yang muncul. Analisis sentimen sebagai pendekatan linguistik dan komputasional memungkinkan peneliti mengidentifikasi emosi, opini, dan sikap publik yang terkandung dalam teks komentar (Liu et al., 2012).

Dalam konteks pengolahan data teks berskala kecil maupun besar seperti komentar Instagram, Natural Language Processing (NLP) merupakan metode yang banyak digunakan untuk melakukan prapemrosesan, ekstraksi fitur, dan analisis makna teks (Jurafsky & Martin, 2021). Natural language processing merupakan pembelajaran mesin yang berfokus pada bahasa manusia dalam memahami, menafsirkan, dan menghasilkan bahasa alami manusia. Natural language processing juga merupakan salah satu bagian dari artificial intelligence (R. Gupta *et al.*, 2021). Pada penelitian modern, pendekatan deep learning semakin banyak diterapkan untuk analisis sentimen karena mampu memahami konteks bahasa secara lebih mendalam. Salah satu model yang populer adalah IndoBERT, yaitu model language representation berbasis Transformer yang dilatih menggunakan korpus berbahasa Indonesia sehingga lebih akurat dalam memahami struktur dan makna teks Indonesia (Koto et al., 2020; Devlin et al., 2019).

Selain itu, keterbatasan ketersediaan dataset berlabel dalam bahasa Indonesia menjadi tantangan tersendiri dalam penelitian analisis

sentimen (Pranata et al., 2024). Untuk mengatasi hal tersebut, penelitian ini menggunakan pendekatan pelabelan semi-otomatis dengan memanfaatkan model IndoBERT sebagai pseudo-labeler untuk menghasilkan dataset berlabel yang kemudian digunakan dalam proses pelatihan model SVM (Setiawan Y & Wulandari, L.A., 2025).

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk melakukan analisis sentimen terhadap komentar Instagram terkait peristiwa Banjir Sumatera 25 November 2025 dengan membandingkan dua pendekatan, yaitu metode deep learning menggunakan IndoBERT dan metode machine learning menggunakan *Support Vector Machine* (SVM). Perbandingan ini dilakukan untuk mengevaluasi kinerja kedua metode dalam mengklasifikasikan sentimen ke dalam kategori positif, negatif, dan netral, serta menganalisis perbedaan dari segi akurasi dan efisiensi komputasi.

Kontribusi dari penelitian ini adalah memberikan analisis komparatif antara metode deep learning dan machine learning dalam konteks analisis sentimen berbahasa Indonesia, khususnya pada data komentar Instagram terkait isu bencana. Hasil penelitian ini diharapkan dapat menjadi referensi dalam pemilihan metode yang tepat sesuai dengan kebutuhan analisis serta keterbatasan sumber daya yang tersedia.

METODE

Penelitian ini menggunakan pendekatan kuantitatif dalam menganalisis sentimen publik pada komentar Instagram terkait bencana banjir di Pulau Sumatera tahun 2025. Data teks diolah menggunakan teknik Natural Language Processing (NLP) seperti pembersihan teks, *case folding*, tokenisasi, dan *stopword removal* untuk meningkatkan kualitas data (Jurafsky & Martin, 2021; Liu, 2012). Pada tahap awal, klasifikasi sentimen dilakukan menggunakan IndoBERT sebagai model

deep learning berbasis *transformer* yang mampu mengidentifikasi sentimen positif, negatif, dan netral serta digunakan sebagai metode pelabelan semi-otomatis (*pseudo-labeling*) (Hidayat, W. A., & Nastiti, V. R. S., 2024). Dataset berlabel yang dihasilkan kemudian digunakan untuk melatih model Support Vector Machine (SVM) dengan fitur TF-IDF sebagai representasi pendekatan *machine learning* tradisional. Selanjutnya, dilakukan perbandingan kinerja antara IndoBERT dan SVM dalam mengklasifikasikan sentimen, serta analisis diperkuat dengan visualisasi WordCloud untuk melihat frekuensi kemunculan kata dominan sebagai pendukung interpretasi opini masyarakat (Wibowo et al., 2024).

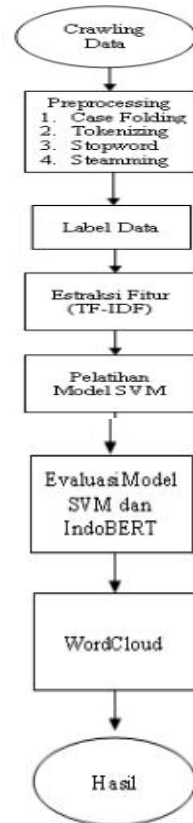
Data yang digunakan dalam penelitian ini berupa komentar publik yang dikumpulkan dari beberapa akun Instagram yang terkait dengan bencana banjir Sumatera 2025 seperti akun berita, pemerintah dan tokoh publik. Pengumpulan data dilakukan dengan metode web scraping selama periode bencana dimulai 24 November 2025 hingga 4 Desember 2025. Jumlah total komentar yang akan dikumpulkan sebanyak 711 komentar dengan rincian seperti tabel data penelitian berikut:

Tabel 1 Data Penelitian

Asal Data Komentar	Jumlah Komentar
Akun Berita 3 Provinsi	328
Akun Pemerintah	209
Tokoh Publik	174

Setelah data dikumpulkan, dilakukan tahap *preprocessing* untuk membersihkan dan menyiapkan teks melalui proses seperti *case folding*, *tokenizing*, dan *stopword removal*. Data yang telah dibersihkan kemudian dilakukan proses pelabelan menggunakan model IndoBERT untuk mengklasifikasikan sentimen komentar terkait bencana banjir Sumatera 2025 ke dalam tiga kategori, yaitu positif, netral, dan negatif. Hasil pelabelan tersebut selanjutnya digunakan sebagai dataset

berlabel untuk proses pelatihan model *Support Vector Machine* (SVM) dengan fitur TF-IDF. Setelah itu, dilakukan evaluasi untuk membandingkan kinerja antara model IndoBERT dan SVM. Adapun alur metodologi penelitian ini ditunjukkan pada Gambar 1.



Gambar 1 Tahapan Penelitian

HASIL DAN PEMBAHASAN

Bab ini menyajikan hasil serta pembahasan dari proses analisis sentimen terhadap komentar masyarakat di Instagram terkait peristiwa banjir Sumatera tahun 2025. Proses penelitian dilakukan secara bertahap mulai dari pengumpulan data, preprocessing, pelabelan sentimen menggunakan model IndoBERT, hingga pelatihan model *Support Vector Machine* (SVM) berbasis fitur TF-IDF sebagai representasi numerik teks. Analisis sentimen sendiri merupakan pendekatan dalam *Natural Language Processing* (NLP) yang digunakan untuk mengidentifikasi opini atau emosi dalam

suatu teks ke dalam kategori tertentu seperti positif, netral, dan negatif (Tarumingkeng, R. C., 2024). Hasil yang diperoleh kemudian dianalisis untuk mengetahui distribusi sentimen yang terbentuk serta mengevaluasi kinerja kedua metode dalam mengklasifikasikan opini masyarakat, sehingga dapat memberikan gambaran mengenai persepsi publik terhadap peristiwa tersebut.

Proses crawling data merupakan tahap awal penelitian yang bertujuan mengumpulkan komentar masyarakat terkait banjir Sumatera melalui Instagram (Haris & D. Eka Ratnawati.,2023). Pengambilan data dilakukan menggunakan tools Python dan Apify, kemudian disimpan dalam format Excel, digabungkan menjadi satu file, dan menghasilkan total 711 komentar yang selanjutnya dikonversi ke format JSON untuk proses berikutnya. Data tersebut berbentuk tabel yang memuat kolom teks berisi opini masyarakat terhadap peristiwa banjir seperti pada Tabel 2 Berikut.

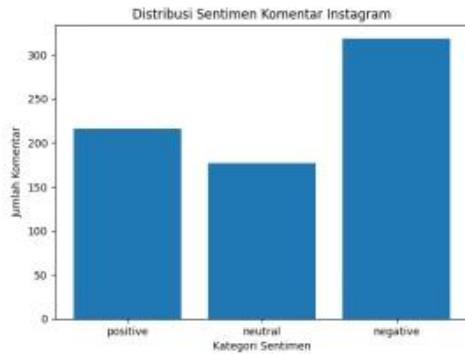
Tabel 2 Bentuk Dataset

Bentuk Data	Data
Text	TPL semakin kaya, rakyat semakin sengsara 😞
Text	Kalau ada audit lingkungan yang jujur, TPL harusnya habis!
Text	Terima kasih relawan dan donatur! Kalian luar biasa ❤️

Selanjutnya, data teks yang telah dikumpulkan diproses pada tahap preprocessing, yaitu tahap untuk membersihkan dan mempersiapkan data agar siap diolah pada proses analisis berikutnya.

Tahapan preprocessing yang dilakukan dalam penelitian ini meliputi case folding untuk menyeragamkan huruf, stopwords removal untuk menghapus kata-kata yang tidak memiliki makna signifikan, tokenizing untuk memecah teks menjadi kata-kata, serta stemming

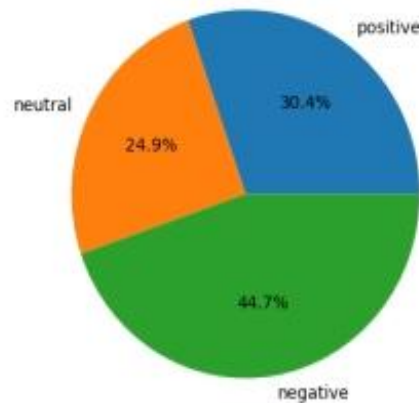
mengidentifikasi topik dominan secara cepat, sedangkan diagram batang dan diagram lingkaran digunakan untuk memberikan gambaran distribusi jumlah dan persentase sentimen secara lebih terstruktur. Berikut Visual dari BarChart dan Piechart.



Gambar 3 Visual BarChart

Berdasarkan diagram batang, dari total 711 data komentar yang dianalisis, diperoleh 318 sentimen negatif, 216 sentimen positif, dan 177 sentimen netral. Hal ini menunjukkan bahwa sentimen negatif mendominasi dibandingkan kategori lainnya. Sementara itu, pada diagram lingkaran terlihat bahwa sentimen negatif memiliki proporsi terbesar sebesar 44,7%, diikuti sentimen positif sebesar 30,4%, dan sentimen netral sebesar 24,9%. Distribusi ini menunjukkan bahwa mayoritas masyarakat cenderung memberikan respon negatif terhadap peristiwa banjir, yang mencerminkan adanya kritik, ketidakpuasan, serta kekhawatiran terhadap kondisi yang terjadi.

Persentase Sentimen Komentar Instagram



Gambar 4 Visualisasi Piechart
SIMPULAN

Penelitian ini berhasil menganalisis sentimen publik terhadap komentar masyarakat di Instagram terkait bencana banjir di Pulau Sumatera tahun 2025 menggunakan pendekatan Natural Language Processing (NLP) dengan model IndoBERT serta metode Support Vector Machine (SVM). Berdasarkan hasil analisis terhadap 711 komentar, diperoleh bahwa sentimen negatif menjadi kategori yang paling dominan dibandingkan sentimen positif dan netral. Hal ini menunjukkan bahwa masyarakat cenderung mengekspresikan kritik, kekhawatiran, dan ketidakpuasan terhadap dampak bencana, khususnya terkait isu kerusakan lingkungan seperti deforestasi dan perkebunan sawit, serta respons pemerintah dalam penanganan bencana. Sementara itu, sentimen netral umumnya berisi informasi kondisi di lapangan, dan sentimen positif mencerminkan dukungan, empati, serta apresiasi terhadap bantuan yang diberikan.

Selain itu, hasil visualisasi WordCloud menunjukkan kata-kata dominan seperti *banjir*, *pemerintah*, *hutan*, *sawit*, dan *bantuan*, yang memperkuat bahwa isu lingkungan dan kebijakan pemerintah menjadi perhatian utama masyarakat. Dari sisi evaluasi model, metode NLP dengan IndoBERT menunjukkan performa yang lebih baik dibandingkan SVM, yang menandakan

bahwa pendekatan berbasis deep learning lebih mampu memahami konteks bahasa dalam analisis sentimen. Dengan demikian, penelitian ini memberikan gambaran mengenai persepsi masyarakat secara real-time melalui media sosial dan dapat menjadi referensi bagi pemerintah serta pihak terkait dalam meningkatkan strategi mitigasi bencana, pengelolaan lingkungan, dan komunikasi publik yang lebih efektif.

DAFTAR PUSTAKA

- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. NAACL-HLT.
- Ellison, & B, N. (2008). *Social Network Sites : Definition , History , and Scholarship*. 13, 210–230. <https://doi.org/10.1111/j.1083-6101.2007.00393.x>
- Firmansyah, & Aditya. (2010). *Situs Jejaring Sosial Menggunakan Elgg*.
- Haris and D. Eka Ratnawati, “Analisis Sentimen berbasis Aspek terhadap Data Ulasan menggunakan Metode K-Nearest Neighbor (Studi Kasus: Aplikasi Olsera POS),” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 7, no. 6, pp. 3041–3046, 2023, [Online]. Available: <http://j-ptiik.ub.ac.id>
- Hidayat, W. A., & Nastiti, V. R. S. (2024). Perbandingan kinerja pre-trained IndoBERT-base dan IndoBERT-lite pada klasifikasi sentimen ulasan TikTok Tokopedia Seller Center dengan model IndoBERT. *JSiI (Jurnal Sistem Informasi)*, 11(2), 13-20.
- Hu, Manikonda, & Kambhampati. (2014). *What We Instagram : A First Analysis of Instagram Photo Content and User Types*. *McCune 2011*, 595–598.
- Jurafsky, D., & Martin, J. H. (2021). *Speech and Language Processing*.
- Julianto, I. T., & Lindawati, L. (2022). Analisis sentimen terhadap sistem informasi akademik Institut Teknologi Garut. *Jurnal Algoritma*, 19(1), 458-468
- Koto, F., Rahimi, A., Lau, J., & Baldwin, T. (2020). IndoLEM and IndoBERT: A Benchmark Dataset and Pre-trained Language Model for Indonesian NLP. COLING.
- Lesmana, Gusti Ngurah Aditya. 2012. “Analisis Pengaruh Media Sosial Twitter Terhadap Pembentukan Brand Attachment.” Universitas Indonesia.
- Liu, K., Li, W., & Guo, M. (2012). *Emoticon Smoothed Language Models for Twitter Sentiment Analysis*. 1678–1684.
- Pranata, J. O. N. I., Agustian, S., & Haerani, E. (2024). Penggunaan Model Bahasa indoBERT pada Metode Random Forest Untuk Klasifikasi Sentimen Dengan Dataset Terbatas. *vol, 6*, 1668-1676
- R. Gupta, D. Srivastava, M. Sahu, S. Tiwari, R. K. Ambasta, and P. Kumar, “Artificial intelligence to deep learning: machine intelligence approach for drug discovery,” *Mol Divers*, vol. 25, no. 3, pp. 1315–1360, Aug. 2021, doi: 10.1007/s11030-021-10217-3
- Setiawan, Y., & Wulandhari, L. A. (2025). Comparative Analysis of IndoBERT and LSTM for Multi-Label Text Classification of Indonesian Motivation Letter. *Jurnal Online Informatika*, 10(2), 260-269.
- Tarumingkeng, R. C. (2024). Natural Language Processing (NLP). *RUDYCT e-PRESS*, no.
- Wibowo, J. A., Mawardi, V. C., & Sutrisno, T. (2024). Visualisasi word cloud hasil analisis sentimen berbasis fitur layanan aplikasi gojek dengan support vector machine. *Jurnal Serina Sains, Teknik dan Kedokteran*, 2(1), 61-70.