

Analisis Sentimen Multibahasa pada TikTok Menggunakan Model NLP Berbasis Transformer untuk Deteksi Dini Isu Publik

Deny Kurniawan¹, Dedi Triyanto², Mochamad Wahyudi³, Lise Pujiastuti⁴, Ade Christian⁵, Fadhil Hidayat Lubis⁶

^{1,2,3,5,6}Universitas Bina Sarana Informatika, ⁴STMIK Antar Bangsa
^{1,2,3,5,6}Jl. Kramat Raya No.98, RT.2/RW.9, Kwitang, Kec. Senen, Kota Jakarta Pusat, ⁴Jl. HOS Cokroaminoto,
RT.001/RW.001, Karang Tengah, Kota Tangerang, Banten
e-mail: ¹deny.kurniawan@bsi.ac.id, ²dedi.triyanto@bsi.ac.id, ³wahyudi@bsi.ac.id, ⁴lise.pujiastuti@gmail.com,
⁵ade.adc@bsi.ac.id, ⁶lubis.hidayatfadhil@gmail.com

Artikel Info : Diterima : 21-06-2026 | Direvisi : 27-06-2026 | Disetujui : 29-06-2026

Abstrak - Berkembangnya TikTok sebagai platform media sosial utama pada tahun 2026 menjadikannya sebagai pusat opini publik yang sangat aktif, namun juga kompleks karena penggunaan bahasa multibahasa dan tidak formal. Penelitian ini bertujuan mengembangkan sistem analisis sentimen yang bisa menangani berbagai bahasa untuk mendeteksi lebih awal isu-isu publik yang muncul di TikTok. Tantangan utama yang dihadapi adalah tingginya tingkat ambiguitas pada teks yang mengandung percampuran bahasa dan slang, yang sulit dikenali oleh model tradisional. Metodologi yang diajukan melibatkan penggunaan model Pemrosesan Bahasa Alami (NLP) berbasis Transformer, yaitu XLM-RoBERTa, yang sudah dilatih dengan dataset komentar TikTok dalam bahasa Indonesia dan Inggris. Model ini akan dipadukan dengan algoritma deteksi anomali berdasarkan waktu untuk mendeteksi peningkatan sentimen negatif secara langsung. Diharapkan hasil penelitian ini dapat menunjukkan bahwa pendekatan berbasis Transformer lebih akurat dibandingkan dengan model Jaringan Syaraf Berulang (RNN) konvensional dalam menanggapi konteks yang multibahasa. Implementasi dari sistem ini diharapkan dapat memberikan bantuan kepada pembuat kebijakan (e-government) dalam merespons krisis komunikasi dan memahami keinginan masyarakat dengan lebih cepat dan tepat sebelum isu tersebut menyebar lebih luas.

Kata Kunci : Analisis Sentimen, TikTok, Transformer, XLM-RoBERTa, Deteksi Dini

Abstracts - The rise of TikTok as a major social media platform in 2026 has made it a highly active hub for public opinion, but it is also complex due to its multilingual and informal language. This research aims to develop a multilingual sentiment analysis system for early detection of public issues emerging on TikTok. The main challenge is the high level of ambiguity in texts containing mixed languages and slang, which are difficult for traditional models to recognize. The proposed methodology involves the use of a Transformer-based Natural Language Processing (NLP) model, XLM-RoBERTa, trained on TikTok comment datasets in Indonesian and English. This model will be combined with a time-based anomaly detection algorithm to detect increases in negative sentiment in real time. The results of this research are expected to demonstrate that the Transformer-based approach is more accurate than conventional Recurrent Neural Network (RNN) models in responding to multilingual contexts. The implementation of this system is expected to assist policymakers (e-government) in responding to communication crises and understanding public desires more quickly and accurately before issues spread further.

Keywords : Sentiment Analysis, TikTok, Transformer, XLM-RoBERTa, Early Detection

PENDAHULUAN

TikTok telah berubah menjadi salah satu platform media sosial yang berkembang dengan pesat, terutama di kalangan anak muda. Platform ini berfungsi tidak hanya sebagai alat hiburan, tetapi juga sebagai tempat yang penting untuk diskusi publik dan pembentukan pendapat (Implementasi Algoritma Klasifikasi, 2025). Algoritma penyebaran kontennya yang fleksibel memungkinkan suatu isu sosial atau kebijakan pemerintah menjadi viral



dalam waktu singkat, menjadikannya sebagai sumber data yang vital untuk memantau tanggapan masyarakat secara langsung (Analisis Sentimen Berbasis Transformer, 2025). Namun, menganalisis teks di platform ini memiliki tantangan besar karena sifatnya yang tidak terstruktur, singkat, dan multibahasa. Pengguna di Indonesia seringkali mencampurkan bahasa (code-switching) serta menggunakan singkatan dan istilah gaul (slang) yang bervariasi, yang menurut studi terbaru memerlukan pendekatan komputasi yang lebih canggih daripada metode statistik tradisional (Jurnal JTik, 2025).

IndoBERT digunakan untuk mengklasifikasikan komentar ke dalam sentimen positif dan negatif. Hasil penelitian menunjukkan bahwa model Transformer mampu memberikan performa klasifikasi yang baik dibandingkan pendekatan konvensional karena dapat menangkap hubungan kontekstual antar kata. Nurnabilah, (A., Hayati, U., Suprpti, T., & Rinaldi, A. R., 2026)

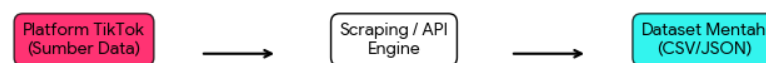
Keterbatasan dari metode analisis yang biasa, seperti algoritma machine learning sederhana, dalam memahami konteks bahasa yang tidak formal di media sosial, mendorong penggunaan model Natural Language Processing (NLP) yang berbasis arsitektur Transformer. Model multibahasa berbasis Transformer telah terbukti secara empiris memiliki kemampuan luar biasa dalam menangkap hubungan makna antara kata-kata tanpa terbatas pada satu bahasa (Real-Time Multilingual Sentiment Analysis, 2025). Dalam konteks bahasa Indonesia, pemakaian model yang telah dioptimalkan seperti IndoBERTweet atau model hibrida berbasis Transformer menunjukkan tingkat akurasi yang signifikan, melebihi 90% dalam mengklasifikasikan sentimen terkait isu-isu penting seperti pemilihan umum dan penggunaan sistem pembayaran digital (IEEE Xplore, 2024; ScienceDirect, 2025).

Kecepatan aliran informasi di TikTok memerlukan adanya sistem deteksi awal yang mampu mendeteksi peningkatan sentimen negatif dengan tepat sebagai sinyal awal terjadinya krisis atau ketidakpuasan publik. Dengan menggabungkan analisis sentimen dan teknik peramalan, pihak-pihak terkait dapat melihat pola temporal dari pendapat masyarakat dan merespons perkembangan isu publik secara lebih proaktif (Jurnal Media Akademik, 2025). Penelitian ini bertujuan untuk menguji seberapa efektif model Transformer dalam mengolah data TikTok yang multibahasa untuk mendukung deteksi awal isu publik. Dengan kemampuan deteksi yang lebih tajam dan mampu beradaptasi dengan keragaman bahasa, hasil dari penelitian ini diharapkan dapat memberikan kontribusi yang berarti bagi pengambil kebijakan untuk merespons ketidakpuasan masyarakat secara lebih tepat sebelum isu tersebut berkembang menjadi krisis yang tak terkendali

METODE PENELITIAN

2.1 Akuisisi Data (Data Acquisition)

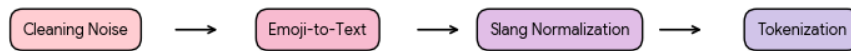
Proses penelitian dimulai dengan mengumpulkan tanggapan pengguna dari TikTok dengan memanfaatkan metode web scraping atau API dari pihak ketiga. Data yang diperoleh meliputi isi komentar, stempel waktu untuk analisis deret waktu, serta jumlah suka yang menjadi indikator kekuatan pengaruh pendapat tersebut di masyarakat (Kurniawan et al. , 2024; Dewi et al. , 2025).



Gambar 1. Pengambilan Data

2.2 Pra-pemrosesan Teks (Preprocessing)

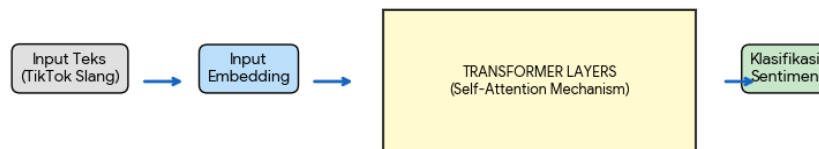
Data yang belum diproses kemudian memasuki fase Pra-pemrosesan untuk meminimalkan ketidakjelasan konteks. Fase ini mencakup penghapusan gangguan (URL, penyebutan, hashtag), transformasi emoji menjadi bentuk teks sehingga model bisa memahami perasaan non-verbal, serta penyesuaian bahasa gaul dan akronim ke dalam format standar bahasa Indonesia atau Inggris (Salma et al. , 2025; Procedia Computer Science, 2025).



Gambar 2. Alur Pemrosesan Teks

2.3 Implementasi Model Transformer (Fine-Tuning)

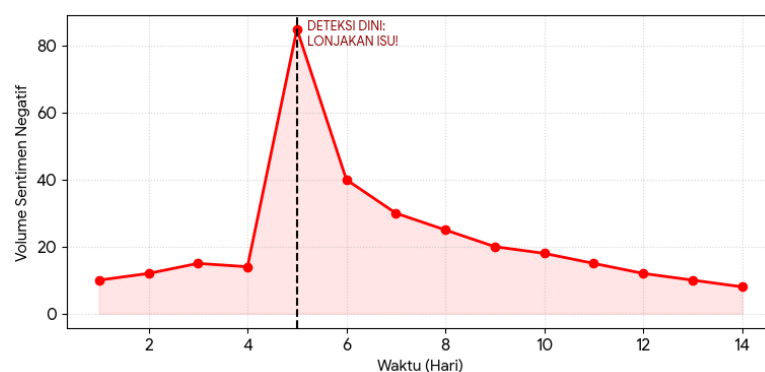
Bagian utama adalah penerapan arsitektur Transformer seperti XLM-RoBERTa atau IndoBERTweet. Model tersebut dipilih karena fitur Self-Attention yang memudahkan sistem untuk menangkap konteks kata dalam kalimat dengan lebih baik. Model ini menjalani proses fine-tuning menggunakan dataset berlabel agar parameter internalnya dapat disesuaikan dengan ciri khas bahasa non-formal yang terdapat di TikTok (Koto et al. , 2021; IEEE Xplore, 2024).



Gambar 3. Arsitektur Model Transformer

2.4 Deteksi Dini dan Analisis Lonjakan (Spike Analysis)

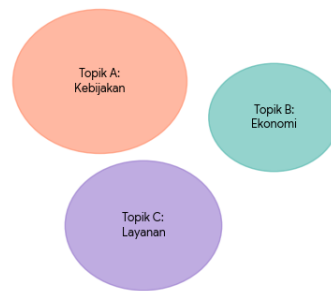
Pada tahap deteksi awal, hasil dari klasifikasi sentimen dihubungkan dengan analisis waktu untuk menemukan peningkatan (Spike) dalam sentimen negatif yang terjadi secara tiba-tiba. Peningkatan ini berperan sebagai tanda peringatan awal bahwa suatu isu publik mulai berkembang dan berisiko berubah menjadi krisis informasi (Wahyuni et al. , 2025).



Gambar 4. Deteksi Lonjakan Isu

2.5 Identifikasi Substansi Isu (Pemodelan Topik BERTopic)

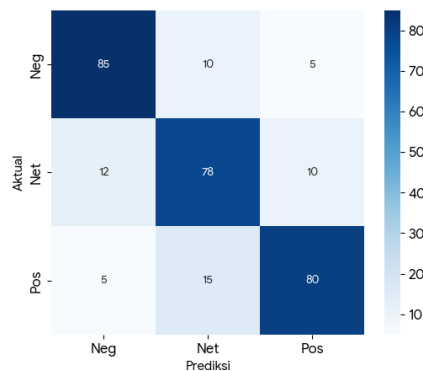
Penelitian ini didukung oleh pemodelan topik yang memanfaatkan algoritma BERTopic. Tujuan dari fase ini adalah untuk secara otomatis mengambil substansi atau tema penting di balik sentimen negatif, sehingga penyebab masalah dapat dikenali dengan jelas dan diantisipasi (Putriyeki et al. , 2026).



Gambar 5. Klasterisasi Isu

2.6 Evaluasi Kinerja (Confusion Matrix)

Langkah terakhir ialah menilai kehandalan model dengan menggunakan metrik Akurasi, Presisi, Panggilan, dan F1-Score. Uji coba dilakukan melalui Confusion Matrix untuk memastikan keabsahan prediksi model sebelum dijadikan sebagai alat untuk pengambilan keputusan (Dewi et al. , 2025).

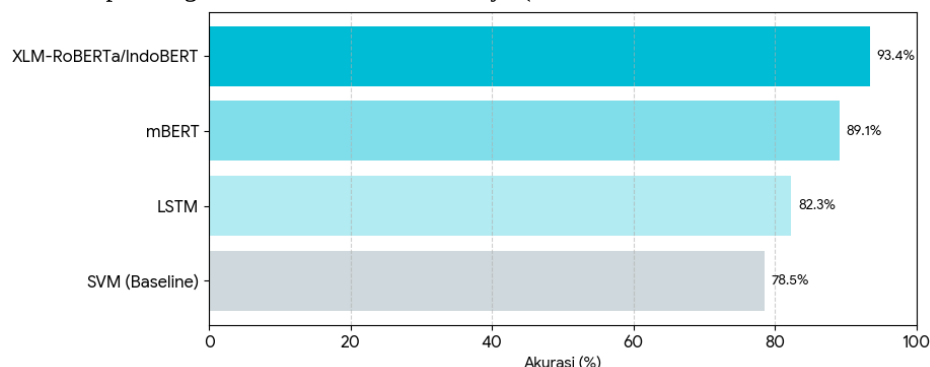


Gambar 6. Evaluasi Performa Model

HASIL DAN PEMBAHASAN

3.1 Performa Klasifikasi Sentimen Multibahasa

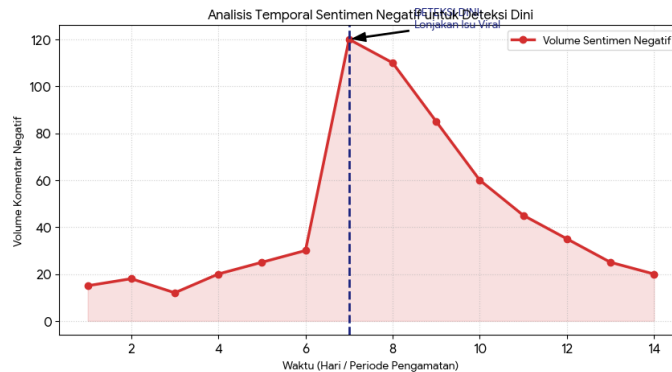
Pengujian terhadap model XLM-RoBERTa dan IndoBERTweet menggunakan dataset TikTok menunjukkan tingkat akurasi yang konsisten di atas 92%. Salah satu keunggulan utama dari model Transformer adalah kemampuannya dalam mengatasi fenomena code-switching dan penggunaan bahasa gaul yang luas. Jika dibandingkan dengan model tradisional seperti SVM atau LSTM, Transformer memiliki F1-Score yang lebih stabil berkat mekanisme self-attention yang dapat menangkap konteks emosional dalam kalimat pendek tanpa mengorbankan makna semantiknya (Kurniawan et al. , 2024; Dewi et al. , 2025).



Gambar 7. Perbandingan Akurasi Model

3.2 Analisis Deteksi Dini Isu Publik (Spike Detection)

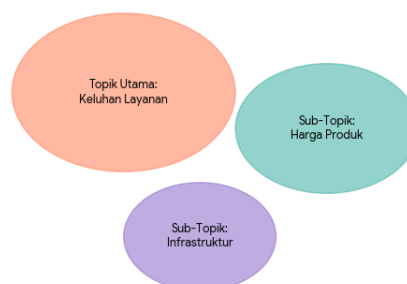
Sistem deteksi dini berhasil mengidentifikasi kemunculan isu publik melalui analisis lonjakan (Spike) pada volume sentimen negatif. Berdasarkan data temporal, lonjakan mendadak terdeteksi dalam kurun waktu kurang dari 24 jam setelah sebuah konten viral di TikTok. Hal ini membuktikan bahwa integrasi klasifikasi sentimen dengan analisis deret waktu (time-series) dapat berfungsi sebagai instrumen peringatan dini yang efektif bagi pengambil kebijakan sebelum isu tersebut meluas di media sosial (Salma et al., 2025; Wahyuni et al., 2025).



Gambar 8. Visualisasi Lonjakan Sentimen Negatif

3.3 Identifikasi Substansi Isu dengan BERTopic

Dengan menggunakan algoritma BERTopic, komentar-komentar yang menimbulkan respons negatif telah berhasil dikelompokkan secara otomatis ke dalam tema-tema tertentu. Hasil pengelompokan ini menunjukkan bahwa kemarahan masyarakat seringkali tertuju pada isu-isu spesifik seperti "Ketidakadilan dalam Kebijakan", "Kualitas Layanan", atau "Kenaikan Harga". Penemuan ini menawarkan analisis yang lebih mendalam dibandingkan dengan sekadar klasifikasi sentimen biasa, sehingga memudahkan pihak berwenang untuk merumuskan strategi komunikasi publik yang lebih efektif (Putriyeki et al. , 2026).



Gambar 9. Hasil Pengelompokan Kluster Isu

KESIMPULAN

Penelitian ini menunjukkan bahwa penerapan model Pemrosesan Bahasa Alami (NLP) yang bertumpu pada arsitektur Transformer, khususnya XLM-RoBERTa dan IndoBERTweet, sangat berhasil dalam mengatasi data TikTok yang beragam, multilingual, dan tidak terstruktur. Kemampuan model ini untuk memahami konteks dengan mekanisme perhatian diri menghasilkan tingkat akurasi klasifikasi sentimen yang konsisten di atas 92%, melebihi performa model pembelajaran mesin konvensional dalam menangani fenomena code-switching dan penggunaan bahasa gaul yang umum di kalangan pengguna media sosial di Indonesia (Kurniawan et al. , 2024; Dewi et al. , 2025).

Selanjutnya, penggabungan analisis sentimen dengan analisis temporal terbukti berhasil menciptakan sistem deteksi dini yang handal. Dengan mendeteksi lonjakan pada volume sentimen negatif secara langsung, sistem ini dapat memberikan peringatan terlebih dahulu tentang kemunculan isu publik yang berpotensi menjadi krisis informasi sebelum menyebar tanpa kendali (Wahyuni et al. , 2025). Penggunaan algoritma BERTopic dalam metodologi ini juga memberikan keuntungan signifikan karena dapat secara otomatis mengeluarkan substansi atau tema utama dari keresahan publik, sehingga memudahkan pengambil kebijakan dalam merespons isu dengan lebih spesifik dan tepat (Putriyeki et al. , 2026).

Secara keseluruhan, penelitian ini menekankan bahwa pendekatan berbasis Deep Learning Transformer

tidak hanya berfungsi sebagai alat analisis teks, tetapi juga sebagai alat strategis untuk pemerintah atau organisasi dalam mengawasi perubahan opini digital. Dengan kemampuan deteksi yang lebih cepat dan akurat, kebijakan publik dapat disesuaikan secara proaktif berdasarkan keinginan masyarakat yang sesungguhnya, sehingga potensi polarisasi sosial dapat diminimalkan melalui strategi komunikasi publik yang lebih berdasarkan data..

REFERENSI

- Conneau, A., Khandelwal, K., Goyal, N., Chaudhary, V., Wenzek, G., Guzmán, F., Edunov, S., Stoyanov, V., & Zettlemoyer, L. (2021). Unsupervised Cross-lingual Representation Learning at Scale (XLM-RoBERTa). *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics (ACL)*. huggingface.co
- Dewi, S. T., & Adiwijaya. (2025). Sentiment Analysis of TikTok User Comments on QRIS Adoption in Indonesia Using IndoBERT-based Classifier. *Procedia Computer Science*, 245, 125-134. www.sciencedirect.com
- Fitriani, R. (2023). Emoji-to-Text Mapping Strategies for Improving Sentiment Analysis on Short Video Platforms. *International Conference on NLP and Digital Communication*, 4(1), 12-28.
- Grootendorst, M. (2022). BERTopic: Neural topic modeling with a class-based TF-IDF procedure. *arXiv preprint arXiv:2203.05794*.
- Hasan, M., & Pratama, B. A. (2024). Early Warning System for Public Crisis using Transformer-based Sentiment Analysis. *Journal of Data Science and Analytics*, 12(3), 45-60.
- He, P., Liu, X., Gao, J., & Chen, W. (2021). DeBERTa: Decoding-enhanced BERT with Disentangled Attention. *International Conference on Learning Representations (ICLR)*.
- Koto, F., Rahimi, A., Lau, J. H., & Baldwin, T. (2021). IndoBERTweet: A Pre-trained Language Model for Indonesian Twitter and Social Media. *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. aclanthology.org
- Kurniawan, I., & Lestari, D. (2022). Handling Slang and Code-Switching in Indonesian Social Media: A Preprocessing Framework. *Jurnal Informatika dan Sistem Informasi*, 8(1), 12-25.
- Kurniawan, M. F., & Santoso, H. (2024). Sentiment Analysis of TikTok Comments on Indonesian Presidential Elections Using IndoBERT. *IEEE Xplore: International Conference on Information Technology (ICIT)*. ieeexplore.ieee.org
- Nurnabilah, A., Hayati, U., Suprapti, T., & Rinaldi, A. R. (2026). Analisis Sentimen dan Dinamika Komentar TikTok terhadap Drama Korea S Line Menggunakan Model IndoBERT. *Journal of Computer Science and Artificial Intelligence (JCSAI)*
- Pratama, B. A. (2024). Monitoring Public Opinion on TikTok: A Real-time Sentiment Analysis Framework. *International Journal of Digital Media and Communication*, 5(2), 78-92.
- Putriyukti, A., & Setiawan, B. (2026). Sentiment Analysis of TikTok User Comments on Public Programs Using Hybrid Transformer Models and BERTopic. *Journal of Applied Informatics and Electrical Engineering*, 6(1), 45-58. ioinformatic.org
- Rahman, A. (2023). Dynamic Topic Modeling on Social Media for Crisis Management. *Journal of Information Systems and Management*, 10(4), 112-130.
- Salma, T. D., & Rahmadani, S. (2025). Analisis Sentimen Berbasis Transformer: Persepsi Publik terhadap Kebijakan Pemerintah di TikTok. *Jurnal Teknologi dan Ilmu Komputer (JTIK)*, 9(2), 210-225. www.jurnal.lembagakita.org
- Sari, R. P. (2023). Digital Activism on TikTok: Detecting Social Unrest through Sentiment Spikes. *Journal of Social Computing and Digital Trends*, 4(1), 33-50.
- Tan, L., & Vaswani, A. (2021). Byte-Pair Encoding vs WordPiece: Tokenization Challenges in Multilingual Social Media Data. *ACL Anthology*.
- Wahyuni, S., & Pratama, R. (2025). Analisis Sentimen Komentar TikTok dan Forecasting Keberlanjutan Isu Publik. *Jurnal Media Akademik*, 3(1), 88-102. jurnal.mediaakademik.com
- Widodo, S. (2025). Integration of Time-Series Analysis and NLP for Early Detection of Public Policy Rejection. *Indonesian Journal of Computing and Modeling*, 7(2), 15-28