DIGIVENTURE
Journal of Applied Digital Business Research

*Vol. 2, No. 1 (2026), hlm. 161-183*
*e-ISSN: 3123-5212*
*Situs web: https://journal.unesa.ac.id/digiventure*

# Analisis Sentimen dan *Topic Modeling* terhadap Promotor Konser K-Pop di Indonesia

Ansheila Destiara Nur Salsabila[1] dan Anita Safitri[2]

[1,2] Program Studi Bisnis Digital, Universitas Negeri Surabaya, Indonesia

*) Corresponding Author (Email: ansheila.22128@mhs.unesa.ac.id)

***ABSTRACT***

*Audience opinions on X (Twitter) reflect public sentiment toward K-pop concert promoters in Indonesia. This study aims to analyze public sentiment, identify the main topics of public conversations, and map these topics based on sentiment polarity. A computational quantitative approach was employed using deep learning-based sentiment analysis and topic modeling. Data were collected through web crawling on X (Twitter), resulting in 32,005 selected tweets, of which 31,799 remained after preprocessing. Sentiment analysis was conducted using BiLSTM and IndoBERT, while topic modeling was performed using Latent Dirichlet Allocation (LDA). The results showed that IndoBERT outperformed BiLSTM, achieving an accuracy of 0.85. Public conversations were dominated by negative sentiment (85.53%), while positive sentiment accounted for 14.47%. Topic modeling identified five positive topics related to concert experience, organization, staff services, and ticket management, and two negative topics concerning ticket prices, concert experience, and refund issues. The findings indicate that positive sentiment is associated with on-site services, whereas negative sentiment mainly concerns administrative and post-concert services. This study demonstrates that integrating sentiment analysis and topic modeling provides insights for evaluating and improving the service quality of K-pop concert promoters in Indonesia.*

***Keywords***: *Sentiment Analysis, BiLSTM, IndoBERT, LDA, Concert Promoters, K-pop.*

## I. PENDAHULUAN

Indonesia menempati peringkat pertama sebagai negara dengan tingkat antusiasme terhadap budaya Korea tertinggi di dunia. Berdasarkan *Overseas Hallyu Survey 2024* yang dipublikasikan oleh *Korea Foundation for International Cultural Exchange* (KOFICE), tingkat antusiasme masyarakat Indonesia terhadap budaya Korea mencapai 86,3% (Irhamni, 2024). Tingginya antusiasme tersebut menunjukkan bahwa Indonesia merupakan salah satu pasar potensial bagi industri K-pop, termasuk penyelenggaraan konser. Hal ini didukung oleh survei Populix (2024) yang

Copyright © 2026 by Authors; this is an open-access article under the CC BY-SA License (https://creativecommons.org/licenses/by-sa/4.0).

menunjukkan bahwa 57% masyarakat Indonesia berencana menghadiri konser musik pada tahun 2024, dengan konser K-pop menjadi pilihan utama pada kategori konser musisi internasional. Tingginya minat tersebut juga tercermin dari cepatnya penjualan tiket konser berskala besar, seperti BLACKPINK *Born Pink World Tour* dan AGUST D *Tour* di Jakarta (Indonesia, 2022, 2023). Selain itu, keterikatan emosional penggemar terhadap idola turut memengaruhi keputusan menghadiri konser meskipun harga tiket relatif tinggi (Putri & Setiawan, 2025).

Meningkatnya penyelenggaraan konser K-pop di Indonesia belum sepenuhnya diikuti oleh kualitas layanan promotor yang memadai. Berbagai permasalahan masih menjadi perhatian publik, seperti perubahan *venue*, kendala sistem penjualan tiket, proses *refund*, serta kurangnya kepastian informasi kepada penonton. Yayasan Lembaga Konsumen Indonesia (YLKI) menyatakan bahwa penyelenggaraan konser di Indonesia masih sering mengalami permasalahan yang berulang sehingga penyelenggara perlu memberikan kepastian kepada konsumen terkait *venue*, jadwal, dan sistem penyelenggaraan konser (F. A. Zahra, 2025). Permasalahan tersebut juga terlihat pada penyelenggaraan konser oleh Mecimapro, iMe Indonesia, dan Dyandra Global Edutainment yang memperoleh berbagai tanggapan mengenai sistem penjualan tiket, proses *refund*, pengelolaan penonton, hingga aspek keamanan selama konser berlangsung (Garditha, 2025; Kumparan, 2025; Stefanie, 2023; Tempo, 2023; Yusron, 2025). Kondisi tersebut menunjukkan bahwa kualitas layanan promotor perlu dievaluasi karena promotor berperan dalam keseluruhan proses penyelenggaraan konser, mulai dari tahap pra-konser hingga pasca-konser (Restiningtyas et al., 2025a).

Berbagai pengalaman penonton selama proses penyelenggaraan konser umumnya diekspresikan melalui media sosial dalam bentuk opini, kritik, maupun apresiasi (Dewi et al., 2024; Rizkina & Hasan, 2023; D. F. Zahra & Carkiman, 2024). Di antara berbagai platform media sosial, X (Twitter) menjadi salah satu media yang banyak digunakan untuk menyampaikan pengalaman, keluhan, dan evaluasi terhadap suatu layanan secara terbuka (Kusumawati et al., 2024; Lutzky, 2024; Uliniansyah et al., 2024; Varnali & Cesmeci, 2021). Karakteristik komunikasi yang berlangsung secara *real-time* menjadikan X (Twitter) sebagai sumber data yang relevan untuk memahami opini publik terhadap kualitas penyelenggaraan konser. Oleh karena itu, penelitian ini memanfaatkan analisis sentimen untuk mengidentifikasi kecenderungan opini publik berdasarkan polaritas sentimen yang terkandung dalam data teks (Chowdhury, 2003; Liu, 2012). Namun, hasil analisis sentimen hanya menunjukkan kecenderungan polaritas dan belum mampu menjelaskan isu-isu yang menjadi pembahasan utama. Oleh karena itu, penelitian ini mengombinasikan analisis sentimen dengan *topic modeling* menggunakan *Latent Dirichlet Allocation* (LDA) untuk mengidentifikasi topik pembahasan yang

mendasari terbentuknya sentimen publik (Nursyahrina et al., 2024; Ugochi et al., 2022).

Karakteristik data media sosial yang singkat, dinamis, dan mengandung berbagai variasi bahasa informal memerlukan metode yang mampu memahami konteks bahasa dengan baik. Sejumlah penelitian telah menerapkan pendekatan *deep learning* untuk analisis sentimen pada data teks. BiLSTM digunakan karena mampu mempelajari hubungan antar kata dalam suatu urutan teks (Alghifari et al., 2022; Huang et al., 2023; Pramana et al., 2024; Xu et al., 2022), sedangkan IndoBERT digunakan karena telah melalui proses *pre-training* menggunakan korpus berbahasa Indonesia sehingga mampu memahami karakteristik bahasa Indonesia dengan lebih baik (Nabiilah et al., 2023; Sayarizki & Nurrahmi, 2024; Wilie et al., 2020). Selain itu, LDA telah dimanfaatkan untuk mengidentifikasi topik utama dalam kumpulan teks (AbuRaed et al., 2024; Hinek & Papp-Váry, 2024; Uthirapathy & Sandanam, 2023; Witarti et al., 2025). Penelitian mengenai konser K-pop juga telah mengkaji pengalaman penonton serta kualitas penyelenggaraan konser (Celuch, 2020; Dewi et al., 2024; Maldonado-López et al., 2025; Restiningtyas et al., 2025a). Namun, berdasarkan telaah terhadap penelitian-penelitian tersebut, belum ditemukan penelitian yang mengintegrasikan analisis sentimen berbasis BiLSTM dan IndoBERT dengan *topic modeling* menggunakan LDA untuk menganalisis sentimen publik terhadap promotor konser K-pop di Indonesia menggunakan data X (Twitter), sekaligus memetakan topik percakapan berdasarkan polaritas sentimen yang dihasilkan.

Berdasarkan kesenjangan tersebut, penelitian ini bertujuan untuk menganalisis sentimen publik terhadap promotor konser K-pop di Indonesia melalui integrasi analisis sentimen berbasis *deep learning* dan *topic modeling.* Analisis sentimen dilakukan dengan membandingkan performa model BiLSTM dan IndoBERT, sedangkan LDA digunakan untuk mengidentifikasi topik utama serta memetakan topik percakapan berdasarkan polaritas sentimen yang dihasilkan. Hasil penelitian ini diharapkan dapat memberikan gambaran mengenai evaluasi penonton terhadap kualitas layanan promotor konser K-pop di Indonesia serta menjadi masukan bagi promotor dalam meningkatkan kualitas layanan dan menyusun strategi komunikasi yang lebih responsif terhadap opini publik di media sosial.

## II. TINJAUAN PUSTAKA

### 2.1 Analisis Sentimen

Analisis sentimen merupakan salah satu tugas dalam *Natural Language Processing* (NLP) yang bertujuan untuk mengidentifikasi, mengekstraksi, dan mengklasifikasikan opini, sikap, atau evaluasi yang terkandung dalam suatu teks terhadap suatu entitas, seperti produk, layanan, organisasi, atau peristiwa (Bird et al., 2009; Chowdhury, 2003; Liu, 2012).

Hasil analisis sentimen umumnya mengelompokkan teks ke dalam polaritas positif, negatif, atau netral sehingga dapat digunakan untuk memahami kecenderungan opini publik. Analisis sentimen memanfaatkan teknik pemrosesan bahasa untuk mengubah data teks yang tidak terstruktur menjadi representasi yang dapat diproses secara komputasional. Menurut Liu (2012), analisis sentimen dapat dilakukan pada tiga tingkat, *yaitu document-level, sentence-level*, dan *aspect-level.* Penelitian ini menggunakan pendekatan *document-level,* di mana setiap *tweet* diperlakukan sebagai satu dokumen dan diklasifikasikan ke dalam sentimen positif atau negatif.

Analisis sentimen dapat diterapkan untuk mengidentifikasi opini pengguna yang disampaikan melalui media sosial terhadap suatu layanan atau penyelenggaraan suatu kegiatan (Rizkina & Hasan, 2023). Pada penelitian ini, analisis sentimen digunakan untuk mengidentifikasi kecenderungan sentimen publik terhadap promotor konser K-pop di Indonesia berdasarkan percakapan di media sosial X (Twitter).

## 2.2 *Deep Learning* untuk Analisis Sentimen

*Deep learning* merupakan pendekatan *machine learning* yang memanfaatkan *deep neural networks* untuk mempelajari representasi data secara otomatis melalui beberapa lapisan *(layers)* (Géron, 2019; Jurafsky & Martin, 2025). Berbeda dengan pendekatan *machine learning* konvensional, *deep learning* mampu mempelajari representasi fitur secara otomatis sehingga dapat menangkap hubungan semantik dan konteks antarkata dengan lebih baik pada berbagai tugas *Natural Language Processing* (NLP), termasuk analisis sentimen (Jurafsky & Martin, 2025; Young et al., 2018).

Perkembangan *deep learning* telah menghasilkan berbagai arsitektur untuk NLP, di antaranya model berbasis sekuensial dan model berbasis transformer (Jurafsky & Martin, 2025). Penelitian ini menggunakan *Bidirectional Long Short-Term Memory* (BiLSTM) dan IndoBERT yang merepresentasikan kedua pendekatan tersebut dalam tugas analisis sentimen.

### a. *Bidirectional Long Short-Term Memory* (BiLSTM)

*Bidirectional Long Short-Term Memory* (BiLSTM) merupakan pengembangan dari *Long Short-Term Memory* (LSTM), yang dikembangkan dari *Recurrent Neural Network* (RNN) untuk mengatasi keterbatasan dalam mempelajari ketergantungan jangka panjang *(long-term dependency)* akibat permasalahan *vanishing gradient* (Jurafsky & Martin, 2025; Young et al., 2018). Berbeda dengan LSTM yang memproses urutan data dalam satu arah, BiLSTM memproses

urutan teks secara bersamaan dari arah maju *(forward)* dan arah mundur *(backward)* sehingga mampu menghasilkan representasi kontekstual yang lebih lengkap (Jurafsky & Martin, 2025).

Kemampuan tersebut menjadikan BiLSTM banyak diterapkan pada analisis sentimen. Xu et al. (2022) menjelaskan bahwa BiLSTM mampu menangkap informasi kontekstual secara lebih baik dibandingkan model satu arah, sedangkan Huang et al. (2023) menunjukkan bahwa mekanisme dua arah menghasilkan representasi semantik yang lebih baik untuk klasifikasi sentimen. Pada teks berbahasa Indonesia, Alghifari et al. (2022) menunjukkan bahwa BiLSTM memberikan performa yang lebih baik dibandingkan LSTM, sementara Pramana et al. (2024) menunjukkan bahwa BiLSTM mampu memanfaatkan informasi kontekstual secara efektif pada tugas *emotion recognition.*

**b. IndoBERT**

*Bidirectional Encoder Representations from Transformers* (BERT) merupakan *pre-trained language model* berbasis arsitektur *Transformer* yang menghasilkan representasi kata secara kontekstual dengan mempertimbangkan konteks sebelum dan sesudah suatu kata secara bersamaan (Devlin et al., 2019). Berbeda dengan model berbasis sekuensial, BERT memanfaatkan mekanisme *self-attention* untuk memahami hubungan antarkata dalam satu kalimat. IndoBERT merupakan model BERT yang dikembangkan khusus untuk pemrosesan bahasa Indonesia dan telah melalui proses *pre-training* menggunakan korpus Indo4B yang mencakup berbagai sumber data berbahasa Indonesia sehingga mampu memahami karakteristik bahasa Indonesia, baik formal maupun informal (Koto et al., 2020; Wilie et al., 2020).

Koto et al. (2020) menunjukkan bahwa IndoBERT memperoleh performa yang baik pada berbagai tugas NLP dalam *benchmark* IndoLEM, termasuk analisis sentimen. Selain itu, Nabiilah et al. (2023) menunjukkan bahwa IndoBERT menghasilkan nilai *F1-score* yang lebih tinggi dibandingkan *multilingual* BERT (mBERT) dan IndoRoBERTa pada klasifikasi *toxic comments,* sedangkan Sayarizki dan Nurrahmi (2024) menunjukkan bahwa IndoBERT memberikan hasil yang baik pada analisis sentimen ulasan aplikasi MyPertamina. Temuan tersebut menunjukkan bahwa IndoBERT memiliki kemampuan yang baik dalam memahami teks berbahasa Indonesia, termasuk variasi bahasa informal di media sosial.

### *2.3 Topic Modeling*

*Topic modeling* merupakan pendekatan *unsupervised learning* dalam *Natural Language Processing* (NLP) yang digunakan untuk mengidentifikasi topik-topik laten dari sekumpulan dokumen tanpa memerlukan data berlabel (Jurafsky & Martin, 2025). Salah satu metode yang umum digunakan adalah *Latent Dirichlet Allocation* (LDA), yang mengasumsikan bahwa setiap dokumen terdiri atas kombinasi beberapa topik, sedangkan setiap topik direpresentasikan sebagai distribusi probabilitas atas kata-kata penyusunnya (Blei et al., 2003). Hasil pemodelan diinterpretasikan berdasarkan kata-kata dengan probabilitas tertinggi pada setiap topik sehingga dapat merepresentasikan tema utama dalam kumpulan dokumen (Blei et al., 2003). Hinek dan Papp-Váry (2024) serta Witarti et al. (2025) menunjukkan bahwa LDA mampu mengidentifikasi tema utama pada data tekstual melalui penentuan jumlah topik yang optimal dan interpretasi kata-kata dominan.

LDA dapat dikombinasikan dengan analisis sentimen untuk mengidentifikasi tema yang mendasari munculnya opini positif maupun negatif. Analisis sentimen menunjukkan kecenderungan polaritas opini, sedangkan LDA mengungkap topik yang menjadi sumber sentimen tersebut. AbuRaed et al. (2024) mengombinasikan analisis sentimen dan LDA dengan menginterpretasikan topik berdasarkan kata dominan dan *representative tweets,* sementara Uthirapathy dan Sandanam (2023) menggunakan LDA bersama BERT untuk menghubungkan topik percakapan dengan hasil klasifikasi sentimen. Kombinasi kedua pendekatan tersebut memungkinkan identifikasi tidak hanya terhadap polaritas sentimen, tetapi juga tema utama yang menjadi perhatian publik.

### 2.4 X (Twitter) sebagai Sumber Data Penelitian

Media sosial menghasilkan *user-generated content* berupa opini, pengalaman, dan pandangan pengguna yang dapat dimanfaatkan untuk memahami evaluasi publik terhadap suatu produk maupun layanan (Patil & Atique, 2015). Di antara berbagai platform media sosial, X (Twitter) memungkinkan pengguna menyampaikan opini, kritik, maupun keluhan secara terbuka sehingga menyediakan data yang bersifat *real-time* dan tidak terstruktur untuk analisis sentimen (Kharde & Sonawane, 2016). Varnali dan Cesmeci (2021) menunjukkan bahwa media sosial dimanfaatkan pelanggan untuk menyampaikan ketidakpuasan dan memberikan umpan balik terhadap layanan, sedangkan Lutzky (2024) serta Kusumawati et al. (2024) menunjukkan bahwa X (Twitter) menjadi sarana penyampaian evaluasi dan interaksi antara pengguna dengan penyedia layanan.

Pada penyelenggaraan acara, data dari media sosial dapat dimanfaatkan untuk mengevaluasi pengalaman pengunjung. Celuch (2020) menggunakan analisis sentimen dan *topic modeling* untuk mengidentifikasi berbagai dimensi pengalaman pelanggan pada proses pembelian tiket acara, sedangkan Zahra dan Carkiman (2024) menerapkan kedua pendekatan tersebut pada percakapan mengenai konser Coldplay di Indonesia dan mengidentifikasi topik yang berkaitan dengan pembelian tiket, pelayanan pelanggan, harga tiket, pengalaman konser, serta lokasi acara. Oleh karena itu, X (Twitter) dipilih sebagai sumber data dalam penelitian ini karena memungkinkan identifikasi kecenderungan sentimen sekaligus tema utama yang menjadi perhatian penonton terhadap penyelenggaraan konser K-pop di Indonesia.

## 2.5 Kualitas Penyelenggaraan Konser K-Pop

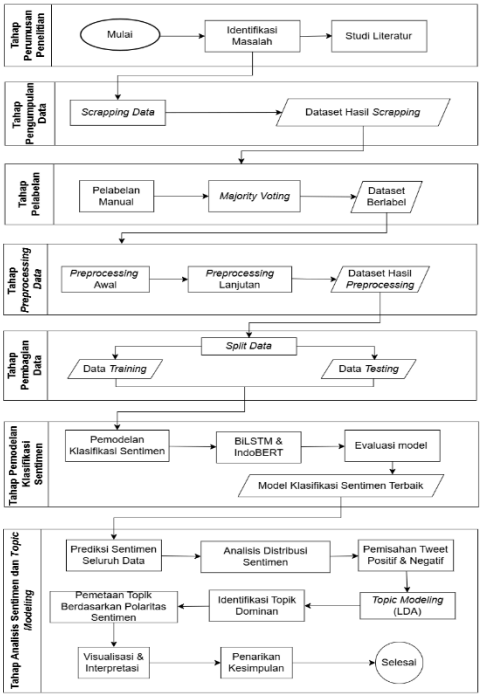
Penyelenggaraan konser merupakan proses penyediaan pengalaman hiburan yang mencakup tahap sebelum, selama, dan setelah acara berlangsung. Dalam industri musik, konser termasuk *experience goods*, yaitu layanan yang kualitasnya baru dapat dievaluasi setelah konsumen mengalaminya secara langsung (Magni, 2023). Oleh karena itu, keberhasilan penyelenggaraan konser tidak hanya ditentukan oleh penampilan artis, tetapi juga oleh kualitas layanan yang diberikan selama seluruh rangkaian acara. Brennan dan Webster (2011) menjelaskan bahwa promotor bertanggung jawab mengelola berbagai aspek penyelenggaraan konser, seperti penjualan tiket, pengelolaan *venue*, fasilitas teknis, keamanan, dan koordinasi operasional. Selain itu, kualitas fasilitas dan pelayanan berperan penting dalam membentuk pengalaman serta evaluasi penonton terhadap penyelenggaraan konser (Dewi et al., 2024; Maldonado-López et al., 2025).

Pada konser K-pop, kualitas penyelenggaraan menjadi semakin penting karena penggemar memiliki keterikatan emosional yang tinggi terhadap artis (Restiningtyas et al., 2025b). Evaluasi penonton terhadap kualitas layanan tersebut juga tercermin melalui opini yang disampaikan di media sosial. Celuch (2020) menunjukkan bahwa analisis ulasan pelanggan mampu mengidentifikasi berbagai aspek pengalaman pengunjung, seperti kualitas pelayanan, kejelasan informasi, biaya tambahan, dan layanan setelah pembelian tiket. Oleh karena itu, opini yang disampaikan penonton di media sosial dapat dimanfaatkan untuk mengevaluasi kualitas penyelenggaraan konser K-pop.

## III. METODE

### 3.1 Alur Penelitian

Penelitian ini dilakukan melalui beberapa tahapan yang dimulai dari pengumpulan data hingga interpretasi dan penarikan kesimpulan hasil analisis. Data penelitian dikumpulkan dari platform X (Twitter) menggunakan teknik *web crawling*, kemudian melalui proses seleksi manual, pelabelan sentimen, dan *preprocessing* sebelum digunakan pada tahap pemodelan. Selanjutnya, dilakukan klasifikasi sentimen menggunakan dua model *deep learning,* yaitu *Bidirectional Long Short-Term Memory* (BiLSTM) dan IndoBERT. Model dengan performa terbaik berdasarkan hasil evaluasi kemudian digunakan untuk memprediksi sentimen pada seluruh dataset. Hasil klasifikasi tersebut selanjutnya dianalisis menggunakan *Latent Dirichlet Allocation* (LDA) untuk mengidentifikasi topik-topik utama yang mendasari sentimen publik. Tahap terakhir adalah interpretasi hasil dan penarikan kesimpulan berdasarkan hasil analisis sentimen dan *topic modeling*. Alur penelitian yang digunakan pada penelitian ini ditunjukkan pada Gambar 1.




Sumber: Diolah oleh Peneliti (2026)

Gambar 1. Diagram Alur Penelitian

### 3.2 Pengumpulan Data

Data penelitian diperoleh dari platform X (Twitter) menggunakan teknik *web crawling*, yaitu metode pengambilan data secara otomatis dengan bantuan perangkat lunak (Agresia & Suryono, 2025). Proses *crawling* dilakukan menggunakan *tweet-harvest* berbasis Node.js yang dijalankan pada Google Colab dengan bantuan Playwright untuk melakukan pencarian *tweet* tanpa menggunakan API. Pengambilan data menerapkan pendekatan *event-based data collection*, yaitu sejak pengumuman konser oleh promotor hingga tujuh hari setelah konser diselenggarakan. Pembatasan periode tersebut didasarkan pada karakteristik X (Twitter) yang bersifat *real-time*, sehingga percakapan publik masih terus berkembang pada periode setelah suatu peristiwa terjadi (Stieglitz & Dang-Xuan, 2013). Selain itu, Gong et al. (2020) juga menggunakan rentang waktu hingga tujuh hari setelah suatu peristiwa untuk menganalisis dinamika aktivitas pengguna di media sosial.

Proses *crawling* dilakukan menggunakan nama akun X (Twitter) masing-masing promotor sebagai kata kunci pencarian. Parameter lang:id digunakan untuk membatasi hasil pencarian pada *tweet* berbahasa Indonesia, sedangkan rentang waktu pengambilan data ditentukan menggunakan parameter *since* dan *until* sesuai periode pengumpulan data setiap promotor. Rincian parameter pengambilan data disajikan pada Tabel 1.

Tabel 1. Parameter Pengambilan Data

| Promotor | Parameter Pengambilan Data | | | |
|---|---|---|---|---|
| | ***Announce-ment* konser** | **H+7 Setelah Konser** | ***Keyword* Pencarian** | **Bahasa** |
| Mecimapro | 13 Mar 2025 | 22 Jun 2025 | “mecimapro” | Indonesia (lang:id) |
| Dyandra Global Edutainment | 13 Mei 2025 | 5 Okt 2025 | “dyandraglobal” | Indonesia (lang:id) |
| iMe Indonesia | 27 Mei 2025 | 9 Nov 2025 | “ime_indonesia” | Indonesia (lang:id) |

Sumber: Diolah oleh Peneliti (2026)

Hasil *crawling* kemudian diseleksi secara manual dengan mempertahankan *tweet* berbahasa Indonesia yang memuat nama akun promotor dan membahas penyelenggaraan konser K-pop, sedangkan *retweet*, konten promosi atau *spam*, *tweet* duplikat, *tweet* yang tidak relevan, serta *tweet* tanpa konten teks dikeluarkan dari dataset. Selanjutnya, sebanyak 1.000 *tweet* dipilih untuk pelabelan manual oleh tiga annotator independen yang terdiri atas peneliti, seorang mahasiswa bidang sastra, dan seorang individu yang berpengalaman menghadiri konser K-pop. Label sentimen

ditentukan ke dalam dua kelas, yaitu positif dan negatif, kemudian ditetapkan menggunakan metode *majority voting* (Uliniansyah et al., 2024)*Tweet* yang mengandung sarkasme atau ironi tidak disertakan dalam data berlabel karena berpotensi menimbulkan ambiguitas dalam penentuan polaritas sentimen (Ardana et al., 2025).

### 3.3 *Preprocessing Data*

Tahap *preprocessing* dilakukan untuk meningkatkan kualitas data teks dengan mengurangi *noise* serta menghasilkan representasi teks yang lebih konsisten sebelum digunakan pada proses klasifikasi sentimen. Pada penelitian ini, preprocessing meliputi case folding, normalisasi, text cleaning, stopword removal, dan stemming. Normalisasi dilakukan menggunakan Kamus Alay Bahasa Indonesia (Salsabila et al., 2018), Kamus *Custom Preprocessing* Analisis Sentimen Indonesia (Zaldin, 2024), serta kamus normalisasi istilah konser K-pop yang disusun oleh peneliti (Salsabila, 2026), untuk menyeragamkan istilah, singkatan, dan kosakata yang bersifat spesifik pada domain penelitian. Selanjutnya, dilakukan *text cleaning* untuk menghapus URL, *mention*, *hashtag*, emoji, simbol, serta karakter lain yang tidak memiliki makna linguistik (Agresia & Suryono, 2025; Ugochi et al., 2022). Setelah itu, dilakukan *stopword removal* menggunakan daftar *stopword* dari Sastrawi, *custom stopwords* dari kamus Zaldin, serta daftar *stopword* tambahan yang disusun oleh peneliti untuk mempertahankan kata-kata yang relevan dalam proses analisis (Ginting et al., 2019; Rizkina & Hasan, 2023).

Tahapan *preprocessing* disesuaikan dengan karakteristik masing-masing model. Pada BiLSTM, dilakukan proses *stemming* menggunakan algoritma Sastrawi untuk mengubah kata berimbuhan menjadi bentuk dasarnya sehingga variasi kosakata dapat dikurangi sebelum proses pembentukan representasi numerik. Sebaliknya, *stemming* tidak diterapkan pada IndoBERT karena model ini menggunakan *tokenizer* berbasis WordPiece yang mampu menghasilkan representasi kontekstual melalui pemecahan kata menjadi *subword*, sehingga perubahan bentuk kata berpotensi menghilangkan informasi linguistik yang dibutuhkan model (Devlin et al., 2019).

### 3.4 Klasifikasi Sentimen

Tahap klasifikasi sentimen diawali dengan pembagian 1.000 *tweet* berlabel menjadi data *training* dan *testing* menggunakan *stratified split* dengan *random state* 42 agar proporsi kelas tetap terjaga dan hasil eksperimen dapat direproduksi (Muraina & Olaniyi, 2022). Penelitian ini menguji empat rasio pembagian data, yaitu 60:40, 70:30, 80:20, dan 90:10 (Ginting et al., 2019; Muraina & Olaniyi, 2022), yang masing-masing

dikombinasikan dengan empat metode penanganan ketidakseimbangan kelas, yaitu *None, Random Oversampling* (ROS), *Random Undersampling* (RUS), dan *Class Weight* (Feng et al., 2021; He & Garcia, 2009). Dengan demikian, diperoleh 16 skenario eksperimen untuk setiap model.

Model BiLSTM dibangun menggunakan *Tokenizer* TensorFlow Keras, sedangkan IndoBERT menggunakan *AutoTokenizer* dan model indobenchmark/indobert-base-p1 dari pustaka *Transformers*. Setiap skenario dievaluasi menggunakan metrik *accuracy, precision, recall,* dan *F1-score.* Konfigurasi dengan performa terbaik kemudian digunakan untuk mengklasifikasikan seluruh dataset sebelum dilakukan proses *topic modeling.* Konfigurasi parameter model BiLSTM disajikan pada Lampiran 1 dan model IndoBERT pada Lampiran 2. Model dengan performa terbaik selanjutnya digunakan untuk melakukan prediksi sentimen terhadap seluruh dataset sebelum dilanjutkan ke tahap *topic modeling.*

### *3.5 Topic Modeling*

Setelah memperoleh label sentimen dari model terbaik, data dipisahkan berdasarkan polaritas positif dan negatif untuk mengidentifikasi topik pada masing-masing kelompok sentimen (Nursyahrina et al., 2024). *Topic modeling* dilakukan menggunakan metode *Latent Dirichlet Allocation* (LDA) dengan pustaka Gensim. Tahap awal meliputi tokenisasi, penghapusan *stopword* khusus LDA, pembentukan *dictionary*, serta representasi dokumen dalam bentuk *Bag-of-Words* (BoW) menggunakan fungsi doc2bow() (Blei et al., 2003; Zhuang et al., 2017).

Jumlah topik ditentukan melalui pengujian konfigurasi 2–5 topik dan dievaluasi menggunakan *coherence score* dengan metrik C_v. Konfigurasi dengan nilai *coherence score* tertinggi dipilih sebagai model terbaik untuk masing-masing kelompok sentimen (Nursyahrina et al., 2024). Hasil pemodelan berupa distribusi topik pada dokumen dan kata dominan pada setiap topik, yang kemudian diinterpretasikan berdasarkan kata dominan serta tweet representatif untuk menentukan label topik (AbuRaed et al., 2024; Harahap et al., 2022).

### 3.6 Interpretasi Hasil dan Penarikan Kesimpulan

Tahap akhir penelitian dilakukan dengan menginterpretasikan hasil klasifikasi sentimen dan *topic modeling* untuk memperoleh gambaran opini publik terhadap kualitas penyelenggaraan konser K-pop oleh promotor di Indonesia. Interpretasi topik dilakukan berdasarkan kata dominan *(top words)* dan *tweet* representatif dari setiap topik LDA untuk menentukan tema utama percakapan publik (AbuRaed et al., 2024; Harahap et al., 2022).

Selanjutnya, topik dipetakan berdasarkan polaritas sentimen positif dan negatif untuk mengidentifikasi isu yang mendasari munculnya sentimen tersebut (Cambria et al., 2013). Kesimpulan penelitian kemudian disusun berdasarkan hasil evaluasi model, distribusi sentimen, dan topik utama yang diperoleh untuk menjawab tujuan penelitian.

## IV. HASIL DAN PEMBAHASAN

### 4.1 Hasil Klasifikasi Sentimen

Penelitian ini menggunakan 31.799 *tweet* hasil *preprocessing* sebagai dataset analisis sentimen. Sebanyak 1.000 *tweet* berlabel digunakan untuk pelatihan dan evaluasi model dengan pembagian data menggunakan *stratified split* agar proporsi kelas tetap terjaga. Pengujian dilakukan terhadap empat rasio pembagian data (60:40, 70:30, 80:20, dan 90:10) serta empat metode penanganan ketidakseimbangan kelas, yaitu *None, Random Oversampling* (ROS), *Random Undersampling* (RUS), dan *Class Weight.*. Berdasarkan hasil tersebut, rasio 90:10 dengan metode ROS dipilih sebagai konfigurasi terbaik untuk pelatihan akhir model BiLSTM dan IndoBERT. Perbandingan kinerja kedua model berdasarkan metrik evaluasi disajikan pada Tabel 2.

Tabel 2. Perbandingan Kinerja Model BiLSTM dan IndoBERT

| Model | *Accuracy* | *Precision Macro* | *Recall Macro* | *F1 Macro* |
|---|---|---|---|---|
| BiLSTM | 0.80 | 0.76 | 0.78 | 0.77 |
| IndoBERT | 0.85 | 0.82 | 0.83 | 0.82 |

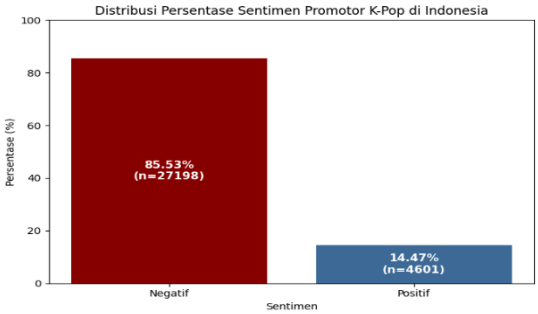
Sumber: Diolah oleh Peneliti (2026)

Berdasarkan Tabel 1, model IndoBERT menunjukkan performa yang lebih baik dibandingkan BiLSTM pada seluruh metrik evaluasi. IndoBERT memperoleh *accuracy* sebesar 0,85, *precision macro* sebesar 0,82, *recall macro* sebesar 0,83, dan *F1 macro* sebesar 0,82, sedangkan BiLSTM memperoleh *accuracy* sebesar 0,80, *precision macro* sebesar 0,76, *recall macro* sebesar 0,78, dan *F1 macro* sebesar 0,77. Hasil tersebut menunjukkan bahwa IndoBERT memiliki kemampuan yang lebih baik dalam mengklasifikasikan kedua kelas sentimen secara seimbang.

Performa IndoBERT yang lebih tinggi dapat dikaitkan dengan kemampuannya dalam membangun representasi kontekstual melalui arsitektur *Transformer*. Berbeda dengan BiLSTM yang memproses teks secara berurutan, IndoBERT memanfaatkan mekanisme *self-attention* sehingga mampu memahami hubungan antarkata berdasarkan keseluruhan konteks kalimat. Kemampuan tersebut membuat IndoBERT lebih efektif dalam mengklasifikasikan teks media sosial berbahasa Indonesia yang umumnya mengandung bahasa informal, singkatan, serta variasi penulisan. Temuan ini

sejalan dengan penelitian Pramana et al. (2024) dan Nabiilah et al. (2023) yang menunjukkan bahwa IndoBERT menghasilkan performa klasifikasi sentimen yang lebih baik dibandingkan model berbasis BiLSTM pada data berbahasa Indonesia.

Berdasarkan hasil evaluasi tersebut, model IndoBERT dipilih sebagai model terbaik untuk melakukan klasifikasi sentimen terhadap seluruh dataset penelitian. Hasil klasifikasi menunjukkan bahwa dari 31.799 *tweet*, sebanyak 27.198 *tweet* (85,53%) diklasifikasikan sebagai sentimen negatif, sedangkan 4.601 *tweet* (14,47%) diklasifikasikan sebagai sentimen positif. Distribusi hasil klasifikasi sentimen disajikan pada Gambar 2.




Sumber: Diolah oleh Peneliti (2026)

Gambar 2. Distribusi Sentimen Promotor K-Pop di Indonesia

## 4.2 Hasil *Topic Modeling*

Setelah klasifikasi sentimen menggunakan IndoBERT, dilakukan *topic modeling* dengan metode LDA untuk mengidentifikasi tema utama pada *tweet* positif dan negatif secara terpisah. Sebelum pemodelan, dilakukan penyaringan menggunakan *stopword* khusus LDA untuk mengurangi kata yang tidak relevan. Jumlah topik optimal ditentukan berdasarkan pengujian *coherence score* pada beberapa konfigurasi, dengan hasil pengujian disajikan pada Tabel 3.

Tabel 3. *Coherence Score Topic Modeling* Sentimen Positif dan Negatif

| Jumlah Topik | *Coherence Score* Sentimen Positif | *Coherence Score* Sentimen Negatif |
|---|---|---|
| 2 | 0,264 | **0,382** |
| 3 | 0,338 | 0,337 |
| 4 | 0,376 | 0,351 |
| 5 | **0,387** | 0,326 |

Sumber: Diolah oleh Peneliti (2026)

Berdasarkan Tabel 2, model LDA pada sentimen positif memperoleh nilai *coherence score* tertinggi sebesar 0,387 pada konfigurasi lima topik.

Sementara itu, pada sentimen negatif nilai *coherence score* tertinggi sebesar 0,382 diperoleh pada konfigurasi dua topik. Oleh karena itu, konfigurasi lima topik untuk sentimen positif dan dua topik untuk sentimen negatif dipilih sebagai model terbaik untuk analisis selanjutnya.

Berdasarkan konfigurasi terbaik tersebut, model LDA menghasilkan kata-kata dominan *(top words)* yang merepresentasikan setiap topik. Kata-kata dominan ini menjadi dasar dalam proses interpretasi tema percakapan pada masing-masing kelompok sentimen. Hasil ekstraksi kata dominan untuk setiap topik disajikan pada Tabel 4.

Tabel 4. Kata Dominan *Topic Modeling* Sentimen Positif dan Negatif

| Sentimen | Topik | Kata Dominan |
|---|---|---|
| Positif | 0 | *venue*, enak, harga, dekat, langsung, sangat, ambil, dengar, beli, jauh |
| | 1 | konser, *venue*, terimakasih, *crew*, ikut, bagus, nct, lupa, selalu, tonton |
| | 2 | minum, air, konser, masuk, *usher*, *staff*, ambil, minta, akhir, habis |
| | 3 | *staff*, terimakasih, tanya, baru, pintu, tas, zona, arah, turut, ramah |
| | 4 | kelas_tiket, nomor_kursi, tunggu, *soundcheck*, *section*, kirim, *presale*, terimakasih, tiket, bagi |
| Negatif | 0 | *venue*, tiket, air, konser, beli, kelas_tiket, minum, harga, jual, gila |
| | 1 | *refund*, duit, bayar, uang, kerja, cepat, balikin, refundnya, utang, minta |

Sumber: Diolah oleh Peneliti (2026)

Secara umum, *hasil topic modeling* menunjukkan bahwa percakapan positif menghasilkan lima kelompok topik yang lebih beragam, sedangkan percakapan negatif terkonsentrasi pada dua topik utama. Interpretasi terhadap setiap topik berdasarkan kata-kata dominan dan konteks *tweet* representatif dibahas pada bagian berikutnya.

### 4.3 Interpretasi Topik dan Pembahasan

Berdasarkan hasil *topic modeling,* percakapan publik mengenai promotor konser K-pop di Indonesia menghasilkan lima topik pada sentimen positif dan dua topik pada sentimen negatif. Interpretasi dilakukan berdasarkan kata dominan *(top words)* hasil LDA serta *tweet* representatif pada setiap topik untuk memastikan kesesuaian konteks pembahasan (AbuRaed et al., 2024; Harahap et al., 2022). Hasil interpretasi menunjukkan bahwa opini penonton mencakup aspek pertunjukan serta layanan promotor sebelum, selama, dan setelah konser. Ringkasan interpretasi topik disajikan pada Tabel 5.

Tabel 5. Interpretasi Topik Hasil LDA

| Sentimen | Topik | Kata Dominan | Interpretasi Topik |
|---|---|---|---|
| Positif | 0 | *venue*, enak, harga, dekat, langsung, sangat, ambil, dengar, beli, jauh | Kenyamanan *venue* dan pengalaman menonton |

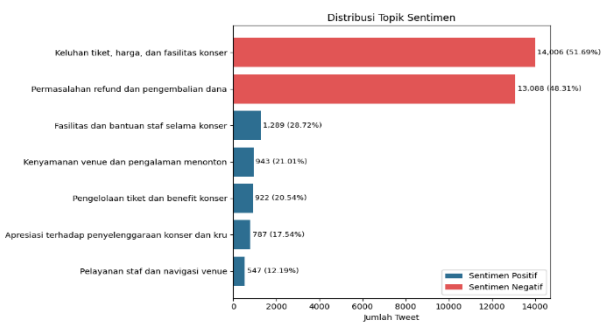
| Sentimen | Topik | Kata Dominan | Interpretasi Topik |
|---|---|---|---|
| Positif | 1 | konser, *venue*, terimakasih, *crew*, ikut, bagus, nct, lupa, selalu, tonton | Apresiasi terhadap penyelenggaraan konser dan kru |
| | 2 | minum, air, konser, masuk, *usher*, *staff*, ambil, minta, akhir, habis | Fasilitas dan bantuan staf selama konser |
| | 3 | *staff*, terimakasih, tanya, baru, pintu, tas, zona, arah, turut, ramah | Pelayanan staf dan navigasi venue |
| | 4 | kelas_tiket, nomor_kursi, tunggu, *soundcheck*, *section*, kirim, *presale*, terimakasih, tiket, bagi | Pengelolaan tiket dan benefit konser |
| Negatif | 0 | *venue*, tiket, air, konser, beli, kelas_tiket, minum, harga, jual, gila | Keluhan tiket, harga, dan fasilitas konser |
| | 1 | *refund*, duit, bayar, uang, kerja, cepat, balikin, refundnya, utang, minta | Permasalahan refund dan pengembalian dana |

Sumber: Diolah oleh Peneliti (2026)

Pada sentimen positif, topik yang terbentuk mencakup kenyamanan *venue*, apresiasi terhadap penyelenggaraan dan kru, fasilitas serta bantuan staf, pelayanan staf, serta pengelolaan tiket dan *benefit* konser. Hal ini menunjukkan bahwa pengalaman positif penonton dipengaruhi oleh kualitas layanan operasional, tidak hanya oleh penampilan artis.

Sebaliknya, sentimen negatif didominasi oleh keluhan terkait tiket, harga, fasilitas konser, serta permasalahan *refund*. Temuan tersebut menunjukkan bahwa ketidaksesuaian antara ekspektasi penonton dan layanan yang diterima menjadi faktor utama munculnya evaluasi negatif terhadap promotor.

Distribusi topik pada masing-masing kelompok sentimen disajikan pada Gambar 3.




Sumber: Diolah oleh Peneliti (2026)

Gambar 3. Distribusi Topik Berdasarkan Sentimen

Berdasarkan Gambar 3, topik fasilitas dan bantuan staf selama konser menjadi topik dominan pada sentimen positif dengan proporsi 28,72% (1.289 *tweet*). Sementara itu, pada sentimen negatif, keluhan terkait

tiket, harga, dan fasilitas menjadi topik terbesar dengan proporsi 51,69% (14.006 *tweet*), diikuti permasalahan *refund* sebesar 48,31% (13.088 *tweet*). Hasil ini menunjukkan bahwa pengalaman positif lebih banyak dipengaruhi oleh kualitas layanan selama konser, sedangkan sentimen negatif berkaitan dengan aspek administrasi dan fasilitas layanan.

Temuan ini sejalan dengan Maldonado-López et al. (2025) yang menyatakan bahwa pengalaman pengunjung dipengaruhi oleh fasilitas, pelayanan, dan organisasi acara. Selain itu, Magni (2023) menjelaskan bahwa konser merupakan experience goods, sehingga kualitas layanan baru dapat dievaluasi setelah penonton mengalami acara secara langsung.

Berdasarkan hasil interpretasi, dilakukan pemetaan topik berdasarkan tahapan layanan promotor untuk mengidentifikasi aspek yang perlu dipertahankan maupun ditingkatkan. Hasil pemetaan disajikan pada Tabel 6.

Tabel 6. Pemetaan Topik Berdasarkan Polaritas Sentimen dan Implikasi bagi Promotor

| Tahapan Layanan Promotor | Topik | Sentimen | Implikasi bagi Promotor |
|---|---|---|---|
| Sebelum konser | Pengelolaan tiket dan benefit konser. | Positif | Mempertahankan sistem penjualan tiket, pembagian *benefit*, dan komunikasi informasi kepada penonton. |
| Sebelum konser | Keluhan tiket, harga, dan fasilitas konser. | Negatif | Mengevaluasi kesesuaian harga tiket dengan fasilitas dan pengalaman yang diterima penonton. |
| Selama konser | Fasilitas dan bantuan staf selama konser. | Positif | Mempertahankan kualitas pelayanan dan responsivitas staf. |
| Selama konser | Kenyamanan *venue* dan pengalaman menonton. | Positif | Mempertahankan kualitas *venue* dan kenyamanan area menonton konser. |
| Selama konser | Apresiasi terhadap penyelenggaraan konser dan kru. | Positif | Mempertahankan koordinasi operasional dan kualitas penyelenggaraan konser. |
| Selama konser | Pelayanan staf dan navigasi *venue*. | Positif | Memastikan informasi, petunjuk arah, dan koordinasi staf mudah dipahami penonton. |
| Setelah konser | Permasalahan *refund* dan pengembalian dana. | Negatif | Meningkatkan kecepatan, transparansi, dan kepastian proses *refund*. |

Sumber: Diolah oleh Peneliti (2026)

Secara umum, sentimen positif didominasi oleh aspek layanan selama konser, seperti kualitas *venue*, pelayanan staf, dan koordinasi

penyelenggaraan. Sebaliknya, sentimen negatif lebih banyak berkaitan dengan layanan sebelum dan setelah konser, terutama harga tiket, fasilitas, dan proses *refund*. Kombinasi analisis sentimen dan *topic modeling* memberikan gambaran yang lebih komprehensif mengenai persepsi penonton serta dapat dimanfaatkan sebagai evaluasi berbasis data *(data-driven evaluation)* untuk meningkatkan kualitas layanan promotor.

## V. KESIMPULAN DAN REKOMENDASI

### 5.1 Kesimpulan

Penelitian ini membandingkan model BiLSTM dan IndoBERT untuk klasifikasi sentimen publik terhadap promotor konser K-pop di Indonesia. Hasil evaluasi menunjukkan bahwa IndoBERT memiliki performa terbaik dengan *accuracy* 0,85 dan *F1-score macro* 0,82, sehingga digunakan untuk klasifikasi seluruh dataset. Dari 31.799 *tweet*, sebanyak 85,53% dikategorikan sebagai sentimen negatif dan 14,47% sebagai sentimen positif, menunjukkan dominasi kritik terhadap penyelenggaraan konser. *Topic modeling* menggunakan LDA mengidentifikasi lima topik pada sentimen positif dan dua topik pada sentimen negatif. Topik yang ditemukan menunjukkan bahwa opini penonton tidak hanya dipengaruhi oleh kualitas pertunjukan, tetapi juga aspek layanan promotor, seperti *venue*, tiket, fasilitas, pelayanan staf, dan proses *refund*. Secara keseluruhan, kombinasi analisis sentimen dan *topic modeling* mampu memberikan gambaran komprehensif mengenai evaluasi publik serta menjadi dasar evaluasi berbasis opini untuk meningkatkan kualitas layanan promotor konser K-pop di Indonesia.

### 5.2 Rekomendasi

Penelitian selanjutnya disarankan menggunakan data dari berbagai platform media sosial agar memperoleh gambaran opini publik yang lebih komprehensif. Selain itu, penelitian dapat membandingkan performa model yang lebih mutakhir, seperti RoBERTa, XLM-RoBERTa, atau *Large Language Models* (LLMs), serta mengeksplorasi metode *topic modeling* yang lebih kontekstual, seperti BERTopic atau Top2Vec. Pengembangan sistem pemantauan sentimen secara *real-time* juga dapat menjadi arah penelitian berikutnya untuk mendukung evaluasi kualitas layanan promotor secara berkelanjutan.

## REFERENSI

AbuRaed, A. G. T., Prikryl, E. A., Carenini, G., & Janjua, N. Z. (2024). Long COVID Discourse in Canada, the United States, and Europe: Topic Modeling and Sentiment Analysis of Twitter Data. *Journal of Medical Internet Research*, *26*, e59425. https://doi.org/10.2196/59425

Agresia, V., & Suryono, R. R. (2025). Comparison of SVM, Naïve Bayes, and Logistic Regression Algorithms for Sentiment Analysis of Fraud and Bots in Purcashing Concert Ticket. *INOVTEK Polbeng - Seri Informatika*, *10*(2), 591–602. https://doi.org/10.35314/npyfdh47

Alghifari, D. R., Edi, M., & Firmansyah, L. (2022). Implementasi Bidirectional LSTM untuk Analisis Sentimen Terhadap Layanan Grab Indonesia. *Jurnal Manajemen Informatika (JAMIKA)*, *12*(2), 89–99. https://doi.org/10.34010/jamika.v12i2.7764

Ardana, M., Mayasari, R., & Maulana, I. (2025). Perbandingan Naïve Bayes dan SVM untuk Analisis Sentimen Ulasan Kompas.id pada Data Tidak Seimbang. *Bulletin of Computer Science Research*, *6*(1), 399–408. https://doi.org/10.47065/bulletincsr.v6i1.929

Bird, S., Klein, E., & Loper, E. (2009). *Natural Language Processing with Python*. O'Reilly.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, *3*(Jan), 993–1022. https://doi.org/10.2307/3050792

Brennan, M., & Webster, E. (2011). Why Concert Promoters Matter. *Scottish Music Review*, *2*(1), 1–25. http://www.scottishmusicreview.org/index.php/SMR/article/view/17/15%3E

Cambria, E., Schuller, B., Xia, Y., & Havasi, C. (2013). New Avenues in Opinion Mining and Sentiment Analysis. *IEEE Intelligent Systems*, *28*(2), 15–21. https://doi.org/10.1109/MIS.2013.30

Celuch, K. (2020). Customers' experience of purchasing event tickets: mining online reviews based on topic modeling and sentiment analysis. *International Journal of Event and Festival Management*, *12*(1), 36–50. https://doi.org/10.1108/IJEFM-06-2020-0034

Chowdhury, G. G. (2003). Natural language processing. *Annual Review of Information Science and Technology*, *37*(1), 51–89. https://doi.org/10.1002/aris.1440370103

Devlin, J., Chang, M.-W., Lee, K., Google, K. T., & Language, A. I. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *Proceedings of the 2019 Conference of the North*, 4171–4186. https://doi.org/10.18653/V1/N19-1423

Dewi, V. T., Supriyadi, T., & Pratama, G. (2024). Pengaruh Fasilitas dan Kualitas Pelayanan terhadap Kepuasan Pengunjung di Konser Musik Blackpink "Born Pink World Tour Jakarta 2023." *Ranah Research : Journal of Multidisciplinary Research and Development*, *6*(6), 2590–2603. https://doi.org/10.38035/rrj.v6i6.1118

Feng, Y., Zhou, M., & Tong, X. (2021). Imbalanced classification: A paradigm-based

review. *Statistical Analysis and Data Mining: The ASA Data Science Journal*, *14*(5), 383–406. https://doi.org/10.1002/sam.11538

Garditha, H. A. (2025). *Promotor Konser K-Pop Mecima Pro Jadi Trending di Topik Indonesia dengan Beberapa Tagar, Ternyata Karena Ini*. Jatim Network. https://www.jatimnetwork.com/trending/4314966439/promotor-konser-k-pop-mecima-pro-jadi-trending-di-topik-indonesia-dengan-beberapa-tagar-ternyata-karena-ini?page=2

Géron, A. (2019). *Hands-On Machine Learning with Scikit-Learn and TensorFlow* (2nd ed.). O'Reilly Media.

Ginting, P. S. B., Irawan, B., & Setianingsih, C. (2019). Hate Speech Detection on Twitter Using Multinomial Logistic Regression Classification Method. *2019 IEEE International Conference on Internet of Things and Intelligence System (IoTaIS)*, 105–111. https://doi.org/10.1109/IoTaIS47347.2019.8980379

Gong, V. X., Daamen, W., Bozzon, A., & Hoogendoorn, S. P. (2020). Crowd characterization for crowd management using social media data in city events. *Travel Behaviour and Society*, *20*(February), 192–212. https://doi.org/10.1016/j.tbs.2020.03.011

Harahap, R. M., Kusumahadi, K., & Nurhazizah, E. (2022). Analysing Service Quality of Mobile Health Platforms Using Text Analytics: A Case Study of Halodoc And Alodokter. *Asian Journal of Research in Business and Management*, *4*(1), 168–182. https://doi.org/10.55057/ajrbm.2022.4.1.14

He, H., & Garcia, E. A. (2009). Learning from Imbalanced Data. *IEEE Transactions on Knowledge and Data Engineering*, *21*(9), 1263–1284. https://doi.org/10.1109/TKDE.2008.239

Hinek, M., & Papp-Váry, Á. (2024). Topic modelling of festival-visitors' opinions: A case study of Sziget Festival. *European Journal of Tourism Research*, *37*(2024), 3703. https://doi.org/10.54055/ejtr.v37i.3037

Huang, Y., Liu, Q., Peng, H., Wang, J., Yang, Q., & Orellana-Martín, D. (2023). Sentiment classification using bidirectional LSTM-SNP model and attention mechanism. *Expert Systems with Applications*, *221*(June 2022), 119730. https://doi.org/10.1016/j.eswa.2023.119730

Indonesia, C. (2022). *Tiket Konser BLACKPINK di Jakarta 11-12 Maret 2023 Sold Out*. CNN Indonesia. https://www.cnnindonesia.com/hiburan/20221115153725-227-874038/tiket-konser-blackpink-di-jakarta-11-12-maret-2023-sold-out

Indonesia, C. (2023). *Cerita Beruntung Menang War Tiket Suga BTS yang "Hectic."* CNN Indonesia. https://www.cnnindonesia.com/hiburan/20230327203420-227-930051/cerita-beruntung-menang-war-tiket-suga-bts-yang-hectic

Irhamni, M. A. F. (2024). *Survei Korean Wave Indonesia jadi negara pusat Hallyu dengan antusiasme budaya Korea tertinggi di dunia.* Goodstats.

https://goodstats.id/article/survei-korean-wave-indonesia-jadi-negara-pusat-hallyu-dengan-antusiasme-budaya-korea-tertinggi-di-dunia-9f0mf

Jurafsky, D., & Martin, J. H. (2025). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition* (3rd ed.). https://web.stanford.edu/~jurafsky/slp3/

Kharde, V. A., & Sonawane, S. (2016). Sentiment Analysis of Twitter Data: A Survey of Techniques. *International Journal of Computer Applications*, *139*(11), 5–15. https://doi.org/10.5120/ijca2016908625

Koto, F., Rahimi, A., Lau, J. H., & Baldwin, T. (2020). IndoLEM and IndoBERT: A Benchmark Dataset and Pre-trained Language Model for Indonesian NLP. *Proceedings of the 28th International Conference on Computational Linguistics*, 757–770. https://doi.org/10.18653/v1/2020.coling-main.66

Kumparan. (2025, October 30). *Deretan Kasus Mecimapro: Dugaan Penggelapan Dana hingga Polemik Refund Tiket | kumparan.com*. https://kumparan.com/kumparank-pop/deretan-kasus-mecimapro-dugaan-penggelapan-dana-hingga-polemik-refund-tiket-2695EvjHxOQ/full

Kusumawati, N., Suhandi, A. T., & Purwanegara, M. S. (2024). Why do Customers Decide to Complain on Social Media? And How to Produce Customer Forgiveness. *Review of Integrative Business and Economics Research*, *13*(3), 235–254. https://buscompress.com/riber-13-3x.html

Liu, B. (2012). Sentiment Analysis and Opinion Mining. In *Sentiment Analysis and Opinion Mining* (Issue April). Morgan & Claypool Publishers. https://doi.org/10.4337/9781800374263.sentiment.analysis

Lutzky, U. (2024). "Doesn't Really Answer My Question . . .": Exploring Customer Service Interactions on Twitter. *International Journal of Business Communication*, *61*(1), 92–114. https://doi.org/10.1177/23294884231200247

Magni, F. (2023). Value creation and appropriation in the live music industry: a population ecology analysis of live music ticket pricing. *Journal of Media Business Studies*, *20*(2), 137–157. https://doi.org/10.1080/16522354.2022.2036440

Maldonado-López, B., Arteaga-Sánchez, R., & Solano-Sánchez, M. Á. (2025). Impact of creative event experiences on the intention to revisit peripheral cross-border destinations: the smuggling festival. *International Journal of Event and Festival Management*, *16*(4), 570–596. https://doi.org/10.1108/IJEFM-07-2024-0085

Muraina, & Olaniyi, I. (2022). Ideal Dataset Splitting Ratios in Machine Learning Algorithms : *In 7th International Mardin Artuklu Scientific Research Conference*, *February*, 496–504.

Nabiilah, G. Z., Prasetyo, S. Y., Izdihar, Z. N., & Girsang, A. S. (2023). BERT base model for toxic comment analysis on Indonesian social media. *Procedia Computer Science*, *216*, 714–721. https://doi.org/10.1016/j.procs.2022.12.188

Nursyahrina, Defit, S., & Sovia, R. (2024). Metode BERTopic dan LDA untuk Analisis Tren Penelitian Bidang Ilmu Komputer. *Jurnal KomtekInfo*, *11*(4), 332–341. https://doi.org/10.35134/komtekinfo.v11i4.580

Patil, H. P., & Atique, M. (2015). Sentiment Analysis for Social Media: A Survey. *2015 2nd International Conference on Information Science and Security (ICISS)*, 1–4. https://doi.org/10.1109/ICISSEC.2015.7371033

Populix. (2024). *Beyond Borders: A Study of Indonesian Concert-Goers' Behavior*. Populix. https://info.populix.co/data-hub/reports/a-study-of-indonesian-concert-goers#

Pramana, R., Jonathan, M., Yani, H. S., & Sutoyo, R. (2024). A Comparison of BiLSTM, BERT, and Ensemble Method for Emotion Recognition on Indonesian Product Reviews. *Procedia Computer Science*, *245*, 399–408. https://doi.org/10.1016/j.procs.2024.10.266

Putri, N. A., & Setiawan, J. H. (2025). Motivasi Konsumen dalam Pembelian Tiket Konser NCT: Studi pada Penggemar K-Pop di Indonesia. *JURNAL SOSIAL EKONOMI DAN HUMANIORA*, *11*(1), 1–12. https://doi.org/10.29303/jseh.v11i1.734

Restiningtyas, A., Alfizi, A., & Putri, P. L. (2025a). The Effect of Fanatism, Price and Location on Buying Interest and Ticket Purchase Decisions for South Korean Band Concerts at The Jawa Pos Arena Building. *Jurnal Manajemen Bisnis*, *12*(2), 262–282. https://doi.org/10.33096/jmb.v12i2.1236

Restiningtyas, A., Alfizi, A., & Putri, P. L. (2025b). The Effect of Fanatism, Price and Location on Buying Interest and Ticket Purchase Decisions for South Korean Band Concerts at The Jawa Pos Arena Building. *Jurnal Manajemen Bisnis*, *12*(2), 262–282. https://doi.org/10.33096/jmb.v12i2.1236

Rizkina, N. Q., & Hasan, F. N. (2023). Analisis Sentimen Komentar Netizen Terhadap Pembubaran Konser NCT 127 Menggunakan Metode Naive Bayes. *Journal of Information System Research (JOSH)*, *4*(4), 1136–1144. https://doi.org/10.47065/josh.v4i4.3803

Salsabila, A. D. N. (2026). *Kamus-Normalisasi-Istilah-Konser-K-pop-by-Ansheila*. https://github.com/ansheilads/Kamus-Normalisasi-Istilah-Konser-K-pop-by-Ansheila

Sayarizki, P., & Nurrahmi, H. (2024). Implementation of IndoBERT for Sentiment Analysis of Indonesian Presidential Candidates. *Journal on Computing*, *9*(2), 61–72. https://doi.org/10.34818/indojc.2024.9.2.934

Stefanie, C. (2023, March 5). *Berguru dari Pengalaman di Konser Mantap NCT Dream*. https://www.cnnindonesia.com/hiburan/20230305183522-227-921213/berguru-dari-pengalaman-di-konser-mantap-nct-dream

Stieglitz, S., & Dang-Xuan, L. (2013). Emotions and Information Diffusion in Social Media—Sentiment of Microblogs and Sharing Behavior. *Journal of Management*

*Information Systems*, *29*(4), 217–248. https://doi.org/10.2753/MIS0742-1222290408

Tempo. (2023, March 13). *Profil iMe, Promotor Konser Blackpink yang Kebanjiran Keluhan Penonton | tempo.co*. https://www.tempo.co/ekonomi/profil-ime-promotor-konser-blackpink-yang-kebanjiran-keluhan-penonton-209477

Ugochi, O., Prasad, R., Odu, N., Ogidiaka, E., & Ibrahim, B. H. (2022). Customer opinion mining in electricity distribution company using twitter topic modeling and logistic regression. *International Journal of Information Technology*, *14*(4), 2005–2012. https://doi.org/10.1007/s41870-022-00890-4

Uliniansyah, M. T., Budi, I., Nurfadhilah, E., Afra, D. I. N., Santosa, A., Latief, A. D., Jarin, A., Gunarso, Jiwanggi, M. A., Hidayati, N. N., Fajri, R., Suryono, R. R., Pebiana, S., Shaleha, S., Ramdhani, T. W., & Sampurno, T. (2024). Twitter dataset on public sentiments towards biodiversity policy in Indonesia. *Data in Brief*, *52*, 109890. https://doi.org/10.1016/j.dib.2023.109890

Uthirapathy, S. E., & Sandanam, D. (2023). Topic Modelling and Opinion Analysis On Climate Change Twitter Data Using LDA And BERT Model. *Procedia Computer Science*, *218*(2022), 908–917. https://doi.org/10.1016/j.procs.2023.01.071

Varnali, K., & Cesmeci, C. (2021). Customer responses to service failures on social media. *Journal of Services Marketing*, *36*(5), 691–709. https://doi.org/10.1108/JSM-11-2020-0484

Wilie, B., Vincentio, K., Winata, G. I., Cahyawijaya, S., Li, X., Lim, Z. Y., Soleman, S., Mahendra, R., Fung, P., Bahar, S., & Purwarianti, A. (2020). IndoNLU: Benchmark and Resources for Evaluating Indonesian Natural Language Understanding. *Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing*, 843–857. https://doi.org/10.18653/v1/2020.aacl-main.85

Witarti, D. I., Sugiyanto, D., Ariesta, A., Ariyani, P. F., & -, R. (2025). LDA-Based Topic Mining for Unveiling the Outstanding Universal Value of Solo Keroncong Music as an Intangible Cultural Heritage of UNESCO. *International Journal of Advanced Computer Science and Applications*, *16*(2), 1001–1010. https://doi.org/10.14569/IJACSA.2025.0160299

Xu, Q. A., Chang, V., & Jayne, C. (2022). A systematic review of social media-based sentiment analysis: Emerging trends and challenges. *Decision Analytics Journal*, *3*(June), 100073. https://doi.org/10.1016/j.dajour.2022.100073

Young, T., Hazarika, D., Poria, S., & Cambria, E. (2018). Recent Trends in Deep Learning Based Natural Language Processing. *IEEE Computational Intelligence Magazine*, *13*(3), 55–75. https://doi.org/10.1109/MCI.2018.2840738

Yusron, A. A. (2025). *Skandal Mecimapro Wake Up Call Industri Konser dan Perlindungan Konsumen*. Detik.Com. https://www.detik.com/pop/korean-wave/d-8250323/skandal-mecimapro-w

ake-up-call-industri-konser-dan-perlindungan-konsumen

Zahra, D. F., & Carkiman, C. (2024). PENGALAMAN PELANGGAN MEMBELI TIKET KONSER COLDPLAY: MENAMBANG ULASAN ONLINE BERDASARKAN PEMODELAN TOPIK DAN ANALISIS SENTIMEN. *Journal of Information System, Applied, Management, Accounting and Research*, *8*(2), 243–260. https://doi.org/10.52362/jisamar.v8i2.1426

Zahra, F. A. (2025). *Viral Promotor Konser DAY6 Disebut Rugikan Konsumen, Apa Kata YLKI?* Kompas. https://amp.kompas.com/tren/read/2025/05/03/133000465/viral-promotor-konser-day6-disebut-rugikan-konsumen-apa-kata-ylki-?utm_source=chatgpt.com

Zhuang, Y., Wang, H., Xiao, J., Wu, F., Yang, Y., Lu, W., & Zhang, Z. (2017). Bag-of-Discriminative-Words (BoDW) Representation via Topic Modeling. *IEEE Transactions on Knowledge and Data Engineering*, *29*(5), 977–990. https://doi.org/10.1109/TKDE.2017.2658571