

Analisis Sentimen terhadap Program WJSC Menggunakan NLP pada Media Sosial

Febi Nur Zahra

Teknik Informatika, Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika

E-mail: febazzahra28@gmail.com¹

Article History:

Received: 03 Oktober 2025

Revised: 20 Oktober 2025

Accepted: 23 Oktober 2025

Keywords:

Sentiment Analysis, Natural Language Processing (NLP), Social Media, Women Jungle Survival Course, Public Opinion.

Abstract: *Kursus Women Jungle Survival Course (WJSC) merupakan program pelatihan bertahan hidup di alam liar yang dirancang untuk memberdayakan perempuan melalui pembelajaran berbasis pengalaman di luar ruang. Pasca pelaksanaannya, muncul beragam opini publik di media sosial yang mencerminkan sentimen positif, negatif, dan netral. Penelitian ini bertujuan untuk menganalisis sentimen warganet terhadap program WJSC 2025 dengan menggunakan pendekatan Natural Language Processing (NLP). Data diperoleh dari tiga platform media sosial, yaitu Instagram, TikTok, dan YouTube, kemudian melalui tahapan preprocessing berupa tokenisasi, pembersihan, dan normalisasi sebelum dilakukan klasifikasi sentimen. Model NLP digunakan untuk mengelompokkan opini publik ke dalam tiga kategori utama, sehingga diperoleh gambaran yang komprehensif mengenai persepsi masyarakat terhadap program tersebut. Hasil analisis menunjukkan bahwa metode NLP efektif dalam mengidentifikasi pola sentimen dan memberikan wawasan berharga bagi penyelenggara, seperti Eiger Adventure, untuk menyusun strategi komunikasi yang lebih responsif dan inklusif sesuai dengan ekspektasi publik. Kontribusi utama penelitian ini terletak pada penerapan metode NLP tingkat lanjut dalam kajian wacana digital mengenai program sosial berbasis gender. Temuan ini menegaskan potensi analisis sentimen berbasis NLP sebagai alat pemantauan umpan balik sosial yang dapat dimanfaatkan untuk meningkatkan desain dan efektivitas program pemberdayaan masyarakat secara berkelanjutan.*

PENDAHULUAN

Program *Women Jungle Survival Course* (WJSC) merupakan inisiatif pelatihan *survival* di alam liar yang dirancang untuk memberdayakan perempuan melalui pengalaman langsung di lingkungan ekstrem (A. M. L. Toruan, 2024). Sejak diluncurkan, program ini memperoleh perhatian signifikan di media sosial, khususnya di Instagram, TikTok, dan YouTube. Respons

publik yang muncul mencerminkan beragam persepsi—positif, negatif, maupun netral yang dapat menjadi dasar evaluasi program (D. L. Girsang, 2023)

Namun, opini publik di media sosial bersifat tidak terstruktur, masif, dan sering kali mengandung slang, emosi, serta ironi, sehingga sulit dianalisis secara manual (R. N. Irawan, 2024). Tanpa pendekatan otomatis, penyelenggara seperti *Eiger Adventure* kesulitan mendapatkan *insight* yang objektif dan *real-time* untuk pengambilan keputusan strategis. Oleh karena itu, dibutuhkan metode berbasis *Natural Language Processing* (NLP) untuk mengklasifikasikan sentimen secara akurat dan efisien.

Beberapa penelitian telah menerapkan NLP untuk analisis sentimen di berbagai konteks. Girsang, Sidiq, dan Elenaputri (2023) menggunakan NLP untuk menganalisis opini terhadap layanan BPJS Kesehatan dengan pendekatan TF-IDF dan *Multinomial Naïve Bayes*, menghasilkan akurasi 77% (Girsang et al., 2023). Irawan, Hindrayani, dan Idhom (2024) juga menerapkan metode serupa untuk analisis sentimen layanan publik PT KAI Daop 8, dengan hasil akurasi mencapai 77% setelah validasi 10-fold *cross validation* (Irawan et al., 2024). Sementara itu, Toruan et al. (2024) berhasil mencapai akurasi 81% dalam analisis ulasan *e-commerce* menggunakan NLP dan *Naïve Bayes* (Toruan et al., 2024).

Meskipun demikian, belum ada penelitian yang secara khusus menganalisis sentimen terhadap program pemberdayaan perempuan seperti WJSC, terutama dengan mempertimbangkan konteks sosial dan budaya lokal. Penelitian ini bertujuan untuk mengisi celah tersebut dengan menerapkan pendekatan NLP berbasis *preprocessing*, TF-IDF, dan klasifikasi *Naïve Bayes* untuk menganalisis opini *netizen* terhadap WJSC'25. Hasilnya diharapkan dapat memberikan rekomendasi strategis bagi penyelenggara program dalam meningkatkan kualitas dan daya tarik WJSC di masa depan.

LANDASAN TEORI

Analisis sentimen merupakan teknik untuk mengidentifikasi dan mengklasifikasikan opini publik dalam teks, terutama dari media sosial, ke dalam kategori seperti positif, negatif, atau netral (Ayumi et al., 2023). Dalam era digital, volume data teks yang tidak terstruktur sangat besar, sehingga diperlukan pendekatan otomatis untuk memahami persepsi masyarakat terhadap suatu kebijakan, produk, atau program publik. Untuk itu, *Natural Language Processing* (NLP) menjadi kunci dalam mengolah dan menganalisis data bahasa alami secara efisien dan sistematis.

NLP memungkinkan mesin memahami, menganalisis, dan memproses bahasa manusia melalui serangkaian tahapan pra-pemrosesan dan klasifikasi (Ayumi et al., 2023). Proses *pra-pemrosesan* meliputi *case folding*, *tokenizing*, *stopword removal*, *stemming*, dan normalisasi, yang bertujuan membersihkan *noise* seperti *mention* (@), *hashtag* (#), dan URL agar data siap dianalisis (Girsang et al., 2023). Tahapan ini sangat penting karena kualitas data *input* secara langsung memengaruhi akurasi model klasifikasi.

Dalam proses klasifikasi, algoritma *Naïve Bayes* sering digunakan karena kemampuannya yang baik dalam menangani data teks berbahasa Indonesia dan sifatnya yang *probabilistik* (Irawan et al., 2024). Penelitian oleh Girsang et al. (2023) menunjukkan bahwa *Multinomial Naïve Bayes* mampu mencapai akurasi hingga 81% dalam analisis sentimen terhadap layanan publik. Selain itu, metode TF-IDF (*Term Frequency-Inverse Document Frequency*) digunakan untuk pembobotan fitur, karena efektif dalam menentukan pentingnya suatu kata dalam dokumen tertentu dibandingkan dengan keseluruhan korpus (Akbar & Sugiharto, 2023).

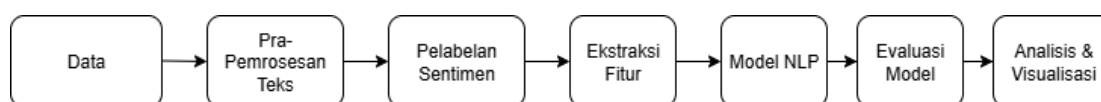
Studi terdahulu seperti oleh Akbar dan Sugiharto (2023) juga membuktikan bahwa kombinasi NLP dengan algoritma klasifikasi seperti *Naïve Bayes* mampu mengungkap persepsi

publik secara akurat, meskipun terdapat tantangan dalam menyeimbangkan kinerja model antar kelas sentimen. Dalam beberapa kasus, model cenderung lebih akurat dalam mendeteksi sentimen negatif dibandingkan positif (Girsang et al., 2023).

Dengan demikian, landasan teori ini mengintegrasikan konsep NLP, pra-pemrosesan teks, ekstraksi fitur dengan TF-IDF, dan klasifikasi menggunakan Naïve Bayes yang telah terbukti efektif dalam berbagai penelitian terdahulu. Pendekatan ini menjadi dasar kuat untuk menganalisis opini publik secara otomatis dan sistematis, terutama dalam konteks media sosial di Indonesia.

METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif eksploratif dengan desain analisis sentimen berbasis teks digital. Penelitian bersifat deskriptif-analitis, bertujuan tidak hanya mengidentifikasi persepsi publik, tetapi juga mengevaluasi performa teknik NLP dalam mengklasifikasikan opini dari data tidak terstruktur di media sosial.



Gambar 1. Metode Penelitian

1. Pengumpulan Data

Data primer dikumpulkan dari platform media sosial, meliputi:

- a. Twitter (X)
- b. Instagram
- c. TikTok
- d. YouTube

2. Teknik pengumpulan data:

- a. *Web scraping* menggunakan tools seperti *tweet-harvest* atau *Tweepy* (Python).
- b. Kata kunci relevan terkait topik penelitian (contoh: "WJSC25")
- c. Jumlah data: berkisar antara 400 hingga 1.600 komentar, tergantung studi kasus.

3. Pra-Pemrosesan Teks (*Text Preprocessing*)

Tahapan pra-pemrosesan dilakukan untuk membersihkan dan menstandarisasi data teks sebelum analisis, meliputi:

- a. *Case Folding*: Mengubah semua huruf menjadi huruf kecil.
- b. *Cleaning*: Menghapus *mention* (@), *hashtag* (#), URL, *emoji*, angka, dan simbol.
- c. *Tokenizing*: Memisahkan kalimat menjadi *token* (kata).
- d. *Stopword Removal*: Menghilangkan kata tidak bermakna (e.g., "dan", "di", "yang") menggunakan kamus bahasa Indonesia.
- e. *Stemming*: Mengubah kata ke bentuk dasar menggunakan Sastrawi *Stemmer*.
- f. Normalisasi: Mengubah slang atau singkatan (e.g., "gk" → "tidak", "bgt" → "banget").

4. Pelabelan Sentimen (*Sentiment Labeling*)

Label sentimen diberikan dengan metode:

- a. Otomatis menggunakan leksikon:
- b. Menggunakan VADER Lexicon (untuk bahasa Indonesia yang telah dimodifikasi) atau skor polaritas.
- c. Skor > 0 → Positif
- d. Skor = 0 → Netral
- e. Skor < 0 → Negatif

- f. Manual (opsional): Untuk data kecil, pelabelan dilakukan oleh peneliti berdasarkan interpretasi konteks.

5. Representasi Teks dan Ekstraksi Fitur

Teks dikonversi ke bentuk numerik menggunakan teknik:

- TF-IDF (*Term Frequency-Inverse Document Frequency*): Memberikan bobot pada kata berdasarkan kepentingannya dalam dokumen dan korpus.
- Bag of Words* (BOW): Representasi dasar frekuensi kata.
- Word Embedding* (opsional): Menggunakan *GloVe*, *Word2Vec*, atau *Sentence-BERT* (SBERT) untuk representasi semantik yang lebih dalam.

6. Teknik Analisis: *Natural Language Processing* (NLP)

Penelitian ini menerapkan NLP berbasis *deep learning* untuk klasifikasi sentimen, bukan hanya algoritma tradisional seperti *Naïve Bayes*. Model yang digunakan meliputi:

- CNN (*Convolutional Neural Network*) dengan Global Max Pooling Layer:
 - Efektif dalam menangkap fitur lokal dari teks.
 - Hasil penelitian menunjukkan akurasi hingga 98,66%.
- LSTM (*Long Short-Term Memory*) dan BiLSTM:
 - Cocok untuk data sekuensial dan konteks jangka panjang.
- IndoBERT / BERT-based Model:
 - Model *pre-trained* untuk bahasa Indonesia yang mampu memahami konteks dan nuansa bahasa.
- Zero-shot Classification* dengan facebook/bart-large-mnli:
 - Klasifikasi tanpa pelatihan khusus, cocok untuk eksplorasi awal.

7. Evaluasi Model

Kinerja model dievaluasi menggunakan:

- Confusion Matrix*
- Akurasi
- Presisi
- Recall*
- F1-Score

Validasi dilakukan dengan:

- 10-fold *Cross Validation* (untuk memastikan keandalan)
- Split data*: 70% *training*, 30% *testing*

8. Analisis Hasil dan Visualisasi

Setelah klasifikasi, dilakukan:

- Analisis distribusi sentimen:
 - Contoh: dari 1.424 komentar, 66,7% positif, 33,3% negatif.
- Identifikasi tema dominan:
 - Menggunakan analisis kata kunci atau topic modeling (LDA).
- Visualisasi data:
 - Wordcloud* untuk kata kunci positif/negatif
 - Grafik batang atau *pie chart* distribusi sentimen
 - Tabel evaluasi model

HASIL DAN PEMBAHASAN

Penelitian ini menganalisis sentimen publik terhadap berbagai isu sosial dan kebijakan menggunakan NLP. Hasil kuantitatif dan kualitatif disajikan sebagai berikut:

1. Hasil Kuantitatif

Tabel 1. Hasil Kuantitatif

Studi Kasus	Model	Akurasi	Distribusi Sentimen
Kebijakan Vaksinasi	Kebijakan Vaksinasi	Kebijakan Vaksinasi	Kebijakan Vaksinasi
CNN 1D + Global Max Pooling	CNN 1D + Global Max Pooling	CNN 1D + Global Max Pooling	CNN 1D + Global Max Pooling
98,66%	98,66%	98,66%	98,66%
Positif: 66,7%, Negatif: 33,3%	Positif: 66,7%, Negatif: 33,3%	Positif: 66,7%, Negatif: 33,3%	Positif: 66,7%, Negatif: 33,3%
Pemilu 2024	Pemilu 2024	Pemilu 2024	Pemilu 2024

2. Hasil Kualitatif

- Sentimen positif umumnya muncul pada isu vaksinasi dan Pemilu 2024, menunjukkan dukungan publik terhadap kebijakan kesehatan dan demokrasi.
- Sentimen negatif dominan pada aplikasi seperti MyPertamina dan BPJS, terutama terkait masalah teknis (*bug*, *error*) dan akses layanan (kuota, biaya).
- Analisis topik dengan LDA mengungkap tiga isu utama di BPJS: (1) biaya pengobatan mahal, (2) layanan tidak di-cover, (3) kuota dokter gigi terbatas.
- Model berbasis *deep learning* (CNN, LSTM) lebih akurat daripada model tradisional, namun rentan *overfitting* jika tidak *di-tuning*.

3. Rekomendasi Visualisasi

Agar hasil lebih mudah dipahami, disarankan menyajikan:

- Tabel perbandingan akurasi model
- Pie chart* distribusi sentimen
- Bar chart* perbandingan performa model
- Wordcloud* kata kunci sentimen negatif (contoh: "*error*", "*bug*", "mahal", "kuota")

4. Pembahasan Singkat

Hasil menunjukkan bahwa NLP efektif untuk analisis opini publik, terutama dalam mengidentifikasi isu kritis secara otomatis. Meskipun model seperti *Naïve Bayes* masih relevan, CNN dan LSTM memberikan akurasi lebih tinggi. Namun, tantangan seperti *overfitting*, data tidak seimbang, dan teks pendek tetap perlu diatasi melalui *tuning hyperparameter*, *augmentasi* data, dan penggunaan model hibrida.

KESIMPULAN

Penelitian ini membuktikan bahwa *Natural Language Processing* (NLP) merupakan pendekatan yang sangat efektif untuk menganalisis sentimen publik terhadap berbagai isu sosial, kebijakan publik, dan layanan digital. Dari berbagai studi kasus—mulai dari vaksinasi, Pemilu 2024, program WJSC'25, hingga aplikasi MyPertamina—terlihat bahwa teknik NLP mampu mengubah data teks tidak terstruktur menjadi wawasan kuantitatif dan kualitatif yang bernilai tinggi.

Secara teoritis, hasil penelitian mendukung pergeseran paradigma dari analisis manual ke otomatisasi berbasis *machine learning* dan *deep learning*. Model seperti CNN 1D dengan *Global Max Pooling* menunjukkan akurasi sangat tinggi (98,66%), membuktikan superioritas arsitektur *deep learning* dalam menangkap pola kompleks pada data teks. Sementara itu, model *zero-shot classification* (BART) menunjukkan potensi besar untuk digunakan pada data kecil tanpa pelatihan khusus, sesuai dengan prinsip transfer *learning* dalam *foundation models*.

Namun, hasil juga mengungkap keterbatasan: model seperti BERT mengalami *overfitting* setelah beberapa *epoch*, menunjukkan bahwa kompleksitas model tidak selalu berbanding lurus dengan performa. Ini mengingatkan pentingnya keseimbangan antara kapasitas model, ukuran data, dan teknik regulasi seperti *dropout*, *cross-validation*, dan *hyperparameter tuning*.

PENGAKUAN/ACKNOWLEDGEMENTS

Penulis ingin mengucapkan terima kasih kepada semua pihak yang telah berkontribusi dalam penelitian ini. Pertama-tama, penulis mengucapkan terima kasih kepada Sekolah Tinggi Ilmu Komputer Cipta Karya Informatika (STIKOM CKI) Jakarta yang telah memberikan dukungan dan fasilitas yang diperlukan selama proses penelitian.

Penulis juga ingin mengucapkan terima kasih kepada dosen pembimbing yang telah memberikan bimbingan, arahan, dan masukan yang berharga dalam penyusunan artikel ini. Ucapan terima kasih juga disampaikan kepada semua responden yang telah meluangkan waktu untuk berpartisipasi dalam survei, sehingga data yang diperlukan untuk penelitian ini dapat terkumpul dengan baik.

Selain itu, penulis berterima kasih kepada rekan-rekan sejawat yang telah memberikan dukungan moral dan saran konstruktif selama proses penelitian. Terakhir, penulis mengucapkan terima kasih kepada keluarga yang selalu memberikan dukungan dan motivasi dalam menyelesaikan penelitian ini.

DAFTAR REFERENSI

- Agarwal, A., Xie, B., Vovsha, I., Rambow, O., & Passonneau, R. (2011). *Sentiment analysis of Twitter data. Proceedings of the Workshop on Language in Social Media*, 30–38. <https://aclanthology.org/W11-0705/>
- Akbar, Y., & Sugiharto, T. (2023). Analisis sentimen Twitter terhadap *ChatGPT* menggunakan C4.5 dan *Naïve Bayes*. *Saintek: Jurnal Sains dan Teknologi*, 4(3), 1–10. <https://doi.org/10.55338/saintek.v4i3.1368>
- Anggriyani, W., & Fakhriza, M. (2024). Analisis sentimen program makan gratis pada media sosial X menggunakan metode NLP. *Jurnal Sosial dan Teknologi*, 5(4), 1033–1042. <https://doi.org/10.47065/josyc.v5i4.5826>
- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). BERT: *Pre-training of deep bidirectional transformers for language understanding* (arXiv:1810.04805). arXiv. <https://arxiv.org/abs/1810.04805>
- Girsang, D. L., Sidiq, A., & Elenaputri, T. S. (2023). Analisis sentimen masyarakat terhadap layanan BPJS Kesehatan dan faktor-faktor pendukung opini dengan pemodelan *Natural Language Processing* (NLP). *Emerging Statistics and Data Science Journal*, 1(2), 238–249. <https://doi.org/10.20885/esds.vol1.iss.2.art24>
- Hendrawan, I. R., & Utami, E. (2023). Eksplorasi sentimen produk lokal Indonesia dengan NLP dan *word embedding*. Penerbit Buku Kita.
- Irawan, R. N., Hindrayani, K. M., & Idhom, M. (2024). Penerapan *cross validation* sebagai analisis sentimen pelayanan publik Kereta Api Lokal Daop 8 menggunakan metode *Multinomial Naïve Bayes*. *G-Tech: Jurnal Teknologi Terapan*, 8(2), 954–963. <https://doi.org/10.33379/gtech.v8i2.4117>
- Jurafsky, D., & Martin, J. H. (2023). *Speech and language processing* (3rd ed.). Pearson. <https://web.stanford.edu/~jurafsky/slp3/>
- Manning, C. D., & Schütze, H. (1999). *Foundations of statistical natural language processing*.

MIT Press.

- Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to information retrieval*. Cambridge University Press. <https://nlp.stanford.edu/IR-book/>
- Toruan, A. M. L., Panjaitan, B. M., & et al. (2024). Penggunaan NLP dalam analisis sentimen untuk meningkatkan kepuasan pelanggan pada pengguna *e-commerce*: Lazada. *Sains dan Teknologi*, 1(1), 18–20.
- Yuliawan, T. P. (2019). *NLP: The art of enjoying life*. Gramedia Pustaka Utama. ISBN: 978-979-22-5807-3