

ANALISIS OPINI PUBLIK TWITTER TERHADAP ISU ENERGI NASIONAL MENGUNAKAN PENDEKATAN INDOBERT DAN NAÏVE BAYES

Jennifer Verty¹⁾, Femmy Johan²⁾, Muhammad Rizky Pribadi³⁾

1. Fakultas Ilmu Komputer, Universitas Multi Data Palembang
email: jenniferverty_2327250040@mhs.mdp.ac.id
2. Fakultas Ilmu Komputer, Universitas Multi Data Palembang
email: femmyjohan_2327250035@mhs.mdp.ac.id
3. Fakultas Ilmu Komputer, Universitas Multi Data Palembang
email: rizky@mdp.ac.id

Abstract

The increase in fuel oil prices and electric vehicles has become an issue that continues to develop on Twitter. This study aims to analyze public sentiment regarding fuel oil price increases and Public Electric Vehicle Charging Stations (SPKLU) in Indonesia, comparing the Naive Bayes and IndoBERT methods in sentiment classification. The research method includes tweet data collection, data preprocessing, public sentiment labeling, model training, and performance evaluation using accuracy, precision, recall, and F1-score metrics. The preprocessing stage consists of case folding, tokenization, stopword removal, and stemming of Indonesian-language text. The data consist of 700 tweets collected from Twitter using the keywords related to fuel oil, electric vehicles, and Public Electric Vehicle Charging Stations (SPKLU). The results show that the IndoBERT method has better performance than Naive Bayes in sentiment classification because it is able to understand the context of the Indonesian language. This study is expected to contribute to the development of Indonesian-language sentiment analysis.

Kata Kunci : Twitter Sentiment Analysis, IndoBERT, Naive Bayes, Electric Vehicle.

A. PENDAHULUAN

Kebijakan penyesuaian harga Bahan Bakar Minyak (BBM) dan dorongan infrastruktur Stasiun Pengisian Kendaraan Listrik Umum (SPKLU) pada tahun 2026 memicu gelombang respon publik yang masif di media sosial X/Twitter. Dinamika opini ini mengandung data bernilai tinggi bagi pengambil keputusan. Namun, ekstraksi sentimen teks berbahasa Indonesia di media sosial memiliki tantangan tersendiri berupa penggunaan bahasa tidak baku, sarkasme, dan konteks yang kompleks.

Penelitian terdahulu oleh Susanto et al. [1] menunjukkan bahwa Naïve Bayes efisien dalam klasifikasi teks, tetapi lemah dalam memahami hubungan semantik antar kata. Di sisi lain, model berbasis *transformer* seperti IndoBERT yang diteliti Alfatah [2] mampu menangkap konteks bahasa Indonesia secara mendalam, tetapi membutuhkan komputasi lebih tinggi. Meskipun kedua metode telah banyak diuji, komparasi eksplisit kinerjanya dalam merespons isu ganda (kebijakan konvensional BBM vs transisi SPKLU) masih sangat terbatas.

Oleh karena itu, penelitian ini bertujuan untuk mengevaluasi dan membandingkan performa model Naïve Bayes dan IndoBERT dalam mengklasifikasikan sentimen publik. Penelitian ini diharapkan memberikan pemahaman yang komprehensif terkait efektivitas model *machine learning* konvensional versus *deep learning transformer* pada domain opini energi nasional.

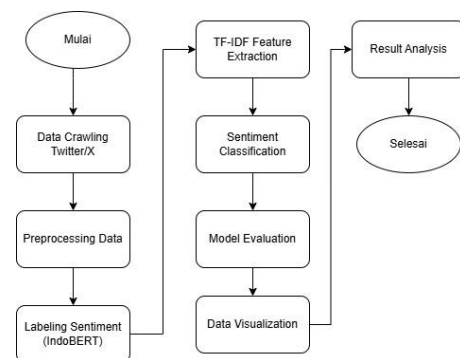
B. METODE PENELITIAN

Metode penelitian dalam penelitian ini terdiri dari beberapa tahapan, yaitu pengumpulan data *tweet* menggunakan *library tweet-harvest* dari *platform* Twitter berdasarkan kata kunci yang berkaitan dengan kenaikan harga BBM, kendaraan listrik, dan SPKLU Indonesia pada tahun 2026 [3]. Data yang diperoleh, dilakukan pemrosesan teks, pembersihan *mention*, *hashtag*, URL, simbol, dan karakter yang tidak diperlukan. Proses analisis sentimen menggunakan model IndoBERT untuk memberikan pelabelan sentimen positif, negatif, dan netral pada data [4]. Data kemudian diproses menggunakan metode Term Frequency–Inverse Document Frequency (TF-IDF) sebagai ekstraksi fitur sebelum dilakukan klasifikasi menggunakan metode Naive Bayes. Hasil klasifikasi, *social network analysis*, grafik *engagement*, dan grafik *top mention* akun Twitter/X untuk membantu memahami pola opini masyarakat terhadap isu.

Metodologi penelitian diawali dengan tahap pengumpulan data dari *platform* X/Twitter menggunakan kata kunci yang berkaitan dengan kenaikan harga Bahan Bakar Minyak (BBM), kendaraan listrik, dan Stasiun Pengisian Kendaraan Listrik Umum (SPKLU) di Indonesia pada tahun 2026. Data *tweet* yang diperoleh kemudian melalui tahap pemrosesan data untuk membersihkan teks dari URL,

simbol, emoji, tanda baca, serta melakukan *case folding*, tokenisasi, *stopword removal*, dan *stemming* sehingga data terstruktur dan siap untuk dianalisis data. Setelah itu, dilakukan pelabelan sentimen menggunakan IndoBERT guna mengklasifikasikan data ke dalam kategori sentimen positif, negatif, dan netral.

Tahap berikutnya adalah Term Frequency–Inverse Document Frequency (TF-IDF) yang digunakan untuk mengubah data teks menjadi representasi numerik sehingga dapat diproses oleh algoritma klasifikasi. Selanjutnya, proses klasifikasi sentimen dilakukan menggunakan metode Naive Bayes dan IndoBERT untuk menentukan sentimen dari *tweet* [5]. Hasil klasifikasi digunakan parameter *accuracy*, *precision*, *recall*, dan *F1-score* untuk membandingkan kinerja kedua metode. Selain itu, penelitian ini juga menggunakan visualisasi data berupa *wordcloud*, grafik sentimen, *social network analysis*, dan grafik *engagement* untuk mengamati pola diskusi dan interaksi masyarakat di media sosial. Tahap akhir penelitian adalah analisis hasil, yaitu menganalisis hasil pengujian dan visualisasi data untuk menarik kesimpulan mengenai respons masyarakat terhadap kenaikan harga BBM dan perkembangan kendaraan listrik di Indonesia. Tahapan metodologi penelitian dapat dilihat pada Gambar 1.



Gambar 1. Tahapan Metodologi Penelitian

C. HASIL DAN PEMBAHASAN

Hasil dari penelitian memberikan berbagai tanggapan terhadap isu kenaikan harga BBM dan perkembangan kendaraan listrik. Berdasarkan hasil analisis sentimen, sebagian besar *tweet* mengandung sentimen positif, *negative* dan netral terhadap kebijakan energi serta perkembangan SPKLU di Indonesia. Visualisasi menunjukkan bahwa kata-kata seperti "BBM", "mobil listrik", dan "SPKLU" menjadi kata yang paling dominan dalam *tweet*. Selain itu, hasil *Social Network Analysis* (SNA) menunjukkan interaksi aktif antara pengguna media sosial dalam membahas isu ini.

Dari hasil pengujian model menunjukkan model IndoBERT memiliki tingkat akurasi yang lebih akurat dibandingkan Naive Bayes karena mampu memahami konteks bahasa Indonesia dengan lebih baik [4]. Sementara, model Naive Bayes memiliki keunggulan dalam kecepatan proses klasifikasi. Oleh karena itu, IndoBERT dinilai lebih efektif untuk analisis sentimen berbahasa Indonesia pada data media sosial.

Hasil pengumpulan data *tweet* proses pengumpulan data dilakukan menggunakan library *tweet-harvest* dengan kata kunci yang berkaitan dengan kenaikan harga BBM, kendaraan listrik, baterai mobil listrik, dan SPKLU Indonesia pada tahun 2026. Data diperoleh dari *platform* X/Twitter dalam periode 1 April 2026 hingga 25 April 2026 [6]. Setelah proses *crawling* selesai, data disimpan dalam format CSV untuk tahap prapemrosesan dan analisis sentimen (Rizkia et al., 2025), sebagaimana ditunjukkan pada Gambar 2.

```
BLENDING SELESAI!
Total Baris Gabungan: 700
=====
```

	full_text	in_reply_to_screen_name	username	Nama_keyord
0	Kenaikan harga BBM non-subsidi kembali jadi pe...	NaN	NaN	Hasil_scrap_dataset_bbm_listrik1 (1)
1	@badgenlus_ perlu di ketahui Subsidy bbm itu y...	badgenlus_	NaN	Hasil_scrap_dataset_bbm_listrik1 (1)
2	13 SPBU Pertamina sudah STOP jual Pertalite!	NaN	NaN	Hasil_scrap_dataset_bbm_listrik1 (1)
3	Menyesuaikan harga BBM non-subsidi dengan harg...	NaN	NaN	Hasil_scrap_dataset_bbm_listrik1 (1)
4	@onyoutcha Diduga harga bbm naik mio sekarang...	onyoutcha	NaN	Hasil_scrap_dataset_bbm_listrik1 (1)

File hasil gabungan disimpan dengan nama: "Hasil_scrap_dataset_bbm_listrik1 (1).csv"

Gambar 2. Proses crawling data *tweet* menggunakan *tweet-harvest*

Berdasarkan hasil *crawling* pada Gambar 2, diperoleh 700 data *tweet* yang digunakan sebagai dataset penelitian. *Tweet* yang dikumpulkan terdiri atas opini masyarakat mengenai kebijakan BBM, kendaraan listrik, dan SPKLU di Indonesia.

Hasil visualisasi data menunjukkan bahwa visualisasi *wordcloud* dari *tweet* yang berkaitan dengan harga BBM dan kendaraan listrik di Indonesia. Kata-kata seperti "BBM", "mobil listrik", "SPKLU", "harga", dan "Indonesia" muncul paling sering dalam *tweet* masyarakat [7]. Menunjukkan bahwa isu mengenai kenaikan harga BBM dan kendaraan listrik menjadi topik yang banyak diperbincangkan di platform X/Twitter.

Hasil visualisasi juga menunjukkan bahwa diskusi masyarakat didominasi oleh topik yang berkaitan dengan kebijakan BBM, kendaraan listrik, dan SPKLU di Indonesia. Hal ini menunjukkan bahwa isu tersebut memperoleh perhatian yang cukup besar dari masyarakat di media sosial.



Gambar 3. Wordcloud Topik Kenaikan Harga BBM dan Kendaraan Listrik

Berdasarkan Gambar 3, kata-kata seperti "BBM", "mobil listrik", "SPKLU", "harga", dan "Indonesia" merupakan kata yang paling sering muncul dalam *tweet* masyarakat. Dominasi kata-kata menunjukkan pembahasan tentang kenaikan harga BBM dan perkembangan kendaraan listrik mendapat perhatian yang cukup besar di platform X/Twitter. Selain itu, sering munculnya istilah yang berkaitan dengan harga BBM mencerminkan perhatian masyarakat terhadap dampak ekonomi, sedangkan pembahasan mengenai kendaraan listrik dan SPKLU menunjukkan meningkatnya minat pendukungnya di Indonesia.



Gambar 4. Wordcloud Sentimen Positif dan Negatif

Berdasarkan Gambar 4, *wordcloud* menunjukkan kata-kata yang paling sering muncul pada kategori sentimen positif dan negatif yang berkaitan dengan pembahasan kenaikan harga BBM dan kendaraan listrik di *platform X/Twitter*. Pada kategori sentimen positif, kata-kata seperti "harga", "BBM", "pemerintah", "subsidi", dan "tetap" lebih sering muncul.

Sementara itu, pada kategori sentimen negatif, kata-kata dominan seperti "harga", "BBM", "naik", dan "yang" menunjukkan adanya kekhawatiran masyarakat serta reaksi negatif terhadap kenaikan harga BBM dan dampak ekonominya. Semakin besar ukuran kata pada visualisasi, semakin sering kata tersebut muncul dalam dataset *tweet*.

Tahap pemrosesan dilakukan untuk membersihkan data teks sebelum proses klasifikasi sentimen. Proses meliputi penghapusan *mention*, *hashtag*, URL, karakter khusus, angka, serta mengubah

seluruh teks menjadi huruf kecil *lowercase*.

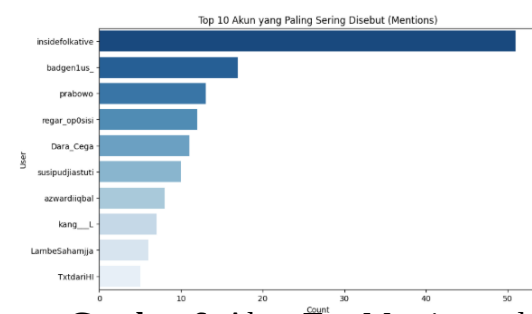
Full Text	Preprocessing Steps	Cleaned/ Final text
"BBM naik lagi @user #hemat energi https://link"	Remove mention, hashtag, URL, punctuation, lowercase	"bbm naik lagi hemat energi"
"Mobil listrik bagus banget!!! #EV2026"	Remove hashtag, lowercase	"mobil listrik bagus banget"
"SPKLU makin banyak di Indonesia"	Remove punctuation, lowercase	"spklu makin banyak di indonesia"

Gambar 5. Hasil Pemrosesan Data Tweet

Berdasarkan Gambar 5, tahap pemrosesan berhasil mengubah *tweet* mentah menjadi teks yang lebih bersih. Penghapusan komponen yang tidak diperlukan seperti *mention*, *hashtag*, URL, dan karakter khusus yang mengurangi *noise* pada dataset sehingga meningkatkan kualitas data dalam proses klasifikasi sentimen.

Setelah tahap pemrosesan selesai dilakukan, data *tweet* dianalisis menggunakan model IndoBERT. Model ini digunakan untuk memberikan label sentimen ke dalam kategori positif, negatif, dan netral [8]. Berdasarkan hasil analisis sentimen, sebagian besar *tweet* mengenai kenaikan harga BBM memiliki sentimen negatif. Sementara itu, sentimen terhadap kendaraan listrik dan SPKLU cenderung positif karena masyarakat mendukung penggunaan energi yang lebih ramah lingkungan.

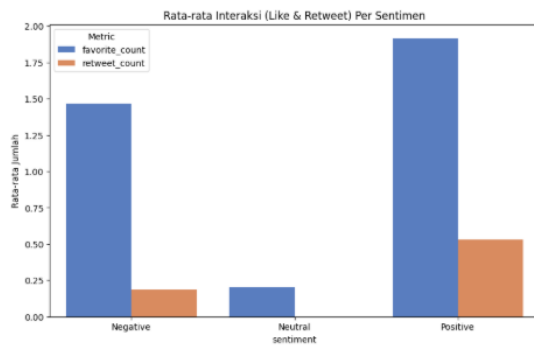
Social Network Analysis (SNA) digunakan untuk mengamati hubungan interaksi antar pengguna Twitter/X berdasarkan aktivitas *mention* pada *tweet*.



Gambar 6. Akun Top Mention pada Diskusi Twitter/X

Berdasarkan Gambar 6, ada 10 akun yang paling sering disebut dalam diskusi terkait kenaikan harga BBM, kendaraan listrik, dan SPKLU di *platform X/Twitter*. Akun yang paling banyak disebut adalah @insidefokative dengan jumlah *mention* tertinggi, diikuti oleh akun @bagdenuls_, @prabowo, dan beberapa akun lainnya. Menunjukkan akun-akun tersebut menjadi rujukan dalam percakapan publik mengenai isu yang dibahas [9].

Analisis *engagement* dilakukan untuk mengetahui rata-rata interaksi pengguna berdasarkan jumlah *like* dan *retweet* pada setiap kategori sentimen.



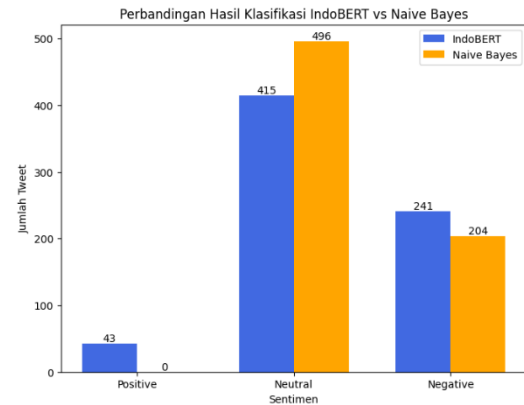
Gambar 7. Rata-rata Engagement Tweet Berdasarkan Kategori Sentimen

Berdasarkan Gambar 7, *tweet* dengan sentimen positif memperoleh rata-rata jumlah *like* (*favorite_count*) dan *retweet* (*retweet_count*) tertinggi, sentimen negatif, sedangkan sentimen netral menunjukkan *engagement* yang paling rendah. Hasil menunjukkan bahwa data *tweet* yang mengandung opini yang lebih kuat cenderung memperoleh interaksi yang lebih tinggi dari pengguna media sosial.

Analisis *top mention* dilakukan untuk mengetahui akun yang paling sering disebut oleh pengguna Twitter/X dalam diskusi mengenai kenaikan harga BBM dan kendaraan listrik. Berdasarkan hasil visualisasi, akun pemerintah, media berita, dan akun otomotif menjadi akun yang paling sering disebut oleh pengguna

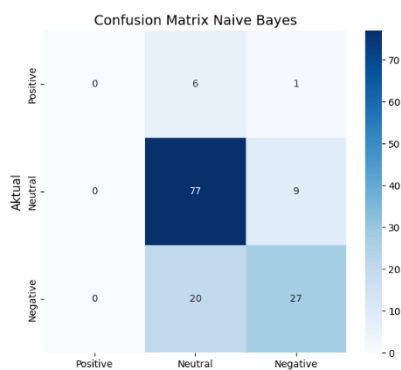
Twitter/X dalam pembahasan mengenai BBM dan kendaraan listrik.

Pada penelitian ini dilakukan perbandingan kinerja antara metode Naive Bayes dan IndoBERT dalam klasifikasi sentimen *tweet* berbahasa Indonesia[10].



Gambar 8. Grafik Perbandingan Akurasi Naive Bayes dan IndoBERT

Berdasarkan Gambar 8, perbedaan hasil klasifikasi sentimen antara model IndoBERT dan Naive Bayes. Pada kategori sentimen positif, IndoBERT mengklasifikasikan 43 *tweet* sebagai positif, sedangkan Naive Bayes tidak menghasilkan *tweet* dengan sentimen positif. Pada kategori netral, Naive Bayes menghasilkan jumlah klasifikasi paling tinggi yaitu 496 *tweet*, lebih banyak dibandingkan IndoBERT yang mengklasifikasikan 415 *tweet* sebagai netral. Pada kategori negatif, IndoBERT mengklasifikasikan 241 *tweet* dan Naive Bayes mengklasifikasikan 204 *tweet*. Hasil ini menunjukkan bahwa IndoBERT cenderung menghasilkan distribusi sentimen yang beragam dan mampu mengidentifikasi sentimen positif, sedangkan Naive Bayes lebih banyak mengelompokkan *tweet* ke dalam kategori netral.



Gambar 9. Confusion Matrix Model Naive Bayes

Confusion Matrix digunakan untuk mengetahui tingkat ketepatan prediksi model Naive Bayes terhadap data aktual [4]. Berdasarkan Confusion Matrix pada Gambar 8, sebagian besar data berhasil diprediksi dengan benar. Namun, masih terdapat beberapa kesalahan, khususnya pada sentimen netral dan negatif, yang disebabkan oleh kemiripan konteks kata dalam *tweet* dan ketidak mampuan Naive Bayes mengenali bahasa Indonesia.

D. KESIMPULAN DAN SARAN

Kesimpulan hasil penelitian, dapat disimpulkan bahwa analisis sentimen terhadap *tweet* yang berkaitan dengan kenaikan harga BBM, kendaraan listrik, dan SPKLU Indonesia berhasil dilakukan menggunakan metode Naive Bayes dan IndoBERT. Penelitian ini, 700 *tweet* yang dikumpulkan dari platform X/Twitter dan menunjukkan bahwa pembahasan mengenai kenaikan harga BBM dan kendaraan listrik isu yang sering dibahas masyarakat.

Hasil evaluasi menunjukkan bahwa IndoBERT mencapai kinerja yang lebih baik dibandingkan Naive Bayes dalam klasifikasi sentimen bahasa Indonesia. Naive Bayes memperoleh nilai *accuracy* sebesar 0,74, *precision* sebesar 0,70, *recall* sebesar 0,74, dan *F1-score* sebesar 0,72, sedangkan IndoBERT memperoleh

nilai 1,00 untuk *accuracy*, *precision*, *recall*, dan *F1-score*. Temuan ini menunjukkan bahwa pendekatan berbasis *transformer* lebih efektif dalam memahami informasi kontekstual dan hubungan semantik pada *tweet* berbahasa Indonesia dibandingkan dengan pendekatan *machine learning* konvensional [5].

Selain itu, hasil visualisasi data menunjukkan bahwa isu terkait kenaikan harga BBM dan kendaraan listrik menarik perhatian publik yang signifikan di platform X/Twitter. Sentimen publik cenderung lebih negatif terhadap kenaikan harga BBM dan lebih positif terhadap pengembangan kendaraan listrik serta infrastruktur SPKLU.

Penelitian ini memberikan kontribusi terhadap pengembangan analisis sentimen berbahasa Indonesia dengan menyediakan evaluasi komparatif antara pendekatan *machine learning* tradisional dan pendekatan berbasis *transformer* dalam menganalisis opini publik terkait isu energi dan transportasi di Indonesia.

Namun, penelitian ini memiliki beberapa keterbatasan, termasuk ukuran dataset yang relatif terbatas, periode pengamatan yang singkat, serta penggunaan hanya dua metode klasifikasi. Oleh karena itu, penelitian selanjutnya disarankan untuk menggunakan dataset yang lebih besar, periode pengamatan yang lebih panjang, serta menambahkan pendekatan *deep learning* lainnya agar diperoleh hasil analisis sentimen yang lebih komprehensif dan andal.

E. REFERENSI

- [1] A. Susanto, M. Atho'il Maula, I. Utomo, W. Mulyono, and K. Sarker, "Sentiment Analysis on Indonesia Twitter Data Using Naïve Bayes and K-Means Method," 2021. [Online]. Available: <http://kateglo.com>.

- [2] D. Alfatah, "Penerapan Model Transformer Untuk Deteksi Sentimen Pada Data Twitter Berbahasa Indonesia Application Of Transformer Model For Sentiment Detection On Indonesian Twitter Data."
- [3] N. Nyoman Aprianti *et al.*, "Public Sentiment Analysis of the Free Nutritious Meals Program (MBG) on Social Media X Using the Naive Bayes Method," 2025. [Online]. Available: <http://jurnal.polibatam.ac.id/index.php/JAIC>
- [4] M. F. Manao and S. Mujiyono, "Sentiment Analysis of the Free Nutritious Meal Program on Twitter Using Naive Bayes and IndoBERT-Based Labeling," *Journal of Information Systems and Informatics*, vol. 8, no. 1, pp. 1302–1327, Mar. 2026, doi: 10.63158/journalisi.v8i1.1345.
- [5] A. Amanda and E. Nurmawati, "ANALISIS SENTIMEN DAN PEMODELAN TOPIK PADA TWEET TERKAIT DATA BADAN PUSAT STATISTIK," *Jurnal Sistem Informasi dan Informatika (Simika) P-ISSN*, vol. 6, no. 2, pp. 2622–6901, 2023.
- [6] A. Mauluvy Senjaya and Y. Sibaroni, "Geospatial Sentiment Analysis of Twitter User (X) on Government Performance in Overcoming Floods in Jabodetabek Using IndoBERT and CNN-LSTM Methods," *Jurnal Pendidikan dan Teknologi Indonesia*, vol. 5, no. 11, pp. 3386–3398, Nov. 2025, doi: 10.52436/1.jpti.1182.
- [7] D. Retno Budiastuti, R. Kartika Sari, and S. Sembiring, "Peran Regulasi Pemerintah Terhadap Penjualan Mobil Listrik pada Agustus 2025," 2025.
- [8] F. Koto, J. H. Lau, and T. Baldwin, "IndoBERTweet: A Pretrained Language Model for Indonesian Twitter with Effective Domain-Specific Vocabulary Initialization," *INDOBERTWEET: A Pretrained Language Model for Indonesian Twitter with Effective Domain-Specific Vocabulary Initialization*, Sep. 2021, [Online]. Available: <http://arxiv.org/abs/2109.04607>
- [9] S. Fitrianti and A. Yudhistira, "Analisis Sentimen Media Sosial Terhadap Calon Pilkada 2024 Dengan Metode Naïve Bayes," *Jurnal Pendidikan dan Teknologi Indonesia*, vol. 5, no. 1, pp. 167–176, Jan. 2025, doi: 10.52436/1.jpti.610.
- [10] A. H. A. Zharifa and E. I. H. Ujianto, "Analisis Sentimen Publik di Twitter Pasca Debat Kelima Pilpres 2024 dengan Naive Bayes," *Edumatic: Jurnal Pendidikan Informatika*, vol. 8, no. 2, pp. 754–763, Dec. 2024, doi: 10.29408/edumatic.v8i2.28048.