

ANALISIS SENTIMEN KOMENTAR YOUTUBE TENTANG BOIKOT TRANS 7 DENGAN METODE *K-Nearest Neighbor* (KNN)

Ah Khafid Nasrul Umam^{1*}, Siti Mujilahwati², Danang Bagus Reknandi³

^{1,2,3} Universitas Islam Lamongan; Jalan Veteran No. 53A, Jetis, Kecamatan Lamongan, Kabupaten Lamongan, Jawa Timur

Keywords:

Analisis Sentimen; Komentar Youtube; Boikot Trans 7; *K-Nearest Neighbor* (KNN).

Correspondent Email:

nasrulumam051203@gmail.com

Abstrak. Perkembangan media sosial menempatkan platform seperti YouTube sebagai barometer opini publik, salah satunya pada fenomena seruan boikot Trans 7 terkait polemik konten Guru Gembul. Penelitian ini bertujuan menganalisis polaritas opini masyarakat serta mengevaluasi kinerja algoritma *K-Nearest Neighbor* (K-NN) berbasis pembobotan fitur IndoRoBERTa untuk mengklasifikasikan komentar ke dalam kategori negatif, netral, dan positif. Dari 3.700 komentar yang dikumpulkan via YouTube API, serangkaian *text preprocessing* menghasilkan 1.856 data valid. Pengujian menunjukkan konfigurasi $K=5$ lebih optimal dibanding $K=3$. Distribusi data didominasi sentimen negatif sebesar 68,53% (1.272 komentar), netral 24,62% (457 komentar), dan positif 6,84% (127 komentar). Evaluasi K-NN ($K=5$) menghasilkan akurasi 64,5%, namun mengalami bias akibat *class imbalance* ekstrem. Model sangat andal mengenali sentimen negatif (presisi 71,9%, *recall* 82,7%), tetapi berkinerja rendah pada netral (presisi 36,8%, *recall* 30,4%) dan positif (*recall* 7,7%). Penelitian ini menyimpulkan bahwa K-NN berbasis IndoRoBERTa mampu memetakan opini publik, tetapi memerlukan teknik penanganan data seperti SMOTE pada riset mendatang.



Copyright © 2026 The Author(s), Published by Jurnal Informatika dan Teknik Elektro Terapan. This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0).

Abstract. The development of social media has positioned platforms like YouTube as a barometer of public opinion, as seen in the boycott call against Trans 7 concerning the controversy surrounding Guru Gembul's content. This study aims to analyze the polarity of public opinion and evaluate the performance of the *K-Nearest Neighbor* (K-NN) algorithm with IndoRoBERTa feature weighting in classifying comments into negative, neutral, and positive categories. From 3,700 comments collected via the YouTube API, a series of *text preprocessing* steps yielded 1,856 valid entries. Parameter testing revealed that a configuration of $K = 5$ was more optimal than $K = 3$. The data distribution was dominated by negative sentiment at 68.53% (1,272 comments), followed by neutral at 24.62% (457 comments), and positive at 6.84% (127 comments). The evaluation of K-NN ($K = 5$) achieved an accuracy of 64.5%, but exhibited strong classification bias due to extreme class imbalance. The model proved highly reliable in identifying negative sentiment (precision 71.9%, recall 82.7%), but performed poorly on neutral sentiment (precision 36.8%, recall 30.4%) and struggled significantly with positive sentiment (recall 7.7%). This study concludes that IndoRoBERTa-based K-NN can map public opinion, but requires data handling techniques such as SMOTE to optimize future research.

1. PENDAHULUAN

Perkembangan teknologi informasi dan komunikasi telah menempatkan media sosial, khususnya YouTube, sebagai barometer opini publik utama bagi masyarakat global dan Indonesia [1]. Kolom komentar YouTube berfungsi sebagai ruang diskusi terbuka yang merepresentasikan kepuasan, kritik, hingga penolakan masyarakat terhadap kinerja entitas media massa konvensional. Salah satu fenomena krusial yang mencuat di ruang digital adalah seruan boikot terhadap stasiun televisi nasional Trans 7. Eskensi polemik ini menghasilkan ribuan data komentar yang masif (big data), tidak terstruktur, dan memiliki kompleksitas bahasa yang tinggi [2]. Menganalisis sentimen publik terhadap krisis reputasi semacam ini secara manual tidak lagi memungkinkan. Oleh karena itu, diperlukan pendekatan komputasional otomatis berbasis Natural Language Processing (NLP) untuk mengukur polaritas opini secara kuantitatif dan objektif [3].

Dalam ranah analisis sentimen dan klasifikasi teks, berbagai algoritma machine learning seperti Naive Bayes, Support Vector Machine (SVM), dan K-Nearest Neighbor (K-NN) telah banyak diterapkan untuk memetakan opini publik [4]. Algoritma K-NN menjadi salah satu metode pembelajaran terawasi (supervised learning) non-parametrik yang populer karena kesederhanaannya serta kemampuannya mengelompokkan data berdasarkan kalkulasi jarak geometris (Euclidean atau Cosine Distance). Di sisi lain, perkembangan model bahasa berbasis Transformer, khususnya IndoRoBERTa, telah menunjukkan performa unggul dalam pemrosesan bahasa alami berbahasa Indonesia. Melalui mekanisme multi-head self-attention dan pre-training pada korpus besar Bahasa Indonesia, IndoRoBERTa mampu mengekstrak representasi fitur kontekstual dan menangkap nuansa linguistik secara lebih mendalam [5].

Meskipun K-NN terbukti efektif pada data dengan pola pengelompokan yang jelas, kinerja K-NN sangat bergantung pada representasi vektor data teks. Sebagian besar penelitian klasifikasi sentimen berbasis K-NN terdahulu masih mengandalkan teknik pembobotan kata statistik tradisional (seperti TF-IDF atau Bag-of-Words). Lemah dalam menangkap konteks semantik dan tata urutan kata, pembobotan

tradisional ini rentan terhadap noise akibat tingginya penggunaan bahasa informal, kata tidak baku, maupun sarkasme pada komentar media sosial, sehingga mengacaukan perhitungan jarak geometris K-NN dan menurunkan akurasi model. Penggunaan representasi fitur canggih dari model Transformer seperti IndoRoBERTa menawarkan solusi untuk menghasilkan vektor fitur yang kaya akan konteks semantik. Namun, eksplorasi mengenai bagaimana sinergi antara pembobotan fitur kontekstual IndoRoBERTa dan algoritma K-NN bekerja khususnya dalam menangani ketidakseimbangan kelas (class imbalance) dan variasi parameter jarak pada kasus krisis reputasi media masih sangat terbatas. Kesenjangan (gap) inilah yang menjadi fokus utama dalam penelitian ini.

Berdasarkan kesenjangan riset tersebut, penelitian ini bertujuan untuk mengintegrasikan teknik pra-pemrosesan teks, representasi fitur kontekstual IndoRoBERTa, dan algoritma K-NN untuk membangun pemodelan klasifikasi sentimen yang akurat terhadap komentar YouTube terkait isu boikot Trans 7.

2. TINJAUAN PUSTAKA

2.1. Analisis Sentimen dan YouTube

Kolom komentar YouTube berfungsi sebagai ruang diskusi publik masif yang memetakan dinamika opini, dukungan, maupun penolakan masyarakat terhadap isu kontroversial secara *real-time* [6]. Analisis sentimen sebagai cabang *Natural Language Processing* (NLP) mengkaji opini, emosi, dan sikap subjektif pengguna dengan mengklasifikasikan data teks tidak terstruktur ke dalam polaritas emosional seperti positif, negatif, atau netral [7]. Secara tipologi, analisis sentimen mencakup *Fine-Grained Sentiment*, *Emotion Detection*, *Intent-Based Analysis*, dan *Aspect-Based Analysis* [8]. Dalam konteks krisis reputasi, penafsiran sentimen ini menjadi instrumen evaluasi strategis bagi organisasi dalam memitigasi isu digital [6].

2.2. K-Nearest Neighbor (K-NN)

K-Nearest Neighbor (K-NN) adalah algoritma *supervised learning* non-parametrik yang bekerja berdasarkan prinsip bahwa data dengan karakteristik serupa berada berdekatan dalam

ruang fitur [9]. Algoritma ini mengklasifikasikan data baru berdasarkan mayoritas label dari K tetangga terdekatnya. Perhitungan kedekatan antar-dokumen umumnya mengandalkan metrik jarak geometris, seperti *Euclidean Distance* (garis lurus) atau *Manhattan Distance* (penjumlahan selisih absolut koordinat). Kinerja K-NN sangat bergantung pada representasi fitur numerik teks serta optimasi parameter, seperti penentuan nilai K dan pemilihan metrik jarak yang sesuai dengan karakteristik dataset.

1. Jarak *Euclidean*:

etode ini digunakan untuk menghitung jarak geometris garis lurus antara dua titik data di dalam ruang bermatra atau n -dimensi.

$$d = \sqrt{\sum_{i=1}^p (x_{2i} - x_{1i})^2}$$

2. Jarak *Manhattan*:

Digunakan untuk menghitung jarak dengan menjumlahkan nilai absolut dari perbedaan koordinatnya.

$$d = \sum_{i=1}^p |x_{2i} - x_{1i}|$$

Keterangan:

x_1 = Data sampel

x_2 = Data uji atau data pengujian

i = Variabel data

d = Nilai jarak

p = Dimensi data

2.3. IndoBERT dan IndoRoBERTa

IndoBERT merupakan *pre-trained language model* berbasis arsitektur *Bidirectional Encoder Representations from Transformers* (BERT) yang dirancang khusus untuk pemrosesan teks Bahasa Indonesia. Menggunakan mekanisme *self-attention* dan *WordPiece tokenizer*, IndoBERT menghasilkan representasi fitur kontekstual dua arah (*bidirectional*). Model ini kemudian dikembangkan lebih lanjut menjadi IndoRoBERTa, yang mengoptimalkan arsitektur BERT dengan mengeliminasi tugas *Next Sentence Prediction* (NSP) serta mengadopsi *Byte-Pair Encoding* (BPE) pada korpus data yang lebih besar, sehingga mampu menghasilkan representasi fitur yang lebih kaya dan presisi untuk tugas klasifikasi teks [10].

1. Sebelum teks diproses oleh lapisan *Transformer*, setiap kata (*token*) diubah

menjadi vektor numerik melalui penjumlahan tiga jenis *embedding*

$$E_{\text{input}} = E_{\text{token}} + E_{\text{position}} + E_{\text{segment}}$$

Keterangan:

E_{input} : Representasi vektor akhir yang masuk ke model.

E_{token} : Vektor identitas dari kata/token tersebut.

E_{position} : Vektor posisi yang menandai urutan kata dalam kalimat.

E_{segment} : Vektor penanda untuk membedakan antar-kalimat.

2. Mekanisme utama dalam *IndoRoBERTa* adalah *Self-Attention*, yang memungkinkan model menghitung seberapa besar keterkaitan antara suatu kata dengan kata-kata lainnya dalam satu kalimat.

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

Keterangan:

Q (*Query*): Vektor proyeksi teks yang sedang dianalisis.

K (*Key*): Vektor proyeksi teks penyeimbang untuk pencocokan.

V (*Value*): Vektor nilai informasi dari teks.

d_k : Dimensi dari matriks K , berfungsi sebagai skala pembagi agar gradien tetap stabil.

softmax: Fungsi aktivasi untuk mengubah nilai matriks menjadi probabilitas bobot perhatian (bernilai 0 hingga 1).

Pada tahap akhir, vektor representasi kontekstual yang dihasilkan oleh lapisan *Transformer* dilewatkan ke lapisan *Linear Classifier* (*Dense Layer*). Fungsi *Softmax* kemudian digunakan untuk mengonversi nilai *logit* menjadi nilai probabilitas untuk setiap kelas sentimen (seperti Positif, Netral, dan Negatif):

$$P(Y = k | x) = \frac{e^{z_k}}{\sum_{j=1}^C e^{z_j}}$$

Keterangan:

$P(Y = k | x)$: Probabilitas teks x masuk ke dalam kategori/label sentimen k .

z_k : Nilai *logit* (skor mentah) untuk kelas ke- k .

C : Total jumlah kelas sentimen yang digunakan dalam klasifikasi.

e : Bilangan Euler (konstanta eksponensial).

2.4. Pra-pemrosesan Teks

Pra-pemrosesan teks merupakan tahap krusial untuk mentransformasi data mentah yang tidak terstruktur dari media sosial menjadi format yang siap diproses oleh algoritma klasifikasi [11]. Tahapan ini meliputi:

Case Folding: Mengubah seluruh huruf menjadi huruf kecil (*lowercase*) serta menghapus angka dan karakter khusus.

Tokenizing: Memecah string teks menjadi potongan kata (*token*) individual.

Normalisasi Kata: Memperbaiki kata tidak baku, *slang*, atau *typo* menjadi bentuk standar sesuai kamus. [12]

Stopword Removal: Menghapus kata-kata umum yang tidak memiliki bobot informasi signifikan terhadap penentuan sentimen.

2.5. Evaluasi Performa

Confusion Matrix digunakan untuk mengukur kinerja model klasifikasi dengan membandingkan prediksi model terhadap label aktual yang terbagi menjadi True Positive (TP), True Negative (TN), False Positive (FP), dan False Negative (FN) [6]. Dari matriks ini, diperoleh empat metrik utama:

Akurasi (Accuracy): Mengukur persentase ketepatan prediksi model secara keseluruhan.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

Presisi (Precision): Mengukur proporsi prediksi positif yang benar-benar bernilai positif.

$$\text{Precision} = \frac{TP}{TP + FP}$$

Sensitivitas (Recall): Mengukur kemampuan model dalam menemukan kembali seluruh data kelas positif aktual.

$$\text{Recall} = \frac{TP}{TP + FN}$$

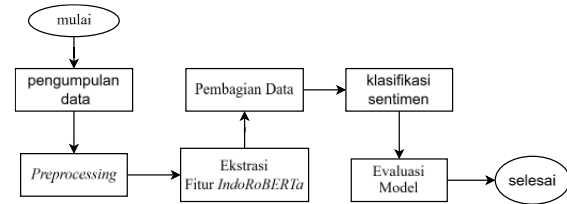
F1-Score: Rata-rata harmonik antara presisi dan recall yang memberikan penilaian seimbang, terutama pada kondisi data tidak seimbang (imbalanced dataset).

$$F_1\text{-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

3. METODE PENELITIAN

Bab ini menjelaskan alur sistematis penelitian yang meliputi pengumpulan data, text preprocessing, ekstraksi fitur IndoRoBERTa, klasifikasi K-NN, hingga evaluasi model. Rancangan tahapan penelitian secara keseluruhan disajikan pada Gambar 1

berikut.



Gambar 1 Alur Penelitian

3.1. Pengumpulan Data (Crawling Data)

Data dikumpulkan melalui *crawling* komentar YouTube pada kanal GEMBULIKUM dan Guru Gembul (periode November 2025) menggunakan Python dan YouTube API. Data mentah berformat CSV kemudian diseleksi dengan menghapus *spam*, duplikasi, serta komentar yang tidak relevan.

3.2. Pra-Pemrosesan Data

Tahap Preprocessing merupakan langkah awal dalam analisis teks untuk menyiapkan data mentah menjadi bentuk yang dapat diproses oleh algoritma. Langkah ini berperan penting karena kualitas hasil klasifikasi sangat dipengaruhi oleh kebersihan dan konsistensi data [14].

3.3. Ekstraksi Fitur Menggunakan IndoRoBERTa

Klasifikasi sentimen dilakukan menggunakan model IndoRoBERTa (w11wo/indonesian-roberta-base-sentiment-classifier). Teks hasil *preprocessing* diubah menjadi *contextual embeddings* untuk menghitung skor probabilitas dan menentukan kelas sentimen (Positif, Netral, atau Negatif) [15].

3.4. Pembagian Data

Dataset penelitian terdiri dari komentar YouTube terkait isu boikot Trans 7 hasil *crawling* via API resmi. Dari 3.700 data awal, diperoleh 1.856 data valid setelah melewati tahap pra-pemrosesan. Dataset ini kemudian dibagi menjadi 80% data latih dan 20% data uji untuk kebutuhan pemodelan.

3.5. Klasifikasi Sentimen

Klasifikasi sentimen K-NN dilakukan dengan menghitung jarak vektor fitur antara data uji dan data latih, lalu menentukan kelas sentimen berdasarkan mayoritas label tetangga terdekat.

3.6. Evaluasi Model

Evaluasi model dilakukan menggunakan *Confusion Matrix* untuk membandingkan hasil prediksi K-NN dengan label sebenarnya pada data uji. Performa klasifikasi sentimen (positif, netral, dan negatif) diukur menggunakan tiga metrik utama, yaitu *Accuracy*, *Precision*, dan *Recall*.

4. HASIL DAN PEMBAHASAN

Bab ini menyajikan hasil analisis sentimen komentar YouTube mengenai isu boikot Trans 7 menggunakan algoritma *K-Nearest Neighbors* (KNN). Pembahasan mencakup aspek antarmuka sistem dan analisis *machine learning*. Dataset sebanyak 3.700 komentar diproses melalui *text preprocessing*, lalu diklasifikasikan ke dalam kategori negatif, netral, dan positif. Kinerja pemodelan dievaluasi menggunakan *Confusion Matrix* untuk mengukur akurasi dan performa klasifikasi pada setiap kelas.

4.1. Hasil Pengumpulan Data

Pengumpulan data dilakukan melalui teknik *crawling* komentar video YouTube terkait isu boikot Trans 7 menggunakan Python dan YouTube API (`googleapiclient.discovery`). Informasi yang diambil mencakup *author*, *comment*, *likes*, dan *published_at*, yang kemudian digabungkan dan disimpan ke dalam format CSV menggunakan library Pandas. Proses ini berhasil mengumpulkan sebanyak 3.700 data komentar.

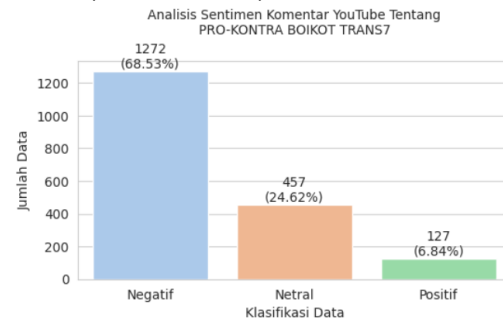
4.2. Hasil Preprocessing

Tahap pra-pemrosesan (*case folding*, tokenisasi, dan *stopword removal*) berhasil menormalisasi dataset dan menyisakan 1.856 data valid dari 3.700 komentar awal. Penyusutan data terjadi akibat penghapusan *noise* murni, seperti emoji, *mention*, dan tautan tanpa teks opini. Data bersih tersebut selanjutnya digunakan untuk klasifikasi sentimen menggunakan algoritma *K-Nearest Neighbor* (K-NN).

4.3. Hasil Klasifikasi Sentimen

Pasca-pemrosesan dan ekstraksi fitur IndoRoBERTa, sebanyak 1.856 komentar diklasifikasikan menggunakan algoritma K-NN. Hasil analisis menunjukkan dominasi kuat sentimen negatif sebesar 68,53% (1.272

komentar) yang mencerminkan kritik dan ketidaksetujuan publik terkait isu boikot TRANS7. Sementara itu, sentimen netral tercatat sebesar 24,62% (457 komentar), dan sentimen positif sangat minim hanya sebesar 6,84% (127 komentar).



Sistem menghasilkan visualisasi Word Cloud untuk menampilkan kata-kata yang paling sering muncul pada setiap kelas sentimen, yaitu positif, netral, dan negatif. Semakin besar ukuran kata, semakin tinggi frekuensi kemunculannya dalam komentar.



Gambar 2 Word Cloud

Visualisasi *Word Cloud* memperlihatkan perbedaan konteks kata pada setiap kelas sentimen. Kelas positif didominasi istilah seperti “trans hebat” dan “sopan santun” yang

mencerminkan apresiasi dan dukungan, sedangkan kelas negatif diwarnai kata “anti kritik” dan “trans tua” yang menunjukkan kritik serta penolakan. Sementara itu, kelas netral didominasi frasa informatif seperti “korupsi dana” dan “kuota haji” tanpa kecenderungan polaritas opini yang jelas

4.4. Hasil Evaluasi

Untuk mengukur kinerja dan efektivitas model *Machine Learning KNN (K-Nearest Neighbors)* dalam mengklasifikasikan sentimen data ke dalam tiga kategori (Negatif, Netral, dan Positif), dilakukan pengujian menggunakan *Confusion Matrix*.

KNN Confusion Matrix

Actual	Negatif	105	22	0
	Netral	31	14	1
	Positif	10	2	1
		Negatif	Netral	Positif
		Predicted		

Dari 186 data uji, model K-NN berhasil memprediksi 120 data secara tepat, mencakup 105 data negatif, 14 netral, dan 1 positif. Sebagian besar kesalahan klasifikasi bergeser ke kelas negatif (31 data netral dan 10 data positif keliru diprediksi sebagai negatif). Hasil ini mengindikasikan bias klasifikasi yang kuat akibat ketidakseimbangan kelas (*class imbalance*), menjadikan model sangat andal mengenali sentimen negatif, tetapi lemah pada netral dan hampir gagal total pada kelas positif. Berdasarkan hasil pengujian menggunakan *Confusion Matrix*, model klasifikasi *K-Nearest Neighbor* (K-NN) menghasilkan tingkat akurasi sebesar 65%.

	precision	recall	f1-score	support
Negatif	0.719	0.827	0.769	127.000
Netral	0.368	0.304	0.333	46.000
Positif	0.500	0.077	0.133	13.000
accuracy	0.645	0.645	0.645	0.645
macro avg	0.529	0.403	0.412	186.000
weighted avg	0.617	0.645	0.617	186.000

Hasil evaluasi klasifikasi menunjukkan bahwa model mencapai akurasi keseluruhan sebesar 64,5% dari total 186 data uji. Performa tertinggi diraih pada kelas Negatif dengan *F1-score* 0,769 (*precision* 0,719; *recall* 0,827), yang membuktikan keandalan model dalam mendeteksi sentimen negatif. Sebaliknya, terjadi penurunan kinerja yang signifikan pada kelas Netral (*F1-score* 0,333) dan kelas Positif (*F1-score* 0,133; *recall* 0,077). Rendahnya performa pada kelas minoritas tersebut tercermin dari nilai *macro avg F1-score* sebesar 0,412, yang mengindikasikan adanya bias klasifikasi akibat ketidakseimbangan data (*class imbalance*).

5. KESIMPULAN

Berdasarkan serangkaian tahapan penelitian, mulai dari pengumpulan data, *preprocessing* hingga evaluasi komparatif perhitungan Analisis Sentimen Komentar Youtube Tentang Boikot Trans 7 Dengan Metode KNN maka dapat ditarik kesimpulan dalam bentuk poin-poin sebagai berikut:

- Dominasi Sentimen Negatif Opini publik terkait isu boikot Trans 7 didominasi secara mutlak oleh sentimen negatif sebesar 68,53% (1.272 komentar), diikuti netral 24,62% (457 komentar), dan positif 6,84% (127 komentar).
- Akurasi Model K-NN Model K-NN (K=5) berbasis ekstraksi fitur IndoRoBERTa berhasil mengklasifikasikan data dengan tingkat akurasi keseluruhan sebesar 64,5% (120 dari 186 data uji berhasil diprediksi tepat).
- Kendala Class Imbalance Ketimpangan jumlah data membuat model mengalami bias: sangat andal mengenali sentimen negatif (Recall 82,7%), tetapi lemah pada netral (Recall 30,4%) dan hampir gagal memprediksi sentimen positif (Recall 7,7%).
- Saran Pengembangan Penelitian selanjutnya memerlukan teknik penanganan ketidakseimbangan data (seperti SMOTE) untuk meningkatkan performa klasifikasi pada kelas minoritas.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada pihak-pihak terkait yang telah memberi dukungan terhadap penelitian ini.

DAFTAR PUSTAKA

- [1] S. Alim dan A. F. Dharma, "YOUTUBE SEBAGAI RUANG PUBLIK ALTERNATIF BAGI ANAK MUDA," *Expo. J. Ilmu Komun.*, vol. 4, no. 1, hal. 1, Apr 2021, doi: 10.33021/exp.v4i1.1437.
- [2] M. D. Taufiqurahman, S. Anraeni, dan H. Darwis, "Analisis Sentimen Komentar Konten Kreator Gaming Menggunakan Metode Naive Bayes dan KNN," *LINIER Lit. Inform. dan Komput.*, vol. 1, no. 4, hal. 317–327, Des 2024, doi: 10.33096/linier.v1i4.2531.
- [3] A. Pinkan Mareta dkk., "Perbandingan Metode Naive Bayes, Decision Tree, Dan Knn Dalam Analisis Sentimen Aplikasi Gojek Di Playstore," *J. Sist. Inf.*, vol. 7, no. 2, hal. 725–734, 2025.
- [4] H. Travaganz Nainggolan dkk., "Klasifikasi Sentimen Menggunakan Algoritma K-Nearest Neighbour (Studi Kasus: Magang Merdeka Belajar)," *J. Inform.*, 2023.
- [5] A. Nafisah dan B. Prasetya, "Klasifikasi Emosi Multikelas Berbasis Teks Bahasa Indonesia Menggunakan IndoRoBERTa," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 9, no. 9, hal. 1–9, 2025, [Daring]. Tersedia pada: <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/15276>
- [6] R. Danyana Umrana, S. Naja Anwar, dan R. Soelistjadi, "ANALISIS SENTIMEN KOMENTAR YOUTUBE TERKAIT KASUS PAGAR LAUT MENGGUNAKAN METODE KNN (K-NEAREST NEIGHBOR)," *J. Inform. Teknol. dan Sains*, vol. 7, no. 3, 2025.
- [7] S. K. Dewi, D. D. Rahmawati, dan A. P. Sari, "Analisis Sentimen Komentar pada Postingan Instagram " StandWithUs " Menggunakan Klasifikasi Naive Baye," *J. Ilmia Inform.*, vol. 2, 2024.
- [8] S. L. M. Sitio dan R. Nadiyahanti, "Analisis Sentimen Kenaikan Harga BBM Pertamina Pada Media Sosial Menggunakan Metode Naive Bayes Classifier," *Build. Informatics, Technol. Sci.*, vol. 4, no. 3, hal. 1224–1231, 2022, doi: 10.47065/bits.v4i3.2311.
- [9] R. Wijaya dan A. Suwandhi, "Sentimen Komentar Universitas Pelita Harapan Pada TikTok Menggunakan Metode K-Nearest Neighbor," *JDMIS J. Data Min. Inf. Syst.*, vol. 2, no. 1, hal. 26–36, Feb 2024, doi: 10.54259/jdmis.v2i1.2418.
- [10] Y. O. Sihombing dan N. V. Situmorang, "Prediksi Sentimen Pada Teks Media Sosial Corporate University," *Pros. Pertem. Ilm. Tah. Nas. Widyaiswara*, hal. 302–316, 2024.
- [11] L. Nursinggah, R. Ruuhwan, dan T. Mufizar, "Analisis Sentimen Pengguna Aplikasi X Terhadap Program Makan Siang Gratis Dengan Metode Naive Bayes Classifier," *J. Inform. dan Tek. Elektro Terap.*, vol. 12, no. 3, 2024, doi: 10.23960/jitet.v12i3.4336.
- [12] E. Hokijuliandy, H. Napitupulu, dan F. Firdaniza, "Analisis Sentimen Menggunakan Metode Klasifikasi Support Vector Machine (SVM) dan Seleksi Fitur Chi-Square," *SisInfo – J. Sist. Inf. dan Inform.*, vol. 5, no. 2, 2024.
- [13] Alfarizi M. R. S., Al-farish, M. Z., Taufiqurrahman, M., Ardiansah, G., dan M. Elgar, "Penggunaan Python Sebagai Bahasa Pemrograman untuk Machine Learning dan Deep Learning," *Karya Ilm. Mhs. Bertauhid (KARIMAH TAUHID)*, vol. 2, no. 1, hal. 1–6, 2023.
- [14] D. N. Larasakti, A. Aziz, dan D. Aditya, "Analisis Sentimen Komentar Video Youtube Dengan Metode K-Nearest Neighbor," *J. Ilm. Wahana Pendidik.*, vol. 9, no. 5, hal. 132–142, 2023, doi: 10.5281/zenodo.7728573.
- [15] A. H. Kaplale, "Rancangan digitalisasi arsip surat masuk dan surat keluar berbasis web dengan Visual Studio Code (studi kasus pada IAIN Ambon)," *J. Teknol. Inf.*, vol. 8, no. 2, hal. 208–217, 2024, [Daring]. Tersedia pada: https://jurnal.una.ac.id/index.php/jurti/article/view/4686?utm_source=chatgpt.com