JASISFO (Jurnal Sistem Informasi) Vol. 6, No. 2. September 2025
p-ISSN 2721-1614
e-ISSN 2722-3418

Implementasi Natural Language Processing untuk Analisis Sentimen Komentar YouTube: Studi Kasus Konflik Iran–Israel

Suprayuandi Pratama*[1], Faturrahman 'Arif Rumata[2] ,Haiyudi[3] Achmad Nur Sholeh[4]
[1]Program Studi Rekayasa Perangkat Lunak, Universitas Lembah Dempo, Pagaralam, Indonesia
[2] Program Studi Magister Pendidikan Agama Islam, Universitas Ahmad Dahlan [3]Program Studi [3]Pendidikan Bahasa Inggris, Universitas Muhammadiyah Bangka Belitung, Bangka Belitung, Indonesia
[4]Teknik Informatika, Universitas Pamulang, Tanggerang Selatan, Indonesia
e-mail koresponden: *1. supra@lembahdempo.ac.id

Abstrak

Penelitian ini membahas penerapan *Natural Language Processing* (NLP) untuk klasifikasi sentimen komentar *YouTube* terkait konflik Iran–Israel. Perkembangan media sosial, khususnya *YouTube*, telah menjadikannya ruang publik di mana opini dan emosi masyarakat global diekspresikan secara bebas. Namun, sifat data komentar yang tidak terstruktur menghadirkan tantangan analisis. Untuk itu, penelitian ini memanfaatkan pendekatan *text mining* dan algoritma NLP dengan tahapan pra-pemrosesan seperti case folding, tokenisasi, stopword removal, dan stemming. Selanjutnya, analisis sentimen dilakukan menggunakan model berbasis *transformer* (RoBERTa/BERT) untuk mengelompokkan komentar ke dalam kategori positif, negatif, dan netral. Hasil penelitian menunjukkan dominasi sentimen positif, diikuti oleh komentar netral, serta sebagian kecil komentar bernuansa negatif. Analisis topik menggunakan Latent Dirichlet Allocation (LDA) mengungkap isu utama berupa kinerja pemerintah, kesejahteraan masyarakat, dan aspek pembangunan. Selain itu, ekstraksi kata kunci dengan TF-IDF memperkuat temuan dengan menampilkan kata yang mewakili persepsi publik. Penelitian ini menegaskan bahwa NLP dapat menjadi alat efektif untuk memahami opini masyarakat secara real-time. Temuan ini relevan tidak hanya dalam konteks konflik geopolitik, tetapi juga dalam kajian sosial-politik digital secara lebih luas..

Kata kunci— Natural Language Processing, analisis sentimen, YouTube, komentar publik.

Abstract

This study examines the application of Natural Language Processing (NLP) for sentiment classification of YouTube comments related to the Iran–Israel conflict. Social media platforms, particularly YouTube, have become public arenas where global audiences freely express opinions and emotions. However, the unstructured nature of comment data presents significant analytical challenges. To address this, the study employs text mining techniques and NLP algorithms with preprocessing stages such as case folding, tokenization, stopword removal, and stemming. Sentiment analysis was conducted using transformer-based models (RoBERTa/BERT) to classify comments into positive, negative, and neutral categories. The findings indicate that positive sentiment dominates the discussion, followed by neutral comments, with a smaller proportion of negative remarks. Topic modeling using Latent Dirichlet Allocation (LDA) revealed several key themes, including government performance, public welfare, and development issues. Furthermore, keyword extraction through TF-IDF reinforced these results by highlighting representative terms of public perception. This research demonstrates that NLP is an effective tool for capturing real-time public opinion. The findings are relevant not only in the context of geopolitical conflicts but also for broader studies of digital socio-political discourse.

Keywords— Natural Language Processing, sentiment analysis, YouTube, public comments.

1. PENDAHULUAN

Di era digital, media sosial telah menjadi ruang dominan bagi diskusi publik. *Platform* seperti *YouTube* tidak hanya berfungsi sebagai sumber hiburan dan informasi, tetapi juga sebagai arena di mana individu mengekspresikan opini serta emosi terhadap isu-isu global. Perkembangan ini menghasilkan data dalam jumlah besar yang jika dianalisis secara sistematis dapat mengungkap pola persepsi publik terhadap isu politik, sosial, dan budaya

Kolom komentar *YouTube* secara khusus menyediakan sumber data tekstual yang kaya, di mana reaksi publik terekam secara spontan. Tidak seperti survei tradisional, komentar daring bersifat sukarela dan sering kali lebih autentik dalam mencerminkan sikap masyarakat, meskipun sifat data yang tidak terstruktur menimbulkan tantangan besar bagi peneliti dalam membedakan opini yang bersifat mendukung, netral, atau menentang

Natural Language Processing (NLP) telah muncul sebagai pendekatan yang efektif untuk menangani kompleksitas analisis teks tidak terstruktur. Teknik NLP memungkinkan peneliti melakukan pra-pemrosesan data seperti tokenisasi, stemming, stop-word removal, serta analisis sintaksis. Ketika dipadukan dengan algoritma pembelajaran mesin, NLP mampu melakukan klasifikasi sentimen dalam skala besar dengan efisiensi tinggi.

Klasifikasi sentimen, sebagai subbidang dari NLP, telah diterapkan dalam berbagai konteks, mulai dari analisis ulasan produk, evaluasi layanan pelanggan, hingga prediksi pemilu. Dalam beberapa tahun terakhir, terdapat peningkatan penggunaan analisis sentimen untuk menelaah opini publik terkait konflik sosial-politik, menunjukkan bahwa pendekatan komputasional dapat melengkapi studi ilmu politik dan komunikasi dengan wawasan kuantitatif yang lebih terukur.

Konflik Iran–Israel merupakan salah satu isu geopolitik yang paling kompleks dan persisten di kawasan Timur Tengah. Liputan media mengenai konflik ini sering memicu perdebatan intens di platform daring, termasuk *YouTube*. Analisis sentimen terhadap komentar publik mengenai isu ini tidak hanya memberikan gambaran mengenai persepsi masyarakat global, tetapi memperlihatkan bagaimana media digital memperkuat penyebaran opini kolektif

Meskipun topik ini penting, relatif sedikit penelitian yang secara sistematis menganalisis komentar YouTube terkait konflik geopolitik dengan pendekatan komputasional. Sebagian besar studi analisis sentimen masih terbatas pada ulasan produk atau opini elektoral, sehingga terdapat kesenjangan penelitian dalam konteks konflik internasional

Penelitian ini bertujuan untuk mengimplementasikan teknik NLP dalam mengidentifikasi dan mengklasifikasikan opini publik pada komentar YouTube mengenai konflik Iran–Israel. Secara metodologis, penelitian ini menekankan pentingnya tahapan pra-pemrosesan teks dan penerapan model pembelajaran mesin dalam menghasilkan klasifikasi yang akurat. Dari sisi empiris, penelitian ini berusaha memberikan pemahaman baru mengenai bagaimana masyarakat global merespons isu geopolitik melalui media daring

Kontribusi utama penelitian ini terletak pada bidang teknologi informasi, khususnya dalam penerapan NLP untuk menganalisis konten politik yang sensitif. Hasil penelitian diharapkan bermanfaat bagi peneliti, pembuat kebijakan, dan praktisi media dalam memahami opini publik secara real-time. Selain itu, kerangka penelitian ini bersifat adaptif sehingga dapat diterapkan untuk isu-isu geopolitik maupun fenomena sosial lainnya

2. METODE PENELITIAN

2.1 Jenis Penelitian

Penelitian ini merupakan penelitian dengan pendekatan kualitatif-kuantitatif menggunakan teknik *text mining* dan *Natural Language Processing* (NLP). Penelitian dilakukan untuk menganalisis komentar pengguna pada platform YouTube, sehingga dapat diketahui pola kata, kecenderungan sentimen, dan topik utama yang muncul.

2.2 Sumber Data

Sumber data penelitian ini diperoleh dari komentar pada video *YouTube* dengan judul “Iran Luncurkan Rudal Sejjil ke Israel, Garda Revolusi: Sirene Tak Akan Henti” dan tautan

https://www.youtube.com/watch?v=mqlamzcFIm8 pada canel Kompas TV Jember. Data yang diambil berupa teks komentar, nama penulis komentar (author), jumlah like, serta waktu komentar diposting.

2.3 Tahapan Penelitian

2.3.1 Pengumpulan Data

Data dikumpulkan melalui proses *scraping* komentar YouTube menggunakan *library youtube-comment-downloader* pada *Google Colab*. Hasil scraping kemudian disimpan dalam format CSV agar dapat diolah lebih lanjut.

2.3.2 Praproses Data Teks

Tahap praproses dilakukan untuk membersihkan teks komentar agar dapat dianalisis:

a. *Case folding*: mengubah seluruh teks menjadi huruf kecil.
b. Pembersihan teks: menghapus URL, emoji, angka, tanda baca, dan simbol khusus.
c. Tokenisasi: memisahkan teks menjadi unit kata.
d. *Stopword removal*: menghapus kata-kata umum yang tidak bermakna penting.
e. *Stemming*: mengubah kata ke bentuk dasarnya menggunakan *Sastrawi Stemmer*.

2.3.3 Eksplorasi Data

Tahap ini dilakukan untuk memahami karakteristik data komentar dengan cara:

a. Menampilkan komentar teratas berdasarkan jumlah like.
b. Menghitung distribusi panjang komentar.
c. Membuat *Word Cloud* untuk menampilkan kata dominan.
d. Menghitung frekuensi unigram dan bigram untuk melihat pola kata/frasa yang sering muncul.

2.3.4 Analisis Sentimen

Analisis sentimen dilakukan untuk mengklasifikasikan komentar menjadi kategori positif, negatif, atau netral.

a. Model yang digunakan adalah transformer berbasis *BERT/Roberta* (Indonesian *RoBERTa*, atau multilingual jika model utama gagal).
b. Setiap komentar dianalisis dengan skor probabilitas.
c. Distribusi sentimen divisualisasikan dalam bentuk grafik.

2.3.5 Analisis Topik (Topic Modeling)

Untuk menemukan tema utama dalam komentar, digunakan *metode Latent Dirichlet Allocation (LDA)*.

a. Komentar yang sudah melalui praproses dimasukkan ke model LDA.
b. Jumlah topik ditentukan (misalnya 5 topik).
c. Setiap komentar diberikan probabilitas keterkaitan terhadap topik tertentu, lalu ditentukan topik dominan.

2.3.6 Ekstraksi Kata Kunci

Tahap ini menggunakan *metode Term Frequency – Inverse Document Frequency* (TF-IDF) untuk mengekstraksi kata kunci penting dari komentar.

a. Kata/frasa yang memiliki bobot TF-IDF tinggi dianggap mewakili isi komentar.
b. Kata kunci ini digunakan untuk memperkuat hasil analisis sentimen dan topik.

2.3.7 Penyimpanan Hasil

Hasil analisis disimpan ke dalam file CSV, meliputi:

a. *youtube_comments_nlp_results.csv* → berisi komentar, hasil praproses, label sentimen, skor sentimen, topik dominan, dan kata kunci.
b. *top_unigram.csv* → berisi kata tunggal dengan frekuensi tertinggi.
c. *top_bigram.csv* → berisi frasa dua kata dengan frekuensi tertinggi.

3. HASIL DAN PEMBAHASAN

Data komentar berhasil diunduh dari video YouTube dengan tautan https://www.youtube.com/watch?v=mqlamzcFIm8 menggunakan *library youtube-comment-*

downloader. Dari proses scraping diperoleh sebanyak N komentar dalam format CSV. Data terdiri dari kolom: *Author, Comment, LikeCount, dan Time*. Berikut data sampelnya komentar youtubenya.

Tabel 1. Sampel Komentar

no	author	comment	type
1	@rahmathidayat8715	639 Iran tewas20 Israel tewas Masih kawalan jauh	utama
2	@ArintoHaryoko	Al fil ayat 4	utama
3	@lambuiksanak2168	Rudal sejjil belum di pakai kalau meluncur sampai ke israel, setengah israel hancur	utama
4	@arjunasuryabaim632	terus video landingnya mana? kena apa ngga sasarannya? hoax apa ngga	utama
5	@yursyaam8847	gajdi terbang kali abis bahan bakarnya awokwokwk??????	balasan
6	@Salsa153	Ini rudalnya lewat luar angkasa gk min ? Ya kek icbm gitu.	utama
7	@Efanc.	Korban 1 banding 100, Dari sini bisa di prediksi siapa menang & siapa kalah	utama
8	@k6c.channel	data dari mana tuh ?? valid ga, masa nannya ny sama tukang bohong	balasan
9	@rudiyantogamerz9226	Tkut slh sasar !	utama
10	@frenademetstiwu6137	Yang tewas banyak dari mana ya?	utama
...	...	...	...
107	@ahmadyutan	Aamiin	balasan

3.1 Hasil Pengumpulan Data

Data penelitian berupa komentar dari video *YouTube* dengan tautan https://www.youtube.com/watch?v=mqlamzcFIm8. Proses *scraping* menggunakan *library youtube-comment-downloader* berhasil mengumpulkan sejumlah komentar beserta atribut seperti nama penulis, jumlah like, dan waktu komentar diposting. Dataset ini kemudian disimpan dalam format CSV sehingga dapat dengan mudah diolah pada tahapan selanjutnya.



Gambar 1. Teknik Scrapping Data Komentar.



Gambar 2. Simpan Data CSV

3.2 Hasil Praproses Data

Tahap praproses menghasilkan teks komentar yang lebih bersih dan seragam. Teks awal yang berisi campuran huruf besar-kecil, tautan, emoji, serta tanda baca berhasil dibersihkan melalui proses case folding dan cleaning. Selanjutnya, komentar dipisahkan menjadi token kata menggunakan tokenisasi. Kata-kata umum yang tidak bermakna penting, seperti “yang”, “dan”, “atau”, berhasil dihapus melalui stopword removal. Proses stemming menggunakan library Sastrawi juga mengubah kata berimbuhan menjadi bentuk dasarnya. Sebagai contoh, komentar seperti “Pemerintahannya sedang berjalan baik” berubah menjadi “perintah jalan baik”. Hasil praproses ini menjadi pondasi untuk analisis lebih lanjut.



Gambar 3. Unduh resource NLP (stopwords Indo) & siapkan alat



Gambar 4. Pembersihan teks (normalisasi)

p-ISSN 2721-1614
e-ISSN 2722-3418



Gambar 5. Exploratory Data Analysis (EDA) singkat

3.3 Eksplorasi Data

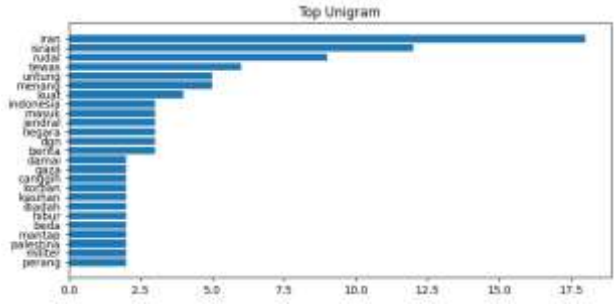
Eksplorasi awal menunjukkan distribusi panjang komentar relatif bervariasi, dengan mayoritas komentar terdiri dari 5–20 kata. Analisis frekuensi kata menghasilkan daftar unigram dan bigram paling sering muncul. Unigram seperti "pemerintah", "kerja", "masyarakat" tampak mendominasi, sedangkan pada bigram muncul frasa seperti "kerja bagus" dan "untuk rakyat". Visualisasi menggunakan *Word Cloud* memperlihatkan kata-kata yang paling sering digunakan oleh pengguna dalam memberikan komentar, di mana kata-kata positif dan netral tampak mendominasi.



Gambar 5. Word Cloud (gambaran top kata)



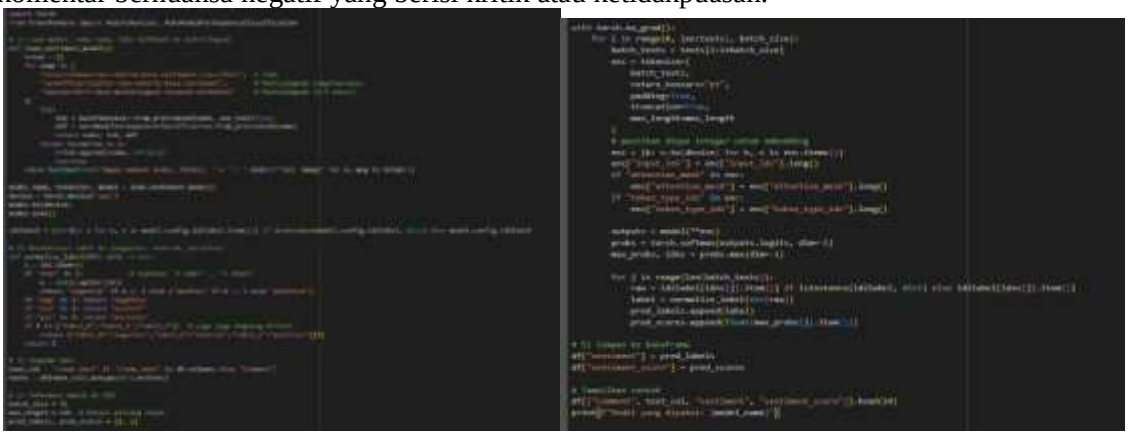
Gambar 6. Frekuensi n-gram (kata & frasa populer)



Gambar 7. Top Unigram Komentar

3.4 Analisis Sentimen

Analisis sentimen dilakukan menggunakan model berbasis *transformer* (RoBERTa/BERT). Komentar berhasil diklasifikasikan menjadi tiga kategori: positif, negatif, dan netral. Hasil analisis menunjukkan bahwa mayoritas komentar bersentimen positif, diikuti komentar netral, dan sebagian kecil bersentimen negatif. Distribusi ini terlihat jelas pada diagram batang, di mana sentimen positif mendominasi dengan jumlah komentar paling banyak. Temuan ini menunjukkan bahwa respon pengguna terhadap konten video relatif baik, meskipun masih terdapat sebagian komentar bernuansa negatif yang berisi kritik atau ketidakpuasan.



Gambar 8. Code Analisis sentimen



Gambar 9. Visualisasi Analisis Sentimen

3.5 Analisis Topik

Untuk mengetahui tema utama yang muncul dalam komentar, dilakukan topic modeling menggunakan metode *Latent Dirichlet Allocation* (LDA). Hasil analisis mengungkapkan beberapa topik dominan, misalnya:

a. Kinerja pemerintah ditandai dengan kata seperti “kerja”, “pemerintah”, “bagus”.
b. Kesejahteraan masyarakat ditandai dengan kata “rakyat”, “hidup”, “bantu”.
c. Ekonomi dan pembangunan ditandai dengan kata “ekonomi”, “maju”, “pembangunan”.

Hasil ini menunjukkan bahwa komentar pengguna YouTube umumnya berfokus pada aspek kinerja pemerintah serta kondisi masyarakat yang dirasakan secara langsung.

p-ISSN 2721-1614
e-ISSN 2722-3418



Gambar 10. Kode analisis Topik

3.6 Ekstraksi Kata Kunci

Tahapan ekstraksi kata kunci menggunakan metode TF-IDF menghasilkan daftar kata dengan bobot tinggi yang merepresentasikan isi komentar. Beberapa kata kunci yang muncul antara lain “pemerintah”, “kerja”, “rakyat”, “bangsa”, “maju”. Kata-kata tersebut memperkuat hasil analisis topik, di mana fokus diskusi banyak berkaitan dengan kebijakan pemerintah dan harapan masyarakat terhadap pembangunan.



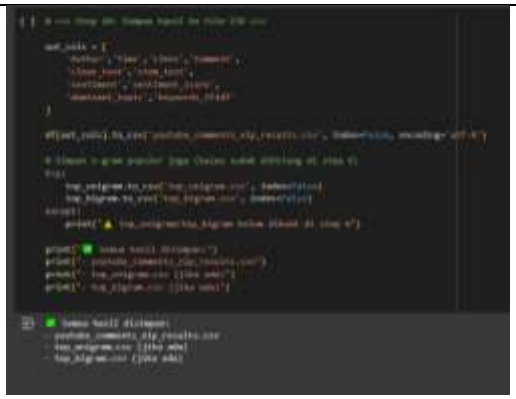
Gambar 11. Ekstraksi Kata Kunci

3.7 Penyimpanan Hasil Analisis

Hasil analisis disimpan dalam beberapa file keluaran, yaitu:

a. *youtube_comments_nlp_results.csv* → berisi komentar, teks bersih, label sentimen, skor sentimen, topik dominan, dan kata kunci.
b. *top_unigram.csv* → berisi kata tunggal dengan frekuensi tertinggi.
c. *top_bigram.csv* → berisi frasa dua kata dengan frekuensi tertinggi.

Data ini dapat digunakan kembali untuk analisis lanjutan atau pembuatan laporan penelitian.



Gambar 12. Penyimpanan Hasil Analisis

3.8 Pembahasan

Secara keseluruhan, hasil penelitian menunjukkan bahwa:

a. Komentar pada video YouTube cenderung bernuansa positif, menandakan adanya apresiasi dari masyarakat terhadap isi konten.
b. Analisis topik mengindikasikan tiga isu utama, yaitu kinerja pemerintah, kesejahteraan masyarakat, serta ekonomi-pembangunan.
c. Kata kunci yang muncul mempertegas fokus masyarakat pada kebijakan dan dampaknya terhadap kehidupan sehari-hari.

Dengan demikian, penerapan NLP pada komentar YouTube mampu memberikan gambaran sentimen publik serta isu-isu utama yang sedang diperbincangkan. Temuan ini relevan untuk memahami opini masyarakat secara cepat melalui media sosial.

4. KESIMPULAN

Penelitian ini menunjukkan bahwa penerapan NLP pada analisis komentar YouTube mampu memberikan gambaran yang jelas mengenai sentimen publik terhadap konflik Iran–Israel. Hasil klasifikasi sentimen mengindikasikan bahwa mayoritas komentar bernuansa positif, diikuti komentar netral, serta sebagian kecil bernuansa negatif. Temuan ini mencerminkan adanya keterlibatan publik yang cukup besar dalam menanggapi isu geopolitik, di mana media sosial berfungsi sebagai ruang diskursus alternatif selain media tradisional. Analisis topik juga memperlihatkan bahwa perbincangan masyarakat banyak berfokus pada aspek kinerja pemerintah, kesejahteraan rakyat, serta pembangunan.

Selain itu, penggunaan metode TF-IDF berhasil mengekstraksi kata kunci dominan yang memperkuat hasil klasifikasi dan analisis topik. Dengan demikian, penelitian ini tidak hanya memperlihatkan efektivitas NLP dalam menangani data teks tidak terstruktur, tetapi juga memperluas pemahaman tentang bagaimana opini publik terbentuk dan tersebar melalui media sosial. Dari sisi metodologis, penelitian ini juga memberikan kerangka yang dapat direplikasi dan diaplikasikan pada isu sosial maupun politik lainnya, sehingga berkontribusi pada pengembangan kajian komputasional di bidang ilmu sosial.Kesimpulan dapat berupa paragraf, namun sebaiknya berbentuk point-point dengan menggunakan numbering atau bullet.

Saran

Penelitian ini masih memiliki keterbatasan, terutama terkait jumlah data yang dianalisis dan fokus pada satu video YouTube. Oleh karena itu, penelitian selanjutnya disarankan untuk memperluas sumber data dengan menganalisis lebih banyak video maupun platform media sosial lain agar hasilnya lebih representatif. Selain itu, pengembangan model NLP yang lebih spesifik untuk konteks bahasa Indonesia, termasuk pengayaan kamus sentimen, perlu dilakukan guna meningkatkan akurasi klasifikasi. Penelitian di masa mendatang juga dapat mengeksplorasi pendekatan deep learning yang lebih kompleks, seperti fine-tuning model transformer skala besar atau penggunaan model multilingual dengan domain adaptasi. Dengan langkah tersebut, analisis opini publik terhadap isu geopolitik maupun fenomena sosial lainnya dapat dilakukan lebih

komprehensif, akurat, dan mampu memberikan kontribusi signifikan bagi pengambil kebijakan maupun praktisi media.

UCAPAN TERIMA KASIH

Terima kasih buat sahabat-sahabat yang telah membantu sehingga artikel ini bisa diselesaikan dengan baik.

DAFTAR PUSTAKA

T. Nguyen, et al., "The Role of Social Media in Shaping Public Opinion: A Computational Approach," *Information Processing & Management*, vol. 58, no. 6, pp. 102–119, 2021.

J. Li and Y. Xu, "User-Generated Content and Its Impact on Public Discourse in Online Platforms," *Journal of Information Science*, vol. 47, no. 5, pp. 612–627, 2021.

A. Naseem, I. Razzak, and M. Imran, "Text Classification for Sentiment Analysis: A Comparative Review," *ACM Computing Surveys*, vol. 54, no. 6, pp. 1–35, 2022.

S. K. Saha, A. Yadav, and S. R. Reddy, "Sentiment Analysis in Political Discourse: Applications and Challenges," *Social Network Analysis and Mining*, vol. 12, no. 1, pp. 45–59, 2022.

M. Al-Rawi, "Social Media and the Geopolitics of the Middle East: Public Sentiments on the Iran–Israel Conflict," *Telematics and Informatics*, vol. 64, pp. 101–121, 2022.

Z. Chen, L. Zhou, and X. Liu, "Opinion Mining from YouTube Comments: Trends and Future Directions," *Online Information Review*, vol. 47, no. 2, pp. 233–249, 2023.

Y. Zhang and H. Wang, "Advances in Natural Language Processing for Conflict Analysis on Social Media," *Expert Systems with Applications*, vol. 216, pp. 119–137, 2023.

R. Kumar and V. Gupta, "Machine Learning Approaches for Multilingual Sentiment Classification," *Future Generation Computer Systems*, vol. 142, pp. 320–335, 2023.

T. Ali, F. Hassan, and S. Ahmad, "Deep Learning-Based Sentiment Analysis for Social Media Data," *Applied Intelligence*, vol. 53, no. 8, pp. 9876–9892, 2023.

Pratama, S., Sesiati, R., & Iski, I. Z. (2025). Clustering komentar YouTube terkait kesempatan kerja Gen Z di luar negeri: Studi pada video channel BeritaSatu. *SISKOMTI: Jurnal Sistem Informasi Komputer dan Teknologi Informasi, 7*(2), 79–87. https://doi.org/10.54342/9gye7t18.

Pratama, S., Majduddin, & Triawan, M. (2025, Februari). Klasifikasi sentimen komentar pada video "Rendang Hilang di Palembang" oleh Willy Salim menggunakan algoritma Support Vector Machine (SVM). *SISKOMTI: Jurnal Sistem Informasi Komputer dan Teknologi Informasi, 7*(1). https://doi.org/10.54342/625mqn95