

ANALISIS SENTIMEN KOMENTAR YOUTUBE BERITA TV ONE TENTANG KASUS SISWA MEROKOK DI SMAN 1 CIMARGA MENGUNAKAN INDOBERT

Muhammad Nabil Arfiansyah^{1*)}, Aryaputra Jagaddatri²⁾, Dwi Arman Prasetya³⁾

^{1,2,3}Sains Data, Universitas Pembangunan Nasional Veteran Jawa Timur

*Email Korespondensi : 23083010074@student.upnjatim.ac.id

Abstrak

Penelitian ini bertujuan menganalisis sentimen masyarakat terhadap pemberitaan kasus siswa merokok di SMAN 1 Cimarga melalui komentar YouTube kanal TV One. Metode yang digunakan adalah pendekatan deskriptif kuantitatif dengan model *Indonesian RoBERTa Sentiment Classifier (IndoRoBERT)*. Data sebanyak 500 komentar dikumpulkan menggunakan *YouTube Data API v3* pada 4 Desember 2025, kemudian melalui tahap pra-pemrosesan dan klasifikasi sentimen. Hasil penelitian menunjukkan sentimen negatif mendominasi opini publik dengan kata kunci "mogok", "siswa", "merokok", dan "salah". Sentimen positif muncul terbatas dengan kata "dukung" dan "bagus", sedangkan sentimen netral didominasi kata informatif seperti "pendidikan" dan "gubernur". Temuan ini mengindikasikan masyarakat cenderung mengkritik perilaku siswa dan penanganan kasus oleh pihak sekolah. Penelitian membuktikan efektivitas model transformer dalam menganalisis sentimen berbahasa Indonesia dan dapat menjadi masukan bagi pemangku kebijakan pendidikan.

Kata kunci: Analisis Sentimen, IndoRoBERT, YouTube, siswa merokok, Natural Language Processing

Abstract

This study aims to analyze public sentiment towards the news coverage of the student smoking case at SMAN 1 Cimarga through YouTube comments on the TV One channel. The method used is a quantitative descriptive approach with the Indonesian RoBERTa Sentiment Classifier (IndoRoBERT) model. Data of 500 comments were collected using the YouTube Data API v3 on December 4, 2025, then went through a pre-processing and sentiment classification stage. The results showed that negative sentiment dominated public opinion with keywords "strike", "students", "smoking", and "wrong". Positive sentiment appeared limited to the words "support" and "good", while neutral sentiment was dominated by informative words such as "education" and "governor". These findings indicate that the public tends to criticize the student's behavior and the school's handling of the case. The study proves the effectiveness of the transformer model in analyzing Indonesian-language sentiment and can provide input for education policymakers.

Keywords: Sentiment Analysis, IndoRoBERT, YouTube, student smoking, Natural Language Processing

PENDAHULUAN

Kebiasaan merokok di kalangan remaja menjadi salah satu masalah kesehatan masyarakat yang memerlukan penanganan segera. Hal ini dipicu oleh faktor lingkungan sosial, kebiasaan bergaul, serta kurangnya pengawasan dari orang tua dan pihak sekolah (Wirawati, 2021). Persoalan ini bukan hanya berpengaruh terhadap kesehatan jasmani

remaja, tetapi juga berkaitan dengan masalah kedisiplinan, pembentukan karakter, dan iklim sekolah. Ketika kasus siswa merokok terjadi di SMAN 1 Cimarga, masyarakat memberikan tanggapan yang luas melalui kolom komentar pada laman *YouTube*, khususnya pada kanal berita nasional seperti *TV One*. Komentar-komentar yang muncul menggambarkan pandangan, persepsi, dan perasaan masyarakat terhadap sikap sekolah, tingkah laku siswa, dan dimensi moralitas dalam pendidikan.

YouTube merupakan salah satu platform media sosial yang banyak dimanfaatkan sebagai wadah diskusi publik. Oleh karena itu, komentar yang diunggah pengguna dapat dijadikan sumber data yang penting untuk memahami pandangan masyarakat terhadap sebuah permasalahan. Berbagai penelitian terdahulu memperlihatkan bahwa pendapat publik dalam komentar *YouTube* dapat dikaji secara sistematis menggunakan metode Natural Language Processing (NLP). Sebagai contoh, (Wicaksono & Nastiti, 2024) menggunakan model LSTM dan Bi-LSTM untuk mengkaji pendapat publik pada acara Indonesia Lawyers Club dan memperoleh tingkat akurasi tinggi mencapai 99,61%. Namun demikian, metode berbasis RNN tersebut memiliki kelemahan dalam memahami konteks yang panjang dan ragam bahasa informal yang umum digunakan di media sosial.

Seiring dengan kemajuan teknologi NLP, model berbasis transformer seperti BERT dan RoBERTa memperlihatkan performa yang lebih baik dalam memahami bahasa natural. (Setiawan et al., 2025) membuktikan bahwa IndoBERT mampu meraih akurasi hingga 95% dalam tugas klasifikasi sentimen Bahasa Indonesia. (Riyadi et al., 2024a) juga memperlihatkan bahwa model IndoBERT lebih efektif dalam mengklasifikasikan komentar pengguna *YouTube* pada ranah hiburan dengan akurasi mencapai 98%. Lebih lanjut, memperlihatkan bahwa model IndoBERT dapat menangkap kecenderungan pendapat masyarakat terhadap kebijakan PPN 12%, yang mana sentimen negatif mendominasi perbincangan publik. Keseluruhan penelitian tersebut menegaskan bahwa model transformer mampu memahami linguistik secara mendalam dan menghasilkan analisis yang lebih akurat dibandingkan dengan metode sebelumnya.

Meskipun telah banyak penelitian tentang analisis sentimen menggunakan model transformer, kajian mengenai tanggapan publik terhadap persoalan pendidikan, terutama perilaku siswa seperti kasus merokok di sekolah, masih sangat terbatas. Padahal, persoalan ini sangat penting untuk dikaji karena berkaitan dengan pandangan masyarakat terhadap kedisiplinan sekolah dan upaya pembentukan karakter peserta didik. Dengan demikian, diperlukan analisis sentimen masyarakat terhadap kasus tersebut menggunakan metode yang lebih modern.

METODE PENELITIAN

A. Jenis Penelitian

Penelitian ini menggunakan pendekatan deskriptif kuantitatif untuk memetakan dinamika sentimen publik terhadap pemberitaan kasus siswa merokok di SMAN 1 Cimarga yang beredar di platform *YouTube*. Pendekatan ini dipilih karena urgensinya dalam mengolah data teks tidak terstruktur (*unstructured text*) yang bervolume besar secara sistematis, sehingga pola opini masyarakat dapat digambarkan secara objektif. Sebagai instrumen klasifikasi utama, penelitian ini mengimplementasikan arsitektur model transformer, khususnya IndoRoBERTa. Pemilihan algoritma ini merujuk pada temuan empiris terdahulu yang membuktikan bahwa model berbasis transformer memiliki superioritas dalam memahami semantik Bahasa Indonesia dibandingkan metode konvensional (Riyadi et al., 2024b).

B. Sumber Data

Data yang digunakan dalam penelitian ini merupakan data tekstual tidak terstruktur (*unstructured text*) yang bersumber dari respons publik di media sosial YouTube, khususnya pada video berita kanal TV One mengenai kasus siswa merokok di SMAN 1 Cimarga. Pemilihan YouTube sebagai sumber data didasari oleh fakta bahwa platform ini diidentifikasi sebagai sumber penting untuk mendapatkan umpan balik masyarakat serta menawarkan data yang kaya untuk memberikan wawasan mengenai aspek-aspek yang dianggap efektif atau perlu diperbaiki dari suatu peristiwa (Firdaus & Trisnawarman, 2025). Data dikumpulkan secara *real time* pada tanggal 4 Desember 2025 pukul 23:00. Guna menjamin validitas dan aksesibilitas data, pengumpulan dilakukan dengan memanfaatkan pustaka YouTube Data API v3 yang memungkinkan penarikan data komentar secara terprogram dan autentik.

Commented [1]: sitasi
<https://garuda.kemdiktisaintek.go.id/documents/detail/4737352>

C. Flowchart Analisis



Gambar 1. Proses Analisis

1. Data Collecting

Tahap pertama dari flowchart analisis adalah pengumpulan data (*data collecting*). Pada tahap ini, komentar YouTube diambil secara otomatis menggunakan YouTube Data API v3. Metode pengumpulan data menggunakan teknik *scraping* berbasis API (*Application Programming Interface*) untuk menghimpun ulasan pengguna secara terprogram ini merujuk pada pendekatan yang diterapkan oleh Hidayat dan Nastiti (2024). Fungsi

comment Threads digunakan untuk mengekstraksi komentar utama dan balasannya dari video terkait pemberitaan kasus siswa merokok di SMAN 1 Cimarga yang diunggah oleh kanal TV One. Proses pengumpulan dilakukan secara *real time*, sehingga data yang diperoleh mencerminkan opini publik aktual pada waktu penelitian berlangsung. Data yang berhasil dihimpun kemudian disimpan dalam bentuk Data Frame untuk mempermudah proses analisis dan pra-pemrosesan. Tahap ini memastikan bahwa penelitian menggunakan data primer berbasis digital yang autentik dan representatif terhadap percakapan publik di platform YouTube.

2. Data Preprocessing

Tahap pra-pemrosesan teks dilakukan untuk menyiapkan komentar YouTube agar berada dalam format yang sesuai untuk dianalisis oleh model transformer. Proses ini penting karena komentar pada media sosial umumnya bersifat informal, mengandung singkatan, emotikon, serta variasi penulisan yang dapat mengganggu kinerja model. Tahapan pra-pemrosesan dalam penelitian ini mencakup beberapa langkah utama. Pertama, dilakukan *case folding* dengan mengubah seluruh karakter menjadi huruf kecil untuk menyeragamkan teks. Kedua, dilakukan penghapusan elemen-elemen non informasi seperti URL, emoji, tanda baca yang tidak relevan, simbol-simbol khusus, serta mention pengguna. Selain itu, tanda pagar (#) juga dihapus agar kata yang menyusul dapat diproses sebagai token biasa.

Selanjutnya, dilakukan normalisasi kata tidak baku menggunakan kamus slang yang disusun secara manual. Normalisasi ini mencakup konversi kata-kata seperti “gk”, “ga”, “tp”, “bgt”, dan “klo” menjadi bentuk baku “tidak”, “tapi”, “banget”, dan “kalau”. Normalisasi penting karena banyak komentar YouTube menggunakan bahasa informal yang tidak tersedia dalam kosakata model. Pada tahap ini tidak dilakukan stemming maupun *stopword removal*, karena model transformer seperti IndoRoBERT bergantung pada kalimat yang utuh. Menghapus *stopword* atau memotong kata dapat mengurangi kualitas representasi dan menurunkan akurasi klasifikasi sentimen.

Namun demikian, untuk keperluan visualisasi, khususnya pembuatan *WordCloud* dan analisis frekuensi kata, *stopword* dihapus secara terpisah setelah proses klasifikasi selesai. *Stopword removal* pada tahap visualisasi dilakukan untuk menghilangkan kata-kata umum seperti “yang”, “ini”, “itu”, “kalau”, dan “saja”, sehingga kata-kata bernuansa sentimen dapat muncul lebih dominan dalam *WordCloud*. Dengan demikian, penelitian ini menerapkan dua jalur pra-pemrosesan: pra-pemrosesan utama untuk model sentimen yang mempertahankan struktur kalimat secara penuh, dan pra-pemrosesan tambahan untuk visualisasi yang menghilangkan *stopword* demi menghasilkan pola kata yang lebih informatif.

3. Klasifikasi Sentimen

Tahap ketiga adalah pemodelan sentimen menggunakan model transformer Bahasa Indonesia, yaitu Indonesian RoBERTa *Sentiment Classifier* (IndoRoBERT). Model ini dipanggil menggunakan pipeline dari *Hugging Face*, yang menggabungkan tokenisasi dan klasifikasi dalam satu alur pemrosesan. Pada tahap ini, setiap komentar yang telah dipra-proses dianalisis oleh model untuk menentukan kategori sentimennya, yaitu positif, negatif, atau netral. Model transformer dipilih karena kemampuannya memahami kalimat secara lebih mendalam dibandingkan metode tradisional seperti Naïve Bayes maupun metode RNN seperti LSTM. Pemodelan dilakukan secara otomatis

4. Analisis Data

HASIL DAN PEMBAHASAN

Penelitian ini berhasil mengumpulkan sebanyak 500 komentar dari video berita TV One yang membahas kasus siswa merokok di SMAN 1 Cimarga. Komentar yang terkumpul terdiri dari komentar utama dan komentar balasan, sehingga mencerminkan keragaman opini publik di media sosial. Data kemudian diproses melalui tahap pra-pemrosesan sehingga menghasilkan teks bersih yang siap dianalisis menggunakan model transformer IndoBERT.

Model Indonesia ROBERTa *Sentiment Classifier* digunakan untuk mengklasifikasikan komentar ke dalam tiga kategori: positif, negatif, dan netral. Hasil klasifikasi menunjukkan bahwa sentimen negatif merupakan kategori paling dominan, disusul sentimen netral, dan kategori positif menjadi yang paling sedikit. Temuan ini menunjukkan bahwa masyarakat cenderung memberikan kritik terhadap kasus yang diberitakan, baik terkait perilaku siswa maupun respons pihak sekolah.

C. Hasil *Wordcloud* dan Frekuensi Data

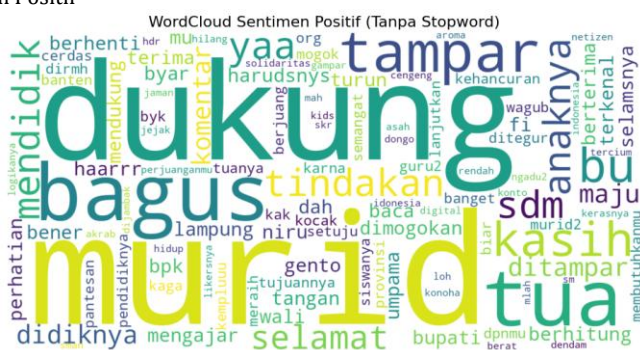
[illegible]

331

WordCloud tanpa stopword menunjukkan bahwa komentar negatif didominasi oleh kata-kata seperti “mogok”, “siswa”, “merokok”, “salah”, “kerja”, “blacklist”, “aturan”, dan “perusahaan”. Kemunculan kata “mogok” dalam frekuensi tertinggi mengindikasikan bahwa sebagian besar komentar menyinggung aksi mogok sekolah yang terjadi setelah kasus viral tersebut. Sementara kata “merokok”, “salah”, dan “aturan” menunjukkan fokus publik pada perilaku siswa dan pelanggaran tata tertib sekolah.

Tabel frekuensi kata memperkuat WordCloud, di mana kata “mogok” muncul sebanyak 62 kali, “siswa” sebanyak 58 kali, dan “merokok” sebanyak 39 kali. Pola ini menegaskan bahwa sentimen negatif banyak berisi kritik terhadap sikap siswa, ketidaksiplinan, dan dampak perilaku tersebut terhadap dunia pendidikan.

b. Sentimen Positif



Gambar 3 WordCloud Sentimen Positif

WordCloud positif memperlihatkan kata-kata seperti “dukung”, “bagus”, “tindakan”, “mendidik”, “kasih”, dan “tampar”. Kata “dukung” merupakan kata dominan, muncul sebanyak 3–4 kali, menunjukkan bahwa sebagian masyarakat mendukung langkah guru atau pihak sekolah dalam mendisiplinkan siswa. Kata “bagus”, “tindakan”, dan “mendidik” juga menunjukkan apresiasi terhadap upaya pendidikan karakter.

Komentar positif umumnya berasal dari pengguna yang memandang bahwa tindakan disipliner yang tegas diperlukan untuk membentuk perilaku siswa di lingkungan sekolah.

c. Sentimen Netral



Gambar 4 WordCloud Sentimen Netral

WordCloud netral memperlihatkan kata-kata seperti “pendidikan”, “gubernur”, “siswa”, “tolong”, “rumah”, “masuk”, “indonesia”, dan “uu”. Pola ini menunjukkan bahwa komentar netral bersifat informatif, mengandung opini ringan, atau sekadar menyampaikan saran tanpa ekspresi emosional tertentu. Tabel frekuensi kata menunjukkan bahwa kata “siswa” muncul 8 kali, sedangkan “merokok”, “biar”, dan “gubernur” masing-masing muncul 3–4 kali. Hal ini menandakan adanya komentar yang lebih bersifat naratif dan tidak memihak.

KESIMPULAN

Penelitian ini menyimpulkan bahwa opini publik terhadap pemberitaan kasus siswa merokok di SMAN 1 Cimarga yang disampaikan melalui platform YouTube didominasi oleh sentimen negatif. Berdasarkan hasil klasifikasi menggunakan model Indonesian RoBERTa Sentiment Classifier, mayoritas komentar menunjukkan kritik terhadap perilaku siswa maupun mekanisme penanganan kasus tersebut oleh pihak sekolah. Hal ini terlihat dari dominasi kata-kata seperti mogok, siswa, merokok, salah, dan perusahaan yang muncul dalam *WordCloud* dan tabel frekuensi kata pada kategori sentimen negatif. Kata-kata tersebut menegaskan bahwa publik banyak menyoroti aspek kedisiplinan, tindakan protes siswa, dan tanggung jawab lembaga pendidikan. Sentimen positif muncul dalam jumlah yang lebih sedikit dan berfokus pada dukungan terhadap tindakan pendisiplinan, tercermin dari kata-kata seperti dukung, bagus, tindakan, dan mendidik. Sementara itu, sentimen netral didominasi oleh kata-kata informatif seperti pendidikan, gubernur, siswa, dan indonesia, yang menunjukkan bahwa sebagian pengguna hanya memberikan tanggapan tanpa ekspresi emosional tertentu. Temuan ini mengindikasikan bahwa peristiwa Cimarga memicu reaksi publik yang kuat dan cenderung bernada negatif, terutama terkait perilaku siswa dan respons institusi sekolah. Secara keseluruhan, penelitian ini membuktikan bahwa model transformer seperti IndoRoBERTa dapat bekerja secara efektif dalam mengklasifikasi sentimen komentar media sosial berbahasa Indonesia dan mampu menggambarkan persepsi publik secara komprehensif, terutama ketika didukung oleh analisis visual seperti *WordCloud* dan tabel frekuensi kata. Hasil penelitian ini diharapkan dapat menjadi masukan bagi pihak sekolah, pemerintah daerah, dan pemangku kebijakan pendidikan dalam memahami dinamika opini masyarakat serta pentingnya pendekatan komunikasi yang lebih baik dalam menangani isu-isu kedisiplinan dan moralitas peserta didik.

UCAPAN TERIMA KASIH

Puji syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa atas segala rahmat dan karunia-Nya sehingga jurnal ini dapat diselesaikan dengan baik. Penyelesaian jurnal ini tidak lepas dari bantuan, bimbingan, dan dukungan berbagai pihak. Oleh karena itu, penulis ingin menyampaikan ucapan terima kasih yang sebesar-besarnya kepada:

1. Bapak Dr.Eng.Ir. Dwi Arman Prasetya, ST., MT., IPU. selaku dosen pembina yang telah memberikan bimbingan, arahan, serta ilmu yang sangat berharga selama proses penyusunan jurnal ini. Dedikasi dan kesabaran beliau menjadi motivasi besar bagi penulis untuk menyelesaikan karya ini dengan sebaik-baiknya.
2. Kedua orang tua tercinta yang senantiasa memberikan doa, dukungan moril dan materiil, serta kasih sayang yang tiada henti. Semangat dan pengorbanan mereka menjadi kekuatan terbesar bagi penulis dalam menghadapi setiap tantangan.
3. Pasangan masa depan penulis yang selalu menemani, memberikan semangat, dan menjadi penyemangat di setiap langkah perjalanan penulis. Kehadirannya memberikan warna dan motivasi tersendiri dalam penyelesaian jurnal ini.

4. Rekan-rekan seperjuangan yang telah berbagi pengalaman, pengetahuan, dan dukungan selama proses penelitian dan penulisan jurnal ini. Kebersamaan dan kerja sama yang terjalin menjadikan perjalanan ini lebih bermakna.

Penulis menyadari bahwa jurnal ini masih jauh dari sempurna. Oleh karena itu, kritik dan saran yang membangun sangat penulis harapkan demi perbaikan di masa mendatang. Semoga jurnal ini dapat memberikan manfaat bagi pembaca dan perkembangan ilmu pengetahuan.

REFERENSI

- Firdaus, M. P., & Trisnawarman, D. (2025). Analisis Sentimen Publik terhadap Program Tabungan Perumahan Rakyat Menggunakan Model IndoBERT Lite pada Komentar YouTube. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 5(1), 359–368. <https://doi.org/10.57152/malcom.v5i1.1744>
- Manoppo, M. R., Kolang, I. C., Nur Fiat, D. N., Mawara, R. M. C., Sumarno, A. D. P., Yusupa, A., & Tarigan, V. (2025). ANALISIS SENTIMEN PUBLIK DI MEDIA SOSIAL TERHADAP KENAIKAN PPN 12% DI INDONESIA MENGGUNAKAN INDOBERT. *Jurnal Kecerdasan Buatan Dan Teknologi Informasi*, 4(2), 152–163. <https://doi.org/10.69916/jkbt.v4i2.322>
- Riyadi, S., Salsabila, L. K., Damarjati, C., & Karim, R. A. (2024a). INTENSIF: Jurnal Ilmiah Penelitian dan Penerapan Teknologi Sistem Informasi 233 Sentiment Analysis of YouTube Users on Blackpink Kpop Group Using IndoBERT 1 *. *INTENSIF*, 8(2), 2549–6824. <https://doi.org/10.29407/intensif.v8n2.22678>
- Riyadi, S., Salsabila, L. K., Damarjati, C., & Karim, R. A. (2024b). INTENSIF: Jurnal Ilmiah Penelitian dan Penerapan Teknologi Sistem Informasi 233 Sentiment Analysis of YouTube Users on Blackpink Kpop Group Using IndoBERT 1 *. *INTENSIF*, 8(2), 2549–6824. <https://doi.org/10.29407/intensif.v8n2.22678>
- Setiawan, D., Utari Iswavigra, D., & Anggiratih, E. (2025). Implementation of IndoBERT for Sentiment Analysis of the Constitutional Court's Decision Regarding the Minimum Age of Vice Presidential Candidates. *Scientific Journal of Informatics*, 12(3). <https://doi.org/10.15294/sji.v12i3.26360>
- Wicaksono, B., & Nastiti, V. R. S. (2024). Analisis Sentimen dalam Opini Publik di Chanel Youtube Indonesia Lawyers Club Tentang Isu Populer dengan Menggunakan Metode LSTM dan Bi-LSTM. *Jurnal Algoritma*, 21(2), 241–251. <https://doi.org/10.33364/algoritma/v.21-2.1696>
- Wirawati, D. (2021). Faktor-faktor yang mempengaruhi perilaku merokok di kalangan remaja. 15(3), 518–524.