

Analisis Sentimen Komentar YouTube pada Video Terkait Insiden Pengemudi Ojek Online dan Anggota Brimob Menggunakan Algoritma Naive Bayes

Zaehol Fatah¹, Lailatus Syarifah²
^{1,2}. Universitas Ibrahimy, Situbondo
Email : lailasyarifahjm@gmail.com

Abstrak

Media sosial telah menjadi ruang ekspresi publik yang dinamis, di mana masyarakat menyampaikan opini terhadap berbagai peristiwa aktual. Penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap insiden antara pengemudi ojek online dan anggota Brimob melalui komentar YouTube. Metode yang digunakan adalah algoritma *Naive Bayes* dengan pendekatan klasifikasi teks. Data dikumpulkan dari 10 video YouTube, menghasilkan 1.143 komentar, yang setelah dibersihkan menjadi 1.121 komentar. Fitur teks dibentuk menggunakan *TF-IDF Vectorizer* dan data dibagi menjadi data latih (896) dan data uji (225). Hasil klasifikasi menunjukkan distribusi sentimen: netral (48.7%), negatif (46.7%), dan positif (4.5%). Evaluasi model menghasilkan akurasi 67%, precision 46%, recall 47%, dan F1-score 45%. Temuan ini menunjukkan bahwa mayoritas komentar bersifat netral dan negatif, serta bahwa *Naive Bayes* cukup efektif dalam klasifikasi opini publik meskipun memiliki keterbatasan dalam menangani data minoritas.

Kata Kunci: *Analisis Sentimen, Komentar YouTube, Naive Bayes, Text Mining, Opini Publik.*

Abstract

Social media has become a dynamic space for public expression, where individuals share opinions on various current events. This study aims to analyze public sentiment regarding the incident between an online motorcycle taxi driver and a member of Brimob through YouTube comments. The method employed is the Naive Bayes algorithm with a text classification approach. Data was collected from 10 YouTube videos, yielding 1,143 comments, which were cleaned down to 1,121 comments for analysis. Text features were constructed using the TF-IDF Vectorizer, and the dataset was split into training data (896 comments) and test data (225 comments). The classification results show sentiment distribution as follows: neutral (48.7%), negative (46.7%), and positive (4.5%). Model evaluation produced an accuracy of 67%, precision of 46%, recall of 47%, and F1-score of 45%. These findings indicate that most comments are neutral and negative, and that Naive Bayes is reasonably effective in classifying public opinion, although it faces challenges in handling minority classes such as positive sentiment.

Keywords: Sentiment Analysis, YouTube Comments, Naive Bayes, Text Mining, Public Opinion.

PENDAHULUAN

Dalam era digital yang semakin terhubung, media sosial telah menjadi ruang ekspresi publik yang sangat aktif. Platform seperti YouTube tidak hanya digunakan untuk berbagi konten hiburan dan edukasi, tetapi juga menjadi wadah

bagi masyarakat untuk menyampaikan opini terhadap berbagai peristiwa aktual (Agarwal, 2023). Kolom komentar pada video YouTube berfungsi sebagai forum terbuka di mana pengguna dapat menanggapi isu sosial, politik, dan budaya secara langsung. Fenomena ini membuka

peluang besar bagi dunia akademik untuk mengkaji dinamika opini publik secara digital. (Linda, 2025)

Salah satu peristiwa yang memicu respons luas dari masyarakat adalah insiden antara pengemudi ojek online dan anggota Brimob. Video yang menampilkan konflik tersebut menyebar dengan cepat dan mengundang ribuan komentar dari pengguna YouTube. Komentar-komentar ini mencerminkan beragam sentimen, mulai dari dukungan terhadap pihak tertentu hingga kritik terhadap institusi keamanan. Namun, volume komentar yang sangat besar menyulitkan proses analisis manual, sehingga dibutuhkan pendekatan berbasis teknologi untuk mengelola dan memahami data tersebut secara efisien.

Melihat kompleksitas opini publik yang muncul di media sosial, diperlukan metode yang mampu mengolah data dalam skala besar secara efisien. Salah satu pendekatan yang terbukti efektif adalah analisis sentimen berbasis machine learning (Hidayat, 2024). Metode ini memungkinkan peneliti untuk mengidentifikasi dan mengklasifikasikan opini publik ke dalam kategori positif, netral, atau negatif secara otomatis. Dalam penelitian ini, pendekatan yang digunakan adalah algoritma Naive Bayes, yang dikenal karena kesederhanaan, efisiensi komputasi, serta efektivitasnya dalam klasifikasi teks. Berbeda dengan metode lexicon-based yang mengandalkan kamus kata, algoritma ini belajar dari data berlabel untuk mengenali pola sentimen secara statistik (Alfin Gerliandeva, 2024).

Kebutuhan akan analisis sentimen berbasis algoritma Naive Bayes telah dibuktikan dalam berbagai studi sebelumnya. Penelitian (Abd Ghofur, 2024) dalam konteks komentar YouTube terhadap tayangan bertema Ponpes Al-Zaytun menunjukkan bahwa algoritma Naive Bayes mampu mengklasifikasikan opini publik dengan akurasi yang cukup tinggi. Studi tersebut menyoroti efektivitas pendekatan machine learning dalam menangani komentar yang bersifat sensitif

dan bernuansa sosial, serta menunjukkan bahwa Naive Bayes dapat bekerja dengan baik meskipun data yang digunakan bersifat informal dan beragam.

Penelitian lain oleh (Sella Adelia, 2023) yang menganalisis komentar YouTube pada tayangan Konferensi Tingkat Tinggi G20 juga membuktikan bahwa algoritma Naive Bayes efektif dalam menangani data komentar yang kompleks dan dinamis. Hasil penelitian tersebut memperkuat relevansi penggunaan Naive Bayes dalam analisis sentimen terhadap isu-isu sosial yang viral di media digital. Kedua studi ini menjadi landasan penting bagi penelitian yang berfokus pada klasifikasi opini publik menggunakan pendekatan machine learning berbasis teks.

Analisis sentimen merupakan pendekatan yang digunakan untuk mengidentifikasi dan mengklasifikasikan opini publik secara otomatis. Algoritma *Naive Bayes* dipilih karena kesederhanaannya, efisiensi komputasi, dan efektivitasnya dalam klasifikasi teks. Penelitian ini bertujuan untuk mengukur distribusi sentimen publik terhadap insiden tersebut dan mengevaluasi kinerja algoritma dalam mengklasifikasikan komentar YouTube. (Yuyun Khanafiyah, 2025)

Dalam penelitian ini, data dikumpulkan dari 10 video YouTube yang membahas insiden ojek online dan Brimob, menghasilkan 1.143 komentar yang setelah dibersihkan menjadi 1.121. Fitur teks dibentuk menggunakan TF-IDF Vectorizer, dan data dibagi menjadi data latih dan data uji. Hasil klasifikasi menunjukkan bahwa mayoritas komentar bersifat netral dan negatif, dengan akurasi model mencapai 67%. Temuan ini menunjukkan bahwa algoritma Naive Bayes cukup efektif dalam mengelompokkan opini publik, meskipun masih menghadapi tantangan dalam menangani data minoritas seperti komentar positif (Triana Putri, 2025).

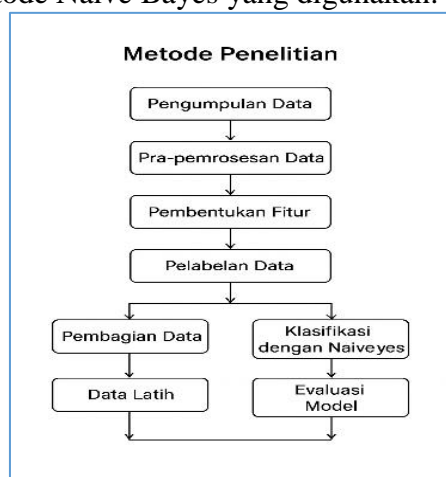
Melalui penelitian ini, diharapkan dapat diperoleh pemahaman yang lebih mendalam mengenai persepsi masyarakat

terhadap insiden sosial yang melibatkan aparat dan warga sipil. Selain itu, hasil analisis dapat menjadi referensi bagi pengambil kebijakan, media, dan akademisi dalam merespons isu-isu publik secara lebih bijak dan berbasis data. Pendekatan ini juga dapat menjadi model penerapan analisis sentimen untuk berbagai kasus sosial lainnya di Indonesia, dengan memanfaatkan kekuatan machine learning dalam mengolah opini publik secara real-time dan akurat.

METODE PENELITIAN

Penelitian ini menerapkan pendekatan *text mining* dengan menggunakan algoritma **Naive Bayes Classifier** untuk melakukan analisis sentimen terhadap komentar-komentar publik pada platform YouTube (Dewi Ayu, 2023). Fokus utama dari metode ini adalah untuk mengidentifikasi polaritas opini masyarakat apakah bersifat *positif*, *negatif*, atau *netral* terhadap insiden yang melibatkan pengemudi ojek online dan anggota Brimob. Dengan memanfaatkan teknik klasifikasi berbasis pembelajaran mesin, penelitian ini bertujuan untuk mengungkap persepsi publik secara sistematis dan terukur.

Berikut adalah tahapan desain metode Naive Bayes yang digunakan:



Gambar 1. Tahapan Penelitian dalam metode Naive Bayes

1. Pengumpulan data
Data dikumpulkan dari sepuluh video

YouTube yang secara khusus membahas insiden antara pengemudi ojek online dan anggota Brimob. Proses pengambilan data dilakukan menggunakan **YouTube API**, yang memungkinkan penarikan komentar secara otomatis dan efisien. Dari proses ini, diperoleh sebanyak **1.143 komentar**. Setelah melalui tahap pembersihan dan seleksi, sebanyak **1.121 komentar** dinyatakan layak untuk dianalisis lebih lanjut.

2. Pra-pemrosesan Data
Tahapan ini bertujuan untuk mengubah teks mentah menjadi format yang siap untuk dianalisis oleh algoritma klasifikasi. Proses pra-pemrosesan meliputi:

- **Case Folding**: Mengubah seluruh huruf dalam komentar menjadi huruf kecil untuk menyamakan format.
- **Cleansing**: Menghapus karakter non-alfabet seperti tanda baca, angka, emoji, dan simbol yang tidak relevan.
- **Tokenizing**: Memecah teks menjadi unit kata atau token untuk memudahkan analisis.
- **Stopword Removal**: Menghilangkan kata-kata umum yang tidak memiliki nilai analitis, seperti "yang", "dan", "di".
- **Stemming**: Mengembalikan kata ke bentuk dasarnya menggunakan **Sastrawi Stemmer**, agar variasi kata dapat disatukan dalam analisis (Kurniawan, 2022).

3. Pembentukan Fitur
Setelah teks diproses, komentar dikonversi ke dalam bentuk representasi numerik menggunakan metode **TF-IDF (Term Frequency-Inverse Document Frequency)**. Teknik ini digunakan untuk mengukur seberapa penting suatu kata dalam dokumen relatif terhadap seluruh korpus. Hasil dari proses ini adalah matriks fitur yang merepresentasikan bobot kata dalam setiap komentar. (Alfin Gerliandeva, 2024)

4. Pelabelan Data

Komentar yang telah dibersihkan kemudian diberi label secara manual berdasarkan analisis konteks dan kata kunci yang terkandung di dalamnya. Tiga kategori sentimen digunakan:

Tabel 1. Pelabelan Text

Positif	Negatif	Netral
Semoga Husnul Khotimah	Brimob jangan emosi	alasan wajar tahan diri hukum lebih ringan bikin masyarakat ga puas adil

Pelabelan dilakukan dengan mempertimbangkan makna kalimat secara keseluruhan, termasuk nada, ekspresi, dan intensitas opini yang disampaikan.

5. Pembagian Data

Dataset yang telah diberi label dibagi menjadi dua bagian untuk keperluan pelatihan dan pengujian model:

- **Data latih:** 80% dari total data (896 komentar)
- **Data uji:** 20% dari total data (225 komentar)

Pembagian ini bertujuan untuk menghindari overfitting dan memastikan bahwa model dapat diuji secara objektif terhadap data yang belum pernah dilihat sebelumnya

6. Klasifikasi dengan Naive Bayes

Proses klasifikasi dilakukan menggunakan algoritma **Multinomial Naive Bayes**, yang dirancang khusus untuk menangani data teks dengan distribusi kata yang bersifat multinomial (Nuraini, 2023). Rumus dasar yang digunakan dalam klasifikasi adalah:

$$P(C|X) = \frac{P(C|X) \cdot P(C)}{P(X)}$$

Dengan penjelasan sebagai berikut:

- $P(C|X)P(C|X)$: Probabilitas suatu komentar termasuk dalam kelas sentimen CC berdasarkan fitur XX
- $P(X|C)P(X|C)$: Probabilitas fitur XX muncul dalam kelas CC
- $P(C)P(C)$: Probabilitas awal (prior) dari kelas CC

- $P(X)P(X)$: Probabilitas keseluruhan dari fitur XX (Rizky, 2024)

Model ini memanfaatkan prinsip probabilistik untuk menentukan kelas sentimen yang paling mungkin sesuai dengan pola kata dalam komentar. (Muhammad Iqbal Ahmadi, 2021)

7. Evaluasi Model

Evaluasi performa model dilakukan dengan menggunakan empat metrik utama:

- **Akurasi:** Mengukur proporsi prediksi yang benar terhadap total data uji.
- **Precision:** Menilai ketepatan prediksi positif terhadap semua prediksi positif yang dihasilkan.
- **Recall:** Mengukur kemampuan model dalam menemukan semua data yang benar-benar positif.
- **F1-score:** Merupakan rata-rata harmonis dari precision dan recall, memberikan gambaran menyeluruh tentang kinerja model (Billy Franko, 2024).

Evaluasi ini bertujuan untuk menilai sejauh mana model mampu mengklasifikasikan komentar secara akurat dan konsisten.

HASIL DAN PEMBAHASAN

Hasil

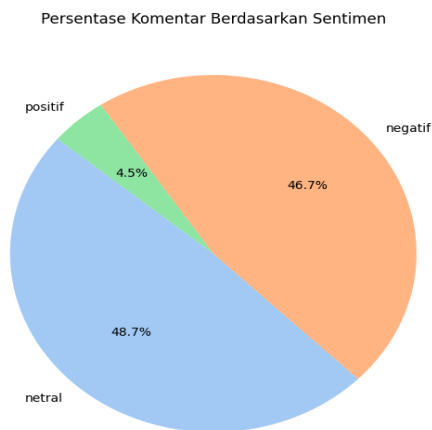
Penelitian ini menghasilkan data berupa klasifikasi sentimen komentar YouTube terhadap insiden antara pengemudi ojek online dan anggota Brimob menggunakan algoritma Naive Bayes. Setelah melalui proses preprocessing dan pelabelan, diperoleh total 1.121 komentar yang terbagi ke dalam tiga kategori sentimen, yaitu positif, netral, dan negatif.

Tabel 2. Distribusi Sentimen Komentar

Kategori	Jumlah	Persentase
Negatif	524	46,7%
Positif	51	4,5%
Netral	546	48,7%

Berdasarkan Tabel 1, sebagian besar komentar yang diberikan oleh pengguna

YouTube bersifat netral (48,7%) dan negatif (46,7%). Hal ini menunjukkan bahwa masyarakat cenderung memberikan tanggapan hati-hati terhadap peristiwa yang melibatkan aparat dan warga sipil, sementara komentar positif relatif sedikit (4,5%).



Gambar 2. Presentase Komentar Berdasarkan Sentimen

Gambar 2 memperlihatkan bahwa sentimen netral mendominasi sebagian besar komentar, diikuti oleh sentimen negatif, sedangkan sentimen positif menjadi kategori dengan jumlah paling sedikit. Distribusi ini menegaskan bahwa opini publik terhadap insiden cenderung bersifat hati-hati dan kritis terhadap pihak yang terlibat.

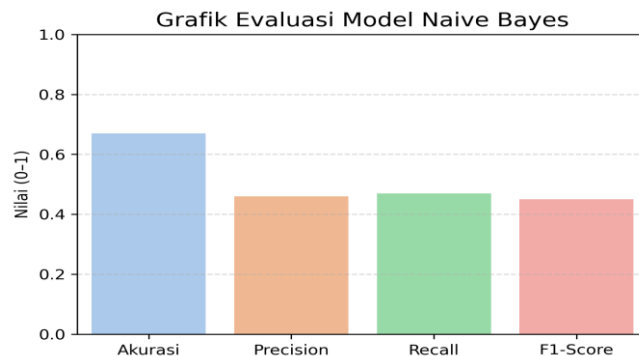
Tabel 3. Hasil Evaluasi Model Naive Bayes

Metrik	Nilai (%)
Akurasi	67
Precision	46
Recall	47
F1-score	45

Model Naive Bayes menunjukkan akurasi sebesar 67%, dengan precision 46%, recall 47%, dan F1-score 45%. Meskipun cukup efektif dalam klasifikasi opini publik, performa model masih dapat ditingkatkan, terutama dalam mendeteksi

komentar positif yang jumlahnya lebih sedikit.

Hasil ini sejalan dengan studi sebelumnya (Ghofur, 2024; Adelia, 2023) yang menunjukkan bahwa Naive Bayes cocok untuk analisis teks pendek seperti komentar YouTube. Ke depan, penggunaan algoritma lain seperti SVM atau teknik penyeimbangan data dapat dipertimbangkan untuk meningkatkan akurasi klasifikasi.



Gambar 3. Grafik Evaluasi Model

Berdasarkan Gambar 2, model Naive Bayes memperoleh nilai akurasi tertinggi sebesar 0,67, yang berarti model mampu mengklasifikasikan sebagian besar data dengan benar. Nilai precision, recall, dan F1-score yang berada di kisaran 0,45–0,47 menunjukkan performa model yang cukup stabil, namun masih dapat ditingkatkan terutama untuk mendeteksi kelas minoritas seperti sentimen positif.

Pembahasan.

Hasil klasifikasi menunjukkan bahwa mayoritas opini publik terhadap insiden antara pengemudi ojek online dan anggota Brimob bersifat netral dan negatif. Dominasi komentar netral menunjukkan bahwa sebagian besar pengguna cenderung menahan diri dan tidak menunjukkan keberpihakan yang ekstrem terhadap salah satu pihak.

Distribusi ini juga memperlihatkan bahwa isu-isu sosial yang melibatkan aparat keamanan sering kali menimbulkan pandangan yang beragam di kalangan masyarakat. Sentimen negatif yang tinggi menunjukkan adanya reaksi emosional

publik terhadap tindakan aparat yang terekam dalam video tersebut, sedangkan komentar positif relatif sedikit karena sebagian besar penonton lebih memilih untuk tidak berpihak.

Dari sisi performa algoritma, Naive Bayes terbukti mampu memberikan hasil klasifikasi yang cukup baik dengan akurasi 67%. Hasil ini sejalan dengan penelitian sebelumnya oleh Abd Ghofur (2024) dan Sella Adelia (2023) yang menggunakan metode serupa untuk analisis komentar YouTube. Nilai precision dan recall yang seimbang menunjukkan kemampuan model dalam mengenali data dengan distribusi kelas yang tidak seimbang.

Secara keseluruhan, hasil penelitian ini membuktikan bahwa Naive Bayes masih relevan digunakan dalam klasifikasi teks pendek seperti komentar YouTube karena kesederhanaannya dan efisiensi waktu komputasinya. Namun, untuk meningkatkan performa, dapat dipertimbangkan penggunaan metode lain seperti Support Vector Machine (SVM) atau Random Forest serta penyeimbangan data menggunakan oversampling atau SMOTE pada penelitian berikutnya.

SIMPULAN (PENUTUP)

Berdasarkan hasil penelitian dan analisis yang telah dilakukan, dapat disimpulkan bahwa algoritma *Naive Bayes* berhasil mengklasifikasikan sentimen komentar YouTube terhadap insiden antara pengemudi ojek online dan anggota Brimob dengan tingkat akurasi sebesar 67%. Mayoritas komentar bersifat **netral (48,7%)** dan **negatif (46,7%)**, sedangkan komentar positif hanya **4,5%**. Hal ini menunjukkan bahwa opini publik terhadap peristiwa tersebut cenderung bersikap hati-hati dan kritis terhadap pihak yang terlibat.

DAFTAR PUSTAKA

Abd Ghofur, M. F. (2024). Analisis Sentimen Komentar Youtube Terhadap Tayangan #Terbaru! Temuan dan Masalah Ahlak Di Ponpes Al-zaytun Menggunakan Metode

Naive Bayes. *G-Tech : Jurnal Teknologi Terapan*, 847-856.

Agarwal, A. &. (2023). Sentiment Analysis of YouTube Comments Using Machine Learning Techniques. *International Journal of Advanced Computer Science and Applications (IJACSA)*, 115–123.

Alfin Gerliandeva, Y. H. (2024). Optimasi Klasifikasi Sentimen pada Komentar Online menggunakan Multinomial Naive Bayes dan Ekstraksi Fitur TF-IDF serta N-grams. *Jurnal Pekommas*, 259-272.

Alfin Gerliandeva, Y. H. (2024). ptimasi Klasifikasi Sentimen pada Komentar Online menggunakan Multinomial Naive Bayes dan Ekstraksi Fitur TF-IDF serta N-grams. *Jurnal_Pekommas*, 259-272.

Billy Franko, N. W. (2024). Analisis Sentimen Terhadap Naturalisasi Pemain pada Youtube Menggunakan Decision Tree dan Naive Bayes. *JURNAL SESSION (Software Development, Digital Business Intelligence, and Computer Engineering)*, 08-16 .

Dewi Ayu, d. (2023). Klasifikasi Sentimen Komentar Youtube Tentang Pembatalan Indonesia Sebagai Tuan Rumah Piala Dunia U-20 Menggunakan Naive Baye. *Jurnal Sistem dan Teknologi Informasi (JSON)*.

Hidayat, R. &. (2024). Implementasi Algoritma Naive Bayes untuk Analisis Sentimen pada Komentar Media Sosial. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 241–249.

Kurniawan, R. (2022). Analisis Sentimen Komentar Di YouTube Tentang Ceramah Ustadz Abdul Somad Menggunakan Algoritma Naive Bayes. *Jurnal Nasional Komputasi dan Teknologi Informasi*, 131-140.

Linda, d. (2025). Analisis Sentimen Komentar Youtube Terhadap Isu Kesehatan Mental Menggunakan

- Algoritma Naïve Bayes dan KNN. *Jurnal Syntax Transformation*.
- Muhammad Iqbal Ahmadi, D. G. (2021). Analisis Sentiment Masyarakat terhadap Kasus Covid-19 pada Media Sosial Youtube dengan Metode Naive bayes. *Jurnal Sains Komputer & Informatika (J-SAKTI)*, 807-814.
- Nuraini, L. &. (2023). Analisis Sentimen Komentar YouTube pada Video Edukasi Menggunakan Pendekatan Text Mining. *Jurnal Ilmiah Informatika dan Komputer (JIK)*, 97–106.
- Rizky, F. A. (2024). Optimasi Klasifikasi Naive Bayes dengan TF-IDF untuk Analisis Sentimen di Media Sosial. *Jurnal Teknologi dan Sistem Informasi (JTSI)*, 50–58.
- Sella Adelia, E. M. (2023). Analisis Sentimen Belajar Programming Pada Media Sosial Youtube Menggunakan Algoritma Klasifikasi Naive Bayes. *Journal of Information Technology Ampera*, 254-264.
- Triana Putri, S. N. (2025). Analisis Sentimen Penggunaan Aplikasi YouTube dengan Metode Naïve Bayes. *UNP Journal of Statistics and Data Science*, 60–66.
- Yuyun Khanafiyah, D. K. (2025). Analisis Sentimen Komentar Youtube Kanal Dirty Vote Menggunakan Metode Naive Bayes Classifier. *e-Proceeding of Engineering* , 6723-6728.