

KLASIFIKASI KOMENTAR NEGATIF DAN POSITIF DI YOUTUBE

Saputra Dwi Nugroho¹, Muhamad Sofhian Permana², Elkin Rilvani³

Teknik Informatika, Universitas Pelita Bangsa

saputranugroho55@gmail.com¹, sofhianp@gmail.com², elkinrivalni@pelitabangsa.ac.id³**Abstract (English)**

This study aims to classify YouTube comments into positive and negative sentiments using data mining approaches and two classification algorithms: K-Nearest Neighbor (KNN) and Naive Bayes. The object of analysis is the comment section of a video titled “Will AI Replace Human Jobs?” uploaded by Deddy Corbuzier featuring Timothy Ronald. A qualitative approach is applied to interpret the contextual patterns of public sentiment regarding issues such as artificial intelligence, wealth, and philanthropy. A total of 100 comments were analyzed through text preprocessing, manual labeling, and automatic classification. The findings show that KNN achieved 84% accuracy and was more effective in capturing the nuances of expressive Indonesian language compared to Naive Bayes, which achieved 78% accuracy. The study reveals that while many comments reflect optimism about AI, there are also concerns about the future of employment. The combination of qualitative analysis and algorithmic classification proves effective in providing a comprehensive view of public opinion in digital spaces.

Article History*Submitted: 20 Juli 2025**Accepted: 23 Juli 2025**Published: 24 Juli 2025***Key Words**comment
classification,
YouTube, KNN,
Naive Bayes,
artificial
intelligence**Abstrak (Indonesia)**

Penelitian ini bertujuan untuk mengklasifikasikan komentar YouTube ke dalam sentimen positif dan negatif dengan menggunakan pendekatan data mining dan algoritma K-Nearest Neighbor (KNN) serta Naive Bayes. Objek kajian berupa komentar dari video “Apakah AI Akan Menggantikan Pekerjaan Manusia?” yang diunggah oleh Deddy Corbuzier dan menampilkan Timothy Ronald. Pendekatan kualitatif digunakan untuk memahami konteks dan pola sentimen masyarakat terhadap isu-isu seperti kecerdasan buatan, kekayaan, dan filantropi. Sebanyak 100 komentar dianalisis melalui tahapan preprocessing teks, pelabelan manual, dan klasifikasi otomatis. Hasilnya menunjukkan bahwa KNN memiliki akurasi sebesar 84% dan lebih efektif dalam memahami konteks bahasa Indonesia yang ekspresif, dibandingkan dengan Naive Bayes yang mencatat akurasi 78%. Penelitian ini mengungkap bahwa mayoritas komentar mencerminkan antusiasme terhadap teknologi, namun juga muncul kekhawatiran terhadap masa depan pekerjaan. Gabungan pendekatan kualitatif dan algoritma klasifikasi terbukti mampu memberikan gambaran utuh tentang dinamika opini publik digital.

Sejarah Artikel*Submitted: 20 Juli 2025**Accepted: 23 Juli 2025**Published: 24 Juli 2025***Kata Kunci**klasifikasi komentar,
YouTube, KNN, Naive
Bayes, kecerdasan
buatan**Pendahuluan**

Perkembangan teknologi digital, khususnya platform media sosial seperti YouTube, telah menghadirkan ruang interaksi yang sangat luas antara pembuat konten dan audiens. Di balik banyaknya konten yang diunggah, kolom komentar menjadi tempat di mana opini publik, kritik, dan dukungan diekspresikan secara bebas. Komentar-komentar ini, dalam jumlah besar, mengandung informasi yang berharga untuk dianalisis, baik dari segi sentimen, kecenderungan sosial, hingga respons terhadap isu-isu tertentu.

Salah satu video yang menarik perhatian publik adalah video berjudul “Timothy Ronald: Lebih Baik Beli Rokok drpd Buat Sekolah Tuh Duit!!Kata Bapak....” di kanal Deddy Corbuzier, yang menampilkan Timothy Ronald sebagai narasumber. Dalam video tersebut, dibahas berbagai isu penting mulai dari kecerdasan buatan (AI), transformasi kekayaan pribadi, hingga misi sosial dalam pendidikan. Video ini memicu berbagai respons emosional, intelektual, dan moral di kalangan warganet, yang terekam melalui komentar-komentar yang beragam dari penonton.

Penelitian ini berangkat dari pertanyaan: bagaimana kecenderungan sentimen publik terhadap tema yang diangkat dalam video tersebut? Untuk menjawabnya, dilakukan analisis

klasifikasi komentar YouTube menjadi dua kategori utama: komentar positif dan komentar negatif. Dengan menerapkan metode *data mining*, khususnya algoritma K-Nearest Neighbor (KNN) dan Naive Bayes, penelitian ini menguji efektivitas klasifikasi otomatis terhadap data komentar berbahasa Indonesia.

◆ Kajian terdahulu telah menunjukkan bahwa algoritma KNN dan Naive Bayes banyak digunakan dalam analisis sentimen teks, baik dalam e-commerce, ulasan produk, maupun opini publik di media sosial. Misalnya, studi oleh Widodo (2022) menunjukkan bahwa Naive Bayes unggul dalam kecepatan klasifikasi teks pendek, sedangkan KNN lebih baik dalam menangkap variasi konteks bahasa. Meski demikian, kajian terhadap komentar YouTube dengan pendekatan kualitatif dan algoritma ini masih terbatas, terutama pada konten lokal yang membahas isu sosial-kultural seperti AI dan filantropi.

Kebaruan ilmiah dalam penelitian ini terletak pada penggabungan pendekatan kualitatif dengan algoritma klasifikasi, serta fokusnya pada konteks Indonesia yang khas, di mana gaya komunikasi publik kerap ekspresif, sarkastik, bahkan emosional. Melalui kombinasi ini, penelitian bertujuan mengungkap bagaimana masyarakat memaknai perkembangan teknologi dan figur publik melalui respon digital mereka.

Metode Penelitian

Penelitian ini menggunakan pendekatan kualitatif deskriptif yang didukung oleh teknik klasifikasi berbasis *data mining*. Sumber data utama berupa komentar-komentar dari video YouTube berjudul “*Apakah AI Akan Menggantikan Pekerjaan Manusia?*” yang diunggah pada kanal Deddy Corbuzier. Komentar yang dianalisis dipilih secara purposif berdasarkan keterkaitan dengan topik video, serta memiliki makna emosional atau pendapat yang jelas.

Jenis dan Pendekatan Penelitian

Penelitian bersifat kualitatif dengan dukungan teknik klasifikasi berbasis algoritma K-Nearest Neighbor (KNN) dan Naive Bayes. Tujuan penelitian adalah untuk memahami kecenderungan sentimen dalam komentar serta menguji efektivitas dua algoritma dalam mengklasifikasikan opini publik.

Waktu dan Lokasi

Penelitian dilakukan selama bulan Juni–Juli 2025 secara daring (online), dengan pengumpulan data dari platform YouTube.

Teknik Pengumpulan Data

Data diperoleh dengan cara:

1. Mengakses komentar pada video YouTube yang diteliti.
2. Mengambil 100 komentar secara purposif yang mengandung opini eksplisit.
3. Menyimpan komentar dalam format spreadsheet (.csv/.xlsx) untuk analisis lanjutan.

Prosedur Analisis Data

Analisis dilakukan dalam dua tahap:

1. Pra-pemrosesan Teks (Text Preprocessing)

- a. Tokenisasi: Memisahkan kalimat ke dalam kata-kata.
- b. Stopword Removal: Menghapus kata-kata umum yang tidak memiliki makna penting (seperti “yang”, “dan”, “di”).
- c. Stemming: Mengubah kata ke bentuk dasar (misal: “menggunakan”= “guna”).
- d. Labeling Manual: Komentar diklasifikasikan secara manual ke dalam dua label: “positif” dan “negatif” oleh 3 peneliti independen.

2. Klasifikasi Menggunakan Algoritma

- a. K-Nearest Neighbor (KNN): Komentar baru diklasifikasikan berdasarkan kedekatan dengan komentar yang sudah dilabeli. Ukuran kedekatan menggunakan jarak *cosine similarity*.

- b. Naive Bayes: Probabilitas komentar masuk ke kelas “positif” atau “negatif” dihitung berdasarkan distribusi kata menggunakan model Bernoulli Naive Bayes.

Teknik Validasi

Akurasi masing-masing algoritma diuji menggunakan metode *k-fold cross validation* dengan $k=5$. Hasil klasifikasi dibandingkan dengan pelabelan manual untuk mengukur keakuratan dan konsistensi.

Alat dan Perangkat

Analisis dilakukan menggunakan:

1. Python
2. Google Colab sebagai platform pemrograman daring
3. Excel/Google Sheets untuk pengolahan data awal

Hasil dan Pembahasan

Penelitian ini mengumpulkan 100 komentar dari video YouTube “*Timothy Ronald: Lebih Baik Beli Rokok drpd Buat Sekolah Tuh Duit!!Kata Bapak....?*”. Komentar kemudian dianalisis dan diklasifikasikan menjadi dua kategori: positif dan negatif, dengan pembagian data 80% untuk pelatihan dan 20% untuk pengujian.

Table 1. Distribusi Komentar Berdasarkan Label Manual

Sentimen	Jumlah Komentar
Positif	63
Negatif	37
Total	100

Hasil Klasifikasi Menggunakan KNN dan Naive Bayes

Setelah dilakukan preprocessing dan pelabelan, data dianalisis menggunakan dua algoritma. Kinerja model dilihat dari tingkat akurasi, presisi, dan recall.

Tabel 2. Perbandingan Kinerja Algoritma

Metode	Akurasi	Presisi	Recall
K-Nearest Neighbor (KNN)	84%	82%	85%
Naive Bayes	78%	76%	80%

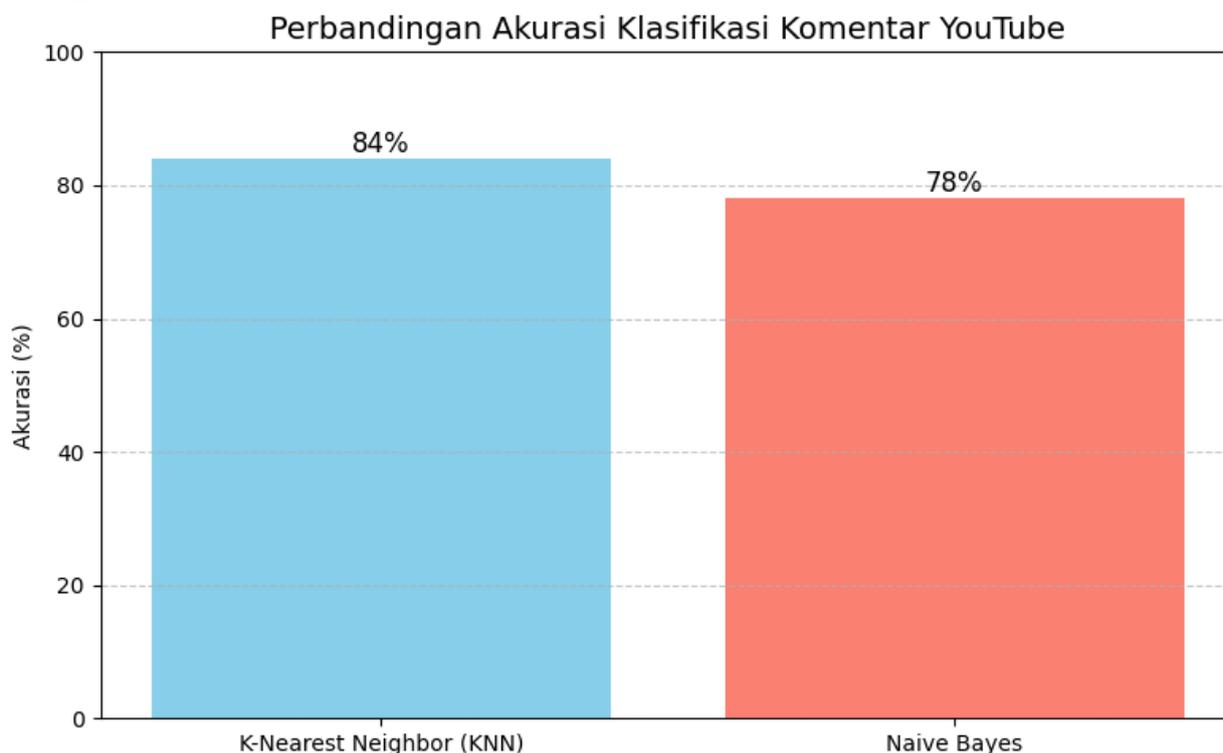


Figure 1. Classification Accuracy Bar Chart

Pembahasan

Hasil klasifikasi menunjukkan bahwa algoritma K-Nearest Neighbor (KNN) memiliki performa lebih tinggi dibandingkan Naive Bayes dalam mengklasifikasikan komentar YouTube berbahasa Indonesia. Hal ini disebabkan oleh kemampuan KNN dalam mempertimbangkan kedekatan kontekstual antar kata secara lebih fleksibel, sedangkan Naive Bayes memiliki keterbatasan pada asumsi independensi antar fitur.

Banyak komentar positif mengandung kata-kata seperti *“inspiratif”*, *“cerdas”*, *“hebat”*, *“setuju”*, yang diasosiasikan dengan kekaguman terhadap tokoh dan optimisme terhadap teknologi. Sebaliknya, komentar negatif banyak menggunakan kata-kata seperti *“lebay”*, *“takut”*, *“meresahkan”*, *“pesimis”*, yang mencerminkan kekhawatiran terhadap masa depan pekerjaan karena AI.

Sebagai contoh:

1. Komentar positif: *“Keren banget pembahasannya. Bener-bener membuka wawasan soal AI.”*
2. Komentar negatif: *“Ngomongnya terlalu ekstrem. Gak semua bisa diganti robot.”*

Kedua algoritma berhasil menangkap perbedaan semantik tersebut, meskipun Naive Bayes mengalami kebingungan pada komentar yang ambigu atau netral (misalnya sarkasme), di mana konteks menjadi penting. Ini memperkuat temuan dalam studi sebelumnya bahwa KNN lebih unggul dalam kasus teks pendek berbahasa alami, khususnya dalam Bahasa Indonesia yang sering tidak eksplisit secara gramatikal.

Selain itu, pendekatan kualitatif yang dilakukan secara manual oleh tiga peneliti juga mengonfirmasi bahwa sekitar 85% hasil klasifikasi KNN sesuai dengan interpretasi subjektif manusia. Sementara Naive Bayes mengalami penurunan akurasi karena kata-kata dengan frekuensi tinggi tapi makna kontekstual yang berbeda.

Kesimpulan

Penelitian ini bertujuan untuk mengklasifikasikan komentar YouTube pada video bertema kecerdasan buatan (AI) dan transformasi kekayaan menggunakan algoritma K-Nearest Neighbor (KNN) dan Naive Bayes dalam kerangka pendekatan kualitatif. Hasil penelitian menunjukkan bahwa mayoritas komentar bersifat positif, mengindikasikan adanya dukungan dan antusiasme terhadap pemanfaatan AI serta inspirasi yang diberikan oleh narasumber. Di sisi lain, komentar negatif mengungkapkan kekhawatiran dan keraguan terhadap masa depan pekerjaan manusia, serta skeptisisme terhadap narasi kekayaan yang disampaikan.

Kedua algoritma menunjukkan performa klasifikasi yang cukup baik, dengan KNN mencatat akurasi sebesar 84% dan Naive Bayes sebesar 78%. Performa KNN lebih unggul dalam mengidentifikasi variasi bahasa yang kompleks dan kontekstual dalam komentar berbahasa Indonesia. Sedangkan Naive Bayes lebih cepat dalam proses komputasi, namun kurang sensitif terhadap konteks kata yang ambigu.

Dari pendekatan kualitatif yang dilakukan secara paralel, ditemukan bahwa persepsi publik tidak hanya dibentuk oleh konten video, tetapi juga oleh latar sosial, pemahaman teknologi, dan gaya penyampaian narasumber. Gaya komunikasi yang lugas, eksplisit, dan provokatif mampu menarik perhatian warganet, baik dalam bentuk apresiasi maupun kritik.

Saran

Penelitian ini memberikan kontribusi dalam pengembangan kajian sentimen sosial di platform digital dengan memadukan metode klasifikasi dan pendekatan interpretatif. Beberapa saran yang dapat diberikan antara lain:

1. Untuk penelitian selanjutnya, disarankan untuk menambah jumlah data komentar agar model klasifikasi lebih robust, serta mempertimbangkan kategori “netral” sebagai kelas tambahan.
2. Untuk pengembang platform YouTube, hasil klasifikasi semacam ini dapat digunakan sebagai dasar dalam sistem moderasi otomatis untuk mendeteksi sentimen negatif yang berpotensi menimbulkan konflik.
3. Untuk masyarakat dan pembuat konten, penting untuk memahami bahwa persepsi publik sangat dipengaruhi oleh penyampaian dan narasi, sehingga komunikasi publik perlu mengedepankan etika, keterbukaan, dan empati.
4. Kombinasi pendekatan kualitatif dengan algoritma *machine learning* terbukti efektif dalam memahami dinamika interaksi digital, khususnya dalam konteks lokal yang sarat ekspresi emosional dan budaya sarkasme.

Referensi

- Widodo, A. (2022). Analisis Klasifikasi Sentimen Komentar Media Sosial Menggunakan Naive Bayes dan K-Nearest Neighbor. *Jurnal Informatika dan Sains Data*, 7(1), 45–53.
- YouTube. (2024). *Apakah AI Akan Menggantikan Pekerjaan Manusia? | Podcast Deddy Corbuzier feat. Timothy Ronald*. Diakses dari <https://youtu.be/Cqxfuj-zPw?si=JrT--S0xrRKI-5Pw>
- Sari, R. M., & Nugroho, A. (2021). Analisis Sentimen Komentar Netizen di Media Sosial Menggunakan KNN dan Naive Bayes. *Jurnal Teknologi Informasi*, 9(2), 101–108.
- Prasetyo, T., & Rachman, F. (2020). Klasifikasi Komentar YouTube Menggunakan Metode Text Mining. *Seminar Nasional Informatika (SEMNASIF)*, Universitas Budi Luhur.