

Analisis Sentimen Menggunakan Metode Naive Bayes Pada Komentar Penonton YouTube Windah Basudara

Adam Putra Berlin¹, Arafat Febriandirza²

^{1,2}Universitas Muhammadiyah Prof. Dr. Hamka, Indonesia

Email: adamputraberlin@gmail.com¹, arafat@uhamka.ac.id²

Abstract

The development of social media has provided users with a space to express their opinions through comments, including on the YouTube platform. One content creator who has a large fanbase and active comment section is Windah Basudara. This study aims to analyze the sentiment of viewer comments on one of Windah Basudara's videos using the Naive Bayes algorithm. This method was chosen due to its effectiveness in text classification and sentiment analysis. The data used consists of comments from the video titled "Mencoba NAMATIN game Keju Joget", which were collected randomly and cleaned through text preprocessing steps such as case folding, tokenizing, stopword removal, and stemming. The comments were classified into two sentiment categories: positive and negative. The analysis results show that the majority of comments carry a positive sentiment, reflecting a favorable response from viewers toward the presented content. The model evaluation demonstrates satisfactory classification results. This study is expected to contribute to understanding audience perception of YouTube content and serve as a reference for further analysis on social media platforms.

Keywords: Analysis Sentiment; Naive Bayes; Windah Basudara; YouTube Comment

Abstrak

Perkembangan media sosial telah memberikan ruang bagi pengguna untuk mengekspresikan pendapat melalui komentar, termasuk pada platform YouTube. Salah satu kreator konten yang memiliki banyak penggemar dan komentar aktif adalah Windah Basudara. Penelitian ini bertujuan untuk menganalisis sentimen komentar penonton terhadap salah satu video Windah Basudara dengan menggunakan algoritma Naive Bayes. Metode ini dipilih karena kemampuannya yang efektif dalam klasifikasi teks dan analisis sentimen. Data yang digunakan berupa komentar dari video berjudul "Mencoba NAMATIN game Keju Joget", yang dikumpulkan secara acak dan dibersihkan melalui proses preprocessing teks seperti case folding, tokenizing, stopword removal dan stemming. Komentar diklasifikasikan menjadi dua kategori sentimen positif dan negative. Hasil analisis menunjukkan bahwa mayoritas komentar memiliki sentimen positif, mencerminkan respon positif penonton terhadap konten yang disajikan. Evaluasi model menunjukkan hasil klasifikasi yang memuaskan. Studi ini diharapkan memberi kontribusi dalam memahami persepsi audiens terhadap konten YouTube serta sebagai referensi dalam pengembangan analisis di media sosial.

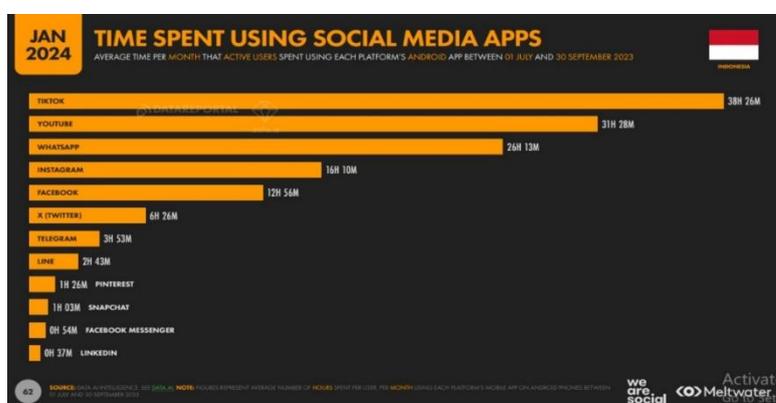
Kata kunci: Analisis Sentimen; Komentar YouTube; Naive Bayes; Windah Basudara.

1. PENDAHULUAN

Di era digital, media sosial telah menjadi platform penting bagi masyarakat untuk berbagi informasi, berinteraksi, dan mengekspresikan diri. Platform-platform seperti YouTube, TikTok Instagram, dan WhatsApp bukan hanya sekedar sarana hiburan, tetapi juga menjadi wadah bagi pengguna media sosial untuk menyuarakan pandangan dan menanggapi berbagai isu sosial [1]. Salah satu platform yang paling populer adalah YouTube, yang menyediakan ruang bagi pengguna media sosial untuk menonton, mengunggah, dan berbagi video dalam

berbagai kategori, mulai dari hiburan, edukasi, hingga konten unik seperti *gaming* [2].

Berdasarkan laporan Digital 2024 Indonesia yang diterbitkan oleh *We Are Social* dan Meltwater, pengguna media sosial di Indonesia menghabiskan waktu yang cukup tinggi di berbagai aplikasi media sosial setiap bulannya. Data terbaru menunjukkan bahwa TikTok menjadi aplikasi yang paling banyak digunakan dengan waktu rata-rata per bulan sebesar 38 jam 26 menit, diikuti oleh YouTube sebesar 31 jam 28 menit. Hal ini menunjukkan bahwa YouTube tetap menjadi platform yang menarik perhatian pengguna media sosial di Indonesia, memungkinkan mereka untuk mengonsumsi dan berinteraksi dengan konten secara intensif[3].



Gambar 1. Data Pengguna Media Sosial di Indonesia

Salah satu fitur penting YouTube yang memungkinkan interaksi antara konten kreator dan audiens adalah kolom komentar. Kolom ini menjadi tempat bagi penonton untuk menyampaikan opini, kritik, saran dan dukungan terhadap konten yang disajikan [4]. Salah satu konten kreator terkenal di Indonesia adalah Windah Basudara, seorang YouTuber yang dikenal dengan gaya bermain *game* yang menghibur dan komunikatif. Salah satu video populernya, “Mencoba NAMATIN Game Keju Joget”, berhasil menarik perhatian dengan jutaan penonton dan puluhan ribu komentar. Komentar-komentar ini tidak hanya mencerminkan reaksi individu, tetapi juga menunjukkan kecenderungan sentiment umum yang dapat berupa apresiasi sampai kritikan dari penontonnya.

Melalui analisis sentimen terhadap komentar-komentar ini, peneliti dan kreator dapat memperoleh wawasan berharga mengenai preferensi penonton dan pengaruh konten terhadap audiens [5]. Sentimen analisis, atau analisis opini, adalah bagian dari *Natural Language Processing* (NLP) yang bertujuan untuk mengidentifikasi dan mengklasifikasikan sentimen positif dan negatif dari suatu teks. Di bidang ini, berbagai metode telah dikembangkan, salah satunya adalah algoritma *Naive Bayes*. Algoritma ini populer digunakan dalam klasifikasi teks karena kemampuannya dalam menangani data besar dan memberikan akurasi yang baik dengan waktu komputasi yang rendah. Dalam konteks ini, algoritma *Naive Bayes* dapat membantu mengklasifikasikan komentar-

komentar penonton terhadap video Windah Basudara menjadi sentimen positif dan negatif.

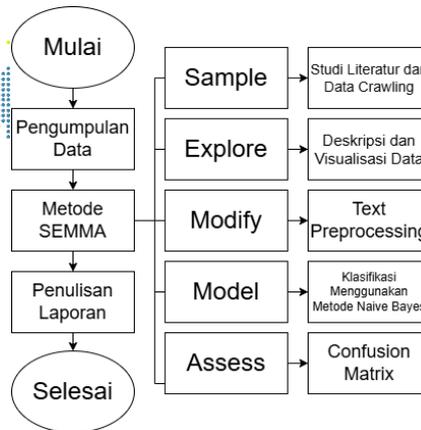
Penelitian yang dilakukan oleh [6] bertujuan untuk menganalisis sentimen masyarakat Indonesia terhadap isu vaksinasi Covid-19 melalui platform Twitter dengan menggunakan metode *Naive Bayes Classifier*. Data yang digunakan merupakan *tweet* berbahasa Indonesia yang dikumpulkan melalui proses *crawling* dengan kata kunci “vaksin covid-19” pada periode Desember 2020 hingga Maret 2021. Setelah dikumpulkan, data diproses melalui tahapan *preprocessing* seperti *case folding*, *tokenizing*, *spelling normalization*, *filtering*, dan *stemming*, dilanjutkan dengan pelabelan sentimen menggunakan kamus *lexicon* dan *negative words*. Selanjutnya, dilakukan pembobotan kata menggunakan TF-IDF dan seleksi fitur dengan Information Gain untuk meningkatkan akurasi model klasifikasi. Dari total 3.406 *tweet* yang dianalisis, sebanyak 1177 tergolong dalam sentimen positif, 664 negatif, dan 1565 netral. Hasil evaluasi model menggunakan *confusion matrix* menunjukkan tingkat akurasi sebesar 78%, recall 80. Penelitian yang dilakukan oleh [7] menganalisis sentiment pada media sosial YouTube terkait serikat pekerja Pertamina menggunakan metode *Naive Bayes*. Hasil penelitian tersebut menunjukkan bahwa data awal, setelah melalui proses *preprocessing*, dikategorikan ke dalam tiga kategori sentimen: netral, positif, dan negatif. Dalam studi ini, penelitian dilakukan menggunakan 300 data untuk pengujian (*testing*) dan 1000 data untuk pelatihan (*training*). Penelitian ini mencapai Tingkat akurasi sebesar 98% untuk ketiga kategori sentimen tersebut. Selain itu, nilai presisi untuk prediksi sentimen positif mencapai 80%, sedangkan prediksi untuk sentimen negatif memperoleh nilai presisi sebesar 66,67%.

Selain itu, fokus pada komentar di YouTube juga penting karena platform ini tidak hanya sebagai media hiburan tetapi juga sebagai tempat diskusi publik di mana orang-orang dari berbagai latar belakang bebas menyuarakan pendapat mereka. Dengan analisis sentimen pada data komentar ini, konten kreator dapat lebih memahami audiens, memperbaiki kualitas konten, dan meningkatkan *engagement* terhadap penonton. Analisis ini juga memiliki relevansi di bidang pemasaran, kajian media, dan ilmu sosial di mana pemahaman respons masyarakat terhadap konten tertentu dapat mendukung pengambilan Keputusan yang lebih baik.

2. METODOLOGI PENELITIAN

2.1. Metode SEMMA

Metode yang digunakan dalam penelitian ini mengacu pada pendekatan SEMMA Data Mining Process [8] dan data diambil dari komentar penonton YouTube Windah Basudara dengan video yang berjudul “Mencoba NAMATIN Game Keju Joget”. Tahapan-tahapan pada penelitian ini digambarkan dalam *flowchart* pada Gambar 2.



Gambar 2. Flowchart Alur Penelitian

a. Sample

Pengumpulan data yang melibatkan data komentar pada video “Mencoba NAMATIN Game Keju Joget” di *channel* YouTube Windah Basudara dengan teknik *web scrapping*, dengan menggunakan *tools* Python melalui media Google Colaboratory. Data yang sudah dikumpulkan disimpan dalam format csv [9].

b. Explore

Pada tahap ini, dilakukan deskripsi data dan visualisasi data. Penjelasan karakteristik dan jumlah komentar yang dianalisis.

c. Modify

Tahapan ini dilakukannya *text preprocessing* yang bertujuan untuk pembersihan data dan pengelolaan data informasi yang tidak lengkap. *Text preprocessing* ini terdiri dari beberapa proses yaitu *cleaning*, *case folding*, *tokenize*, *stopword removal* dan *stemming*.

d. Model

Klasifikasi data berdasarkan kelas untuk menentukan sentimen penonton yang berada di dalam komentar merupakan opini positif atau negatif berdasarkan kamus *lexicon based*. Setelah itu, data dibagi menjadi dua bagian, yaitu data *training* dan data *testing*, untuk diuji dengan metode *Naive Bayes* [10].

e. Assess

Tahap evaluasi terhadap pemodelan yang telah dibuat dengan mengevaluasi hasil prediksi model dengan membandingkannya terhadap data uji yang sudah memiliki label sentimen. yang telah diberi label. Hasil evaluasi dihitung berdasarkan besaran dari *precision*, *recall* dan akurasi.

2.2. Naive Bayes

Naive Bayes adalah salah satu metode yang dapat digunakan untuk mengklasifikasi data dengan mencari probabilitas yang paling tinggi. Klasifikasi Bayes adalah pengklasifikasian statistik untuk menentukan prediksi suatu kelas anggota probabilitas [11]. *Naive Bayes* merupakan metode klasifikasi yang didasarkan pada Teorema Bayes, dengan ciri asumsi utama yang kuat terhadap independen satu sama lain secara kondisional dan didasarkan pada teori penyederhanaan ketika nilai atribut secara kondisional diberikan nilai output

Tahap berikutnya adalah *stopword removal*, yaitu menghapus kata-kata umum yang tidak memiliki kontribusi signifikan terhadap analisis sentimen menggunakan *library* Sastrawi. Tahap terakhir adalah *stemming*, yang berfungsi mengubah kata berimbuhan menjadi bentuk dasarnya agar hasil klasifikasi lebih akurat. Semua proses ini membantu meningkatkan kualitas data teks sebelum dimasukkan ke dalam algoritma klasifikasi *Naive Bayes*.

3.4. Model

Tahap pemodelan dalam penelitian ini menggunakan pendekatan *lexicon-based* untuk melakukan labeling sentimen pada komentar YouTube, yang diklasifikasikan menjadi dua kategori yaitu positif dan negatif berdasarkan kecocokan kata dalam kamus sentimen. Hasilnya, sebanyak 2.181 komentar termasuk dalam sentimen positif dan 537 komentar termasuk sentimen negatif. Data yang telah dilabeli kemudian dibagi untuk proses klasifikasi menggunakan metode *Naive Bayes* dengan variasi pembagian data latih dan data uji, yaitu 90:10, 80:20, dan 70:30.

3.5. Assess

Tahap *assess* dalam penelitian ini bertujuan mengevaluasi performa model klasifikasi *Naive Bayes* dengan mengukur nilai *accuracy*, *precision*, dan *recall* menggunakan *confusion matrix*. Evaluasi dilakukan pada tiga rasio data uji, yaitu 90:10, 80:20, dan 70:30. Hasil menunjukkan bahwa model memiliki kinerja yang konsisten dengan *accuracy* sebesar 85% pada ketiga rasio. Untuk rasio 90:10, *precision* mencapai 88% dan *recall* 85%. Pada rasio 80:20 dan 70:30, *precision* stabil di angka 87% dan *recall* tetap 85%. Selain itu, pada *confusion matrix* rasio 90:10, model mencatat 219 *true positive* dan 13 *true negative*, dengan 40 *false positive* dan tanpa *false negative*.

Untuk mengoptimalkan evaluasi, penelitian ini menerapkan *k-fold cross validation* dengan nilai $k = 10$. Hasil validasi silang menunjukkan performa model yang stabil, dengan nilai *accuracy* berkisar antara 83%–86%, *precision* 88%–93%, dan *recall* 80%–88%. Hal ini memperkuat keandalan model dalam menangani variasi data uji.

Tabel 1. Hasil Evaluasi Model

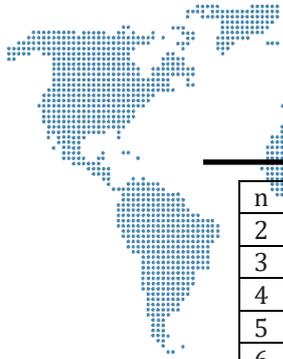
Rasio	Accuracy	Precision	Recall
90:10	85%	88%	85%
80:20	85%	87%	85%
70:30	85%	87%	85%

Tabel 2. Confusion Matrix (Rasio 90:10)

	Prediksi Negatif	Prediksi Positif
Aktual Negatif	13	40
Aktual Positif	0	219

Tabel 3. Cross Validation (k = 10)

n	Accuracy	Precision	Recall
1	84%	92%	80%



n	Accuracy	Precision	Recall
2	83%	91%	86%
3	86%	90%	85%
4	86%	92%	84%
5	83%	91%	86%
6	85%	88%	81%
7	85%	92%	81%
8	86%	93%	84%
9	84%	92%	88%
10	86%	93%	86%

4. SIMPULAN

Penelitian ini menyimpulkan bahwa metode *Naive Bayes* efektif dan efisien dalam mengklasifikasikan sentimen pada komentar YouTube, terutama ketika digunakan dengan *preprocessing* yang optimal dan pendekatan *lexicon-based*. Mayoritas komentar terhadap video Windah Basudara bersifat positif, yang menunjukkan keterlibatan dan apresiasi audiens terhadap kontennya. Dari total 17.136 komentar yang telah diproses, model *Naive Bayes* berhasil mengklasifikasikan 2.181 komentar sebagai sentimen positif dan 537 sebagai sentimen negatif. Evaluasi model menunjukkan konsistensi akurasi sebesar 85% dengan precision hingga 88% dan recall 85% pada berbagai rasio pembagian data. Validasi silang *10-fold* memperkuat hasil ini dengan rentang akurasi 83%–86%. Konsistensi performa model pada berbagai skenario menunjukkan bahwa pendekatan ini cocok digunakan untuk analisis opini publik di media sosial. Ke depannya, model dapat dikembangkan lebih lanjut menggunakan pendekatan *hybrid* atau metode *deep learning* untuk memperoleh hasil yang lebih optimal dan generalisasi yang lebih baik terhadap beragam data teks.

DAFTAR PUSTAKA

- [1] J. Pendidikan And D. Kebudayaan, “Revitalisasi Budaya Di Era Digital Dan Eksplorasi Dampak Media Sosial Terhadap Dinamika Sosial-Budaya Di Tengah Masyarakat,” Vol. 3, No. 3, Pp. 172–184, 2023, Doi: 10.55606/Jurdikbud.V3i3.
- [2] A. A. Ningtyas, A. Solichin, And R. Pradana, “Analisis Sentimen Komentar Youtube Tentang Prediksi Resesi Ekonomi Tahun 2023 Menggunakan Algoritme Naive Bayes,” 2023.
- [3] T. Wahyu Amalia And A. Prihandari Satvikadewi, “Personal Branding Content Creator Arif Muhammad (Analisis Visual Pada Akun Youtube @Arif Muhammad),” 2020.
- [4] E. Tohidi, R. Perdana Herdiansyah, And E. Wahyudin, “Analisa Sentimen Komentar Video Youtube Di Channel Tvonenews Tentang Calon Presiden Prabowo Subianto Menggunakan Support Vector Machine,” 2024. [Online]. Available: https://www.youtube.com/watch?v=Wu_Wqhgl1yk
- [5] H. Faisal, A. Febriandirza, And F. N. Hasan, “Analisis Sentimen Terkait Ulasan Pada Aplikasi Pln Mobile Menggunakan Metode Support Vector Machine,” 2024.
- [6] F. Septianingrum, J. H. Jaman, And U. Enri, “Analisis Sentimen Pada Isu Vaksin Covid-19 Di Indonesia Dengan Metode Naive Bayes Classifier,” *Jurnal Media Informatika Budidarma*, Vol. 5, No. 4, P. 1431, Oct. 2021, Doi: 10.30865/Mib.V5i4.3260.
- [7] M. Soekarno Putra And S. D. Wati, “Analisis Sentimen Serikat Pekerja Pertamina Tolak Ahok Pada Media Sosial Youtube Menggunakan Algoritma Naive Bayes,” 2021.

- [8] M. Dwison Alizah, A. Nugroho, U. Radiyah, W. Gata, P. I. Komputer, And N. Mandiri, "Sentimen Analisis Terkait Lockdown Pada Sosial Media Twitter," *Ijse-Indonesian Journal On Software Engineering*, Vol. 6, No. 2, Pp. 223–229, 2020.
- [9] N. Azpiranda, A. A. Suprianto, N. Y. Setiawan, E. Suryawati, R. S. Yuwana, And A. Febriandirza, "Sentiment Analysis On Customer Reviews Using Support Vector Machine And Usability Scoring Using System Usability Scale," 2021. [Online]. Available: [Www.Jitecs.Ub.Ac.Id](http://www.jitecs.ub.ac.id)
- [10] A. N. Ihsan And S. Tresnawati, "Analisis Sentimen Pada Platform X Terhadap Layanan Provider Tri Menggunakan Naïve Bayes Dan Support Vector Machine," *Jurnal Informatika Dan Teknik Elektro Terapan*, Vol. 12, No. 3s1, Oct. 2024, Doi: 10.23960/Jitet.V12i3s1.5264.
- [11] B. Samodera, K. Kartini, And M. M. Al Haromainy, "Implementasi Majority Vote Pada Metode Naive Bayes Dan Support Vector Machine(Studi Kasus: Kenaikan Pajak Hiburan)," *Jurnal Informatika Dan Teknik Elektro Terapan*, Vol. 12, No. 3, Aug. 2024, Doi: 10.23960/Jitet.V12i3.4799.
- [12] B. B. Suherman, "Sistem Pakar Diagnosa Penyakit Dan Hama Pada Tanaman Jagung Menggunakan Metode Naive Bayes," *Jurnal Informatika Dan Rekayasa Perangkat Lunak (Jatika)*, Vol. 2, No. 3, Pp. 390–398, 2021, [Online]. Available: [Http://Jim.Teknokrat.Ac.Id/Index.Php/Informatika](http://jim.teknokrat.ac.id/index.php/informatika)
- [13] F. I. Wibowo And A. Febriandirza, "Analisis Sentimen Ulasan Pengguna Game Pubg Di Google Play Store Menggunakan Algoritma Naïve Bayes," *Jurnal Sistem Komputer Dan Informatika (Json)*, Vol. 5, No. 3, P. 590, Mar. 2024, Doi: 10.30865/Json.V5i3.7264.