

Analisis Sentimen Opini Publik Mengenai Demo Santri Trans7 pada Platform Tiktok Menggunakan Algoritma Naïve Bayes

Shafira Agustina Putri¹, Ilma Ma'rifatul Qur'ana²

^{1,2}Teknik Informatika, Fakultas Teknik dan Ilmu Komputer, Universitas Nusantara PGRI Kediri

E-mail: ¹shafiraagustina38@gmail.com, ²ilmamarifatulqurana@gmail.com

Abstrak – Analisis sentimen merupakan salah satu pendekatan dalam pengolahan bahasa alami yang digunakan untuk memahami opini atau persepsi pengguna terhadap suatu produk atau layanan. Penelitian ini bertujuan untuk menerapkan metode Naïve Bayes dalam mengklasifikasikan sentimen opini publik berbasis teks dengan memanfaatkan aplikasi berbasis web menggunakan framework Streamlit. Metode yang digunakan mengacu pada penelitian analisis sentimen ulasan pengguna aplikasi Mobile Legends di Google Play Store yang menggunakan algoritma Naïve Bayes Classifier. Tahapan penelitian meliputi pengumpulan data, preprocessing teks (case folding, cleaning, tokenisasi, stopword removal, dan stemming), ekstraksi fitur menggunakan TF-IDF, serta proses klasifikasi sentimen menjadi positif, negatif, dan netral. Model yang telah dilatih kemudian diimplementasikan ke dalam aplikasi Streamlit untuk melakukan analisis sentimen secara interaktif. Hasil penelitian menunjukkan bahwa metode Naïve Bayes mampu mengklasifikasikan sentimen dengan cukup baik dan dapat diimplementasikan dalam aplikasi berbasis web untuk membantu pengguna memahami kecenderungan opini publik secara cepat dan visual. Penelitian ini diharapkan dapat menjadi referensi pengembangan sistem analisis sentimen berbasis web di masa mendatang.

Kata Kunci — Analisis sentimen media sosial, Naïve bayes, Komentar tiktok.

1. PENDAHULUAN

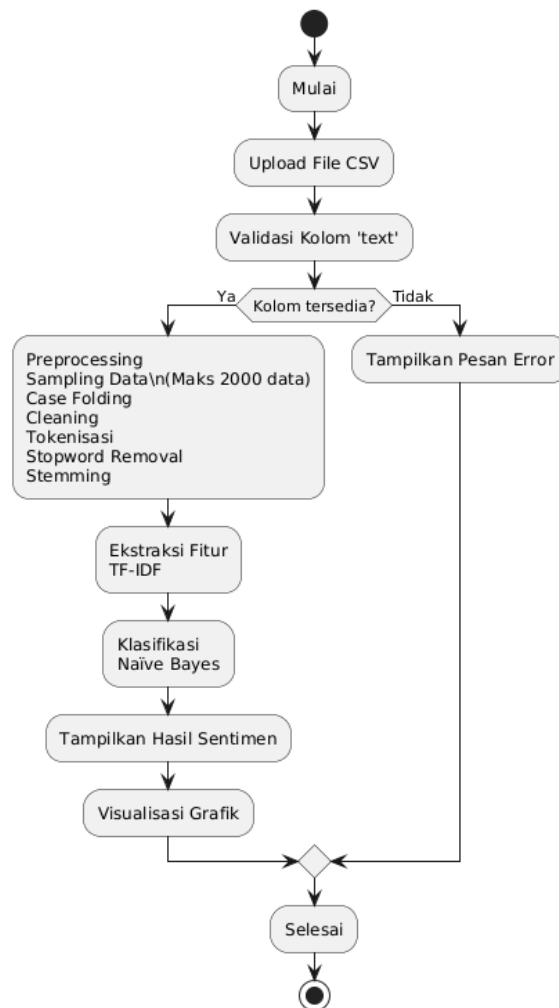
Analisis sentimen merupakan salah satu teknik dalam text mining dan natural language processing (NLP) yang digunakan untuk mengidentifikasi serta mengklasifikasikan opini atau kecenderungan sikap yang terkandung dalam data teks ke dalam kategori tertentu, seperti positif, negatif, dan netral [1]. Teknik ini berperan penting dalam memahami persepsi publik terhadap suatu produk, layanan, maupun isu tertentu berdasarkan teks yang dihasilkan oleh pengguna di lingkungan digital [2].

Pesatnya perkembangan teknologi informasi menyebabkan meningkatnya produksi data teks dalam jumlah yang sangat besar melalui berbagai platform digital, seperti media sosial, portal berita daring, forum diskusi, serta kolom ulasan aplikasi [3]. Dalam konteks berita dan isu terkini, opini masyarakat yang tersebar dalam bentuk komentar digital sering kali menjadi indikator utama dalam menilai kualitas layanan, tingkat kepuasan pengguna, hingga keberhasilan suatu produk digital. Namun, besarnya volume data tersebut menjadikan proses analisis secara manual tidak lagi efisien, sehingga diperlukan metode otomatis yang mampu mengolah dan menganalisis sentimen secara cepat dan sistematis [4].

Salah satu metode yang banyak diterapkan dalam analisis sentimen adalah Naïve Bayes Classifier, yaitu algoritma klasifikasi berbasis probabilistik yang memiliki konsep sederhana dan efisiensi komputasi yang tinggi [5]. Penelitian yang dilakukan oleh Kurniawan et al. menganalisis ulasan pengguna aplikasi Mobile Legends di Google Play Store menggunakan metode Naïve Bayes dan menunjukkan bahwa algoritma tersebut mampu mengklasifikasikan sentimen pengguna dengan tingkat akurasi yang cukup baik [6]. Penelitian tersebut juga menerapkan tahapan preprocessing seperti case folding, cleaning, tokenisasi, stopword removal, dan stemming, serta ekstraksi fitur menggunakan TF-IDF sebagai dasar pembentukan model klasifikasi.

Meskipun demikian, penelitian sebelumnya masih berfokus pada pengujian performa model secara eksperimental dan belum mengintegrasikan hasil analisis sentimen ke dalam sebuah sistem atau aplikasi yang dapat digunakan secara langsung oleh pengguna umum. Oleh karena itu, penelitian ini dikembangkan sebagai bentuk penyempurnaan dengan mengimplementasikan model Naïve Bayes ke dalam sebuah aplikasi berbasis web menggunakan Streamlit. Implementasi ini diharapkan mampu mengatasi keterbatasan penelitian terdahulu dengan menyediakan sistem analisis sentimen yang bersifat interaktif, mudah digunakan, serta relevan untuk menganalisis opini publik terhadap isu-isu terkini di era digital [7].

2. METODE PENELITIAN



Berikut penjelasan dari gambar diagram alur analisis sentimen.

- a. Upload File CSV
Proses memasukkan dataset ulasan pengguna ke dalam sistem. Dataset yang digunakan berbentuk file CSV yang berisi kumpulan teks ulasan.
- b. Validasi Kolom ‘text’
Untuk memastikan bahwa dataset memiliki kolom yang berisi teks ulasan.
- c. Sampling Data
Dilakukan untuk membatasi jumlah data yang dianalisis agar proses pengolahan lebih cepat, terutama pada tahap pelatihan dan pengujian model Naïve Bayes [8].
- d. Preprocessing
Tahapan utama dalam *text* preprocessing meliputi cleaning, case folding, tokenisasi, stopwords, dan stemming.
 - 1) Cleaning
Penghapusan karakter yang tidak diperlukan seperti tanda baca, angka, URL, dan simbol lainnya yang tidak berpengaruh terhadap analisis sentimen [9].
 - 2) Case Folding
Proses mengubah seluruh teks menjadi huruf kecil. Proses ini bertujuan untuk menyeragamkan bentuk teks sehingga kata dengan makna yang sama tidak dianggap berbeda akibat perbedaan huruf kapital [10].
 - 3) Tokenisasi
Proses memecah kalimat menjadi kata-kata tunggal.

- 4) Stopword
Untuk menghilangkan kata-kata umum yang tidak memiliki makna sentimen yang signifikan [11].
- 5) Stemming
Proses mengubah kata ke bentuk dasarnya agar variasi kata dapat direduksi dan meningkatkan akurasi klasifikasi.
- e. TF_IDF
Penelitian ini proses ekstraksi fitur dilakukan menggunakan Term Frequency-Inverse Document Frequency (TF-IDF) guna mengubah teks komentar netizen menjadi numerik berdasarkan kata yang sering muncul [12]. Pembobotan dilakukan dengan mengalikan frekuensi kemunculan kata dengan invers frekuensi dokumen.
- f. Klasifikasi Naïve Bayes
Sistem melakukan proses pengelompokan sentimen berdasarkan probabilitas kemunculan kata pada masing-masing kelas sentimen, seperti positif dan negatif [13]. Algoritma Naïve Bayes dipilih karena memiliki performa yang baik dalam pengolahan teks serta efisien dalam proses komputasi.

$$P(H | X) = \frac{P(X | H) P(H)}{P(X)}$$

Keterangan :

H = Hipotesis kelas sentimen (Positif, Negatif, atau Netral)

X = Data teks yang akan diklasifikasikan

$P(H | X)$ = Probabilitas data X termasuk ke dalam kelas H

$P(X | H)$ = Probabilitas kemunculan kata-kata pada data X terhadap kelas H

$P(H)$ = Probabilitas awal (prior) dari kelas H

$P(X)$ = Probabilitas data X

- g. Tampilan Hasil Sentimen
Menunjukkan kategori sentimen dari setiap ulasan.
- h. Visualisasi Grafik
Menyajikan hasil analisis sentimen dalam bentuk grafik agar lebih mudah dipahami oleh pengguna atau peneliti. Proses analisis diakhiri pada tahap Selesai, yang menandakan bahwa seluruh rangkaian analisis sentimen telah berhasil dijalankan.

3. HASIL DAN PEMBAHASAN

3.1 Dataset Komentar

Pengumpulan dataset menggunakan bantuan melalui website Epifi. Pengumpulan dataset tersebut berjumlah 2000 data komentar yang kemudian diexport ke dalam bentuk csv.

3.2 Preprocessing

Dari dataset tersebut dilakukan proses preprocessing data. Meliputi proses Cleaning, Case Folding, Tokenisasi, Stopword dan Stemming. Sehingga menghasilkan dataset yang bersih.

Berikut contoh dari dataset yang sudah dipreprocessing :

Tabel 1. Preprocessing

	Original Text	Cleaning & Case Folding	Tokenisasi	Stopword Removal
1428	LAH SLH TEMPAT ANJ	lah slh tempat anj	lah, slh, tempat, anj	lah, slh, tempat, anj
1531	👤:IQ itu apa? 😂😂👏	iq itu apa	iq, itu, apa	iq, apa

1465	ngakak sih ya 🙄 kan trans7 yang ngatain kok transmart yang di sholawat in 🙄🙄🙄🙄🙄	ngakak sih yakan trans yang ngatain kok transmart yang di sholawat in	ngakak, sih, yakan, trans, yang, ngatain, kok, transmart, yang, di, sholawat, in gak, bisa, bayanginnnnnnn, dunia, tanpa, anak, santriii, karna, anak, santri, ilmu, pegangan, kokoh, dunia, yang, akhir, zaman, iniiii	ngakak, sih, yakan, trans, ngatain, kok, transmart, kok, sholawat, in gak, bayanginnnnnnn, dunia, anak, santriii, karna, anak, santri, ilmu, pegangan, kokoh, dunia, akhir, zaman, iniiii
354	Gak bisa bayanginnnnnnn dunia tanpa anak santriii, karna anak santri ilmu pegangan kokoh dunia yang akhir zaman iniiii	gak bisa bayanginnnnnnn dunia tanpa anak santriii karna anak santri ilmu pegangan kokoh dunia yang akhir zaman iniiii	gak, bisa, bayanginnnnnnn, dunia, tanpa, anak, santriii, karna, anak, santri, ilmu, pegangan, kokoh, dunia, yang, akhir, zaman, iniiii	gak, bayanginnnnnnn, dunia, anak, santriii, karna, anak, santri, ilmu, pegangan, kokoh, dunia, akhir, zaman, iniiii
824	LUCU BANGET ANJER AWOAKWOK	lucu banget anjer awoakwok	lucu, banget, anjer, awoakwok	lucu, banget, anjer, awoakwok

3.2 Labelling Data

Selanjutnya untuk pelabelan data tidak dilakukan secara langsung. Dikarenakan proses pelabelan telah dilakukan sebelumnya pada tahap pelatihan model, dan hasil disimpan dalam bentuk file model_sentimen.pkl. Dengan demikian, data yang diunggah pengguna akan langsung diberi label sentimen oleh model yang telah dilatih.

No.	Text	Source / Video id	CreaTime	UserId
1	"Padahal BigBoss mereka Korupsi Kuota Haji... mereka pada diem bae 🙄🙄🙄"	7562126818099186951	2025-10- 17T11:00:33.000Z	amrio_amrio
2	"ibarat yg salah yanto, yg dilabrak yanti"	7562126818099186951	2025-10- 17T12:27:03.000Z	hamaniskuuuu
3	KAMPUNGNYA ITU LOH NGAMPUNG BANGET 🙄	7562126818099186951	2025-10- 18T01:21:50.000Z	alizarincleora
4	ngerugiin bgt anj	7562126818099186951	2025-10- 18T01:53:28.000Z	machtalalumu26
5	mereka ni pikirannya apasih 🙄	7562126818099186951	2025-10- 17T23:25:31.000Z	ccyb3rr0

3.3 Pembagian Dataset

Seluruh data yang telah diunggah pengguna digunakan sebagai data uji, karena model Naïve Bayes telah melalui proses pelatihan dan evaluasi sebelumnya. Sesuai dengan tujuan aplikasi, yaitu menyediakan analisis sentimen secara cepat dan interaktif.

3.4 Klasifikasi dan Evaluasi

Proses klasifikasi sentimen pada aplikasi dilakukan menggunakan algoritma Naïve Bayes yang sama dengan penelitian rujukan. Data teks yang telah melalui preprocessing diekstraksi menggunakan metode TF-IDF, kemudian diklasifikasikan ke dalam kelas sentimen positif, negatif, atau netral melalui fungsi model.predict() pada file app.py.

Dengan pendekatan ini, mekanisme klasifikasi tetap konsisten dengan artikel Mobile Legends, namun diimplementasikan dalam bentuk aplikasi berbasis web yang siap digunakan oleh pengguna umum. Berikut hasil masing-masing pengujian dapat dilihat pada gambar 3.1.

```
=== AKURASI ===
0.8079470198675497

=== CLASSIFICATION REPORT ===
```

	precision	recall	f1-score	support
negatif	0.78	0.90	0.84	136
netral	0.88	0.72	0.79	99
positif	0.79	0.75	0.77	67
accuracy			0.81	302
macro avg	0.82	0.79	0.80	302
weighted avg	0.81	0.81	0.81	302

Gambar 3. 1 Hasil Klasifikasi

Berikut penjelasan dari gambar hasil klasifikasi :

Dalam gambar diatas model akurasi memperoleh 81%, akan tetapi untuk hasil dari :

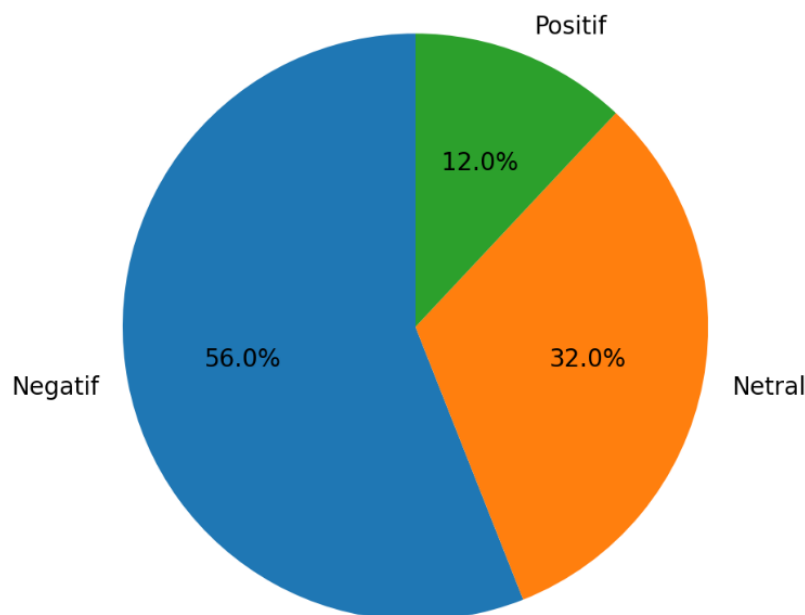
- Negatif: Model sangat baik dalam mengenali data negatif, terbukti dari nilai recall yang tinggi (90%). Artinya, sebagian besar data negatif berhasil dikenali dengan benar.
- Netral: Precision tinggi (88%) menunjukkan bahwa prediksi netral jarang keliru, namun recall-nya lebih rendah (72%), menandakan masih ada data netral yang tidak terdeteksi.
- Positif: Meskipun precision cukup baik (79%), nilai recall untuk kelas positif hanya 75%. Ini berarti sekitar 25% data yang sebenarnya positif tidak berhasil dikenali oleh model. Hal ini bisa berdampak pada analisis sentimen yang membutuhkan sensitivitas tinggi terhadap opini positif, misalnya dalam evaluasi kepuasan pelanggan atau ulasan produk.

Jadi hasil dari tabel ini menunjukkan performa yang solid secara umum, terutama dalam mendeteksi sentimen negatif.

3.5 Visualisasi

Pada artikel Mobile Legends, evaluasi model dilakukan secara kuantitatif menggunakan confusion matrix dan perhitungan metrik evaluasi, yang menunjukkan bahwa Naïve Bayes memiliki tingkat akurasi yang cukup baik, meskipun terdapat perbedaan performa antar kelas sentimen.

Sebaliknya, aplikasi app.py tidak menampilkan evaluasi model secara eksplisit. Evaluasi model diasumsikan telah dilakukan pada tahap pelatihan sebelumnya. Fokus aplikasi adalah pada penyajian hasil klasifikasi, yang ditampilkan dalam bentuk tabel sentimen, grafik distribusi sentimen, dan visualisasi frekuensi kata dominan. Pendekatan ini menekankan aspek kegunaan sistem (usability) dibandingkan pengujian performa model. Setelah melakukan proses klasifikasi selanjutnya melakukan visualisasi dengan tujuan agar melihat seberapa banyak sentimen positif, negative, dan netral dari hasil proses klasifikasi. Berikut gambar dari hasil visualisasi, dapat dilihat pada gambar 3.2 di bawah.



Gambar 3. 2 Visualisasi Sentimen

Dari gambar 3.2 ini menunjukkan hasil distribusi sentimen dalam bentuk pie chart, yang menggambarkan bahwa :

- Negatif:** 56.0% (≈ 169 data) pada tanda ini berwarna biru
- Netral:** 32.0% (≈ 97 data) untuk netral berwarna orange
- Positif:** 12.0% (≈ 36 data) dan untuk positif berwarna hijau

Distribusi ini menunjukkan bahwa data yang digunakan untuk pelatihan atau pengujian didominasi oleh sentimen negatif. Kelas positif memiliki jumlah data paling sedikit, hanya sekitar seperdelapan dari total.

4. KESIMPULAN

Aplikasi analisis sentimen opini publik berbasis web berhasil dikembangkan dengan menerapkan algoritma Naïve Bayes. Sistem mampu melakukan pengolahan data komentar TikTok secara otomatis, mulai dari preprocessing teks hingga klasifikasi sentimen ke dalam kategori positif, negatif, dan netral. Hasil analisis disajikan dalam bentuk tabel dan visualisasi grafik sehingga memudahkan pengguna dalam memahami kecenderungan opini publik secara cepat dan objektif.

5. SARAN

Pengembangan selanjutnya dapat dilakukan dengan menambahkan algoritma klasifikasi lain untuk meningkatkan akurasi hasil analisis. Selain itu, sistem dapat dikembangkan dengan menambah jumlah dataset, integrasi data secara real-time, serta penyempurnaan fitur visualisasi agar informasi yang disajikan menjadi lebih lengkap dan informatif.

DAFTAR PUSTAKA

- [1] M. Nahdhudin, N. C. H. Wibowo, M. R. Handayani, and K. Umam, “Klasifikasi Sentimen Masyarakat Terhadap Aplikasi Tiktok Menggunakan Algoritma Naive Bayes,” *Jurnal Teknologi informasi dan Ilmu Komputer*, vol. 1, no. 3, pp. 107–112, 2025, doi: 10.65258/jutekom.v1.i3.13.
- [2] A. Prihanum and D. Fadillah, “Analisis Sentimen Tanggapan Publik di Media Sosial Instagram Terhadap Program CSR ‘ESG Existence (EXIST)’ PT Telkom,” *Jurnal Audiens*, vol. 5, no. 4, pp. 565–580, 2024, doi: 10.18196/jas.v5i4.469.

-
- [6] “View of ANALISIS SENTIMEN BERDASARKAN ULASAN PENGGUNA APLIKASI MYPERTAMINA PADA GOOGLE PLAYSTO.pdf.”
- [7] D. S. Nurrochmah, N. Rahaningsih, R. D. Dana, and C. L. Rohmat, “Penerapan Algoritma Naive Bayes dalam Analisis Sentimen Ulasan Aplikasi KitaLulus di Google Play Store,” *Jurnal Informatika Terpadu*, vol. 11, no. 1, pp. 1–11, 2025, doi: 10.54914/jit.v11i1.1544.
- [8] T. Arifqi, N. Suarna, and W. Prihartono, “Penggunaan Naive Bayes Dalam Menganalisis Sentimen Ulasan Aplikasi Mcdonald’S Di Indonesia,” *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 2, pp. 1949–1956, 2024, doi: 10.36040/jati.v8i2.8740.
- [9] N. M. Damayanti, “Analisis Sentimen Publik Pada Tagar #Btscomeback Di Platform X Menggunakan Indobertweet,” *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 13, no. 3, 2025, doi: 10.23960/jitet.v13i3.7176.
- [10] M. A. Amirullah, Nadila Devi Tsabita, F. Rahmawati, and M. Afrizal, “Penggunaan Huruf Miring dan Kapital dalam Novel ‘Hati Suhita’ Karya Khilma Anis,” *Morfologi : Jurnal Ilmu Pendidikan, Bahasa, Sastra dan Budaya*, vol. 3, no. 4, pp. 225–237, 2025, doi: 10.61132/morfologi.v3i4.1972.
- [11] R. Rahmadani, A. Rahim, and R. Rudiman, “Analisis Sentimen Ulasan ‘Ojol the Game’ Di Google Play Store Menggunakan Algoritma Naive Bayes Dan Model Ekstraksi Fitur Tf-Idf Untuk Meningkatkan Kualitas Game,” *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 12, no. 3, 2024, doi: 10.23960/jitet.v12i3.4988.
- [12] M. H. Mahendra, D. T. Murdiansyah, and K. M. Lhaksmana, “Analisis Sentimen Tweet COVID-19 menggunakan K-Nearest Neighbors dengan TF-IDF dan Ekstraksi Fitur CountVectorizer,” *DIKE : Jurnal Ilmu Multidisiplin*, vol. 1, no. 2, pp. 37–43, 2023, doi: 10.69688/dike.v1i2.35.
- [13] A. Karimah, G. Dwilestari, and M. Mulyawan, “Analisis Sentimen Komentar Video Mobil Listrik Di Platform Youtube Dengan Metode Naive Bayes,” *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 1, pp. 767–737, 2024, doi: 10.36040/jati.v8i1.8373.
- [14] H. Sakdiyah, “Analisis sentimen customer review brand Kopi Kenangan menggunakan metode Naive Bayes,” *Fakultas Sains dan Teknologi*, 2023.