



## Analisis Sentimen Pengguna YouTube terhadap Game Mobile menggunakan Metode *Naïve Bayes*

Aura Rahayu Aksa Radiana<sup>1\*</sup>, Fathoni Mahardika<sup>2</sup>, Dani Indra Junaedi<sup>3</sup>

<sup>1-3</sup>Informatika, Universitas Sebelas April, Indonesia

\*Penulis Korespondensi: [220660121094@student.unsap.ac.id](mailto:220660121094@student.unsap.ac.id)

**Abstract.** *This study aims to develop a sentiment classification method for YouTube user comments related to the game Love and Deepspace using the Naïve Bayes algorithm, focusing on improving the text data processing and understanding user perceptions. Comment data were collected through scraping from YouTube videos, followed by preprocessing including text cleaning, normalization, stopword removal, stemming, and translation into English. Initial labeling was conducted using TextBlob, then the data were randomly sampled for training the Naïve Bayes model. Evaluation involved comparing sentiment distributions and visualization using Word Cloud and bar charts. The Naïve Bayes model achieved an accuracy of 77.36% in sentiment classification. The sentiment distribution shows differences between TextBlob (positive: 1,011, neutral: 1,312, negative: 575) and Naïve Bayes (positive: 901, neutral: 1,627, negative: 370), with Naïve Bayes being more conservative. The Word Cloud visualization identifies dominant words such as "bang," "game," and "main," while the bar chart shows the largest proportion of neutral sentiment. Naïve Bayes is effective for sentiment classification on informal comment data, with significant differences from rule-based methods like TextBlob. This research contributes to the development of text data processing techniques and user perception analysis, as well as opening up optimization opportunities with other algorithms like SVM for better accuracy.*

**Keywords:** *Love and Deepspace; Naïve Bayes; Sentiment Analysis; TextBlob; YouTube Comments.*

**Abstrak.** Penelitian ini bertujuan untuk mengembangkan metode klasifikasi sentimen pada komentar pengguna YouTube terkait game Love and Deepspace menggunakan algoritma Naïve Bayes, dengan fokus pada peningkatan proses pengolahan data teks dan pemahaman persepsi pengguna. Data komentar dikumpulkan melalui scraping dari video YouTube, diikuti pra-pemrosesan meliputi pembersihan teks, normalisasi, penghilangan stopwords, stemming, dan penerjemahan ke bahasa Inggris. Labeling awal dilakukan dengan TextBlob, kemudian data disampling secara acak untuk pelatihan model Naïve Bayes. Evaluasi melibatkan perbandingan distribusi sentimen dan visualisasi menggunakan Word Cloud serta diagram batang. Model Naïve Bayes mencapai akurasi 77,36% dalam klasifikasi sentimen. Distribusi sentimen menunjukkan perbedaan antara TextBlob (positif: 1.011, netral: 1.312, negatif: 575) dan Naïve Bayes (positif: 901, netral: 1.627, negatif: 370), dengan Naïve Bayes lebih konservatif. Visualisasi Word Cloud mengidentifikasi kata-kata dominan seperti "bang", "game", dan "main", sementara diagram batang memperlihatkan proporsi sentimen netral terbesar. Naïve Bayes efektif untuk klasifikasi sentimen pada data komentar informal, dengan perbedaan signifikan dari metode rule-based seperti TextBlob. Penelitian ini berkontribusi pada pengembangan teknik pengolahan data teks dan analisis persepsi pengguna, serta membuka peluang optimasi dengan algoritma lain seperti SVM untuk akurasi yang lebih baik.

**Kata kunci:** Analisis Sentimen; Bayes Naif; Cinta dan Antariksa; Komentar YouTube; TextBlob.

### 1. LATAR BELAKANG

Dalam era digital saat ini, platform media sosial seperti YouTube telah menjadi sumber utama bagi pengguna untuk berbagi pengalaman dan opini mereka, termasuk terhadap produk hiburan seperti game mobile (Yuna Burnama et al., 2024). Analisis sentimen terhadap komentar pengguna di YouTube dapat memberikan wawasan mendalam tentang persepsi publik, yang sangat relevan untuk industri game yang berkembang pesat. YouTube bukan hanya platform video terbesar di dunia dengan lebih dari 2 miliar pengguna aktif bulanan, tetapi

juga menjadi wadah bagi ulasan, *gameplay*, dan diskusi game yang autentik, di mana pengguna sering mengungkapkan sentimen mereka melalui komentar yang tidak terstruktur (Shaikh et al., 2023). Komentar pengguna mencerminkan reaksi spontan dan beragam dari komunitas global, yang berbeda dari survei terstruktur, sehingga memberikan data sentimen yang lebih organik dan *real-time*.

Love and Deepspace merupakan salah satu judul game populer dalam genre romansa dan interaksi sosial, yang dikembangkan oleh *Papergames* yang merupakan sebuah perusahaan yang berasal dari Tiongkok, dengan basis pemain yang luas terutama di kalangan remaja dan dewasa muda. Popularitasnya yang meningkat, ditandai oleh unduhan jutaan kali dan diskusi intens di *platform* digital, menjadikannya objek ideal untuk mengkaji tren sentimen terkini. Game mobile seperti Love and Deepspace sering kali bergantung pada umpan balik pengguna untuk iterasi pengembangan, di mana sentimen positif, negatif ataupun netral dapat memengaruhi retensi pemain, monetisasi, dan reputasi merek. Dengan memahami sentimen, pengembang dapat mengidentifikasi kekuatan (seperti mekanik *gameplay* yang menarik) dan kelemahan (seperti *bug* atau konten yang kurang menarik), sehingga meningkatkan kualitas produk (Danish et al., 2025).

Dalam analisis sentimen, berbagai metode telah dikembangkan, termasuk pendekatan berbasis aturan (*rule-based*), seperti menggunakan kamus sentimen (*lexicon-based*) seperti VADER atau SentiWordNet, yang mengandalkan skor kata-kata untuk menentukan polaritas. Metode lain yang umum adalah berbasis pembelajaran mesin (*machine learning*), seperti Support Vector Machine (SVM), Random Forest, atau Long Short-Term Memory (LSTM), yang melatih model pada dataset berlabel untuk klasifikasi sentimen. Selain itu, ada pendekatan hibrida yang menggabungkan *lexicon* dengan *machine learning* untuk meningkatkan akurasi. Di antara ini, metode Naïve Bayes dipilih karena efisiensinya dalam klasifikasi sentimen berbasis teks, dengan asumsi probabilistik yang sederhana dan cepat untuk diproses, terutama pada dataset besar seperti komentar YouTube, serta kemampuannya memberikan akurasi yang baik tanpa memerlukan sumber daya komputasi tinggi (P. R. Sari et al., 2024).

Pemilihan Naïve Bayes didukung oleh sejumlah penelitian terdahulu yang menunjukkan efektivitasnya, algoritma ini terbukti efisien untuk klasifikasi sentimen dalam skala data kecil dan relevan sebagai metode yang ringan dan transparan untuk memproses teks pendek, terutama dalam mendeteksi sentimen dominan (Loca & Abdullah, 2025). Selain itu, dalam studi komparasi, Naïve Bayes menunjukkan kinerja yang kompetitif dan efisiensi waktu, bahkan memperlihatkan akurasi yang lebih unggul dibandingkan metode lain pada skenario pembagian data uji yang lebih besar (Armaeni et al., 2024).

Secara keseluruhan, penelitian ini bertujuan untuk menganalisis sentimen pengguna YouTube terhadap game mobile Love and Deepspace menggunakan metode Naive Bayes, guna mengidentifikasi pola sentimen dominan, faktor pendorong, dan implikasinya bagi pengembangan game. Hasilnya diharapkan dapat memberikan rekomendasi praktis untuk pengembang game dan kontribusi akademik dalam bidang analisis sentimen berbasis teks digital, menjawab kebutuhan mendalam untuk memahami sentimen dalam industri game mobile yang kompetitif.

## 2. KAJIAN TEORITIS

Perkembangan teknologi informasi dan media sosial telah menghasilkan jumlah data teks yang sangat besar setiap harinya. Data tersebut berasal dari berbagai platform digital seperti media sosial, forum diskusi, ulasan pelanggan, maupun komentar pada platform video. Banyaknya data teks yang bersifat tidak terstruktur mendorong munculnya berbagai teknik dalam bidang *Natural Language Processing* (NLP) dan *Machine Learning* untuk mengolah serta memahami opini masyarakat secara otomatis. Salah satu teknik yang banyak digunakan adalah analisis sentimen, yaitu proses menganalisis teks digital untuk mengetahui kecenderungan emosi atau opini penulis, seperti positif, negatif, maupun netral. Analisis sentimen juga dikenal sebagai *opinion mining* karena bertujuan untuk memahami sikap, pandangan, dan emosi subjektif seseorang terhadap suatu topik tertentu (Arsi & Waluyo, 2021).

Popularitas analisis sentimen terus meningkat karena memiliki banyak manfaat praktis, seperti memantau reputasi merek, menganalisis tren opini masyarakat, hingga mendukung pengambilan keputusan pada bidang bisnis dan pemerintahan. Dalam penerapannya, analisis sentimen membutuhkan metode klasifikasi yang mampu mengolah data teks secara cepat dan efisien. Salah satu metode yang sering digunakan adalah Naive Bayes. Metode ini merupakan algoritma klasifikasi probabilistik dalam *Machine Learning* yang menggunakan Teorema Bayes untuk memprediksi probabilitas suatu kelas berdasarkan fitur input dengan asumsi bahwa setiap fitur bersifat independen satu sama lain.

Naive Bayes bekerja dengan menghitung probabilitas posterior berdasarkan data pelatihan, kemudian menentukan kelas dengan probabilitas tertinggi. Algoritma ini memiliki beberapa varian, seperti Gaussian Naive Bayes, Multinomial Naive Bayes, dan Bernoulli Naive Bayes. Proses dasarnya meliputi pelatihan model menggunakan data historis, perhitungan probabilitas prior dan likelihood, serta proses klasifikasi terhadap data baru (Hartono et al., 2023). Metode ini dikenal sederhana, cepat, dan efisien sehingga banyak digunakan dalam klasifikasi teks, filtering spam email, maupun analisis sentimen.

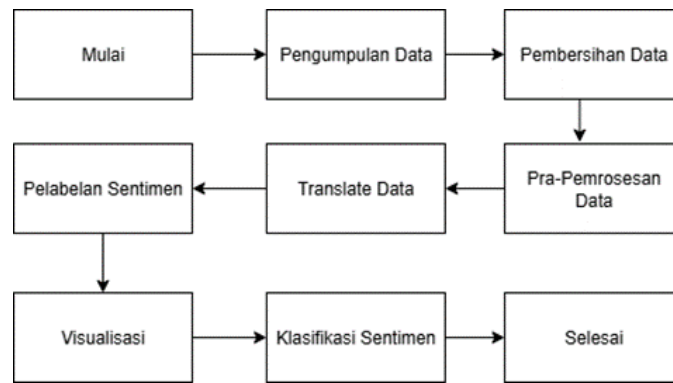
Secara historis, Naive Bayes berasal dari Teorema Bayes yang diperkenalkan oleh Thomas Bayes pada tahun 1763 dan kemudian dikembangkan oleh Pierre-Simon Laplace untuk kebutuhan inferensi statistik. Dalam perkembangan awal Machine Learning, metode ini telah digunakan pada sistem diagnosis medis dan klasifikasi teks sejak tahun 1960-an (Widiastuti et al., 2023). Pada klasifikasi teks, khususnya Multinomial Naive Bayes, proses klasifikasi dilakukan dengan menghitung frekuensi kemunculan kata pada setiap kelas. Metode ini efektif digunakan pada dataset kecil, namun memiliki kelemahan berupa masalah *zero-frequency*, yaitu probabilitas bernilai nol ketika terdapat kata baru yang tidak ditemukan pada data pelatihan. Selain itu, asumsi independensi antar fitur sering kali tidak sepenuhnya sesuai dengan kondisi nyata sehingga dapat memengaruhi tingkat akurasi ketika terdapat korelasi tinggi antar fitur.

Penerapan Naive Bayes dalam analisis sentimen telah banyak dilakukan pada berbagai penelitian terdahulu, salah satunya penelitian yang dilakukan oleh Nhadya Vita Loca dan Dedy Abdullah pada tahun 2025 dengan judul “Analisis Sentimen Terhadap Dedy Mulyadi Berdasarkan Komentar YouTube Menggunakan Metode Naïve Bayes”. Penelitian tersebut menggunakan metode Naive Bayes untuk menganalisis sentimen masyarakat terhadap Dedy Mulyadi berdasarkan komentar pada platform YouTube. Dataset yang digunakan terdiri dari 600 komentar sehingga termasuk dalam dataset berskala kecil. Penelitian tersebut menunjukkan bahwa metode Naive Bayes dapat digunakan secara efektif untuk mengklasifikasikan sentimen positif, negatif, maupun netral pada data komentar media sosial.

Bagian ini menguraikan teori-teori relevan yang mendasari topik penelitian dan memberikan ulasan tentang beberapa penelitian sebelumnya yang relevan dan memberikan acuan serta landasan bagi penelitian ini dilakukan. Jika ada hipotesis, bisa dinyatakan tidak tersurat dan tidak harus dalam kalimat tanya.

### 3. METODE PENELITIAN

Penelitian ini menggunakan pendekatan sistematis, yang terdiri dari beberapa tahapan kunci. Tujuan dari pendekatan ini adalah untuk melakukan analisis sentimen pada komentar salah satu live streaming video di akun YouTube Windah Basudara. Proses metodologi ini digambarkan dalam diagram alur metode penelitian.



**Gambar 1.** Tahapan - tahapan Penelitian

### **Pengumpulan Data**

Data komentar video YouTube dikumpulkan melalui *scraping* atau *API*, mencakup berbagai komentar pengguna dari satu atau beberapa video, kemudian disimpan dalam format terstruktur seperti *CSV* untuk memudahkan proses analisis selanjutnya pengumpulan data komentar (Yuyun Khanafiyah & Kartika Sari, 2025). Tidak menggunakan *API* resmi YouTube, pustaka python *youtube-comment-downloader* digunakan untuk mengumpulkan data komentar pada video review game di kanal YouTube Windah Basudara dengan judul "AKU INGIN MENCoba GAME TOP 1 SIMULASI DI PLAYSTORE INI... Love and Deepspace.". Setelah itu, setiap komentar yang diterima dinormalisasi menjadi dictionary yang mengandung informasi seperti ID komentar, isi teks, waktu komentar, dan data penulis, termasuk tautan profil dan gambar, jumlah suara dan balasan, serta metadata tambahan seperti penanda balasan dan waktu pengambilan data. Proses ini menawarkan metode yang mudah dan gratis untuk mengumpulkan komentar YouTube secara lokal tanpa batasan akses *API*, tetapi tidak mendukung pengambilan data dari *chat live* secara *real-time*.

### **Pembersihan Data**

Pada tahap ini, data dibersihkan untuk meningkatkan kualitas kumpulan data. Ini termasuk menghilangkan komentar duplikat dan komentar yang mengandung mention, spasi berlebih, hashtag, *URL*, dan karakter non-alfanumerik. Setiap komentar diubah menjadi huruf kecil untuk konsistensi dan kemudahan analisis. Sebuah studi *preprocessing* data teks telah menyarankan bahwa tahap pembersihan ini sangat penting untuk menghilangkan suara dan memperbaiki kualitas data input.

### **Pra-pemrosesan Data**

Pra-pemrosesan data adalah tahap penting dalam analisis sentimen, dan tujuan dari tahap ini adalah untuk menyiapkan data teks untuk analisis yang lebih akurat dan efisien (Hakim, 2021). Berikut ini langkah-langkah pra-pemrosesan data yang dilakukan. Normalisasi

teks adalah proses menyamakan format teks agar seragam, misalnya mengubah semua huruf menjadi huruf kecil dan menghilangkan variasi penulisan yang tidak konsisten untuk meningkatkan konsistensi data analisis (Aufar et al., 2023).

*Stopwords removal* adalah tahap menghapus kata-kata yang tidak memiliki makna signifikan seperti kata sambung dan kata fungsi, yang berfungsi mengurangi noise sehingga model fokus pada kata bermakna (Audina et al., 2025). Tokenisasi berarti memecah teks atau kalimat menjadi unit-unit kecil berupa kata atau token individual yang memudahkan analisis model terhadap tiap bagian teks. Ini penting agar mesin dapat memahami unit dasar teks untuk dianalisis satu per satu (Setiawan et al., 2025). *Stemming* mengubah kata ke bentuk dasarnya dengan menghilangkan imbuhan agar mengurangi variasi kata yang berhubungan dan meningkatkan performa pemodelan (Pardede & Darmawan, 2025). Secara keseluruhan, tahap pra-pemrosesan sangat penting untuk meningkatkan kualitas dan konsistensi data teks sebelum melakukan analisis sentimen. Ini membuat hasil yang dihasilkan lebih akurat dan signifikan (Laia et al., 2025).

### **Pra-pemrosesan Data**

Proses penerjemahan data teks dilakukan dengan menggunakan pustaka *deep\_translator* dalam bahasa pemrograman *python* (N. K. R. Sari et al., 2023). Data teks yang kosong atau tidak valid terlebih dahulu dibersihkan dengan mengganti nilai kosong menjadi string kosong. Fungsi *safe\_translate* dibuat untuk menerjemahkan setiap kalimat dari bahasa sumber ke bahasa Inggris secara otomatis, dengan penanganan kesalahan agar jika terjadi kegagalan penerjemahan, teks asli tetap dikembalikan. Selanjutnya, setiap baris data teks diterjemahkan secara berurutan dengan penambahan jeda sebanyak 2 detik setiap 50 baris untuk menghindari batasan kecepatan (*rate limit*) API penerjemah.

### **Pelabelan Data (*Labeling*)**

Pelabelan data dilakukan dengan menggunakan pustaka *TextBlob* yang berbasis *lexicon* untuk mendapatkan nilai *polarity* dan *subjectivity* dari teks komentar berbahasa Inggris yang sudah disiapkan, kemudian diklasifikasikan menjadi sentimen positif, negatif, dan netral (Putri et al., 2024). Setiap nilai kosong pada kolom ini diisi dengan string kosong agar proses analisis tidak mengalami *error*. Selanjutnya, setiap komentar diolah secara bergantian untuk mengukur nilai polaritas sentimen yang dapat bernilai positif, netral (nol), atau negatif. Berdasarkan nilai polaritas yang diperoleh, komentar tersebut diklasifikasikan ke dalam tiga kategori yaitu "Positif" jika polaritas lebih besar dari nol, "Netral" jika polaritas sama dengan nol, dan "Negatif" jika polaritas kurang dari nol. Kategori sentimen ini kemudian disimpan dalam sebuah daftar status untuk setiap komentar sebagai hasil labeling. Selain itu, jumlah total

komentar dan distribusi dari masing-masing kategori sentimen juga dihitung dan dilaporkan. Proses ini memungkinkan pemberian label sentimen otomatis untuk dataset teks yang telah diterjemahkan dan dipersiapkan untuk analisis lebih lanjut.

### **Visualisasi Data**

Visualisasi data membantu menyederhanakan informasi kompleks, mengungkap pola dan tren tersembunyi, serta memfasilitasi pengambilan keputusan yang lebih cepat dan akurat. Visualisasi juga meningkatkan keterlibatan audiens serta mendukung *storytelling* data yang efektif. Informasi yang disajikan secara visual cenderung lebih mudah diingat dibandingkan dengan teks atau angka saja, sehingga meningkatkan retensi informasi. Visualisasi data teks akan dilakukan melalui dua metode utama untuk mendukung analisis sentimen (Nadia et al., 2024). Pertama, Penggunaan *Word Cloud* dalam analisis sentimen telah terbukti sebagai metode yang efektif untuk menggali persepsi publik. *Word Cloud*, yang merupakan representasi visual dari frekuensi kata dalam sebuah teks, memungkinkan para peneliti untuk dengan cepat mengidentifikasi tema dan topik utama dalam data teks yang besar. Dalam konteks analisis sentimen, *Word Cloud* membantu dalam memahami penyebaran dan frekuensi kata-kata yang bermuatan emosional positif atau negatif dalam percakapan publik (Agusia et al., n.d.). Kedua, distribusi sentimen hasil analisis dikomunikasikan melalui diagram batang (*bar chart*) yang dibuat dengan menggunakan pustaka *matplotlib* dan *seaborn*. Visualisasi menggunakan diagram batang digunakan untuk menampilkan total keseluruhan dari setiap kelas sentimen (positif, negatif, netral) dalam dataset sehingga memudahkan pemahaman tentang bagaimana sentimen tersebar dan memberikan gambaran tren sentimen secara kuantitatif dan visual (Silaban & Gumay, 2025). Kombinasi kedua visualisasi ini memberikan gambaran komprehensif baik dari aspek konten kata yang paling dominan dalam komentar maupun alokasi komentar berdasarkan sentimen yang dianalisis.

### **Klasifikasi Sentimen**

Setelah dilakukan proses pelabelan data secara otomatis menggunakan pustaka *TextBlob*, hasil pelabelan tersebut digunakan sebagai data latih untuk melatih model *Naïve Bayes*. Metode *Naive Bayes Classifier* merupakan salah satu metode yang terbaik untuk pelatihan domain-domain dan hasil klasifikasinya memiliki tingkat akurasi yang tinggi. Semakin banyak data yang diujikan maka tingkat akurasi dari hasil klasifikasi akan semakin tinggi. Hal ini juga dipengaruhi oleh jumlah data latih yang digunakan pada metode *Naive Bayes Classifier* (Suryani et al., 2019). Model *Naïve bayes* dilatih dengan menggunakan data latih yang telah diberi label, lalu diuji menggunakan data uji untuk mengukur performa

klasifikasi yang dilakukan. Algoritma Naïve Bayes menghitung probabilitas suatu teks termasuk ke dalam kategori sentimen tertentu menggunakan rumus:

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

Dengan  $C$  adalah kelas sentimen dan  $X$  adalah fitur kata pada teks (Ardiani et al., 2020). Perhitungan likelihood fitur menggunakan asumsi independensi antar fitur, sehingga probabilitas total adalah hasil perkalian probabilitas fitur individu. Evaluasi model dilakukan dengan mengukur akurasi, yaitu persentase prediksi yang benar terhadap total data uji, dengan rumus:

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} \dots\dots\dots(ii)$$

Di mana  $TP$ ,  $TN$ ,  $FP$ , dan  $FN$  adalah jumlah true positive, true negative, false positive, dan false negative pada hasil klasifikasi (Paabanan Simanjuntak et al., 2024). Pengujian model menunjukkan bahwa metode ini efektif dalam mengklasifikasikan sentimen teks dengan akurasi yang cukup baik, sehingga dapat digunakan sebagai alat bantu dalam memahami opini publik secara otomatis. Bagian ini memuat rancangan penelitian meliputi disain penelitian, populasi/ sampel penelitian, teknik dan instrumen pengumpulan data, alat analisis data, dan model penelitian yang digunakan. Metode yang sudah umum tidak perlu dituliskan secara rinci, tetapi cukup merujuk ke referensi acuan (misalnya: rumus uji-F, uji-t, dll). Pengujian validitas dan reliabilitas instrumen penelitian tidak perlu dituliskan secara rinci, tetapi cukup dengan mengungkapkan hasil pengujian dan interpretasinya. Keterangan simbol pada model dituliskan dalam kalimat.

#### 4. HASIL DAN PEMBAHASAN

Data komentar yang dikumpulkan terdiri dari total 3437 baris data. Data tersebut menunjukkan adanya variasi waktu posting komentar yang tersebar dalam jangka waktu beberapa bulan hingga beberapa hari terakhir.

	text	time	scraped_at
0	Thank You Girl! Btw aku juga main game horor ...	8 bulan yang lalu (diedit)	2025-10-19T04:21:35.485362Z
1	Giw sebagai penyuka game lads yang ga bisa inst...	5 hari yang lalu	2025-10-19T04:21:36.344067Z
2	Wah, jujur ga baru tau bang indah pernah main...	6 hari yang lalu	2025-10-19T04:21:37.211235Z
3	gelesah banget latin pas topup 🤔 🤔 🤔	6 hari yang lalu	2025-10-19T04:21:38.077374Z
4	terimakasih bangg udh main game ini ❤️ respect ...	7 hari yang lalu	2025-10-19T04:21:38.943948Z
--	--	--	--
3432	Bang aku cinta kamu	8 bulan yang lalu	2025-10-19T05:14:08.207631Z
3433	pertama	8 bulan yang lalu	2025-10-19T05:14:09.169040Z
3434	fort	8 bulan yang lalu	2025-10-19T05:14:10.042240Z
3435	Sama-sama dreinaaaa	8 bulan yang lalu	2025-10-19T05:14:11.283417Z
3436	hallow dreina	8 bulan yang lalu	2025-10-19T05:14:12.294366Z

3437 rows × 3 columns

**Gambar 2.** Data mentah

Data komentar berhasil diekstraksi dan direpresentasikan dalam sebuah DataFrame yang terdiri dari tiga kolom utama. Kolom *text* berisi isi komentar dari pengguna yang beragam mulai dari apresiasi hingga ekspresi perasaan terkait topik yang dibahas. Dataset ini juga memuat kolom waktu relatif komentar (*time*), yang menunjukkan seberapa lama komentar tersebut diposting, misalnya "8 bulan yang lalu" atau "5 hari yang lalu". Selain itu, kolom *scraped\_at* mencatat waktu pengambilan data secara otomatis dalam format *timestamp ISO 8601* yang menggunakan zona waktu *UTC*, sehingga memberikan konteks temporal yang akurat untuk analisis. Hasil ini menggambarkan bahwa data yang diperoleh cukup kaya dan memadai untuk dianalisis lebih lanjut dalam konteks analisis sentimen, tren komunikasi pengguna, atau penelitian perilaku interaksi di platform digital. Pendekatan ini juga memastikan bahwa dataset yang digunakan memiliki atribut lengkap serta metadata waktu yang dapat meningkatkan kedalaman analisis.

Proses pembersihan teks dilakukan dengan menggunakan fungsi khusus yang menghilangkan elemen-elemen tidak relevan seperti mention pengguna, *hashtag*, tautan *URL*, dan karakter khusus yang kerap muncul pada komentar media sosial. Fungsi ini menerapkan pola regular expression untuk menghapus username yang diawali simbol '@', *hashtag*, serta seluruh tautan yang berformat *http* atau *https*. Selain itu, fungsi juga menghapus semua karakter selain huruf, angka, dan spasi, lalu menyederhanakan spasi ganda menjadi satu spasi saja serta menghilangkan spasi berlebih di awal dan akhir teks.

	text	time	scraped_at
0	thank you girls btw aku juga main game horor r...	8 bulan yang lalu (diedit)	2025-10-19T04:21:35.485362Z
1	gw sebagai penyuka game lads yang ga bisa inst...	5 hari yang lalu	2025-10-19T04:21:36.344067Z
2	wah jujur gw baru tau bang windah pernah main ...	6 hari yang lalu	2025-10-19T04:21:37.211235Z
3	gelisah banget liatin pas topup	6 hari yang lalu	2025-10-19T04:21:38.077374Z
4	terimakasih bangg udh main game ini respect s...	7 hari yang lalu	2025-10-19T04:21:38.943949Z
...	...	...	...
3432	bang aku cinta kamu	8 bulan yang lalu	2025-10-19T05:14:08.207631Z
3433	pertama	8 bulan yang lalu	2025-10-19T05:14:09.169040Z
3434	fisrt	8 bulan yang lalu	2025-10-19T05:14:10.042240Z

**Gambar 3.** Tampilan Data Bersih

Setelah pembersihan, kolom teks dalam dataset hanya berisi komentar yang sudah bersih dan lebih relevan untuk analisis lebih lanjut seperti analisis sentimen atau pemodelan topik. Contoh teks yang bersih meliputi kalimat seperti "thank you girls btw aku juga main game horor romantis" dan "gw sebagai penyuka game lads yang ga bisa *install*", serta kalimat pendek seperti "bang aku cinta kamu". Dengan demikian, hasil ini menunjukkan bahwa data komentar yang sudah dibersihkan dan terstruktur siap untuk tahap analisis lanjutan tanpa gangguan dari *noise* teks yang biasa ditemukan pada platform media sosial. Proses normalisasi teks dilakukan untuk menyamakan variasi penggunaan bahasa gaul dan singkatan yang umum ditemukan dalam komentar pengguna agar data menjadi lebih konsisten dan mudah dianalisis. Hal ini dicapai dengan mendefinisikan sebuah kamus pengganti (dictionary) yang memetakan kosakata tidak baku seperti "yg" menjadi "yang". Fungsi normalisasi kemudian menerapkan penggantian ini secara literatif pada setiap teks komentar dalam dataset.

	text	time	scraped_at
0	thank you girls btw aku juga main game horor r...	8 bulan yang lalu (diedit)	2025-10-19T04:21:35.485362Z
1	gw sebagai penyuka game lads yang tidak bisa i...	5 hari yang lalu	2025-10-19T04:21:36.344067Z
2	wah jujur aku baru tau bang windah pernah main...	6 hari yang lalu	2025-10-19T04:21:37.211235Z
3	gelisah banget liatin pas topup	6 hari yang lalu	2025-10-19T04:21:38.077374Z
4	terimakasih bangg sudah main game ini respect...	7 hari yang lalu	2025-10-19T04:21:38.943949Z
...	...	...	...
3432	bang aku cinta kamu	8 bulan yang lalu	2025-10-19T05:14:08.207631Z
3433	pertama	8 bulan yang lalu	2025-10-19T05:14:09.169040Z
3434	fisrt	8 bulan yang lalu	2025-10-19T05:14:10.042240Z
3435	samasama dreinaaaa	8 bulan yang lalu	2025-10-19T05:14:11.283417Z
3436	hallow dreina	8 bulan yang lalu	2025-10-19T05:14:12.294366Z

2898 rows x 3 columns

**Gambar 4.** Data yang sudah di normalisasi



```

0      [thank, you, girls, btw, aku, main, game, horo...
1      [gw, penyuka, game, lads, tidak, install, karn...
2      [wah, jujur, aku, baru, tau, bang, windah, per...
3      [gelisah, banget, liatin, pas, topup]
4      [terimakasih, bangg, main, game, inii, respect...
...
3432      [bang, aku, cinta, kamu]
3433      [pertama]
3434      [fisrt]
3435      [samasama, dreinaaaa]
3436      [hallow, dreina]
Name: text, Length: 2898, dtype: object

```

**Gambar 6.** Data hasil Tokenisasi

Hasil tokenisasi menghasilkan data dalam bentuk daftar token untuk setiap baris komentar, memungkinkan analisis kata secara lebih detail pada tiap unit linguistik. Misalnya, komentar “*thank you girls btw aku main game horror...*” diubah menjadi token-token seperti [*thank, you, girls, btw, aku, main, game, horror...*], begitu pula dengan komentar-komentar lainnya yang telah diurai menjadi kumpulan kata-kata terpisah. Tokenisasi ini merupakan langkah penting dalam pengolahan bahasa alami (*Natural Language Processing*) karena memudahkan analisis lebih lanjut seperti pembuatan model teks, analisis frekuensi kata, dan pengembangan algoritma klasifikasi. Dengan data dalam bentuk token, proses analisis dapat dilakukan pada tingkat kata yang lebih granular, meningkatkan akurasi dan interpretabilitas hasil penelitian. Proses *stemming* dilakukan untuk mengubah kata-kata dalam teks komentar menjadi bentuk dasarnya atau akar kata, sehingga variasi kata yang memiliki makna sama namun berbentuk berbeda dapat distandarisi. Pada penelitian ini, *stemming* menggunakan pustaka Sastrawi yang khusus dirancang untuk bahasa Indonesia. Fungsi *stemming* diterapkan pada setiap token hasil tokenisasi dengan memanfaatkan objek *stemmer* yang secara iteratif memproses setiap kata dalam komentar.

```

thank you girls btw aku main game horor romantis hantu ganteng judul gamenya homicipherin
scord paling rame indonesia buat cari temen festival motionime fest coming soon 2025
gw suka game lads tidak install karna batas simpan hp sungguh sangat berterimakasih bang
wah jujur aku baru tau bang windah pernah main lads
gelisah banget liatin pas topup
terimakasih bangg main game ini respect sama bg windahh bikin part 2 dong banggg
kalian jijik ama game cw ngeliat lu sebaliknyangeri bngt org gk pernah ngaca
gue cowo ken main karna gameplay grafik

```

**Gambar 7.** Data hasil proses Stemming

Hasil *stemming* menunjukkan bahwa kata-kata dalam komentar berubah menjadi bentuk yang lebih sederhana dan konsisten, misalnya kata kerja berimbuhan, kata jamak, dan variasi kata lain dikonversi ke bentuk dasar. Contoh keluaran *stemming* memperlihatkan kumpulan kata-kata yang telah distem seperti “*thank you girl btw aku main game horor*”

romantis hantu ganteng judul game", "gw suka game lads tidak install karna batas simpan hp", hingga "live opening nya sampe 3x wkwwk". Langkah ini sangat penting untuk mengurangi redundansi kata dan meningkatkan keakuratan analisis teks selanjutnya, seperti pemodelan topik dan analisis sentimen, karena kata yang memiliki makna sama tetapi berbentuk berbeda tidak diperlakukan sebagai entitas berbeda. Dengan demikian, *stemming* membantu menyempurnakan preparasi data teks dalam penelitian ini. Dalam penelitian ini, data komentar awal yang sebagian besar menggunakan bahasa Indonesia dan dialek informal terlebih dahulu diterjemahkan ke dalam bahasa Inggris menggunakan pustaka *GoogleTranslator* dari modul *deep\_translator*. Proses terjemahan ini dilakukan secara otomatis dengan penanganan error untuk menjaga keberlanjutan proses tanpa berhenti saat terjadi kesalahan.

Terjemahan ke bahasa Inggris dilakukan dengan tujuan untuk meningkatkan akurasi pelabelan dan klasifikasi sentimen. Bahasa Inggris dipilih karena sebagian besar alat analisis sentimen dan model pembelajaran mesin yang terdapat dalam pustaka populer memiliki performa lebih optimal pada teks berbahasa Inggris dibandingkan bahasa lokal atau informal. Dengan menerjemahkan data ke bahasa Inggris terlebih dahulu, model sentimen dapat memanfaatkan kamus, leksikon, dan algoritma yang lebih matang sehingga hasil klasifikasi menjadi lebih valid dan representatif.

Unnamed: 0	text	text_english
0	0 thank you girls btw aku main game horor romant...	thank you girls btw I'm playing a romantic hor...
1	1 gw suka game lads tidak install karna batas si...	I like the game, lads, I didn't install it bec...
2	2 wah jujur aku baru tau bang windah pernah main...	Wow, to be honest, I just found out that Winda...
3	3 gelisah banget liatin pas topup	I'm really nervous looking at it when I top it up
4	4 terimakasih bangg main game ini respect sama b...	Thank you, proud to play this game. Respect th...

**Gambar 8.** DataFrame setelah proses penerjemahan

Hasil terjemahan disimpan dalam kolom baru bernama *text\_english*, yang memuat versi terjemahan dari komentar asli. Contoh data menunjukkan bahwa teks komentar informal dan campuran bahasa berhasil dikonversi menjadi kalimat yang lebih baku dan mudah dipahami, misalnya dari "gw suka game lads tidak install karna batas simpan hp" menjadi "*I like the game, lads, I didn't install it because of the storage limit on my phone.*" Pendekatan ini memperkuat kualitas data sebelum analisis sentimen sehingga memungkinkan penelitian untuk memperoleh hasil yang lebih akurat dan bermakna dalam memahami persepsi pengguna terhadap produk atau layanan yang dikaji. Proses ini merupakan tahap pelabelan sentimen otomatis terhadap Data yang sebelumnya telah diterjemahkan ke dalam bahasa Inggris. Dengan menggunakan analisis polaritas dari pustaka *TextBlob*, setiap komentar diberi label sentimen berdasarkan

nilai polaritasnya: komentar dengan polaritas positif diberi label "Positif", polaritas nol diberi label "Netral", dan polaritas negatif diberi label "Negatif".

Hasil Analisis Data:

Positif = 1011

Netral = 1312

Negatif = 575

Total Data : 2898

**Gambar 9.** Hasil labeling Textblob

Hasil pelabelan menunjukkan terdapat 1.011 komentar berlabel Positif, 1.312 komentar berlabel Netral, dan 575 komentar berlabel Negatif dalam dataset. Label sentimen ini kemudian ditambahkan ke dalam dataset pada kolom baru klasifikasi, sehingga tiap baris data kini memuat informasi teks asli, terjemahan dalam bahasa Inggris, serta kategori sentimen yang melekat.

Unnamed: 0	text	text_english	Klasifikasi
0	thank you girls btw aku main game horor romantL	thank you girls btw i'm playing a romantic hor...	Positif
1	ge suka game lads tidak install karna batas sl	I like the game. lads. I didn't install it bec...	Negatif
2	wah jujur aku baru tau bang windah pernah mainL	Wow, to be honest, I just found out that Wind...	Positif
3	geliish benget latin pas topup	I'm really nervous looking at it when I top it up	Positif
4	terimakasih bang main game ini respect sama bL	Thank you. proud to play this game. respect th...	Positif
...	...	...	...
2893	bang aku cinta kamu	bro I love you	Positif
2894	pertama	First	Positif
2895	flirt	flirt	Netral
2896	samarata dreinaaaa	yellow dreinaaaa	Netral
2897	hallow dreina	hallow dreina	Netral

2898 rows x 4 columns

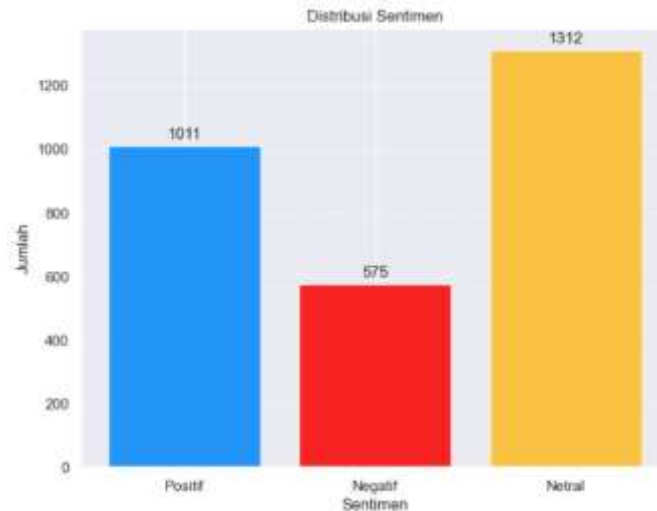
**Gambar 10.** Data Hasil labeling TextBlob

Contoh data memperlihatkan bahwa komentar beragam dari ucapan pujian, kritik, hingga ungkapan netral telah berhasil diberi label yang sesuai, mendukung proses klasifikasi yang konsisten. Data berlabel ini dapat digunakan selanjutnya untuk pelatihan model pembelajaran mesin atau analisis statistik guna memahami distribusi dan pola sentimen dalam dataset.

Dengan kata lain, proses ini hanya melakukan pelabelan sentimen otomatis menggunakan model *rule-based*, tanpa melakukan pelatihan atau evaluasi model, sehingga berfungsi sebagai dasar awal untuk analisis sentimen lebih lanjut.

*Wordcloud* dalam analisis sentimen adalah representasi visual yang menampilkan frekuensi kata dalam sebuah teks secara mudah dipahami, di mana kata yang sering muncul





**Gambar 12.** Diagram batang hasil labeling TextBlob

Selanjutnya, distribusi sentimen komentar dikaji secara kuantitatif dengan menggunakan diagram batang. Dari total komentar yang dianalisis, mayoritas berada dalam kategori netral (1.312 komentar), diikuti oleh komentar positif (1.011 komentar), sementara komentar negatif merupakan jumlah terkecil (575 komentar). Penyajian data tersebut menggunakan kode warna berbeda memudahkan pemahaman perbandingan volume sentimen positif, negatif, dan netral secara visual. Label jumlah komentar yang akurat di atas masing-masing batang memperkuat kejelasan informasi. Secara keseluruhan, kedua visualisasi ini saling melengkapi untuk mendukung interpretasi hasil penelitian, dimana *Word Cloud* menggali tema dan kosa kata dominan, dan grafik batang menunjukkan proporsi sentimen yang melekat pada komentar pengguna. Pendekatan ini membantu memberikan insight mendalam tentang persepsi dan respons audiens terhadap konten yang dibahas, sehingga memperkaya analisis sentimen dan perilaku pengguna secara menyeluruh.

Pada tahap persiapan data untuk analisis lebih lanjut, dataset yang telah berisi label sentimen, dipisahkan menjadi tiga kelompok berdasarkan kategori sentimen yaitu Positif, Negatif, dan Netral. Proses pemisahan ini bertujuan untuk mengelompokkan data secara eksplisit sehingga memudahkan pengolahan data yang seimbang. Selanjutnya, dilakukan pengambilan sampel acak sebesar 50% dari masing-masing kelompok menggunakan metode *random sampling* tanpa pengulangan. Pengambilan sampel ini bertujuan untuk menyeimbangkan jumlah data dari tiap kategori sehingga tidak terjadi ketidakseimbangan (*imbalanced data*) yang dapat mempengaruhi kinerja model analisis selanjutnya.

```
from textblob.classifiers import NaiveBayesClassifier

cl = NaiveBayesClassifier(train_set)
print("Akurasi Test:", cl.accuracy(dataset))
```

Akurasi Test: 0.7736369910282954

### Gambar 13. Hasil Akurasi Naïve Bayes

Pada tahap pemodelan, dilakukan pelatihan model klasifikasi sentimen menggunakan algoritma Naive Bayes yang diterapkan melalui pustaka TextBlob. Data pelatihan (*train\_set*) yang telah diseimbangkan berdasarkan kategori sentimen Positif, Negatif, dan Netral digunakan untuk melatih model tersebut. Model kemudian diuji dengan seluruh dataset asli untuk mengukur akurasi klasifikasi. Hasil evaluasi menunjukkan akurasi model yang diperoleh mencapai nilai tertentu (nilai akurasi dicetak pada output print, yang menunjukkan tingkat keberhasilan model dalam mengklasifikasikan sentimen komentar berdasarkan data yang diberikan).

Hasil yang diperoleh menunjukkan akurasi sebesar 0.7736 atau sekitar 77,36%, yang berarti model dapat mengklasifikasikan sentimen dengan benar pada sekitar 77 dari 100 contoh komentar. Nilai ini menunjukkan kinerja model yang cukup baik dalam mengenali sentimen berdasarkan teks komentar yang diberikan. Secara singkat, proses ini mengilustrasikan bahwa Naive Bayes merupakan metode klasifikasi yang efektif dan sederhana dalam konteks analisis sentimen, dengan tingkat akurasi yang layak sebagai dasar untuk analisis lebih lanjut atau pengembangan model yang lebih kompleks. Klasifikasi sentimen terhadap data yang telah diterjemahkan ke dalam bahasa Inggris menggunakan model klasifikasi Naïve Bayes yang sebelumnya sudah dilatih. Proses klasifikasi dilakukan dengan menggunakan fungsi `classify()` dari pustaka TextBlob yang menerapkan model `cl` pada setiap komentar secara berurutan. Setiap komentar dikategorikan ke dalam salah satu dari tiga label sentimen: Positif, Netral, atau Negatif berdasarkan hasil klasifikasi model. Selanjutnya, jumlah komentar pada masing-masing kategori dihitung untuk memberikan gambaran distribusi sentimen dalam dataset.

Hasil Analisis Data:

Positif = 901

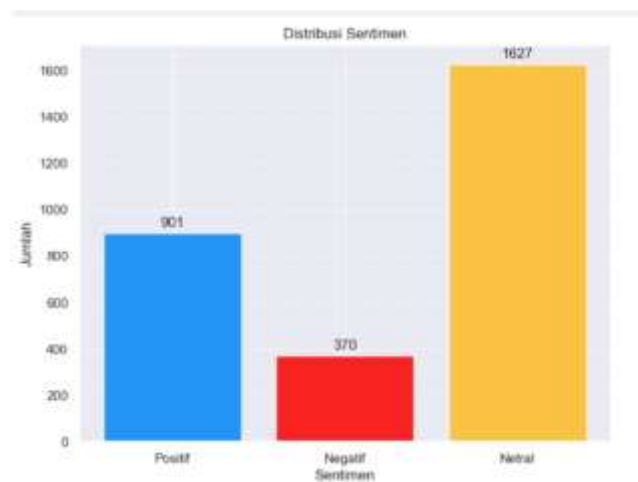
Netral = 1627

Negatif = 370

Total Data : 2898

**Gambar 14.** Hasil Labeling menggunakan Naïve Bayes

Hasil klasifikasi menunjukkan bahwa dari total 2.898 komentar, sebanyak 901 komentar (31,1%) diklasifikasikan sebagai Positif, 1.627 komentar (56,2%) sebagai Netral, dan 370 komentar (12,8%) sebagai Negatif. Distribusi ini mengindikasikan bahwa sebagian besar komentar memiliki sentimen Netral, sementara komentar positif juga cukup signifikan, dan komentar negatif merupakan bagian terkecil dari keseluruhan data. Hasil ini memperkuat pemahaman mengenai persepsi dan sikap pengguna terhadap topik yang dianalisis, serta menunjukkan efektivitas model Naive Bayes dalam mengklasifikasikan sentimen komentar secara otomatis.



**Gambar 15.** Diagram Batang hasil labeling Naïve Bayes

Grafik ini memperlihatkan bahwa sebagian besar komentar tergolong sentimen netral, diikuti oleh komentar positif dengan jumlah yang cukup signifikan, sementara komentar negatif berjumlah paling sedikit. Hal ini mengindikasikan bahwa mayoritas respon pengguna bersikap netral terhadap topik yang dianalisis, dengan proporsi sentimen positif lebih banyak dibanding negatif. Selain itu, grafik juga dilengkapi dengan grid horizontal berupa garis putus-putus yang memudahkan pembacaan dan perbandingan jumlah komentar antar kategori. Label sumbu X dan Y serta judul "Distribusi Sentimen" juga memperjelas konteks data yang

divisualisasikan. Hasil analisis sentimen dari dua metode berbeda, yakni pendekatan rule-based TextBlob dan model klasifikasi Naive Bayes, menunjukkan perbedaan distribusi dan klasifikasi yang mencolok pada dataset komentar yang sama. Labeling menggunakan TextBlob cenderung memberikan label sentimen positif atau negatif lebih banyak dibandingkan dengan Naive Bayes, yang lebih konservatif dan cenderung mengklasifikasikan komentar ke kategori netral. Pada level data, TextBlob mengklasifikasikan 1.011 komentar sebagai positif dan 575 sebagai negatif, dengan sekitar 1.312 komentar netral. Di sisi lain, Naive Bayes menghasilkan jumlah komentar netral yang jauh lebih banyak, yakni 1.627, sementara komentar positif berkurang menjadi 901 dan komentar negatif berkurang menjadi 370. Perbedaan ini disebabkan oleh cara kerja kedua metode. TextBlob menggunakan analisis polaritas langsung berdasarkan kamus dan aturan linguistik, sehingga cenderung lebih sensitif terhadap kata-kata bermuatan positif atau negatif. Sebaliknya, Naive Bayes mempelajari pola dari data pelatihan yang telah diberikan label, sehingga dapat mengenali konteks lebih kompleks namun mungkin lebih berhati-hati dalam mengkategorikan sentimen ekstrem.

```
Text: aku download butuh waktu 2 jam loading nya 3 jam
Classifier: Netral
Classifier Bayes: Positif

Text: emangg zayne kek sunghoon enhyphen cakep pol mennn
Classifier: Negatif
Classifier Bayes: Netral

Text: 22514 lucu banget ekspresi nya bang windahhh23223 ahahaha lagi
Classifier: Positif
Classifier Bayes: Netral
```

### Gambar 16. Perbedaan Hasil Labeling

Perbedaan klasifikasi ini menegaskan bahwa TextBlob sebagai metode berbasis aturan lebih responsif terhadap kata bermuatan sentimen secara eksplisit, sedangkan Naive Bayes sebagai model pembelajaran mesin mengandalkan konteks dan pola yang dipelajari dari data pelatihan, sehingga menghasilkan labeling yang lebih berhati-hati dan proporsi sentimen netral lebih besar. Dengan demikian, dalam pemilihan metode analisis sentimen, penting mempertimbangkan karakteristik data dan tujuan penelitian, serta melakukan validasi menyeluruh terhadap hasil untuk mendapatkan pemahaman yang lebih komprehensif mengenai pola sentimen yang ada di dalam dataset.

## 5. KESIMPULAN DAN SARAN

Penelitian ini menegaskan bahwa penerapan algoritma Naïve Bayes mampu memberikan klasifikasi sentimen yang efektif pada data komentar pengguna YouTube terkait game mobile. Penelitian ini berhasil meningkatkan proses pengolahan data teks dengan melalui tahap pra-pemrosesan, penerjemahan, dan pelabelan sentimen, yang kemudian diolah menggunakan model Naïve Bayes dengan hasil akurasi yang cukup baik. Temuan penelitian menunjukkan adanya perbedaan yang signifikan antara metode rule-based (TextBlob) dan model pembelajaran mesin (Naïve Bayes) dalam hal sensitivitas terhadap konteks dan distribusi sentimen, dimana Naïve Bayes cenderung memberikan hasil yang lebih konservatif dan proporsi sentimen netral yang lebih besar.

Penelitian ini memberikan langkah konkret dalam pengembangan metode klasifikasi sentimen berbasis machine learning yang dapat dioptimalkan lebih lanjut, misalnya dengan menambahkan dan membandingkan algoritma lain seperti Support Vector Machine (SVM). Hal ini penting untuk meningkatkan akurasi dan reliabilitas sistem analisis sentimen, yang pada gilirannya dapat membantu pengambil keputusan dalam industri game mobile maupun platform digital lainnya dalam memahami opini dan preferensi pengguna dengan lebih akurat. Dengan pendekatan yang sistematis ini, penelitian membuka peluang bagi pengembangan aplikasi analisis sentimen yang lebih canggih dan aplikatif dalam berbagai konteks industri yang bergantung pada feedback pengguna secara digital. Kesimpulan ini menggarisbawahi bahwa optimasi metode machine learning khususnya dalam klasifikasi sentimen merupakan area yang esensial untuk pengembangan penelitian lebih lanjut, sehingga dapat mendukung peningkatan teknologi pengolahan bahasa alami dan analisis opini yang semakin akurat dan dapat dipercaya.

## DAFTAR REFERENSI

- Agusia, P., Uli, M., Manurung, A., Calista, V., & Mawardi, V. C. (n.d.). Pemanfaatan *word cloud* pada analisis sentimen dalam menggali persepsi publik. *Seminar Nasional Corisindo*.
- Ardiani, L., Sujaini, H., & Tursina, T. (2020). Implementasi *sentiment analysis* tanggapan masyarakat terhadap pembangunan di Kota Pontianak. *Jurnal Sistem dan Teknologi Informasi (JUSTIN)*, 8(2), 183. <https://doi.org/10.26418/justin.v8i2.36776>
- Armaeni, P. P., Wiguna, I. K. A. G., & Parwita, W. G. S. (2024). Sentiment analysis of YouTube comments on the closure of TikTok Shop using Naïve Bayes and Decision Tree method comparison. *Jurnal Galaksi*, 1(2), 70–80. <https://doi.org/10.70103/galaksi.v1i2.15>

- Audina, D., Purnamasari, A. I., Bahtiar, A., & Tohidi, E. (2025). Peningkatan model klasifikasi sentimen pengguna aplikasi Tomoro Coffee menggunakan algoritma Naive Bayes. *Jurnal Informatika & Rekayasa Elektronika*, 8(1). <http://e-journal.stmiklombok.ac.id/index.php/jire>
- Aufar, A. F., Rosid, M. A., Eviyanti, A., & Astutik, I. R. I. (2023). Optimizing text preprocessing for accurate sentiment analysis on e-wallet reviews. *JICTE (Journal of Information and Computer Technology Education)*, 7(2), 42–50. <https://doi.org/10.21070/jicte.v7i2.1650>
- Danish, A., Amin, B. M., Toh, M. Z., Bhuiyan, M. M., Nafis, N. M., & Kamarudin, S. (2025). Sentiment analysis on YouTube comments using machine learning techniques based on video games content.
- Hakim, B. (2021). Analisa sentimen data *text preprocessing* pada data mining dengan menggunakan *machine learning*. *JBASE: Journal of Business and Audit Information Systems*, 4(2). <https://doi.org/10.30813/jbase.v4i2.3000>
- Laia, N. A., Barus, S. P., Kunci, K., & Sentimen, A. (2025). Analisis sentimen pengguna YouTube pada video berjudul “10 Tahun Jokowi Jadi Presiden.” *JIKA*, 9(2).
- Loca, N. V., & Abdullah, D. (2025). Analisis sentimen terhadap Dedy Mulyadi berdasarkan komentar YouTube menggunakan metode Naïve Bayes. *Progresif: Jurnal Ilmiah Komputer*, 21(2), 874. <https://doi.org/10.35889/progresif.v21i2.3102>
- Nadia, P. H., Wahyuni, S. S., & Sari, A. P. (2024). Penerapan Naive Bayes Classifier untuk analisis sentimen ulasan pelanggan pada Frenz Accessories Handphone. *JAMASTIKA*, 3.
- Paaban Simanjuntak, S., Sandino Berutu, S., & Setyawan, G. C. (2024). Implementasi metode CNN pada klasifikasi sentimen terhadap pelaksanaan Piala Dunia U-17. *Journal of Engineering and Emerging Technology*, 2(1). [www.jeet.unram.ac.id](http://www.jeet.unram.ac.id)
- Pardede, J., & Darmawan, D. (2025). Perbandingan algoritma stemming Porter, Sastrawi, Idris, dan Arifin & Setiono pada dokumen teks bahasa Indonesia. *Jurnal Teknologi Informasi dan Ilmu Komputer*, 12(1), 69–76. <https://doi.org/10.25126/jtiik.2025128860>
- Putri, G. R., Maulana, M. A., & Bahri, S. (2024). Perbandingan algoritma Naïve Bayes dan TextBlob untuk mendapatkan analisis sentimen masyarakat pada sosial media. *Teknika*, 13(2), 213–218. <https://doi.org/10.34148/teknika.v13i2.815>
- Sari, N. K. R., Suarjaya, I. M. A. D., & Buana, P. W. (2023). Perbandingan *translation library* pada Phyton: Studi kasus analisis sentimen penyakit menular di Indonesia. <https://doi.org/10.24843/JTRTI.2021.v02.i03.p21>
- Sari, P. R., Indah, D. R., Rasywir, E., Firdaus, M. A., & Athalina, G. (2024). Comparison of Naive Bayes and SVM algorithms for sentiment analysis of PUBG Mobile on Google Play Store. *Sistemasi: Jurnal Sistem Informasi*. <http://sistemasi.ftik.unisi.ac.id>
- Setiawan, K., Apriyanto, K. J., & Karya Informatika, C. (2025). Pengembangan model analisis sentimen berbasis Naive Bayes terhadap isu kepemimpinan Presiden Prabowo di TikTok. *Journal of Information Technology and Computer Science (INTECOMS)*, 8(4).
- Shaikh, A. R., Alhoori, H., & Sun, M. (2023). YouTube and science: Models for research impact. *Scientometrics*, 128(2), 933–955. <https://doi.org/10.1007/s11192-022-04574-5>

- Silaban, N., & Gumay, M. G. (2025). Analisis sentimen pengguna aplikasi pinjaman online melalui media sosial X (Twitter) menggunakan metode Support Vector Machine. *Jurnal Minfo Polgan*, 14(2), 2116–2130. <https://doi.org/10.33395/jmp.v14i2.15300>
- Suryani, P. S. M., Linawati, L., & Saputra, K. O. (2019). Penggunaan metode Naïve Bayes Classifier pada analisis sentimen Facebook berbahasa Indonesia. *Majalah Ilmiah Teknologi Elektro*, 18(1), 145. <https://doi.org/10.24843/mite.2019.v18i01.p22>
- Yuna Burnama, Z., Rosid, M. A., & Azizah, N. L. (2024). Analisis sentimen pada komentar YouTube dalam turnamen MPL Season 13 dengan metode *ensemble machine learning*.
- Yuyun Khanafiyah, S. T., & Sari, D. K. (2025). Analisis sentimen komentar YouTube kanal Dirty Vote menggunakan metode Naive Bayes Classifier. *Agustus*, 12(4).