



SENTIMENT ANALYSIS OF YOUTUBE COMMENTS ON THE VINA CIREBON CASE USING GBM AND LOGISTIC REGRESSION

Asro¹, John Chaidir², Patah Herwanto³

^{1,2}Fakultas Ekonomi dan Bisnis, Universitas Primagraha, Kota Serang, Banten 42111, Indonesia
³Fakultas Teknologi Informasi, STMIK IM, Bandung Indonesia

*Email: ¹6110122023@upg.ac.id, ²johnchaidir@primagraha.ac.id, ³pherwanto@stmik-im.ac.id

Genesis : (Article received: 3 Desember 2024; Revision: 8 Desember 2024; published: 17 Desember 2024)

Abstract

This research aims to analyze public sentiment regarding the Vina Cirebon case on YouTube by using Gradient Boosting Machine (GBM) and Logistic Regression methods. The study leverages comments from YouTube to evaluate public opinion through precision, recall, F1-score, and accuracy metrics derived from confusion matrices. The results indicate that Logistic Regression outperforms GBM in terms of accuracy (88.29% vs. 81.48%), precision (0.854 vs. 0.739), recall (0.602 vs. 0.352), and F1-score (0.706 vs. 0.476). This suggests that Logistic Regression is more effective in capturing the nuances of public sentiment on this issue. The analysis highlights the predominance of negative sentiments, reflecting widespread public disapproval and concern. This study provides valuable insights into the public's reaction to the Vina Cirebon case, demonstrating the efficacy of sentiment analysis in understanding social issues through social media data.

Keywords: sentiment analysis, Vina Cirebon case, YouTube comments, GBM, Logistic Regression, public opinion

ANALISIS SENTIMEN KOMENTAR YOUTUBE TENTANG KASUS VINA CIREBON MENGGUNAKAN GBM DAN REGRESI LOGISTIK

Abstrak

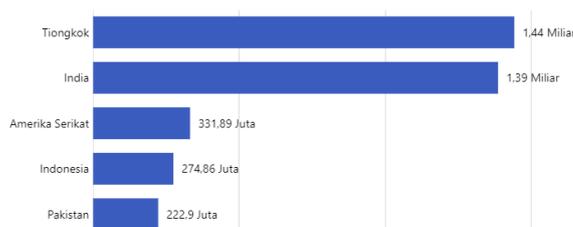
Penelitian ini bertujuan untuk menganalisis sentimen publik terhadap kasus Vina Cirebon di YouTube dengan menggunakan metode *Gradient Boosting Machine* (GBM) dan Regresi Logistik. Studi ini memanfaatkan komentar-komentar di YouTube untuk mengevaluasi opini publik melalui metrik precision, recall, F1-score, dan akurasi yang diperoleh dari confusion matrix. Hasil penelitian menunjukkan bahwa Regresi Logistik mengungguli GBM dalam hal akurasi (88,29% vs. 81,48%), precision (0,854 vs. 0,739), recall (0,602 vs. 0,352), dan F1-score (0,706 vs. 0,476). Hal ini mengindikasikan bahwa Regresi Logistik lebih efektif dalam menangkap nuansa sentimen publik terkait isu ini. Analisis menunjukkan dominasi sentimen negatif, mencerminkan ketidaksetujuan dan keprihatinan publik yang luas. Penelitian ini memberikan wawasan berharga tentang reaksi publik terhadap kasus Vina Cirebon, menunjukkan efektivitas analisis sentimen dalam memahami isu sosial melalui data media sosial.

Kata Kunci: analisis sentimen, kasus Vina Cirebon, komentar YouTube, GBM, Regresi Logistik, opini publik

1. PENDAHULUAN

Indonesia, sebagai negara dengan jumlah penduduk terbesar keempat di dunia, memainkan peran penting dalam dinamika sosial dan politik global. Menurut DataBox Katadata dan Kusumawati, Indonesia memiliki populasi sekitar 274 juta jiwa. Pada tahun 2021, jumlah pengguna aktif media sosial mencapai 170 juta [1], menjadikannya negara dengan penduduk terbanyak keempat setelah Tiongkok, India, dan Amerika Serikat. Platform media sosial seperti YouTube, Facebook, dan Twitter telah menjadi sarana utama bagi publik untuk mengekspresikan opini, berbagi informasi, dan berinteraksi satu sama lain.

Namun, platform ini juga menjadi tempat penyebaran misinformasi dan hoaks, yang dapat mempengaruhi persepsi publik secara signifikan.



Gambar 1. Indoensia Peringkat ke 4 di Dunia

Kasus Vina Cirebon adalah salah satu isu sosial yang telah menarik perhatian publik secara luas dan

menghasilkan diskusi yang meluas di platform media sosial. Memahami sentimen publik seputar kasus ini sangat penting, karena mencerminkan reaksi dan opini masyarakat yang lebih luas yang dapat mempengaruhi dinamika sosial. Metode Gradient Boosting Machine (GBM) *XGBoost*, *short for eXtreme Gradient Boosting*, has become a top ensemble learning algorithm, enhancing traditional gradient boosting with various innovative features.[2] dikenal karena kemampuannya dalam meningkatkan akurasi prediksi melalui penguatan bertahap dari model-model dasar yang lemah. Di sisi lain, Regresi Logistik Logistic regression adalah metode yang sering digunakan dalam menganalisis data dengan satu atau lebih variabel prediktor yang menggambarkan variabel respon. [3] merupakan metode yang sederhana namun efektif untuk klasifikasi biner dan sering digunakan dalam berbagai aplikasi analisis data karena kemampuannya dalam menangani dataset yang besar dan kompleks. Data komentar YouTube dikumpulkan dan dianalisis secara bulanan dari Januari hingga Juni 2024, menghasilkan 12,150 komentar.

Distribusi komentar per bulan adalah sebagai berikut:

Tujuan utama dari penelitian ini adalah untuk:

- 1) Menganalisis sentimen publik terhadap kasus Vina Cirebon berdasarkan komentar-komentar di YouTube.
- 2) Membandingkan performa metode GBM dan Regresi Logistik dalam klasifikasi sentimen.
- 3) Memberikan wawasan tentang pola sentimen yang dominan serta implikasinya terhadap persepsi publik dan kebijakan.

Penelitian ini memiliki beberapa manfaat

1.1 Tinjauan Pustaka

Penelitian sebelumnya yang terkait dengan analisis sentimen telah menunjukkan berbagai pendekatan dan metode yang digunakan untuk memahami opini publik. Naive Bayes, Support Vector Machine (SVM), dan Logistic Regression adalah beberapa metode yang sering digunakan. Naive Bayes dikenal karena kesederhanaannya dan efektivitasnya dalam banyak aplikasi, sementara SVM dan Logistic Regression digunakan karena kemampuannya dalam menangani dataset yang kompleks dan memberikan hasil yang akurat.

The XGBoost algorithm is a prominent boosting technique. It is a widely recognized supervised machine learning (ML) algorithm[4]. Dan juga metode pembelajaran mesin yang telah terbukti

efektif dalam meningkatkan akurasi prediksi melalui penguatan bertahap dari model-model dasar yang lemah. Penelitian telah menunjukkan bahwa GBM dapat memberikan hasil yang lebih baik dibandingkan dengan metode lain dalam berbagai aplikasi, termasuk analisis sentimen. Namun, dalam konteks analisis sentimen, kombinasi dan perbandingan metode yang berbeda sering kali memberikan wawasan yang lebih komprehensif tentang data dan model yang paling efektif. *The Logistic Regression method is the appropriate choice for analyzing sentiment* [5], di sisi lain, merupakan metode yang sederhana namun sangat efektif untuk klasifikasi biner. Metode ini sering digunakan dalam berbagai aplikasi analisis data karena kemampuannya dalam menangani dataset yang besar dan kompleks. Regresi Logistik telah digunakan secara luas dalam analisis sentimen untuk mengklasifikasikan opini publik berdasarkan teks yang dihasilkan dari berbagai platform media sosial. Dengan semakin banyaknya pengguna internet yang menggunakan media sosial, dikembangkanlah sebuah sistem yang dapat menganalisis opini terkait suatu topik atau bahasan, yang dikenal sebagai opinion mining [6]. Dan kemudian Hasil akhir penelitian ini menunjukkan persentase sentimen positif dan negatif dari ulasan komentar pengguna yang diperoleh melalui metode regresi logistic dan GBM [7].

2. METODE PENELITIAN

Penelitian ini bertujuan untuk menganalisis sentimen publik terhadap kasus Vina Cirebon di YouTube dengan menggunakan dua metode pembelajaran mesin, yaitu Gradient Boosting Machine (GBM) dan Regresi Logistik. Proses penelitian melibatkan beberapa tahapan utama, mulai dari pengumpulan data, pra-pemrosesan data, penerapan model, hingga evaluasi hasil. Setiap tahapan dirancang untuk memastikan bahwa data yang digunakan berkualitas tinggi dan model yang diterapkan memberikan hasil yang akurat dan dapat diandalkan.

2.1 Tahapan Metode Penelitian

1. Pengumpulan Data

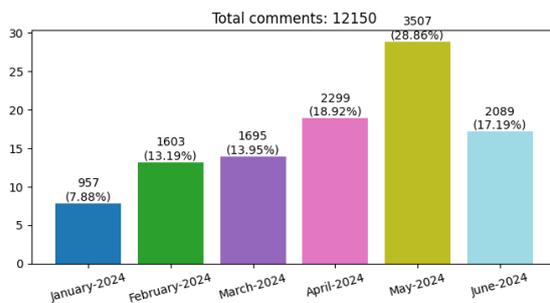
Data yang digunakan dalam penelitian ini berupa komentar-komentar yang diambil dari YouTube terkait kasus Vina Cirebon. Komentar-komentar ini dikumpulkan selama periode Januari hingga Juni 2024. Proses pengumpulan data dilakukan menggunakan teknik scraping yang diimplementasikan melalui Python dan Google Client

API di dalam software Visual Studio untuk mendapatkan komentar-komentar yang relevan dari berbagai video yang membahas kasus Vina Cirebon. Total komentar yang dikumpulkan berjumlah 12,150. Dan Berikut adalah hasil dari proses crawling data:

Tabel 1. Hasil proses crawling data

Tahun	Bulan	Total	Persen
2024	January	957	7.88%
	February	1603	13.19%
	March	1695	13.95%
	April	2299	18.92%
	May	3507	28.86%
	June	2089	17.19%
Total		12,150	100%

Berikut adalah Grafik berikut menunjukkan distribusi jumlah komentar yang dikumpulkan setiap bulan:



Gambar 2. Hasil Setiap Bulan

Grafik ini menunjukkan bahwa dari total 12,150 komentar yang dikumpulkan, sebagian besar komentar berasal dari bulan Mei 2024 (28.86%), diikuti oleh April (18.92%) dan Juni (17.19%). Ini menunjukkan bahwa minat publik terhadap topik kasus Vina Cirebon meningkat seiring waktu.

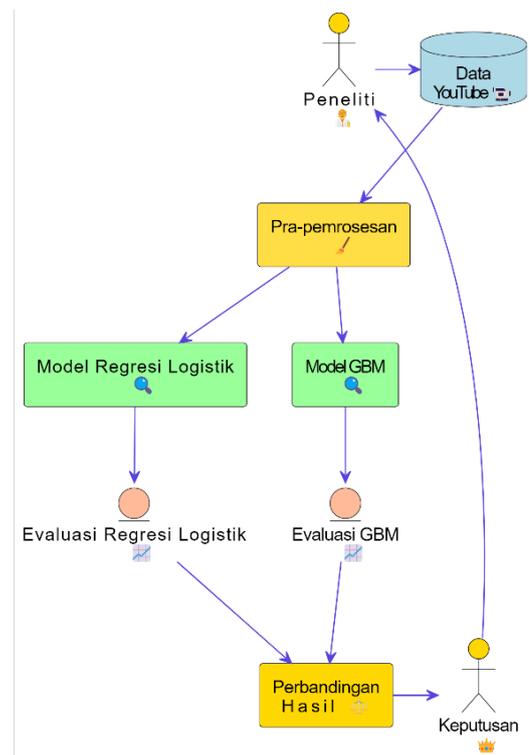
2. Pra-pemrosesan Data

Pra-pemrosesan data merupakan langkah penting untuk memastikan kualitas data yang digunakan dalam analisis sentimen. Tahapan pra-pemrosesan meliputi:

- **Penghapusan Tanda Baca dan Angka:** Tanda baca dan angka dihapus untuk menghindari gangguan dalam analisis teks.
- **Lowercasing:** Merubah huruf text dataset menjadi huruf kecil semua [8].
- **Penghapusan Stop Words:** Kata-kata umum yang tidak memiliki makna khusus dalam analisis dihapus.
- **Stemming:** Kata-kata diubah menjadi bentuk dasar mereka untuk memudahkan analisis.
- **Tokenisasi:** Teks dipecah menjadi token-token yang lebih kecil untuk analisis yang lebih mendalam.

3. Pemisahan Data

Setelah pra-pemrosesan, data dibagi menjadi dua bagian: data latih dan data uji dengan rasio 80:20 untuk memastikan model dapat dievaluasi dengan data yang belum pernah dilihat sebelumnya. *The stages of this research were carried out in 9 stages* [9]. tahapan. Berikut adalah gambar flowchart yang menggambarkan tahapan-tahapan penelitian:



0Gambar 3. Metode Penelitian

4. Penerapan Model

Dua model pembelajaran mesin diterapkan pada data yang telah diproses:

- **Gradient Boosting Machine (GBM):** GBM adalah teknik ensemble yang membangun model prediksi dalam bentuk urutan model-model yang lebih lemah, biasanya decision trees. Model-model ini dibangun secara bertahap, dan masing-masing model baru mengoreksi kesalahan dari model sebelumnya. *The predictive model is built by combining M additive tree models $(f_0, f_1, f_2, \dots, f_M)$ to predict the outcomes (Equation 1)[10].*

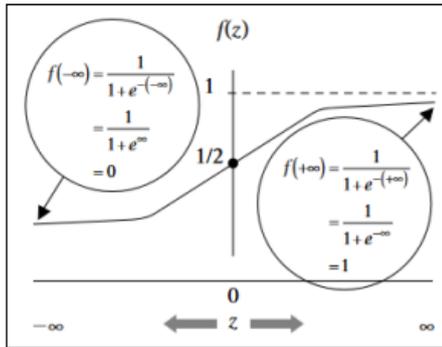
$$f(x) = \sum_{m=0}^M f_m(x)$$

Model ensemble pohon ini dioptimalkan dengan mengurangi kesalahan generalisasi yang diharapkan, L , sesuai dengan Persamaan 2:

$$L = \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

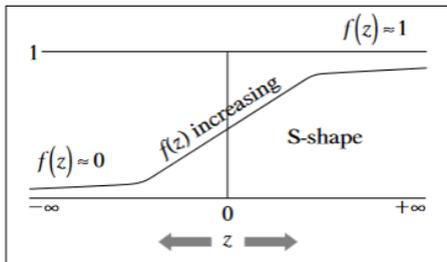
▪ **Regresi Logistik:**

Regresi Logistik populer digunakan karena hasil dari fungsi logistik $f(z)$ berkisar antara 0 dan 1. Model ini dirancang untuk menjelaskan probabilitas yang selalu berada di antara nilai 0 dan 1. Salah satu contohnya adalah probabilitas seseorang terjangkit suatu penyakit (David G. Kleinbaum & Mitchel Klein, 1994, p.5-6). Pada "Gambar 1. Fungsi Logistik" di bawah, dapat dilihat bahwa hasil yang diberikan berkisar antara 0 dan 1, meskipun nilai masukan untuk variabel z bervariasi (David G. Kleinbaum & Mitchel Klein, 1994, p.5-6)[11]



Gambar 4. Logistic Function

Sumber Gambar: David G. Kleinbaum and Mitchel Klein, Logistic Regression: A Self-Learning Text, 2nd Edition, (Springer, New York, NY, (1994)), h.6.



Gambar 5. Slope Logistic Function

Sumber Gambar: David G. Kleinbaum and Mitchel Klein, Logistic Regression: A Self-Learning Text, 2nd Edition, (Springer, New York, NY, (1994)), h.6.

5. **Evaluasi Model**

Untuk mengevaluasi kinerja model, digunakan beberapa metrik evaluasi yang diperoleh dari confusion matrix:

- **Precision:** Mengukur proporsi prediksi positif yang benar dari semua prediksi positif. Rumus untuk precision ditunjukkan dalam persamaan berikut [12].

$$Precision = \frac{TP}{TP + FP}$$

- **Recall:** Mengukur proporsi prediksi positif yang benar dari semua data yang benar-benar positif. Berikut adalah rumus Recall.

$$Recal = \frac{TP}{TP + FN}$$

- **F1-score:** Harmonic mean dari precision dan recall, memberikan gambaran kinerja model secara keseluruhan. Berikut adalah Rumus F1-score.

$$F1 - score = 2 \times \left(\frac{Precision \times Recall}{Precision + Recall} \right)$$

- **Akurasi:** Mengukur proporsi prediksi yang benar dari keseluruhan data. Rumus untuk menghitung accuracy ditunjukkan dalam persamaan berikut:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

6. **Visualisasi Hasil**

Hasil analisis dan evaluasi model divisualisasikan menggunakan berbagai alat visualisasi seperti matplotlib dan seaborn untuk mempermudah interpretasi hasil. Grafik distribusi sentimen dan metrik kinerja model disajikan untuk memberikan wawasan yang lebih jelas tentang efektivitas masing-masing model dalam menangani analisis sentimen.

3. **HASIL DAN PEMBAHASAN**

3.1 **Hasil Pra-Pemrosesan Data**

Hasil pra-pemrosesan data terdiri dari beberapa tahapan: case folding, pembersihan teks, penghapusan stop words, normalisasi teks, dan stemming. Berikut adalah hasil dari setiap tahapan:

1) **Case Folding**

Tahapan ini mengubah semua huruf menjadi huruf kecil untuk konsistensi.

```
comment(Sebelum) \
Akibat pergaulan bebas ya tuailah hasilnya
Kata2 nya seperti cerita asli lagi ngelihat 😊😄
Sebelum para pelaku mati mereka tdk akan perna...
Kisah mengenaskn \nkn jg orang tua vina ngk m...
Bpngkar&kawal smpe tuntas kasus ini..Tangkap 3...
```

```
Case_Folding(Sesudah)
akibat pergaulan bebas ya tuailah hasilnya
kata2 nya seperti cerita asli lagi ngelihat 😊😄
sebelum para pelaku mati mereka tdk akan perna...
kisah mengenaskn \nkn jg orang tua vina ngk m...
bpngkar&kawal smpe tuntas kasus ini..tangkap 3...
```

Gambar 6. Screenshot Case Folding

2) **Clean Text**

Pada tahap ini, teks dibersihkan dari tanda baca, angka, dan simbol lain yang tidak relevan.

Case Folding (Sebelum) \

akibat pergaulan bebas ya tuailah hasilnya
kata2 nya seperti cerita asli lagi ngelihat 😊 😊
sebelum para pelaku mati mereka tdk akan perna...
kisah mengenaskn \nknj jg orang tua vina ngk m...
bpngkar&kawal smpe tuntas kasus ini..tangkap 3...

Cleaned Text (Sesudah)

akibat pergaulan bebas ya tuailah hasilnya
kata nya seperti cerita asli lagi ngelihat
sebelum para pelaku mati mereka tdk akan perna...
kisah mengenaskn knj jg orang tua vina ngk mer...
bpngkarkawal smpe tuntas kasus initanangkap pela...

Gambar 7. Screenshot *Cleant text*

3) *Penghapusan Stop words dan tokenizing*

Stopwords adalah kata-kata umum yang tidak memberikan makna signifikan dalam analisis sentimen, sehingga dihapus dari teks. Teks diubah menjadi token (kata-kata terpisah). Stopwords dari NLTK dan file eksternal digabungkan. Token satu karakter juga dihapus. Setelah itu, baris kosong dan duplikat dihapus dari DataFrame. Hasilnya ditampilkan dalam tabel teks sebelum dan sesudah penghapusan stopwords.

Cleaned_Text (Sebelum)

akibat pergaulan bebas ya tuailah hasilnya
kata nya seperti cerita asli lagi ngelihat
sebelum para pelaku mati mereka tdk akan perna...
kisah mengenaskn knj jg orang tua vina ngk mer...
bpngkarkawal smpe tuntas kasus initanqkap pela...

Text_Tokenizing (Sesudah)

[akibat, pergaulan, bebas, ya, tuailah, hasilnya]
[kata, nya, seperti, cerita, asli, lagi, ngeli...]
[sebelum, para, pelaku, mati, mereka, tdk, aka...]
[kisah, mengenaskn, knj, jg, orang, tua, vina,...]
[bpngkarkawal, smpe, tuntas, kasus, initanangkap...]

Filtered_Text (Final)

[akibat, pergaulan, bebas, ya, tuailah, hasilnya]
[nya, cerita, asli, ngelihat]
[pelaku, mati, tdk, tenang, dlm, hidupnya]
[kisah, mengenaskn, knj, jg, orang, tua, vina,...]
[bpngkarkawal, smpe, tuntas, initanangkap, pelak...]

Gambar 8. Screenshot *Stop words dan tokenizing*

4) *Normalisasi Teks*

Normalisasi teks memperbaiki kesalahan penulisan pada kata yang dilakukan untuk mengubah kata-kata yang disingkat atau tidak baku menjadi bentuk standar. Dan berikut adalah hasil gambar yang di terapkan seleh di normaliasaikan [13].

Penghapusan Tanda Baca dan Simbol: Tanda baca dan simbol yang tidak relevan dihapus untuk memastikan bahwa hanya terdiri dari kata-kata yang signifikan.

Penggantian Kata Singkatan: Kata-kata yang sering disingkat dalam teks informal, seperti "tdk" menjadi "tidak" atau "krn" menjadi "karena", diubah ke dalam bentuk lengkapnya.

Perbaikan Ejaan: Kesalahan ejaan yang umum diperbaiki untuk memastikan bahwa semua kata dalam teks sesuai dengan bentuk standar yang diakui dalam bahasa tersebut.

Pengubahan ke Bentuk Dasar: Kata-kata yang memiliki variasi bentuk, seperti kata kerja dengan imbuhan atau kata benda jamak, dikembalikan ke bentuk dasarnya. Misalnya, "makanannya" menjadi "makanan".

Penghapusan Stop Words: Kata-kata umum yang tidak memberikan kontribusi signifikan dalam analisis, seperti "dan", "atau", "tetapi", dihapus dari teks.

Penggantian Sinonim: Kata-kata yang memiliki makna sama atau mirip diubah menjadi satu bentuk standar untuk mengurangi variasi yang tidak perlu. Misalnya, "indah" dan "cantik" mungkin diubah menjadi salah satu dari keduanya tergantung pada konteksnya.

Filtered_Text (Sebelum)

[akibat, pergaulan, bebas, ya, tuailah, hasilnya]
[nya, cerita, asli, ngelihat]
[pelaku, mati, tdk, tenang, dlm, hidupnya]
[kisah, mengenaskn, knj, jg, orang, tua, vina,...]
[bpngkarkawal, smpe, tuntas, initanangkap, pelak...]

TextNormalization (Sesudah)

[akibat, pergaulan, bebas, ya, tuailah, hasilnya]
[nya, cerita, asli, melihat]
[pelaku, mati, tidak, tenang, dalam, hidupnya]
[kisah, mengenaskn, kenapa, juga, orang, tua, ...]
[bpngkarkawal, sampai, tuntas, initanangkap, pel...]

Gambar 9. Screenshot *Normalisasi Teks*

5) *Stemming*

Berikut tahapan selanjutnya setelah melakukan teks tokenizing dan filtered maka langkah selanjutnya adalah stemming yang akan mengubah kata-kata mendaji bentuk dasarnya. Dan berikut adalah hasil tabe screenshot stemming tersebut.

```

TextNormalization
[akibat, pergaulan, bebas, ya, tuailah, hasilnya]
[nya, cerita, asli, melihat]
[pelaku, mati, tidak, tenang, dalam, hidupnya]
[kisah, mengenaskn, kenapa, juga, orang, tua, ...]
[bpngkarkawal, sampai, tuntas, initangkap, pel...

TextStemming
[akibat, gaul, bebas, ya, tuai, hasil]
[nya, cerita, asli, lihat]
[laku, mati, tidak, tenang, dalam, hidup]
[kisah, mengenaskn, kenapa, juga, orang, tua, ...]
[bpngkarkawal, sampai, tuntas, initangkap, lak...
    
```

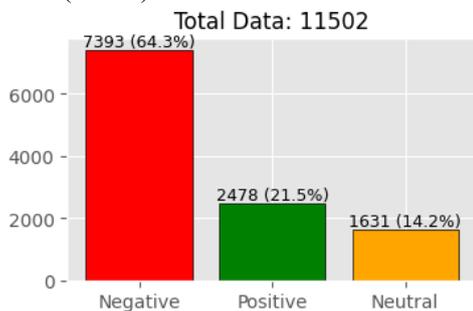
Gambar 10. Screenshot *Stemming*

3.2 Hasil analisis Sentimen

Setelah teks diproses, analisis sentimen dilakukan untuk mengategorikan komentar ke dalam sentimen positif, negatif, dan netral. Berikut adalah hasil analisis sentimen:

Kategori	Total
Negatif	7,393
Netral	1,631
Positif	2,478
Total	11,502

Berikut adalah gambar visualisasi hasil jumlah total pelabelan analisis sentimen, yang menunjukkan persentase masing-masing kategori [14]. Seperti dijelaskan pada Tabel 1 di atas, visualisasi ini menampilkan hasil analisis sentimen dengan total 11.502 data. Sentimen negatif memiliki jumlah total 7.393 (64.3%), sentimen positif memiliki jumlah total 2.478 (21.5%), dan sentimen netral memiliki jumlah total 1.631 (14.2%).

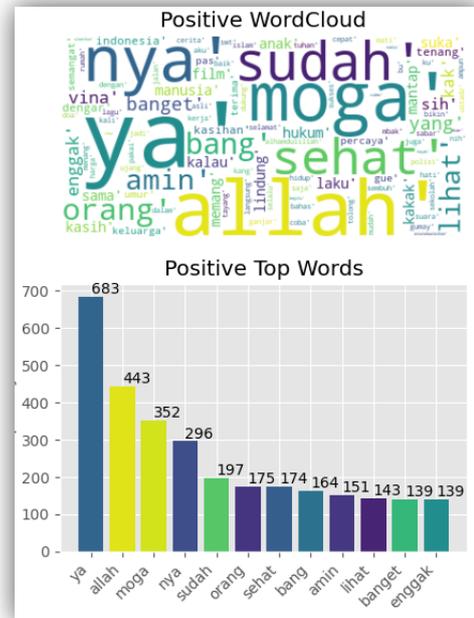


Gambar 11. Visualisasi sentimen

Kata-kata yang muncul dalam semua data ulasan ditampilkan bentuk wordcloud [15] dan Frekuensi Kata adalah alat untuk memberikan gambaran visual dari kata-kata yang paling sering muncul dalam setiap kategori sentimen, seperti positive wordcloud, negative wordcloud, neutral wordcloud, positive top words, negative top words, serta neutral top words. Berikut adalah WordCloud dan grafik frekuensi kata untuk setiap sentimen:

1) Positive Sentiment

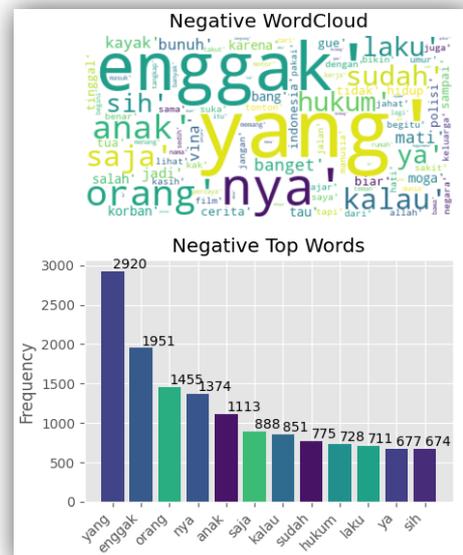
- **Word Cloud:** Kata-kata menonjol seperti "nya," "ya," "sudah," "moga," "sehat," dan "allah."
- **Frequency:** Kata "ya" memiliki frekuensi tertinggi (683), diikuti oleh "sudah" (443), "moga" (352), "nya" (296), dan "allah" (197).



Gambar 12. Positive Visualisasi wordcloud dan frekuensi

2) Negative Sentiment

- **Word Cloud:** Kata-kata menonjol seperti "enggak," "yang," "nya," dan "anak."
- **Frequency:** Kata "yang" memiliki frekuensi tertinggi (2920), diikuti oleh "enggak" (1951), "nya" (1455), "orang" (1133), dan "anak" (1113).



Berikut adalah rincian confusion matrix untuk model regresi logistik:

1. True Negative (TN): 1,444
2. False Positive (FP): 27
3. False Negative (FN): 175
4. True Positive (TP): 286
5. Neutral (Predicted as Neutral): 65

Rumus dan hasil evaluasi untuk Regresi Logistik:

1. $ccuracy = \frac{286+1444}{286+1444+27+175} \times 100\% = 81.48\%$
2. $Precision = \frac{286}{286+27} = 0.854$
3. $Recal = \frac{286}{286 + 175} = 0.602$
4. $F1 - score = 2 \times \left(\frac{0.854 \times 0.602}{0.854+0.602} \right) = 0.706$

4. DISKUSI

Dari hasil evaluasi kedua model, dapat disimpulkan bahwa model Regresi Logistik mengungguli model GBM dalam hal akurasi, precision, recall, dan F1-score. Hal ini mengindikasikan bahwa Regresi Logistik lebih efektif dalam menangkap nuansa sentimen publik terkait kasus Vina Cirebon. Analisis menunjukkan dominasi sentimen negatif, mencerminkan ketidaksetujuan dan keprihatinan publik yang luas terhadap kasus ini. Visualisasi WordCloud dan grafik frekuensi kata juga memperkuat temuan ini, dengan kata-kata seperti "enggak", "yang", dan "orang" sering muncul dalam komentar negatif. Penelitian ini memberikan wawasan berharga tentang reaksi publik terhadap kasus Vina Cirebon, menunjukkan efektivitas analisis sentimen dalam memahami isu sosial melalui data media sosial. Selain itu, hasil ini dapat memberikan panduan bagi pembuat kebijakan dalam merespons isu-isu yang mendapat perhatian publik secara lebih efektif dan tepat sasaran.

Tabel 5. Hasil Evaluasi Model

Model	Akurasi	Precision	Recall	F1-score
(GBM)	81.48%	0.739	0.352	0.476
Regresi Logistik	88.29%	0.854	0.602	0.706

Penjelasan Hasil Evaluasi

- 1) **Akurasi:** Regresi Logistik (88.29%) lebih tinggi dari GBM (81.48%).
- 2) **Precision:** Regresi Logistik (0.854) lebih tinggi dari GBM (0.739).
- 3) **Recall:** Regresi Logistik (0.602) lebih tinggi dari GBM (0.352).
- 4) **F1-score:** Regresi Logistik (0.706) lebih tinggi dari GBM (0.476).

Confusion matrix menunjukkan Regresi Logistik lebih baik dalam mengklasifikasikan sentimen negatif dan positif dengan lebih sedikit kesalahan dibandingkan GBM.

5. KESIMPULAN

Berdasarkan hasil evaluasi kinerja yang telah dilakukan, diperoleh kesimpulan bahwa Regresi Logistik menunjukkan performa yang lebih baik dibandingkan *Gradient Boosting Machine* (GBM) dalam menganalisis sentimen publik terhadap kasus Vina Cirebon di YouTube. Regresi Logistik memiliki nilai akurasi sebesar 88.29%, lebih tinggi dibandingkan GBM yang mencapai 81.48%. Precision dan recall Regresi Logistik masing-masing sebesar 0.854 dan 0.602, juga lebih unggul dibandingkan GBM yang memiliki precision 0.739 dan recall 0.352. F1-score Regresi Logistik sebesar 0.706 lebih tinggi dibandingkan GBM yang hanya mencapai 0.476. Hal ini mengindikasikan bahwa Regresi Logistik lebih efektif dalam menangkap nuansa sentimen publik terkait kasus Vina Cirebon. Analisis sentimen menunjukkan dominasi sentimen negatif dalam komentar-komentar YouTube terkait kasus ini, mencerminkan ketidaksetujuan dan keprihatinan publik yang luas. Penelitian ini memberikan wawasan berharga tentang reaksi publik terhadap kasus Vina Cirebon, menunjukkan efektivitas analisis sentimen dalam memahami isu sosial melalui data media sosial. Hasil ini dapat memberikan panduan bagi pembuat kebijakan dalam merespons isu-isu yang mendapat perhatian publik secara lebih efektif dan tepat sasaran.

6. DAFTAR PUSTAKA

- [1] N. D. Kusumawati, S. Al Faraby, And M. Dwifabri, "Analisis Sentimen Komentar Beracun Pad Media Sosial Menggunakan Word2vec Dan Support Vectore Machine (Svm)." [Online]. Available: [Http://J.Mp/Ktenok](http://J.Mp/Ktenok)
- [2] A. F. L. Ptr, M. M. Siregar, And I. Daniel, "Analysis Of Gradient Boosting, Xgboost, And Catboost On Mobile Phone Classification," *Journal Of Computer Networks, Architecture And High Performance Computing*, Vol. 6, No. 2, Pp. 661–670, Apr. 2024, Doi: 10.47709/Cnahpc.V6i2.3790.
- [3] R. A. Husen, R. Astuti, L. Marlia, R. Rahmaddeni, And L. Efrizoni, "Analisis

- Sentimen Opini Publik Pada Twitter Terhadap Bank Bsi Menggunakan Algoritma Machine Learning,” *Malcom: Indonesian Journal Of Machine Learning And Computer Science*, Vol. 3, No. 2, Pp. 211–218, Oct. 2023, Doi: 10.57152/Malcom.V3i2.901.
- [4] I. S. Siva Rao, P. Rajya Lakshmi, D. N. V Syma Kumar, A. Yugandhara Reddy, J. Karthik, And B. Bhavana, “International Journal Of Intelligent Systems And Applications In Engineering An Approach For Product Recommendation Using Light Gbm.” [Online]. Available: www.ijisae.org
- [5] A. J. Abimanyu, M. Dwifabri, And W. Astuti, “Sentiment Analysis On Movie Review From Rotten Tomatoes Using Logistic Regression And Information Gain Feature Selection,” *Building Of Informatics, Technology And Science (Bits)*, Vol. 5, No. 1, Jun. 2023, Doi: 10.47065/Bits.V5i1.3595.
- [6] Y. Pratama, D. Triantoro Murdiansyah, And K. Muslim Lhaksana, “Jurnal Media Informatika Budidarma Analisis Sentimen Kendaraan Listrik Pada Media Sosial Twitter Menggunakan Algoritma Logistic Regression Dan Principal Component Analysis,” 2023, Doi: 10.30865/Mib.V7i1.5575.
- [7] P. Studi Informatika, F. Matematika Dan Ilmu Pengetahuan Alam, J. Raya Kampus Udayana, B. Jimbaran, K. Selatan, And B. Indonesia, “Analisis Sentimen Aplikasi Zenius Menggunakan Metode Logistic Regression I Made Juniandika A1 , Ida Bagus Made Mahendra A2.”
- [8] S. A. Assaidi And F. Amin, “Analisis Sentimen Evaluasi Pembelajaran Tatap Muka 100 Porsen Pada Pengguna Twitter Menggunakan Metode Logistic Regression.”
- [9] A. Robi Padri, A. Asro, And I. Indra, “Classification Of Traffic Congestion In Indonesia Using The Naive Bayes Classification Method,” *Journal Of World Science*, Vol. 2, No. 6, Pp. 877–888, Jun. 2023, Doi: 10.58344/Jws.V2i6.285.
- [10] F. Alzamzami, M. Hoda, And A. El Saddik, “Light Gradient Boosting Machine For General Sentiment Classification On Short Texts: A Comparative Evaluation,” *Ieee Access*, Vol. 8, Pp. 101840–101858, 2020, Doi: 10.1109/Access.2020.2997330.
- [11] E. Darmaja, V. C. Mawardi, And N. J. Perdana, “Review Sentimen Analisis Aplikasi Sosial Media Di Google Playstore Menggunakan Metode Logistic Regression,” 2021.
- [12] E. R. Lidinillah, T. Rohana, And A. R. Juwita, “Analisis Sentimen Twitter Terhadap Steam Menggunakan Algoritma Logistic Regression Dan Support Vector Machine,” *Teknosains: Jurnal Sains, Teknologi Dan Informatika*, Vol. 10, No. 2, Pp. 154–164, Jul. 2023, Doi: 10.37373/Tekno.V10i2.440.
- [13] M. Alfyanto, “Perbandingan Algoritma Random Forest Dan Logistic Regression Untuk Analisis Sentimen Ulasan Aplikasi Tumbuh Kembang Anak Di Play Store,” *Jurnal Sistem Informasi Dan Ilmu Komputer*, Vol. 2, No. 1, Pp. 77–86, 2024, Doi: 10.59581/Jusiik-Widyakarya.V2i1.2262.
- [14] A. Robi Padri, “Hci Dan Media Sosial: Studi Kasus Analisis Sentimen Pilpres 2024 Di Twitter Menggunakan Naive Bayes Classifier,” *Jurnal Simetris*, Vol. 14, No. 2, 2023.
- [15] B. Kholifah, I. Thoib, N. Sururi, And N. D. Kurnia, “Analisis Sentimen Warganet Terhadap Isu Layanan Transportasi Online Berbasis Inset Lexicon Menggunakan Logistic Regression,” *Kumpulan Jurnal Ilmu Komputer (Klik)*, Vol. 11, No. 1, 2024.
- [16] Azizah. N. S, “Public Sentiment Analysis on the 2024 Presidential Election Using Naive Bayes Classifier (NBC) and Support Vector Machine (Svm) On Social Media Data,” *Prosidingseminarinternasional*
- [17] A. Sulaiman, “Analisis Sentimen tentang Transformasi Program Makan Siang menjadi Makan Bergizi Gratis menggunakan Logistik Regression pada laman Youtube,” *Jurnal ICT: Information Communication & Technology*, vol. 24, no. 1, 2024,. Available: <https://ejournal.ikmi.ac.id/index.php/jict-ikmi>