

Analisis Sentimen Komentar Youtube Rencana Pemerintah terhadap Pemblokiran Roblox Menggunakan Model Naïve Bayes

Firdiawati^{1*}, Salvia Nabillah Syifa², Harun Al Rosyid³

Universitas Negeri Surabaya

DOI:

<https://doi.org/10.53697/jkomitek.v6i1.3454>

*Correspondence: Firdiawati

Email: firdiawati.22026@mhs.unesa.ac.id

Received: 02-04-2026

Accepted: 12-05-2026

Published: 28-06-2026



Copyright: © 2026 by the authors. Submitted for open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

Abstrak: Perkembangan teknologi digital telah meningkatkan popularitas platform permainan daring seperti Roblox. Namun, rencana pemerintah untuk memblokir platform tersebut memunculkan beragam respons publik yang terlihat pada kolom komentar YouTube. Penelitian ini bertujuan untuk mengidentifikasi kecenderungan sentimen masyarakat terhadap isu pemblokiran Roblox. Data penelitian terdiri dari 2.500 komentar YouTube yang diperoleh melalui YouTube Data API, kemudian melalui tahapan preprocessing dan pelabelan sentimen berbasis lexicon hingga tersisa 2.427 komentar yang layak dianalisis. Data selanjutnya diubah menjadi fitur TF-IDF dan diklasifikasikan menggunakan metode Naïve Bayes algoritma Multinomial Naïve Bayes dengan skema pembagian data 90:10 dan 80:20. Hasil terbaik diperoleh pada skema 80:20 dengan akurasi 71,19%, precision 73,31%, recall 71,19%, dan f1-score 70,23%. Hasil yang didapat dari penelitian menunjukkan bahwa opini publik didominasi sentimen netral dengan kecenderungan negatif. Penelitian ini menyimpulkan bahwa Multinomial Naïve Bayes efektif dalam mengklasifikasikan sentimen komentar berbahasa Indonesia terkait isu pemblokiran Roblox

Katakunci: Analisis Sentimen, Roblox, Naive Bayes, YouTube, TF-IDF

Abstract: The development of digital technology has increased the popularity of online gaming platforms such as Roblox. However, the government's plan to block the platform has generated various public responses that can be observed in YouTube comment sections. This study aims to identify public sentiment tendencies regarding the Roblox blocking issue. The research data consist of 2,500 YouTube comments collected using the YouTube Data API, followed by preprocessing and lexicon-based sentiment labeling, resulting in 2,427 comments eligible for analysis. The data were then transformed into TF-IDF features and classified using the Naïve Bayes method, specifically the Multinomial Naïve Bayes algorithm, with 90:10 and 80:20 train-test split schemes. The best results were obtained using the 80:20 split, achieving an accuracy of 71.19%, precision of 73.31%, recall of 71.19%, and an f1-score of 70.23%. The findings indicate that public opinion is predominantly neutral with a tendency toward negativity. This study concludes that the Multinomial Naïve Bayes algorithm is effective in classifying Indonesian-language comments related to the Roblox blocking issue.

Keywords: Sentiment Analysis, Roblox, Naive Bayes, YouTube, TF-IDF

Pendahuluan

Roblox merupakan platform permainan daring yang sangat populer di Indonesia saat ini, khususnya di kalangan remaja maupun anak-anak (Yuliastika et al., 2023). Munculnya isu rencana pemerintah untuk memblokir Roblox karena dugaan keterkaitan dengan perjudian online memicu perhatian masyarakat. Isu tersebut kemudian menimbulkan berbagai reaksi dan opini publik. Kolom komentar YouTube menjadi salah

satu ruang yang aktif digunakan untuk menanggapi persoalan ini. Popularitas Roblox menjadikannya salah satu media hiburan digital yang banyak digunakan setiap hari.

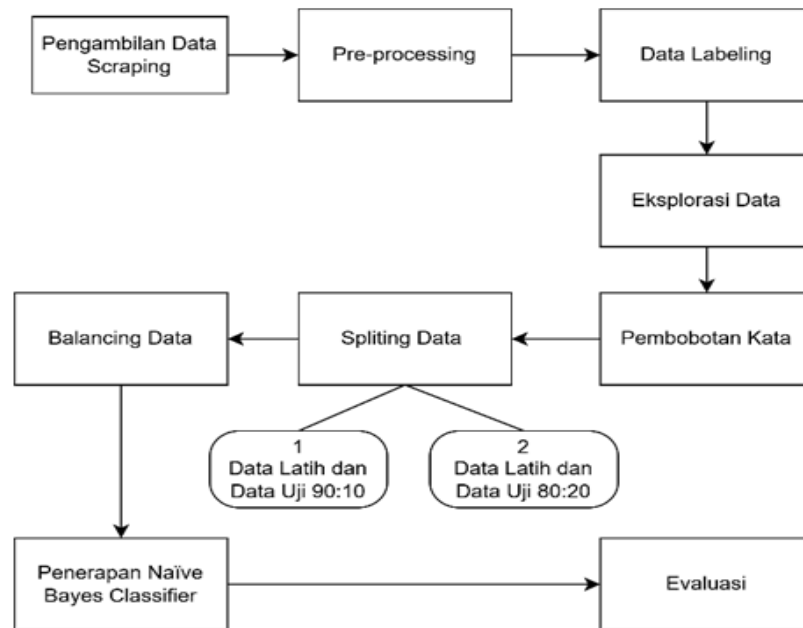
Kolom komentar YouTube menjadi ruang diskusi yang aktif dalam merespons isu pemblokiran Roblox. Komentar yang muncul berisi berbagai opini publik yang mencerminkan sudut pandang masyarakat. Karena itu, data komentar YouTube dapat dimanfaatkan untuk menganalisis persepsi pengguna. Penggunaan media sosial sebagai sumber analisis opini kini menjadi pendekatan umum dalam penelitian berbasis masyarakat digital (Rochadiani & Setiawan, 2023). Pendekatan ini memungkinkan peneliti memperoleh data secara cepat dan dalam jumlah besar.

Analisis sentimen diperlukan untuk mengukur respons masyarakat secara objektif dengan memanfaatkan teknik text mining (García-Méndez, 2025). Komentar pengguna kemudian dianalisis untuk mengidentifikasi sentimen positif, negatif, atau netral (Huwaida et al., 2024). Text mining sendiri berfungsi mengolah data tidak terstruktur menjadi informasi yang terukur (Hassani et al., n.d.). Dalam konteks isu pemblokiran Roblox, hasil analisis sentimen dapat memberikan gambaran sikap publik terhadap kebijakan tersebut. Pendekatan ini memanfaatkan pemrosesan bahasa alami untuk mengekstraksi makna dalam teks komentar.

Analisis sentimen merupakan bagian dari text mining yang berfungsi mengklasifikasikan opini berdasarkan isi teks (Wicaksono & Nata, 2023). Untuk mengolah data teks dalam jumlah besar, algoritma seperti Naïve Bayes sering digunakan karena sederhana dan memiliki performa yang baik. Variannya, Multinomial Naïve Bayes (MNB), dinilai lebih sesuai untuk data komentar YouTube yang sering memiliki ketidakseimbangan kelas (Misrun et al., 2024). Sejumlah penelitian juga menerapkan preprocessing seperti pembersihan teks, tokenisasi, dan vektorisasi TF-IDF untuk meningkatkan hasil klasifikasi (Fatah & Syarifah, 2025). Oleh karena itu, MNB menjadi pilihan yang efektif untuk mengukur kecenderungan sentimen secara otomatis pada data komentar dalam jumlah besar.

Penelitian ini bertujuan mengumpulkan komentar YouTube terkait isu pemblokiran Roblox melalui tahapan preprocessing dan pelabelan sentimen. Penelitian juga membangun model klasifikasi menggunakan Multinomial Naïve Bayes dan mengevaluasi performanya dengan metrik akurasi, precision, recall, dan f1-score (Hofmann, 2025). Hasil yang diperoleh diharapkan memberikan gambaran kuantitatif mengenai persepsi publik terhadap isu tersebut. Temuan ini dapat menjadi acuan akademis dalam pengembangan sistem analisis sentimen. Secara keseluruhan, penelitian memberikan manfaat akademis, praktis, dan teknologis bagi pihak yang ingin memahami opini masyarakat secara lebih objektif (Zulfikar et al., 2023).

Metode Penelitian



Gambar 1. Alur kerja

Penelitian ini menggunakan pendekatan kuantitatif dengan teknik text mining untuk menganalisis sentimen pada komentar YouTube terkait rencana pemblokiran game Roblox. Sebanyak 2.500 komentar diperoleh melalui proses scrapping menggunakan Python di Google Colab dengan YouTube Data API, dan setelah pembersihan awal tersisa 2.427 komentar yang valid. Dataset dalam format CSV serta seluruh kode tersedia jika dibutuhkan untuk penelitian lanjutan.

Pengolahan data dilakukan melalui tahapan cleansing, case folding, normalisasi, tokenisasi, penghapusan stopword, dan stemming menggunakan Sastrawi dan NLTK. Pelabelan sentimen dilakukan secara lexicon-based yang disesuaikan dengan konteks komentar serta mempertimbangkan jumlah like. Data kemudian diubah menjadi fitur numerik menggunakan TF-IDF dan dibagi menjadi data latih dan uji dengan skema 90:10 serta 80:20. Model klasifikasi dibangun menggunakan Multinomial Naïve Bayes (Su, 2024), lalu dievaluasi menggunakan confusion matrix dan metrik akurasi, presisi, recall, serta f1-score. Karena penelitian ini memakai data publik YouTube dan tidak melibatkan subjek manusia secara langsung, persetujuan etis tidak diperlukan.

Hasil dan Pembahasan

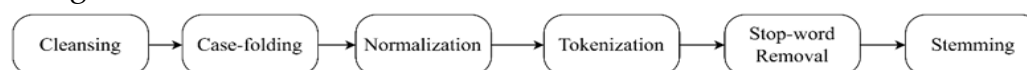
Pengambilan Data (*Scraping*)

Data Scraping merupakan metode yang memainkan peran penting untuk solusi pengambilan data dalam jumlah besar dengan waktu relatif singkat dan akurat secara otomatis dari sumber daring tertentu (Anugrah et al., 2025). Penelitian ini berhasil mengumpulkan 2.500 komentar YouTube menggunakan metode scraping melalui Python di Google Colab. Data diambil menggunakan YouTube API dan diproses hingga diperoleh

2.427 kpubomentar yang layak dianalisis setelah duplikasi dan karakter irrelevant dihapus. Dataset disimpan dalam format CSV sebagai dasar analisis berikutnya.

Preprocessing (Pembersihan Data Teks)

Tahap ini bertujuan agar teks siap untuk dikonversi ke bentuk numerik. Tahapan preprocessing yang dilakukan dalam penelitian ini meliputi beberapa langkah seperti gambar, yaitu cleansing, Normalization, case-folding, tokenization, stop-word removal, dan stemming.



Gambar 2. Alur Preprocessing

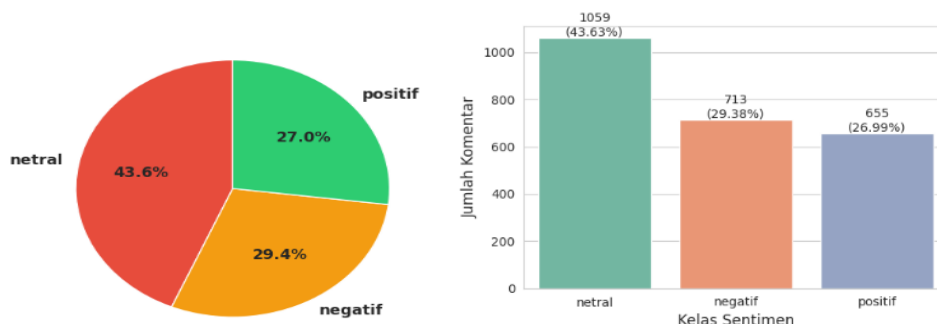
- Cleansing: menghapus URL, emoji, angka, dan karakter tidak penting.
- Case Folding: mengubah seluruh huruf menjadi kecil.
- Normalization dilakukan dengan menyelaraskan kata-kata yang tidak baku,
- Tokenizing: memecah kalimat menjadi kata.
- Stopword Removal: menghapus kata-kata umum yang tidak memiliki makna penting.
- Stemming: mengembalikan kata ke bentuk dasarnya untuk mengurangi variasi kata yang sebenarnya memiliki akar yang sama.

Data Labelling

Dataset awal berjumlah 2.500 komentar, kemudian melalui proses pre-processing untuk menghapus duplikasi dan elemen tidak relevan, sehingga tersisa 2.427 komentar. Data bersih tersebut dilabeli menggunakan metode keyword labeling dengan mencocokkan kata-kata yang merepresentasikan sentimen negatif, netral, atau positif. Hasil pelabelan menunjukkan distribusi sentimen sebagai berikut: sentimen netral menjadi yang paling dominan (1.059 komentar), diikuti sentimen negatif (713 komentar), dan sentimen positif sebagai yang paling sedikit (655 komentar).

Eksplorasi Data

Pada Gambar 3. Diagram Pie dan Batang Sentiment Distributor menunjukkan distribusi sentimen dari komentar YouTube terlihat bahwa sentimen netral merupakan kategori yang paling dominan dengan porsi sekitar 43,6% atau 1059 komentar. Selanjutnya, sentimen negatif berada di posisi kedua dengan 29,4% atau 713 komentar, sedangkan sentimen positif menjadi yang paling sedikit dengan 27,0% atau 655 komentar. Distribusi menggambarkan bahwa sebagian besar pengguna memberikan komentar bersifat netral, sementara komentar negatif dan positif memiliki proporsi yang relatif seimbang.

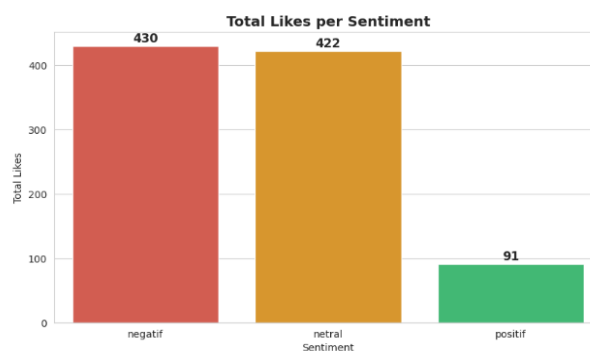


Gambar 3. Diagram Pie dan Batang Sentiment Distribution

Pada gambar 4. menampilkan WordCloud (Kata Kunci) untuk setiap kelas sentimen Pada sentimen **positif**, kata paling dominan seperti *roblox*, *blokir*, *anak*, dan *keras*, yang mencerminkan dukungan untuk pemerintah memblokir roblox. WordCloud **negatif** didominasi kata seperti *roblox*, *game*, *urusin*, *judi* yang menandakan adanya keluhan, kritik, serta penolakan terhadap isu pemblokiran Roblox. Sementara itu, pada sentimen **netral**, terlihat banyak kata informatif seperti *perintah*, *main*, dan *aja* yang bersifat informasi, ekspresi umum, tidak menyerang, tidak mendukung kuat.



Gambar 4. Kata Kunci Sentimen Positif, Negatif, dan Netral



Gambar 5. Total Like per Sentiment

Berdasarkan Gambar 5. yang menampilkan total like komentar untuk setiap kategori sentimen, dapat dilihat bahwa komentar bersentimen negatif memperoleh jumlah like tertinggi, yaitu 430 like. Komentar bersentimen netral berada pada posisi kedua dengan 422 like, selisih yang tidak terlalu jauh dari sentimen negatif. Sementara itu, komentar bersentimen positif menunjukkan jumlah like paling rendah, yaitu hanya 91 like. Hasil ini menunjukkan bahwa komentar negatif dan netral lebih banyak menarik perhatian

pengguna dibandingkan komentar positif, kemungkinan karena topik yang dibahas memicu respons kritis atau informatif.

Pembobotan Kata

Metode TF-IDF digunakan untuk mengonversi teks ke dalam bentuk angka dengan memberikan bobot pada setiap kata (Chakravarthi & Muralidaran, 2020). Bobot tersebut dihitung berdasarkan seberapa sering kata muncul dalam satu dokumen, kemudian disesuaikan dengan frekuensi kemunculannya di seluruh dokumen lain. Dengan cara ini, model dapat menilai tingkat pentingnya suatu kata: kata yang lebih spesifik dan relevan terhadap isi dokumen akan memperoleh bobot lebih tinggi, sedangkan kata yang terlalu umum akan diberi bobot rendah (Maulana & Anshori et al., 2025). Rumus perhitungan bobot untuk setiap kata (term- t) dalam dokumen d ditunjukkan pada Persamaan (1) dibawah ini.

$$w_{t,d} = tf_{t,d} \times \log\left(\frac{N}{df_t}\right) \quad (1)$$

Keterangan:

$w_{t,d}$ = bobot TF-IDF

$tf_{t,d}$ = banyaknya t dalam dokumen d

N = frekuensi dokumen

df_t = frekuensi dokumen yang mengandung kata t

Dalam tahap pembobotan, metode TF-IDF diterapkan menggunakan `TfidfVectorizer` dari pustaka `scikit-learn`. Tools ini menyediakan berbagai opsi konfigurasi seperti normalisasi dan pembatasan jumlah fitur sehingga proses penghitungan bobot dapat disesuaikan dengan kebutuhan analisis. Karena bagian IDF menurunkan skor kata yang terlalu sering muncul pada banyak dokumen, teknik ini efektif mengurangi dominasi kata-kata umum dan menonjolkan istilah yang lebih relevan serta informatif.

Splitting Data

Pembagian data merupakan tahap penting dalam proses pembangunan model (Jain et al., 2022). Dataset dipisahkan menjadi data latih dan data uji untuk memastikan model dapat belajar dari sebagian data serta dievaluasi secara objektif pada bagian data lainnya. Pada penelitian ini digunakan dua skenario pembagian, yaitu 90:10 dan 80:20, untuk menilai konsistensi performa model pada jumlah data latih yang berbeda. Proses pemisahan dilakukan dengan stratified splitting agar proporsi kelas sentimen tetap seimbang. Hasilnya, skenario 90:10 menghasilkan 2.184 data latih dan 243 data uji, sedangkan skenario 80:20 menghasilkan 1.941 data latih dan 486 data uji. Pembagian dilakukan menggunakan fungsi `train_test_split` dengan `random_state` untuk memastikan hasil yang konsisten.

Balancing Data

Hasil penelitian menunjukkan bahwa distribusi data pada setiap kelas sentimen tidak merata. Pada data latih, komentar dengan sentimen netral merupakan yang paling dominan, yakni sebanyak 953 komentar pada skenario 90:10 dan 847 komentar pada skenario 80:20. Sentimen positif dan negatif berada pada jumlah yang lebih rendah, yaitu masing-masing 589 dan 462 komentar pada skenario 90:10, serta 524 dan 570 komentar pada

skenario 80:20. Ketidakeimbangan distribusi ini membuat model lebih condong pada kelas mayoritas. Untuk mengatasi hal ini, dilakukan penyeimbangan data menggunakan SMOTE yang menghasilkan distribusi kelas yang sama pada kedua skenario (953 data untuk 90:10 dan 847 data untuk 80:20). Dengan data yang lebih seimbang, model dapat belajar lebih merata dan memberikan prediksi yang lebih stabil serta akurat.

Penerapan Metode Naive Bayes Classifier

Naive Bayes Classifier adalah metode yang digunakan untuk klasifikasi sentimen karena kesederhanaan, efektivitas, kemampuan interpretasi, dan efisiensi komputasi (Afuan et al., 2025). Penelitian ini menerapkan algoritma Naïve Bayes, MultinomialNB, untuk mengklasifikasikan sentimen komentar YouTube mengenai isu pemblokiran Roblox. Metode Naïve Bayes Classifier diterapkan melalui algoritma MultinomialNB, yang dilatih menggunakan data TF-IDF hasil preprocessing dan balancing SMOTE. Kinerja model dievaluasi menggunakan accuracy, precision, recall, dan F1-score pada skema pembagian data 90:10 dan 80:20, serta divisualisasikan melalui classification report dan confusion matrix.

Classification Report:

	precision	recall	f1-score	support
Negative	0.66	0.80	0.72	71
Neutral	0.77	0.52	0.62	106
Positive	0.67	0.86	0.75	66
accuracy			0.70	243
macro avg	0.70	0.73	0.70	243
weighted avg	0.71	0.70	0.69	243

Gambar 4. Classification Report 90:10

Classification Report:

	precision	recall	f1-score	support
Negative	0.64	0.83	0.72	143
Neutral	0.81	0.52	0.63	212
Positive	0.71	0.90	0.79	131
accuracy			0.71	486
macro avg	0.72	0.75	0.72	486
weighted avg	0.73	0.71	0.70	486

Gambar 5. Classification Report 80:20

Evaluasi

Hasil pengujian menggunakan metode Naïve Bayes Classifier divisualisasikan dalam bentuk confusion matrix, yang dimana bentuk ini memberikan perbandingan hasil klasifikasi yang dibantu oleh metode pembobotan TF-IDF.

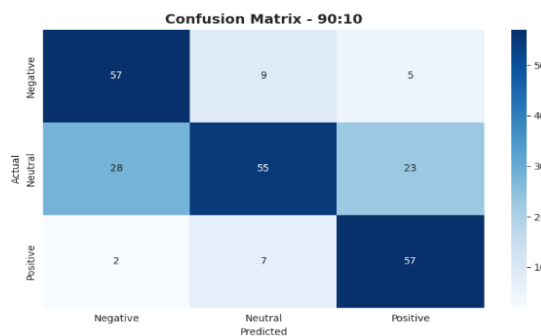
		Kelas Prediksi	
		1	0
Kelas Sebenarnya	1	TP	FN
	0	FP	TN

Keterangan:

- TP (True Positive): jumlah data dari kelas 1 yang benar diprediksi sebagai kelas 1.
- TN (True Negative): jumlah data dari kelas 0 yang benar diprediksi sebagai kelas 0.
- FP (False Positive): jumlah data dari kelas 0 yang salah diprediksi sebagai kelas 1.
- FN (False Negative): jumlah data dari kelas 1 yang salah diprediksi sebagai kelas 0

Metode evaluasi dalam data mining yang paling sering digunakan adalah mencari nilai accuracy, precision, recall dan f1-score. Beberapa nilai yang dihitung pada tahap evaluasi tersebut antara lain adalah.

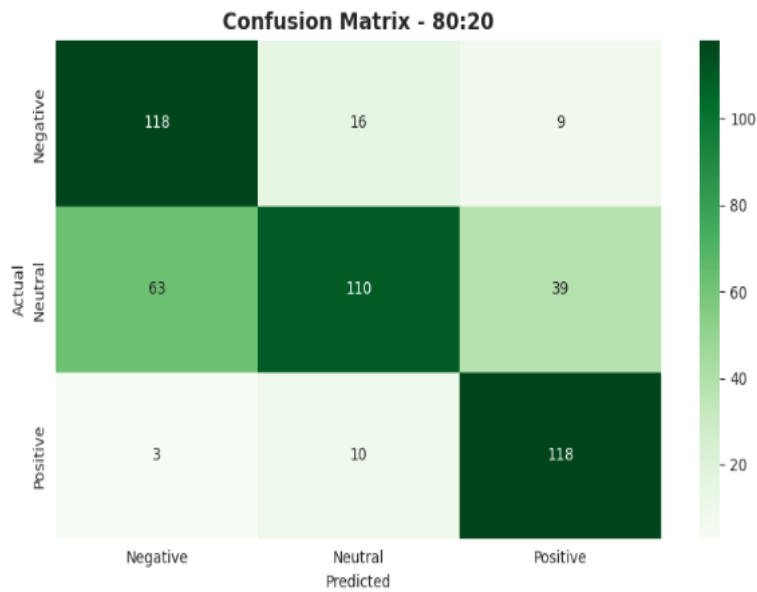
1. Perhitungan Accuracy: $\frac{TP_{total}}{Total\ Sample} = \frac{170}{243} = 0.6955 = 69.55\%$
2. Perhitungan Precision : $\frac{\sum(Precision_i \times Support_i)}{\sum Support} = \frac{(0.66 \times 71) + (0.77 \times 106) + (0.67 \times 66)}{243} = \frac{172.70}{243} = 0.7115 = 71.15\%$
3. Perhitungan Recall: $\frac{\sum(Recall_i \times Support_i)}{\sum Support} = \frac{(0.80 \times 71) + (0.52 \times 106) + (0.86 \times 66)}{243} = \frac{168.68}{243} = 0.6949 = 69.55\%$
4. Perhitungan F1-score: $\frac{\sum(F1_i \times Support_i)}{\sum Support} = \frac{(0.72 \times 71) + (0.62 \times 106) + (0.75 \times 66)}{243} = \frac{166.34}{243} = 0.6870 = 68.70\%$



Gambar 6. Confusion Matrix 90:10

Pada skema pembagian data 90:10, model memperoleh akurasi sebesar 69.55%, dengan nilai precision 71.15%, recall 69.55%, dan f1-score 68.70%. Hasil ini menunjukkan bahwa ketika data latih jauh lebih besar, model mampu mempelajari pola sentimen dengan lebih baik dan menghasilkan prediksi yang lebih stabil. Berdasarkan laporan klasifikasi, kelas Positive memiliki performa terbaik dengan f1-score 0.75, disusul kelas Negative dengan f1-score 0.72, sedangkan kelas Netral menjadi yang terendah dengan f1-score 0.52. Temuan ini sejalan dengan hasil confusion matrix yang menunjukkan bahwa model cukup baik dalam mengenali sentimen positif dan negatif, namun masih sering keliru ketika membedakan kelas netral, terutama karena adanya kemiripan pola bahasa yang muncul pada data.

- a) Perhitungan Accuracy: $\frac{TP_{total}}{Total\ Sample} = \frac{346}{486} = 0.7119 = 71.19\%$
- b) Perhitungan Precision : $\frac{\sum(Precision_i \times Support_i)}{\sum Support} = \frac{(0.64 \times 143) + (0.81 \times 212) + (0.71 \times 131)}{486} = \frac{356.25}{486} = 0.7331 = 73.31\%$
- c) Perhitungan Recall: $\frac{\sum(Recall_i \times Support_i)}{\sum Support} = \frac{(0.83 \times 143) + (0.52 \times 212) + (0.90 \times 131)}{486} = \frac{346}{486} = 0.7119 = 71.19\%$
- d) Perhitungan F1-score: $\frac{\sum(F1_i \times Support_i)}{\sum Support} = \frac{(0.72 \times 143) + (0.63 \times 212) + (0.79 \times 131)}{486} = \frac{340.99}{486} = 0.7023 = 70.23\%$



Gambar 7. Confusion Matrix - 80:20

Pada skema pembagian 80:20, kinerja model menurun dibandingkan 90:10, dengan akurasi 71.19% serta f1-score tertinggi pada kelas Positive (0.79), disusul Negative (0.72), dan Netral yang kembali menjadi terendah (0.62). Confusion matrix juga menunjukkan bahwa kelas netral memiliki kesalahan prediksi paling banyak. Pola ini konsisten dengan skema 90:10, di mana kelas positif selalu paling mudah dikenali, sementara kelas netral tetap menjadi tantangan utama karena sifatnya yang ambigu dan kurang memiliki kata beremosi kuat.

Secara keseluruhan, kedua skema menghasilkan pola kinerja serupa model Naive Bayes lebih mampu mengenali sentimen positif dan negatif, tetapi kesulitan pada sentimen netral, dipengaruhi oleh tumpang tindih bahasa serta distribusi data yang kurang seimbang. Oleh karena itu, peningkatan akurasi dapat dilakukan melalui penyeimbangan data, penambahan fitur, atau penggunaan model yang lebih kompleks.

Kesimpulan

Dari 2.427 komentar YouTube tentang rencana pemblokiran Roblox, sentimen publik terbagi cukup merata, dengan sentimen netral sebagai kategori terbesar, diikuti negatif, dan positif sebagai yang paling sedikit. Multinomial Naïve Bayes menghasilkan akurasi terbaik 71,19% pada skema 80:20. Kelas positif memiliki f1-score tertinggi, negatif berada di tengah, sementara kelas netral paling rendah karena sifat komentarnya yang ambigu. Ketidakseimbangan data awal memengaruhi akurasi, terutama pada kelas netral. Setelah balancing, performa model menjadi lebih stabil meski tantangan pada kelas netral tetap terlihat. Secara umum, metode ini efektif untuk analisis sentimen komentar berbahasa Indonesia.

Saran

Penelitian selanjutnya perlu data dari platform lain, pelabelan yang lebih akurat, serta eksplorasi algoritma lain seperti SVM, Random Forest, atau BERT. Teknik balancing

yang lebih optimal dan analisis topik juga dapat memperkuat hasil dan mendukung pengambilan kebijakan terkait platform digital.

Daftar Pustaka

- Afuan, L., Khanza, M., & Hasyati, A. Z. (2025). *Enhancing Sentiment Analysis Of The 2024 Indonesian Presidential Inauguration On X Using Smote-Optimized Naive Bayes Classifier Peningkatan Analisis Sentimen Pelantikan Presiden Ri Tahun 2024 Pada X Menggunakan Naive Bayes Classifier Yang Dioptimalkan*. 6(1), 325–333.
- Anugrah, L. R., Pandega, L. R., Haiqal, D., & Munawwar, R. F. (2025). *Pemanfaatan Big Data dan Teknik Data Scraping untuk Analisis Kepuasan Pengguna Aplikasi Tangerang Live Berdasarkan Ulasan Pengguna*. 5, 3480–3489.
- Huwaida, S. F., Kusumawati, R., & Isnaini, B. (2024). *Analisis sentimen komentar youtube terhadap pemindahan ibu kota negara menggunakan metode Naive Bayes*. 6(1), 26–39. <https://doi.org/10.37905/jji.v6i1.24718>
- Jain, E., Neeraja, J., Banerjee, B., Bengal, W., & Ghosh, P. (n.d.). *A D Iagnostic A Pproach To A Ssess The Q Uality Of D Ata*.
- Maulana, M., & Anshori, Y. (2025). *Implementasi pembobotan tf-idf pada chatbot telegram untuk sistem layanan informasi*. 10(3), 1869–1877.
- Misrun, C. A., Haerani, E., Fikry, M., & Budianita, E. (2024). *Analisis sentimen komentar youtube terhadap Anies Baswedan sebagai bakal calon presiden 2024 menggunakan metode naive bayes classifier*. 4(1), 207–215.
- Yuliastika, M. (2023). *Motif Penggunaan Game Online Roblox pada Anak Usia Sekolah Tivani Yuliastika 1 , Mayasari 2 , Ana Fitriana Poerana 3 Universitas Singaperbangsa Karawang*. 9(9), 364–371.
- Chakravarthi, B. R., & Muralidaran, V. (2020). *Corpus Creation for Sentiment Analysis in Code-Mixed Tamil-English Text*.
- Fatah, Z., & Syarifah, L. (2025). *Analisis Sentimen Komentar YouTube pada Video Terkait Insiden Pengemudi Ojek Online dan Anggota Brimob Menggunakan Algoritma Naive Bayes*. 4, 90–96.
- García-Méndez, S. (2025). *Advancements in Natural Language Processing, Semantic Networks, and Sentiment Analysis*.
- Hassani, H., Beneki, C., Unger, S., & Mazinani, M. T. (n.d.). *Text Mining in Big Data Analytics*. 1–34. <https://doi.org/10.3390/bdcc4010001>
- Hofmann, S. (2025). *Hate Speech and Sentiment of YouTube Video Comments From Public and Private Sources Covering the Israel-Palestine Conflict*. *Wirtschaftsinformatik*, September 2024.
- Rochadiani, T. H., & Setiawan, R. (2023). *Sentiment Analysis of YouTube Comments Toward Chat*. 21(1), 60–68.
- Su, V. (2024). *COVID-19 on YouTube: A Data-Driven Analysis of Sentiment , Toxicity , and Content Recommendations*.
- Wicaksono, D., & Nata, I. A. (2023). *SENTIMENT ANALYSIS TASKS DENGAN METODE KLASIFIKASI MACHINE LEARNING (STUDY KASUS REPOSITORIREVIEW*

MOVIE)DENGANWEKA. 4(4), 432–446.

Zulfikar, W. B., Atmadja, A. R., & Pratama, S. F. (2023). *Sentiment Analysis on Social Media Against Public Policy Using Multinomial Naive Bayes*. 10(1), 25–34.
<https://doi.org/10.15294/sji.v10i1.39952>