

Analisis Sentimen Komentar Youtube Tentang Pemblokiran Aplikasi Roblox Menggunakan Metode *Support Vector Machine* Dan *Naive Bayes (SVMNB)*

Muhammad Zikri Maulidi^{1*}, Destiarini², Joko Kuswanto³

Jurusan : Teknik dan Komputer Universitas Baturaja

Jl.Ratu Penghulu No.02301 Karang Sari Baturaja OKU SUMSEL 23115

* E-mail: muhammadzikrimaulidi45@gmail.com

ABSTRAK

Perkembangan teknologi informasi telah mengubah cara Masyarakat mengekspresikan pendapat dan berinteraksi secara daring. Salah satu sarana ekspresi publik yang paling sering digunakan adalah kolom komentar di platform YouTube. Studi ini bertujuan untuk menganalisis pandangan publik terhadap isu pemblokiran aplikasi Roblox oleh pemerintah Indonesia melalui respons pengguna YouTube. Metode yang diterapkan dalam penelitian ini adalah Support Vector Machine (SVM) dan Naive Bayes (NB) yang digunakan untuk mengkategorikan komentar ke dalam tiga jenis sentimen: positif, negatif, atau netral. Data penelitian diambil dari tiga saluran YouTube: Kompas TV, Kompas Jawa Timur, dan Liputan6, dengan total 300 komentar yang diperoleh melalui platform ExportComments. Prosedur analisis meliputi langkah-langkah pra-pemrosesan, penandaan data, TF-IDF, dan klasifikasi dengan dua algoritma: Support Vector Machine (SVM) dan Naive Bayes, diakhiri dengan evaluasi model. Hasil penelitian menunjukkan bahwa metode Support Vector Machine (SVM) mencapai akurasi 73,33%, sedangkan metode Naive Bayes mencapai akurasi 67%. Hasil ini menunjukkan bahwa SVM mengungguli Naive Bayes dalam mengklasifikasikan sentimen dari komentar YouTube.

Kata Kunci: Analisis Sentimen, YouTube, Pemblokiran Roblox, Support Vector Machine, Naive Bayes

ABSTRACT

The development of information technology has changed the way people express their opinions and interact online. One of the most frequently used means of public expression is the comments section on the YouTube platform. This study aims to analyze public views on the issue of the blocking of the Roblox application by the Indonesian government through YouTube user responses. The methods applied in this study are Support Vector Machine (SVM) and Naive Bayes (NB) which are used to categorize comments into three types of sentiment: positive, negative, or neutral. The research data was taken from three YouTube channels: Kompas TV, Kompas Jawa Timur, and Liputan6, with a total of 300 comments obtained through the ExportComments platform. The analysis procedure includes pre-processing steps, data tagging, TF-IDF, and classification with two algorithms: Support Vector Machine (SVM) and Naive Bayes, ending with model evaluation. The results show that the Support Vector Machine (SVM) method achieves an accuracy of 73.33%, while the Naive Bayes method achieves an accuracy of 67%. These results indicate that SVM outperforms Naive Bayes in classifying sentiment from YouTube comments.

Keywords: Sentiment Analysis, YouTube, Roblox Blocking, Support Vector Machine, Naive Bayes

PENDAHULUAN DAN TINJAUAN PUSTAKA

Perkembangan teknologi informasi dan komunikasi selama dua dekade terakhir telah mengubah cara manusia berinteraksi, menyebarkan informasi, dan membentuk opini publik. Selain menyediakan hiburan, situs media sosial seperti *YouTube*, *Twitter*, *Facebook*, dan platform game online seperti Roblox telah muncul sebagai tempat utama untuk wacana publik, pembentukan opini, dan mobilisasi sosial. Para peneliti, legislator, dan profesional bisnis dapat mempelajari banyak hal tentang emosi publik, preferensi, dan respons terhadap peristiwa atau undang-undang dari konten yang dihasilkan pengguna seperti unggahan, ulasan, dan komentar (Amrullah, M. S., & Pane, S. F 2023).

Untuk menganalisis sentimen ini, dua algoritma pembelajaran mesin digunakan: *Support Vector Machine (SVM)* dan *Naïve Bayes*. Dalam penelitian analisis sentimen, kedua metode kategorisasi teks ini sering digunakan. Karena dapat mengidentifikasi pemisah optimal antara kelas data, *Support Vector Machine (SVM)* terkenal karena akurasi yang sangat baik saat menangani data skala besar, termasuk teks komentar. *Naïve Bayes*, di sisi lain, merupakan metode yang cocok untuk perbandingan karena keunggulannya dalam hal kesederhanaan, ekonomi komputasi, dan kinerja yang andal pada dataset yang besar. Salah satu platform game online yang terkenal adalah aplikasi Roblox. Roblox lebih dari sekadar aplikasi game biasa, ini adalah lingkungan digital di mana pengguna, khususnya anak-anak dan remaja, dapat membuat, berbagi, dan memainkan game yang dibuat oleh komunitas mereka sendiri. Dengan memungkinkan pengguna untuk membuat dan memainkan game yang dibuat oleh pengguna lain, Roblox telah menarik banyak perhatian. Dengan jutaan pengguna aktif setiap hari, Roblox telah tumbuh menjadi sensasi di seluruh dunia. Pemblokiran aplikasi Roblox adalah salah satu contoh kontroversi seputar popularitas ini (Hidayat, H 2024).

Kebijakan pemblokiran ini secara normatif bertujuan untuk melindungi anak-anak dari paparan konten yang berpotensi berbahaya, termasuk risiko kekerasan dan perundungan siber yang dapat muncul melalui penggunaan platform tersebut. Namun, implementasi kebijakan ini tidak lepas dari kontroversi. Beberapa pihak memandang pemblokiran sebagai tindakan pencegahan untuk menjaga keamanan digital.

Kontroversi mengenai Pemblokiran aplikasi Roblox telah memicu perdebatan publik yang sengit, ditandai dengan munculnya pandangan yang saling bertentangan. Media sosial, khususnya platform seperti YouTube, berfungsi sebagai ruang wacana publik tempat orang-orang mengekspresikan pendapat mereka, baik yang mendukung maupun menentang kebijakan tersebut. Kebijakan pemblokiran ini secara normatif bertujuan untuk melindungi anak-anak dari paparan konten yang berpotensi berbahaya, termasuk risiko kekerasan dan perundungan siber yang dapat muncul melalui penggunaan platform tersebut. Pada tanggal 6 dan 7 Agustus 2025, Menteri Negara Prasetyo Hadi menyatakan bahwa larangan dapat terjadi jika konten tersebut terbukti mengandung unsur kekerasan atau negatif terhadap aplikasi Roblox.

1. Rumusan Masalah

Berdasarkan latar belakang tersebut, maka rumusan permasalahan dalam penelitian ini adalah:

1. Bagaimana persepsi masyarakat terhadap pemblokiran Aplikasi Roblox yang tercermin melalui komentar-komentar di *YouTube*?
2. Bagaimana proses penerapan metode *Support Vector Machine dan Naïve Bayes (SVMNB)* dalam melakukan analisis sentimen terhadap komentar *YouTube* yang membahas pemblokiran Aplikasi Roblox?
3. Seberapa baik tingkat akurasi metode *Support Vector Machine dan Naïve Bayes (SVMNB)* dalam mengklasifikasikan sentimen komentar pengguna *YouTube* ke dalam kategori positif, negatif atau netral?

2. Tujuan Penelitian

Tujuan penelitian ini adalah untuk menganalisis sentiment tentang Pemblokiran Aplikasi Roblox menggunakan metode *Support Vector Machine dan Naïve Bayes (SVMNB)*. Penelitian ini bertujuan antara lain:

1. Mengumpulkan dan mengolah data komentar *YouTube* yang membahas pemblokiran Aplikasi Roblox. Komentar-komentar tersebut digunakan sebagai data untuk mengidentifikasi opini publik apakah bersifat positif, negatif, atau netral.
2. Menerapkan metode *Support Vector Machine dan Naïve Bayes (SVMNB)* untuk menentukan kategori sentimen dari setiap komentar *Youtube* yang membahas pemblokiran aplikasi Roblox. Penggunaan *Support Vector Machine dan Naïve Bayes (SVMNB)* diharapkan dapat memberikan hasil klasifikasi yang akurat karena metode ini efektif dalam mengolah data teks berdimensi tinggi dan mampu membedakan pola opini dengan baik.
3. Mengklasifikasikan komentar menjadi kategori sentimen positif, negatif, dan netral.

3. Analisis Sentimen

Analisis sentimen merupakan teknik pada bidang kecerdasan protesis dan pemrosesan bahasa alami (NLP) yang dipakai buat mengidentifikasi dan mengukur opini, emosi, atau perilaku subjektif yang terkandung pada teks. Tujuan utama dari analisis sentimen adalah buat mengidentifikasi dan mengklasifikasikan kesamaan opini yang masih ada pada sebuah teks, baik secara eksplisit juga implisit, guna memilih apakah sentimen tersebut bersifat positif, negatif, atau netral menurut konteks, struktur kalimat, dan pemilihan istilah yang dipakai sang penulis pada membicarakan pandangannya. Tujuan berdasarkan analisis sentimen dalam dasarnya merupakan buat tahu bagaimana sikap, opini, juga emosi seorang atau grup terhadap suatu objek eksklusif yang terekspresikan melalui teks (Safitri, R., & Alfira, N 2021).

4. Aplikasi Roblox

Aplikasi Roblox adalah game online gratis yang memungkinkan pengguna menciptakan dan menikmati game menurut pengguna lain, menggunakan lebih 150 juta pengguna aktif pada lebih 200 negara. Ini lebih terkenal daripada *Minecraft* yang mempunyai 90 juta pengguna aktif bulanan. Roblox menunjukkan aneka macam permainan tanpa biaya tambahan dan bisa diakses melalui aneka macam perangkat misalnya komputer, ponsel, dan konsol game, dan mendukung platform misalnya *Windows, iOS, MacOS, Xbox One, dan Android*(Fauzi, A., & Yunial, A. H 2024).

5. Machine Learning

Machine Learning (Pembelajaran Mesin) merupakan salah satu pilar utama dalam bidang kecerdasan buatan (*Artificial Intelligence/AI*) yang berhubungan dengan pengembangan algoritma yang mampu belajar melakukan tugas otomatis atas dasar sejumlah besar seperti (data), tanpa adanya program sebelum (Novia, E. A., Rahayu, W. I., & Prianto, C. 2020).

Machine Learning bertujuan adalah membangun sistem komputer yang mampu belajar dan meningkatkan kinerjanya dari data tanpa diprogram secara eksplisit, sehingga mampu mengotomatiskan tugas berulang, menemukan pola dan wawasan baru, membuat prediksi akurat, dan mengoptimalkan pengambilan keputusan untuk memecahkan masalah kompleks dan mendorong pertumbuhan bisnis (Widagdo, A. S., & Saputro, F. E. N 2024).

6. Support Vector Machine (SVM)

Support Vector Machine (SVM) adalah metode dari *supervised learning* untuk menganalisis data dan mengidentifikasi pola pada pengelompokan data. *Support Vector Machine (SVM)* mengkonversi sebuah teks menjadi vektor sebelum digunakan pada klasifikasi. Ide dasar dari *Support Vector Machine* ini yaitu mendapatkan area keputusan optimal (*Hyperlane*) untuk setiap titik data (Riwanto, M. H 2025).

Support Vector Machine (SVM) juga unggul dalam menangani data berdimensi tinggi dan mampu mengatasi permasalahan yang terkait dengan dimensionalitas. *Support Vector Machine (SVM)* digunakan untuk membuat model klasifikasi sehingga dapat digunakan untuk memprediksi sentimen teks secara lebih akurat berdasarkan pola yang dipelajari dari data train. Salah satu kelebihan dari *SVM* adalah algoritma ini kebal terhadap adanya *overfitting* (kondisi yang terlalu rumit dan terlalu sesuai dengan data training) atau dapat dikatakan mempunyai risiko *overfitting* rendah, terlebih saat dipergunakan pada proses normalisasi. Sedangkan kekurangan dari *SVM* adalah membutuhkan waktu lama untuk proses pembelajaran (learning), khususnya pada dataset yang mempunyai ukuran sangat besar dan non-linier yang lebih kompleks (Purnamasari, D 2023).

7. Youtube

YouTube adalah sebuah situs *web* video sharing (berbagi video) yang populer dimana para pengguna dapat memuat, menonton, dan berbagi klip video secara gratis. *YouTube* digunakan sebagai sumber data utama karena memiliki interaksi yang sangat tinggi dan menampung opini publik secara langsung terhadap isu tertentu. Didirikan pada bulan februari 2005 oleh 3 orang mantan karyawan PayPal, yaitu Chad Hurley, Steve Chen dan Jawed Karim. Umumnya video-video di *YouTube* adalah video klip film, TV, serta video buatan para penggunanya sendiri (Narulita, S 2025).

8. Naïve Bayes

Naive Bayes adalah metode klasifikasi yang sering digunakan dalam analisis sentimen karena kesederhanaan dan efektivitasnya. Kelebihan dari algoritma ini termasuk kemampuannya untuk menangani data besar dan *imbalanced data*, meskipun ada beberapa kekurangan, seperti asumsi independensi antar fitur. Perbandingan *Naive Bayes* dengan algoritma lain, seperti KNN dan SVM, menunjukkan bahwa *Naive Bayes* sering kali memberikan hasil yang kompetitif dalam konteks analisis sentimen (Urva, G. 2023).

9. Export Comment

Export Comment adalah salah satu *website* yang berguna untuk mengambil data pada sosial media seperti *Instagram*, *Youtube*, *Twitter* dll. Data diambil dari komentar pada sosial media (*Youtube*), yang bisa di download dalam format *Excel* dan *CSV*(Arif, E., Suherman, S., & Widodo, A. P 2025).

10. Python

Python adalah salah satu bahasa pemrograman yang populer digunakan oleh perusahaan besar serta para pengembang untuk menciptakan berbagai jenis aplikasi, termasuk aplikasi berbasis *desktop*, *web*, dan *mobile*. Bahasa pemrograman ini awalnya diciptakan oleh Guido van Rossum di Belanda pada tahun 1990 dan mengambil nama dari acara televisi favorit Guido, yaitu *Monty Python's Flying Circus*. Awalnya, *Python* dikembangkan oleh Van Rossum sebagai hobi pribadinya, namun seiring berjalannya waktu, bahasa pemrograman ini menjadi sangat berpengaruh dalam industri dan pendidikan. *Python* dikenal karena sintaksisnya yang sederhana, fleksibel, dan memiliki banyak pustaka pendukung untuk analisis data, seperti *pandas*, dan *scikit-learn*(Martono, A., Henderi, & Padelis 2025).

METODE/EKSPERIMEN

1. Metode Support Vector Machine (SVM)

Metode *Support Vector Machine (SVM)* bekerja dengan mencari *hyperplane* optimal yang memisahkan kelas-kelas dalam data secara maksimal. *Support Vector Machine* sangat efektif dalam menangani data berdimensi tinggi dan sering digunakan untuk analisis data yang kompleks. *Support Vector Machine* menggunakan *kernel trick* untuk memetakan data *non-linear* ke ruang yang lebih tinggi, sehingga kelas dapat dipisahkan dengan lebih baik. *Support Vector Machine* sering kali unggul dibandingkan algoritma lain dalam analisis sentimen, terutama ketika data memiliki pola yang kompleks atau distribusi yang tidak seimbang. Algoritma ini dikenal karena keakuratannya yang tinggi dalam berbagai jenis dataset, termasuk yang memiliki dimensi tinggi atau distribusi kelas yang tidak merata(Mola, S. A. S., Djawa, S. N. R., & Mauko, A. Y 2025).

2. Metode Naïve Bayes

Metode *Naive Bayes* merupakan algoritma klasifikasi berbasis *Teorema Bayes* yang menggunakan pendekatan probabilistik. *Naive Bayes* mengasumsikan bahwa setiap fitur (kata) dalam teks bersifat independen satu sama lain. Dalam analisis sentimen, metode ini menghitung peluang suatu teks termasuk dalam kelas positif, negatif, atau netral berdasarkan frekuensi kemunculan kata tertentu(Maulana, B. A., Fahmi, M. J., Imran, A. M., & Hidayati, N 2024).

3. Sumber Data

Sumber data dalam penelitian ini berasal dari komentar pengguna pada platform *YouTube* yang membahas isu pemblokiran aplikasi *Roblox* oleh pemerintah Indonesia. Data komentar diambil dari tiga kanal berita, yaitu *Kompas TV*, *Kompas Jawa Timur*, dan *Liputan6*, yang mengunggah video terkait topik tersebut. Pengambilan data dilakukan pada periode 20 Agustus 2025 sampai dengan 20 September 2025 dengan memanfaatkan platform *ExportComments* sebagai alat bantu pengumpulan data. Dalam penelitian analisis sentimen ini, untuk mendapatkan

300 data (komentar) peneliti mengambil keputusan untuk menganalisis 100 data (komentar) dari setiap postingan (Hendrawan, I. R., & Utami, E 2023).

4. Metode Pengumpulan Data

Penelitian ini, menggunakan beberapa metode dalam pengumpulan data, seperti dibawah ini;

1.1. Studi Literatur

Studi Literatur adalah referensi ilmiah yang digunakan untuk memperoleh pengetahuan tertentu. Dilakukan sebagai langkah awal untuk memahami dasar teori, metode dan meninjau hasil penelitian terdahulu yang relevan dengan topik analisis sentimen serta pemblokiran aplikasi Roblox (Asri, Y., & Fajri, M 2024).

1.2. Data Sekunder

Metode utama dalam penelitian analisis sentimen adalah pengumpulan data sekunder yang bersumber dari media digital seperti:

- 1) Media sosial: *YouTube*
- 2) Berita: Kompas TV, Kompas Jawa Timur dan Liputan6

Langkah-langkahnya:

1. Menentukan sumber data

- a. Opini publik di komentar Youtube tentang “Kebijakan Pemblokiran Aplikasi Roblox” seperti di Kompas TV, Kompas Jawa Timur dan Liputan6.
- b. Buku dan Literatur ilmiah sebagai rujukan akademik

2. Menentukan kata kunci (*Keywords*) atau hashtag dan menentukan periode waktu pengambilan data

- a. “Pemerintah blokir Roblox KompasTV Jawa Timur dan #komdigi
- b. 20 Agustus 2025 – 20 September 2025

3. Mengambil Data menggunakan

- a. *ExportComment*
- b. *Dataset Publik* dari *Kaggle*, *Google Dataset Search*, atau sumber terbuka lainnya.

1.3. Data Primer

- a. Kuesioner Online (*Google Form*, *Microsoft Form*, *Survey Monkey* dan lain-lain), untuk mengumpulkan opini singkat yang akan diklasifikasikan sebagai positif, negatif, atau netral.

Komentar pada *YouTube*, untuk menggali ekspresi atau persepsi pengguna terhadap topik pada pemblokiran aplikasi Roblox.

1.4. *Crawling Data*

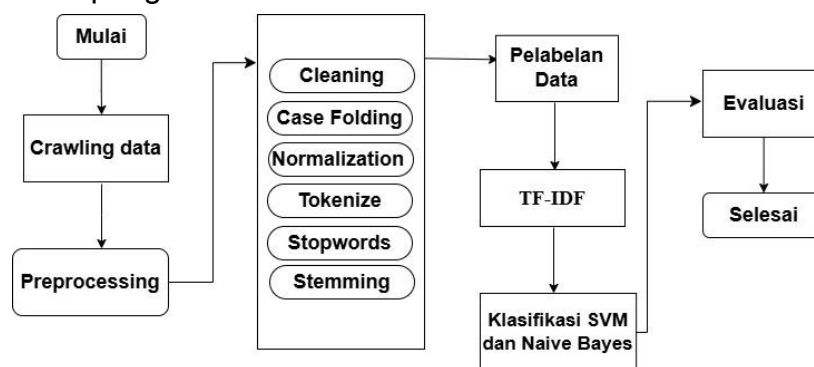
Crawling data adalah proses otomatis yang dilakukan oleh program komputer atau bot untuk mengumpulkan data dari berbagai sumber, terutama dari situs *web*. Proses ini dimulai dengan program tersebut mengunjungi satu atau beberapa URL awal yang sudah ditentukan, kemudian secara sistematis mengikuti tautan-tautan yang ada di halaman tersebut untuk menjelajahi halaman-halaman *web* lain yang terhubung (Setiawan, Z 2023).

Dalam penelitian analisis sentimen ini, untuk mendapatkan 300 data (komentar) peneliti mengambil keputusan untuk menganalisis 100 data (komentar) dari setiap postingan. Peneliti menetapkan jumlah sampel 300 data (komentar) untuk penentuan

jumlah ini mengacu pada teori pemilihan sampel berdasarkan relevansi dan kesesuaian data dengan topik penelitian. Jumlah 300 data (komentar) juga sejalan dengan teori ukuran sampel pada penelitian teks (*text mining*), yang menyebutkan bahwa 200 - 500 data (komentar) sudah memadai untuk menghasilkan pola sentimen yang stabil dan dapat dilatih menggunakan metode *Support Vector Machine (SVM)* dan *Naïve Bayes* (Mola, S. A. S., 2025).

5. Tahapan Penelitian

Tahapan penelitian merupakan serangkaian langkah-langkah yang bertujuan untuk memberikan gambaran tentang proses alur yang akan dilakukan secara sistematis dan Terukur. Adapun alat bantu yang digunakan pada penelitian ini yaitu menggunakan *google collab* yang berfungsi untuk mendukung berbagai tahapan penelitian dalam pengolahan sebuah dataset.



Gambar 1. Tahapan Penelitian

Pada gambar 1. menunjukkan langkah-langkah yang dilakukan pada saat penelitian yang meliputi pengumpulan data (*Crawling Data*), *Preprocessing*, *Labeling*, *TF-IDF* (*Term Frequency-Inverse Document Frequency*), *Klasifikasi Support Vector Machine dan Naïve Bayes* dan *Evaluasi*.

Berikut langkah-langkah Tahapan penelitian yang digunakan:

- 1) *Pengumpulan Data (Crawling Data)*
Data diperoleh dari kolom komentar *YouTube* pada tiga kanal berita, yaitu Kompas TV, Kompas Jawa Timur, dan Liputan6, yang membahas tentang isu pemblokiran aplikasi Roblox, dimana tanggal pengambilan data terhitung dari 20 Agustus 2025 sampai dengan 20 September 2025.
- 2) *Preprocessing Data*
Data komentar yang sudah diperoleh masih mengandung banyak unsur yang tidak relevan, seperti tanda baca, emoji, URL, dan kata tidak baku. Oleh karena itu dilakukan beberapa tahap pembersihan, antara lain:
 - a. *Cleaning*: Membersihkan teks dari simbol, tanda baca, angka, dan emoji.
 - b. *Case Folding*: Mengubah seluruh huruf menjadi huruf kecil
 - c. *Tokenize*: merupakan proses memotong setiap kata dalam kalimat dengan pemisah tanda petik (').
 - d. *Stopword Removal*: Menghapus kata umum yang tidak memiliki makna penting (misalnya "dan", "yang", "di").
 - e. *Stemming*: mengubah setiap kata ke bentuk dasarnya agar seragam, misalnya "dilakukan" menjadi "laku".
 - f. *Normalization*: Mengubah kata-kata tidak baku menjadi yang baku.
- 3) *Pelabelan Data*

Proses pelabelan data yaitu sebuah tahap untuk melakukan pemberian label positif, negatif atau netral. Dalam hal ini label positif menunjukkan dukungan terhadap konten, label negatif menunjukkan ketidaksetujuan, dan label netral menunjukkan netralitas terhadap konten.

4) *TF-IDF*

TF-IDF (Term Frequency-Inverse Document Frequency) adalah sebuah teknik yang digunakan untuk mengukur sejauh mana suatu kata penting dalam sebuah dokumen jika dibandingkan dengan keseluruhan kumpulan dokumen. *TF-IDF* terdiri dari dua komponen utama:

- a. *Term Frequency (TF)*: Ini adalah frekuensi kemunculan sebuah kata (term) dalam sebuah dokumen. Semakin sering sebuah kata muncul dalam dokumen, semakin tinggi nilai *TF*-nya.
- b. *Inverse Document Frequency (IDF)*: Ini adalah ukuran seberapa penting sebuah kata dalam seluruh kumpulan dokumen. Kata-kata yang jarang muncul diseluruh dokumen akan memiliki nilai *IDF* yang lebih tinggi, karena dianggap lebih penting/khas.

5) Klasifikasi Sentimen dengan Algoritma *Support Vector Machine* dan *Naïve Bayes*

Data yang sudah dikonversi kemudian dianalisis menggunakan dua metode, yaitu:

- a. *Support Vector Machine (SVM)*: Digunakan untuk mencari garis pemisah terbaik antara sentimen positif, negatif, atau netral.
- b. *Naïve Bayes*: Digunakan sebagai pembanding karena sederhana namun efektif dalam menangani data teks besar.

6) Evaluasi Model

Untuk mengukur performa algoritma *Support Vector Machine* dan *Naïve Bayes*, digunakan beberapa metrik evaluasi seperti:

- a. *Accuracy* (akurasi): seberapa banyak prediksi yang benar dibandingkan keseluruhan data.
- b. *Precision* (ketepatan): seberapa tepat model dalam memprediksi suatu kelas sentimen.
- c. *Recall* (daya tangkap): seberapa baik model mengenali seluruh data dari kelas tertentu.
- d. *F1-Score*: rata-rata harmonik antara *precision* dan *recall* untuk menilai keseimbangan model (Nugraha, F. A. 2020).

HASIL DAN PEMBAHASAN

1. Pengumpulan Data

Data diambil dari komentar pada sosial media (*Youtube*) melalui teknik crawling data menggunakan situs website *exportcomments*. Komentar dikumpulkan dari tiga kanal berita, yaitu Kompas TV, Kompas Jawa Timur, dan Liputan6, yang mengunggah video terkait topik tersebut. Pengambilan data dilakukan pada periode 20 Agustus 2025 sampai dengan 20 September 2025. Data yang diperoleh kemudian disimpan dalam format CSV untuk selanjutnya diproses dan dianalisis.

2. Preprocessing Data

Adapun tahapan-tahapan pemrosesan *preprocessing data* sebagai berikut:

a) *Cleaning Data*

Proses *Cleaning Data* merupakan tahapan awal dalam pengolahan data yang bertujuan untuk meningkatkan kualitas data dengan cara menghilangkan elemen-elemen yang tidak relevan. Pada tahap ini, teks komentar dibersihkan dari tanda baca, simbol, angka, tautan (*URL*), emoji, serta spasi berlebih. Tahapan ini dilakukan untuk menghasilkan data teks yang lebih terstruktur, konsisten, dan siap digunakan dalam proses analisis sentimen.

Hasil pembersihan data di bawah ini menghasilkan teks komentar yang lebih bersih, terstruktur, dan siap untuk diproses pada tahap selanjutnya, seperti *case folding*, *tokenizing*, dan ekstraksi fitur. Data yang telah dibersihkan ini diharapkan dapat meningkatkan kualitas analisis dan akurasi model klasifikasi sentimen yang digunakan dalam penelitian.

Tabel 1. Hasil *Cleaning*

Sebelum	Sesudah
Jangan diblokir pak 🙏🙏🙏🙏🙏🙏 itu game satu satunya loh dari 2024 pls pak jangan diblokir aku nanti kangen roblox 🙏🙏🙏🙏🙏🙏	Jangan diblokir pak itu game satu satunya loh dari pls pak jangan diblokir aku nanti kangen roblox
judi gak di blokir koruptor gak diblokirrobox diblokirapa maksud nyaaaa Roblox diblokirJalanan rusak dibiarin 😞👉	judi gak di blokir koruptor gak di blokir roblox di blokir apa maksut nyaaaa Roblox diblokir jalanan rusak dibiarin

Berdasarkan hasil pembersihan data (*cleaning*) di atas ini yang ditampilkan pada tabel di atas ini, dapat dijelaskan bahwa proses *cleaning* berhasil menyederhanakan dan merapikan data komentar tanpa menghilangkan makna utama dari setiap komentar.

b) Case Folding

Pada tahap *Case Folding*, data komentar yang telah melalui proses *cleaning* selanjutnya diproses dengan cara mengubah seluruh huruf menjadi huruf kecil. Tahap ini bertujuan untuk menyeragamkan bentuk teks, sehingga tidak ada perbedaan antara huruf besar dan huruf kecil yang dapat memengaruhi hasil analisis sentimen.

Tabel 2. Hasil *case_folding*

Sebelum	Sesudah
woyyyy knp koruptor dibiarkan giliran game jadi masalah besar Oh aman aja gw maen roblox dpr gk di urus hhhh	woyyyy knp koruptor dibiarkan giliran game jadi masalah besar oh aman aja gw maen roblox dpr gk di urus hhhh

c) Normalisasi

Tahap *normalisasi* berfungsi untuk mengubah kata-kata tidak baku, singkatan, atau bahasa gaul menjadi bentuk kata yang lebih baku dan standar. Tujuan dari tahap ini adalah agar teks menjadi lebih mudah dipahami oleh sistem dan meningkatkan akurasi dalam proses analisis sentimen.

Tabel 3. Hasil *Normalisasi Kata*

Sebelum	Sesudah
roblox kalo ngomong kasar bisa di sensor broooo	roblox kalau mengomong kasar bisa di sensor brooo
Judi onlen juga hapus	judi onlen juga hapus
woyyyy knp koruptor dibiarkan giliran game jadi masalah besar	woyyyy kenapa koruptor dibiarkan giliran game jadi masalah besar

d) Tokenize

Pada tahap *tokenize*, sistem membaca sebuah kalimat atau teks utuh dan kemudian memisahkan setiap katanya menjadi bagian-bagian yang berdiri sendiri, supaya komputer bisa memahami dan memprosesnya dengan lebih mudah. Misalnya, jika ada kalimat “*Saya sedang belajar pemrosesan bahasa alami*”, maka proses *tokenisasi* akan mengubahnya menjadi daftar kata seperti ['Saya', 'sedang', 'belajar', 'pemrosesan', 'bahasa', 'alami']. Dengan bentuk seperti ini, komputer dapat mengenali setiap kata secara terpisah dan melakukan analisis yang lebih terarah.

Tabel 4. Hasil *Tokenize*

Sebelum	Sesudah
Oh aman aja gw maen roblox dpr gk di urus hhhh	[oh,aman,saja,gue,main,roblox,dpr,tidak,di,urus,hhhh]
Roblox diblokirJalanan rusak dibiarin 🙄👉	[roblox,diblokirjalanan,rusak,dibiarkan]
Jangan diblokir pak 🙄🙄🙄🙏🙏🙏 itu game satu satunya loh dari 2024 pls pak jangan diblokir aku nanti kangen roblox 🙄🙄🙄🙄🙄🙏🙏	[jangan,diblokir,pak,itu,game,satu,satunya,l oh,dari,pls,pak,jangan,diblokir,aku,nanti,kangen,roblox]

e) Removal_stopwords

Proses *stopword removal* adalah tahap dalam pengolahan teks yang berfungsi untuk menghapus kata-kata umum (sering muncul) yang dianggap tidak memiliki makna penting atau tidak memberikan kontribusi signifikan terhadap analisis isi teks.

Tabel 5. Hasil *Stopwords Removal*

Sebelum	Sesudah
[woyyyy,kenapa,koruptor,dibiarkan,giliran,game,jadi,masalah,besar]	[woyyyy,koruptor,dibiarkan,giliran,game]
[jangan,diblokir,pak,itu,game,satu,satunya,loh,dari,pls,pak,jangan,diblokir,aku,nanti,kangen,roblox]	[diblokir,game,satunya,loh,pls,diblokir,kangen,roblox]

f) Proses Stemming Data

Proses *Stemming Data* merupakan tahap penting dalam pengolahan teks (atau *text preprocessing*) yang bertujuan untuk mengembalikan setiap kata ke bentuk dasarnya atau bentuk akarnya (*root word*), dengan cara menghapus

Berdasarkan Gambar diatas, hasil pelabelan data menunjukkan bahwa mayoritas komentar yang dianalisis memiliki kecenderungan sentimen negatif, yang menggambarkan adanya banyak kritik dan keluhan dari pengguna terhadap topik yang diteliti. Hasil pelabelan ini menjadi dasar penting dalam proses klasifikasi sentimen, karena label yang diberikan akan digunakan untuk melatih dan mengevaluasi kinerja model analisis sentimen pada tahap selanjutnya.

5. Ekstrak Fitur Menggunakan *TF-IDF*

Gambar di bawah ini menunjukkan hasil dari proses representasi data teks menggunakan metode *TF-IDF* (*Term Frequency–Inverse Document Frequency*), yaitu salah satu teknik yang digunakan dalam pengolahan bahasa alami (*Natural Language Processing*) untuk mengubah teks menjadi bentuk numerik agar dapat diproses oleh komputer atau algoritma pembelajaran mesin.

```

=== Matriks TF-IDF ===
  analisis belajar berguna berkaitan dan data learning machine \
0  0.00  0.443  0.00  0.00  0.00  0.000  0.293  0.293
1  0.44  0.000  0.44  0.00  0.00  0.347  0.230  0.230
2  0.00  0.347  0.00  0.00  0.00  0.000  0.230  0.230
3  0.00  0.000  0.00  0.44  0.44  0.347  0.230  0.230

  membuat menjadi mudah python saling sangat saya science suka \
0  0.00  0.00  0.00  0.00  0.00  0.00  0.562  0.00  0.562
1  0.00  0.00  0.00  0.00  0.00  0.44  0.000  0.00  0.000
2  0.44  0.44  0.44  0.44  0.00  0.00  0.000  0.00  0.000
3  0.00  0.00  0.00  0.00  0.44  0.00  0.000  0.44  0.000

  untuk
0  0.00
1  0.44
2  0.00
3  0.00

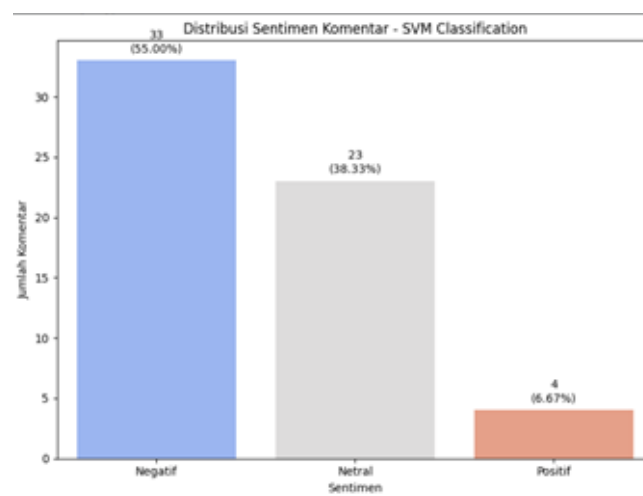
```

Gambar 4. Hasil Representasi Data Teks Menggunakan *Matriks TF-IDF*

Melalui representasi matriks *TF-IDF* ini, komputer dapat memahami teks dalam bentuk angka, sehingga selanjutnya data ini bisa digunakan untuk proses analisis sentimen, klasifikasi teks, atau pemodelan pembelajaran mesin lainnya.

6. Klasifikasi *Support Vector Machine* (SVM)

Grafik di bawah ini menampilkan distribusi sentimen dalam tiga kategori, yaitu negatif, netral, dan positif, berdasarkan jumlah komentar yang termasuk ke dalam masing-masing klasifikasi.



Gambar 5. Hasil Klasifikasi SVM

Berdasarkan grafik, dapat dilihat bahwa sentimen negatif merupakan kategori yang paling dominan, dengan total 33 komentar, atau sekitar 55% dari keseluruhan data. Dominasi komentar negatif ini menunjukkan bahwa sebagian besar pengguna memberikan respon yang tidak puas atau bernada kritik terhadap isu yang sedang dibahas. Kategori kedua yang paling banyak muncul adalah sentimen netral, yaitu sebanyak 23 komentar atau sekitar 38,33%. Komentar-komentar dalam kategori ini umumnya tidak menunjukkan emosi tertentu, bersifat informatif, atau hanya berupa tanggapan yang tidak mengarah pada opini positif maupun negatif.

Sementara itu, sentimen positif menjadi kategori dengan jumlah paling sedikit, yaitu 4 komentar, atau sekitar 6,67%. Persentase yang relatif rendah ini menunjukkan bahwa hanya sebagian kecil pengguna yang memberikan respon bernada dukungan atau apresiasi.

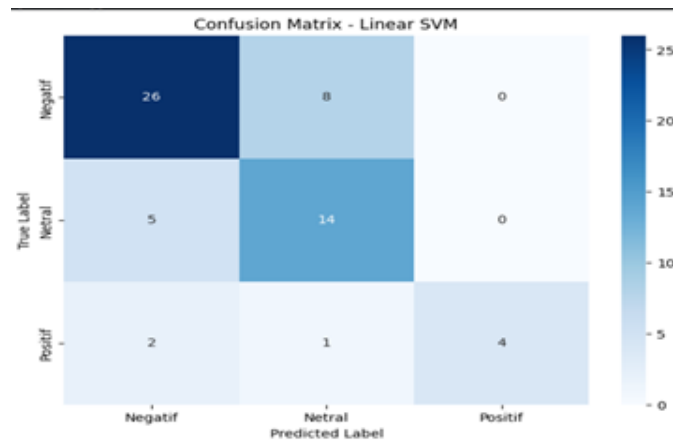
Secara keseluruhan, distribusi ini menggambarkan bahwa persepsi pengguna cenderung lebih banyak mengarah ke ketidakpuasan atau penolakan terhadap konteks yang dianalisis. Dominasi sentimen negatif ini dapat mengindikasikan adanya pemblokiran dianggap merugikan, tidak sesuai harapan, atau memicu reaksi emosional dari para pengguna.

7. Evaluasi Model SVM

Tabel 7. Evaluasi Model *Support Vector Machine* (SVM)

Kelas	Precision	Recall	F1-Score	Support
Negatif	0.79	0.76	0.78	34
Netral	0.61	0.74	0.67	19
Positif	1.00	0.57	0.73	7
Accuracy	-	-	0.73	60
Macro avg	0.80	0.69	0.72	60
Weighted avg	0.76	0.73	0.74	60

Hasil Evaluasi kinerja model diatas dilakukan menggunakan *metrik precision, recall, dan f1-score* untuk masing-masing kelas sentimen, yaitu negatif, netral, dan positif. Hasil evaluasi menunjukkan bahwa model memiliki kinerja yang lebih optimal dalam mengklasifikasikan sentimen negatif dan netral, yang tercermin dari nilai *precision* dan *recall* yang relatif lebih tinggi dibandingkan kelas lainnya. Pada kelas negatif, model mampu mengidentifikasi data dengan cukup baik, yang ditunjukkan oleh nilai *precision* sebesar 0,79 dan *recall* sebesar 0,76. Hal ini menunjukkan bahwa model tidak hanya mampu memprediksi sentimen negatif dengan tepat, tetapi juga mampu menemukan sebagian besar data negatif yang ada. Pada kelas netral, model menunjukkan performa yang cukup stabil dengan nilai *precision* sebesar 0,61 dan *recall* sebesar 0,74. Hasil ini mengindikasikan bahwa model cukup konsisten dalam mengenali sentimen netral, meskipun masih terdapat potensi kesalahan dalam proses klasifikasi. Sementara itu, pada kelas positif, model menunjukkan karakteristik yang berbeda. Meskipun nilai *precision* mencapai nilai maksimum (1,00), nilai *recall* yang lebih rendah (0,57) mengindikasikan bahwa model masih belum sepenuhnya mampu mendeteksi seluruh data positif yang terdapat dalam dataset. Hal ini menunjukkan bahwa model cenderung berhati-hati dalam memberikan label positif.



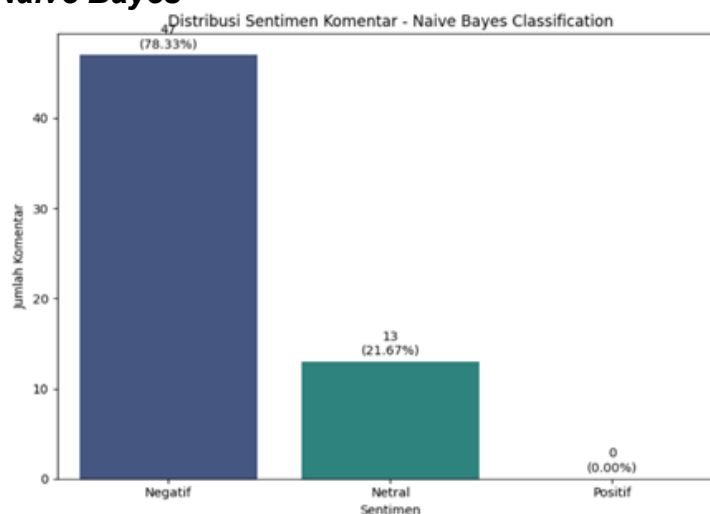
Gambar 6. Hasil *Confusion Matrix* Kernal Linear SVM

Dari gambar diatas hasil *confusion matrix*, terlihat bahwa *Support Vector Machine* (SVM) bekerja cukup baik, terutama pada sentimen negatif. Dari total 60 data uji, model berhasil memprediksi 26 komentar negatif dengan benar. Namun, ada juga 8 komentar negatif yang justru terbaca sebagai netral. Hal ini menunjukkan bahwa beberapa komentar bernada negatif memiliki sifat yang tidak terlalu keras, sehingga model terkadang menangkapnya sebagai kalimat netral. Untuk sentimen netral, model juga berhasil 14 komentar dengan tepat, tetapi masih terjadi salah klasifikasi sebanyak 5 komentar yang dianggap sebagai negatif. Kesalahan ini biasanya terjadi karena komentar netral sering mengandung kata-kata yang cenderung bernada kritik atau penolakan sehingga dianggap sebagai negatif oleh model. Pada kelas positif, hasilnya tidak sekuat dikelas lainnya. Model hanya berhasil mengenali 4 komentar positif secara benar, sementara beberapa komentar positif lainnya salah terbaca sebagai negatif atau netral. Jika dihitung menggunakan rumus akurasi,

$$Akurasi = \frac{\text{Jumlah Prediksi Benar}}{\text{Total Data}} \times 100\% = (26 + 14 + 4) / 60 \times 100\% = 73.33\%.$$

Dari nilai akurasi tersebut dapat disimpulkan bahwa *Support Vector Machine* (SVM) mampu mengklasifikasi komentar dengan tingkat ketepatan sekitar 73.33%, yang berarti performanya cukup baik

8. Klasifikasi *Naïve Bayes*



Gambar 7. Hasil Klasifikasi *Naïve Bayes*

Berdasarkan grafik diatas ini, terlihat bahwa sentimen negatif mendominasi dengan jumlah 47 komentar, atau sekitar 78,33% dari total data. Proporsi ini menunjukkan bahwa sebagian besar komentar yang dianalisis oleh model mengarah pada ketidakpuasan, kritik, atau penolakan terhadap pemblokiran yang dibahas. Dominasi sentimen negatif juga mengindikasikan bahwa persepsi publik cenderung kuat ke arah negatif.

Menariknya, pada hasil klasifikasi menggunakan *Naïve Bayes*, tidak ditemukan komentar dengan sentimen positif (0 komentar atau 0%). Tidak adanya komentar positif ini dapat menunjukkan dua kemungkinan:

- 1) Model cenderung sangat sensitif terhadap kata-kata bermakna negatif sehingga komentar ambigu diarahkan ke kategori negatif atau netral.
- 2) Memang sebagian besar komentar pada dataset tidak berisi dukungan atau rasa puas, sehingga model tidak menemukan pola kuat untuk sentimen positif.

Secara keseluruhan, hasil klasifikasi ini memperlihatkan bahwa persepsi pengguna dalam dataset cenderung sangat negatif, lebih ekstrem dibandingkan hasil analisis menggunakan metode lainnya. Hal ini juga dapat menjadi indikator bahwa *Naïve Bayes* memiliki kecenderungan tertentu dalam membedakan polaritas, khususnya pada data berbahasa informal seperti komentar media sosial.

9. Evaluasi Model *Naïve Bayes*

Berdasarkan Evaluasi Model *Naïve Bayes*, Nilai tertinggi yang muncul adalah recall sebesar 0.91 pada model *Naive Bayes*, yang menunjukkan bahwa model sangat baik dalam menemukan dan mengenali hampir seluruh data yang termasuk ke dalam kategori negatif. Sedangkan nilai yang paling rendah terdapat pada kelas sentimen Positif. Hal ini terlihat dari nilai *precision*, *recall*, dan *f1-score* yang semuanya bernilai 0,00 pada kelas positif. Nilai tersebut menunjukkan bahwa model tidak mampu mengenali ataupun memprediksi komentar positif dengan benar. Dengan kata lain, tidak ada satu pun data komentar positif yang berhasil diklasifikasikan secara tepat oleh model. Nilai *f1-score* untuk kelas Netral adalah sebesar 0,56. Nilai ini menunjukkan bahwa kemampuan model dalam mengklasifikasikan sentimen netral berada pada tingkat cukup, karena masih terdapat kesalahan dalam proses prediksi.

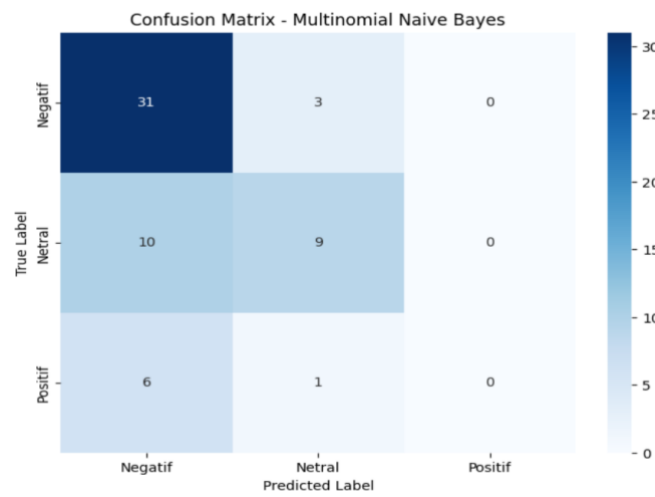
Tabel 8. Evaluasi Model *Naïve Bayes*

Kelas	Precision	Recall	F1-Score	Support
Negatif	0.66	0.91	0.77	34
Netral	0.69	0.47	0.56	19
Positif	0.00	0.00	0.00	7
Accuracy	-	-	0.67	60
Macro avg	0.45	0.46	0.44	60
Weighted avg	0.59	0.67	0.61	60

Pada hasil Evaluasi model, Algoritma *Multinomial Naïve Bayes* digunakan untuk memprediksi sentimen komentar yang telah melalui tahap *preprocessing*. Hasil evaluasi ditampilkan dalam bentuk *confusion matrix* yang memperlihatkan bagaimana model membaca komentar sebagai sentimen negatif, netral, atau positif.

Berdasarkan *confusion matrix* dibawah ini, model *Naïve Bayes* menunjukkan kinerja yang cukup baik dalam mengenali komentar negatif. Dari keseluruhan data, model berhasil memprediksi 31 komentar negatif dengan benar. Namun, masih terdapat 3 komentar negatif yang diprediksi sebagai netral. Ini berarti bahwa walaupun

model cukup mampu menangkap pola kata bernada negatif, tetapi tetap ada beberapa komentar yang bentuk bahasanya dianggap lebih mendekati netral oleh model. Untuk sentimen netral, model memprediksi 9 komentar secara benar, tetapi masih melakukan kesalahan pada 10 komentar netral yang justru diklasifikasikan sebagai negatif. Hal ini menunjukkan bahwa komentar netral sering kali memiliki pilihan kata yang mirip dengan komentar negatif, sehingga model cenderung salah menempatkannya. Kecenderungan ini juga menandakan bahwa *Naïve Bayes* lebih kuat dalam mengidentifikasi komentar negatif dibandingkan sentimen lainnya.



Gambar 8. Hasil *Confusion Matrix Multinomial Naïve Bayes*

Untuk sentimen positif, model tidak mampu mengidentifikasi satu pun komentar positif dengan benar. Dari total komentar positif, 6 komentar positif diprediksi sebagai negatif, dan 1 komentar diprediksi sebagai netral. Tidak adanya prediksi positif yang tepat menunjukkan bahwa model kesulitan mengenali kata-kata yang merepresentasikan sentimen positif. Hal ini biasanya terjadi karena jumlah data positif sangat sedikit, sehingga model tidak memiliki cukup contoh untuk belajar pola ciri khas kalimat positif. Jika dihitung menggunakan rumus akurasi,

$$Akurasi = \frac{\text{Jumlah Prediksi Benar}}{\text{Total Data}} \times 100 \% = \frac{30 + 9 + 0}{60} \times 100 \%$$

$$= \frac{31 + 3 + 0 + 10 + 9 + 0 + 6 + 1 + 0}{60} \times 100 \% = \frac{40}{60} \times 100 \% = 66.67\%$$

10. Rumusan Evaluasi Model

Evaluasi model dilakukan menggunakan *confusion matrix* dan *classification report*. *Confusion matrix* adalah alat yang digunakan untuk menilai kinerja model klasifikasi dengan membandingkan prediksi model dengan label sebenarnya, dari *confusion matrix* seperti akurasi, presisi, *recall*, dan *F1-score* dapat dihitung untuk menilai efektivitas model. Rumus perhitungannya sebagai berikut:

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

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

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

sebesar 73.33%, yang lebih tinggi dibandingkan dengan akurasi model *Naïve Bayes* sebesar 67%.

Perbedaan kinerja ini mengindikasikan bahwa algoritma SVM memiliki kemampuan yang lebih baik dalam mempelajari dan merepresentasikan pola-pola kompleks pada data teks. Sebaliknya, algoritma *Naive Bayes* menunjukkan keterbatasan dalam mengklasifikasikan data, khususnya pada kelas sentimen positif, yang berdampak pada penurunan performa secara keseluruhan.

Secara keseluruhan, hasil penelitian ini menunjukkan bahwa model SVM *Linear* lebih efektif dan lebih reliabel dibandingkan dengan model *Naive Bayes* dalam melakukan klasifikasi sentimen, sehingga SVM dapat direkomendasikan sebagai metode yang lebih optimal untuk digunakan dalam penelitian ini.

PENUTUP

1. Kesimpulan

Berdasarkan hasil penelitian mengenai analisis sentimen komentar *YouTube* terkait pemblokiran aplikasi Roblox menggunakan metode *Support Vector Machine (SVM)* dan *Naïve Bayes*, maka dapat ditarik beberapa kesimpulan sebagai berikut:

- a. Persepsi masyarakat terhadap pemblokiran aplikasi Roblox cenderung didominasi sentimen negatif. Dari hasil klasifikasi, sebagian besar komentar menunjukkan penolakan, kritik, atau ketidaksetujuan terhadap rencana pemblokiran. Hal ini mengindikasikan bahwa kebijakan tersebut dianggap merugikan bagi sebagian pengguna, terutama mereka yang melihat Roblox sebagai media hiburan dan kreativitas anak-anak.
- b. Metode *Support Vector Machine (SVM)* memiliki performa yang lebih baik dibandingkan *Naïve Bayes*. Proses evaluasi model menunjukkan bahwa SVM memiliki akurasi sebesar 73.33%, sedangkan *Naive Bayes* hanya mencapai 67%. Perbedaan ini memperlihatkan bahwa SVM lebih mampu mengenali pada bahasa komentar *Youtube* yang cenderung tidak baku dan memiliki struktur kalimat kompleks.

2. Saran

Meskipun memiliki hasil yang baik, penelitian ini masih memiliki beberapa kekurangan yang dapat menjadi perhatian untuk penelitian berikutnya.

- a. Jumlah data yang digunakan masih relatif sedikit, yaitu 300 komentar. Jumlah ini belum sepenuhnya merepresentasikan seluruh opini publik yang tersebar di media sosial, sehingga model masih berpotensi menghasilkan bias terhadap data tertentu.
- b. Bahasa yang digunakan dalam komentar *YouTube* sering kali tidak baku, mengandung singkatan, emoji, atau sarkasme yang sulit ditangkap oleh model, sehingga dapat menurunkan tingkat akurasi klasifikasi. Model belum mempertimbangkan analisis konteks atau kalimat kompleks seperti ironi, humor, atau kalimat ganda, yang sebenarnya sering muncul di media sosial.

UCAPAN TERIMAKASIH

Penulis mengucapkan terima kasih yang sebesar-besarnya kepada Ibu Destiarini dan Pak Joko Kuswanto sebagai pembimbing saya. Kesabaran dan dedikasi anda dalam membimbing saya sungguh luar biasa setiap pertemuan bimbingan adalah pelajaran berharga yang akan selalu saya ingat.

DAFTAR PUSTAKA

- Amrullah, M. S., & Pane, S. F. (2023). *Analisis sentimen masyarakat terhadap kebijakan polisi tilang manual di Indonesia*. Penerbit Buku Pedia.
- Hidayat, H., Santoso, F., & Lidimillah, L. F. (2024). *Analisis sentimen pengguna YouTube tentang Rohingya menggunakan algoritma SVM (Support Vector Machine)*. G-Tech: Jurnal Teknologi Terapan, 8(3), 1729–1738.
- Safitri, R., Alfira, N., Tamitadiani, D., Dewi, W. W. A., & Febriani, N. (2021). *Analisis sentimen: Metode alternatif penelitian big data*. Universitas Brawijaya Press.
- Fauzi, A., & Yunial, A. H. (2024). *Analisis sentimen (sentiment analysis): Evaluasi sentimen layanan dataset Twitter US Airline*. CV. Bintang Semesta Media.
- Novia, E. A., Rahayu, W. I., & Prianto, C. (2020). *Sistem perbandingan algoritma K-Means dan Naïve Bayes untuk memprediksi prioritas pembayaran tagihan rumah sakit berdasarkan tingkat kepentingan*. Kreatif Industri Nusantara.
- Widagdo, A. S., & Saputro, F. E. N. (2024). *Buku monograf analisis sentimen mobil listrik berbasis ulasan YouTube*. PT Sonpedia Publishing Indonesia.
- Riwanto, M. H., Ardhiyansyah, P., Adiansyah, R. A., Alfiansyah, A., Waek, G., & Fahlapi, R. (2025). *Analisis sentimen komentar YouTube terkait penerapan makan bergizi gratis menggunakan model algoritma SVM*. Kohesi: Jurnal Multidisiplin Saintek, 6(12).
- Purnamasari, D., Aji, A. B., Wulandari, D. A. P., Reza, F. A., Safrila, M. O., Yanda, N., & Hidayati, U. (2023). *Pengantar Metode Analisis Sentimen*. Depok, Penerbit: Gunadarma.
- Narulita, S., Sekarlangit, S., & Novianingrum, M. P. (2025). *Deteksi alergen pada produk pangan menggunakan algoritma Support Vector Machines (SVM)*. Bridge: Jurnal Publikasi Sistem Informasi dan Telekomunikasi, 3(1), 64–76. <https://doi.org/10.62951/bridge.v3i1.393>
- Urva, G., Desyanti, Albanna, I., Sungkari, M. S., Gunawan, I. M. A. O., Adhicandra, I., Ramadhan, S., Rahardian, R. L., Herlawati, Rahmadya, Trias Handayanto, A. G. B. A., Hartatik, & Atika, P. D. (2023). *Penerapan data mining di berbagai bidang: Konsep, metode, dan studi kasus*. PT Sonpedia Publishing Indonesia.
- Arif, E., Suherman, S., & Widodo, A. P. (2025). *Revolusi prediksi saham: Pemanfaatan machine learning dan analisis sentimen dalam dunia keuangan*. Greenbook Publisher.
- Mola, S. A. S., Djawa, S. N. R., & Mauko, A. Y. (2025). *Text mining: Analisis sentimen dengan Naïve Bayes*. Kaizen Media Publishing.
- Maulana, B. A., Fahmi, M. J., Imran, A. M., & Hidayati, N. (2024). *Analisis sentimen terhadap aplikasi Pluang menggunakan algoritma Naive Bayes dan Support Vector Machine (SVM)*. MALCOM: Indonesian Journal of Machine Learning and Computer Science, 4(2), 375–384.
- Hendrawan, I. R., & Utami, E. (2023). *Natural language processing: Eksplorasi sentimen masyarakat dalam evaluasi produk lokal Indonesia menggunakan algoritma bag of words, TF-IDF, Word2Vec, dan Doc2Vec*. Penerbit Andi.
- Asri, Y., & Fajri, M. (2024). *Implementasi lexicon vader dan naïve bayes pada aplikasi PLN mobile*. Uwais Inspirasi Indonesia.
- Setiawan, Z., Fajar, M., Priyatno, A. M., Putri, A. Y. P., Aryuni, M., Yuliyanti, S., Widiputra, H., Meilani, B. D., Ibrahim, R. N., Azdy, R. A., Junaidi, S., & Wijaya, A. (2023). *Buku ajar data mining*. PT Sonpedia Publishing Indonesia.

- Mola, S. A. S., Ohe Roma, R. V. K. I., & Widiastuti, T. (2025). *Text mining: Analisis sentimen dengan lexicon*. Kaizen Media Publishing.
- Nugraha, F. A., Harani, N. H., & Habibi, R. (2020). *Analisis sentimen terhadap pembatasan sosial menggunakan deep learning*. Kreatif.
- Martono, A., Henderi, & Padeli. (2025). *Unlock big data & machine learning: Menggunakan 14 algoritma unggulan untuk bisnis dan industri 5.0*. CV. Bintang Semesta Media.