

Analisis Sentimen Masyarakat Terhadap Pembatasan Akses Media Sosial Bagi Anak Pada *Platform* YouTube Menggunakan Metode *Support Vector Machine* (SVM)

Indah Sartika^{1*}, Imam Saputra²

^{1,2}, Sekolah Tinggi Ilmu Manajemen Sukma

*E-mail : indhsrtk@gmail.com

Information Article

History Article

Submission: 01-05-2026

Revision: 10-05-2026

Published: 14-05-2026

DOI Article:

10.62421/jibema.v3i4.251

ABSTRAK

Isu pembatasan akses media sosial bagi anak menimbulkan berbagai respons masyarakat di YouTube. Namun, kecenderungan sentimen masyarakat terhadap isu tersebut belum diketahui secara jelas karena komentar yang muncul memiliki bentuk opini yang beragam dan tidak terstruktur. Penelitian ini dilakukan dengan tujuan mengkaji kecenderungan tanggapan publik terhadap suatu isu melalui pendekatan analisis sentimen berbasis algoritma *Support Vector Machine* (SVM). Sumber data penelitian berasal dari seribu komentar pengguna YouTube yang diperoleh secara daring menggunakan metode *scraping*. Data yang telah terkumpul kemudian melalui beberapa tahapan pengolahan, seperti pembersihan dan normalisasi teks, pemberian bobot kata menggunakan metode TF-IDF, serta penyeimbangan distribusi data dengan teknik *oversampling* agar proses klasifikasi dapat berjalan lebih optimal. Hasil penelitian menunjukkan bahwa sentimen negatif menjadi yang paling dominan, yaitu 581 komentar atau 58,1%, diikuti sentimen netral sebanyak 229 komentar atau 22,9%, dan sentimen positif sebanyak 190 komentar atau 19,0%. Pengujian menunjukkan SVM memperoleh akurasi 0,81 serta cukup mampu mengenali sentimen positif dan netral, meskipun masih kesulitan membedakan komentar negatif dan netral.

Kata Kunci: Analisis sentimen, Youtube, SVM, text mining, media sosial

ABSTRACT

The issue of restricting social media access for children has sparked various responses from the public on YouTube. However, the trend of public sentiment towards this issue is not yet clearly known because the comments that appear take diverse and unstructured forms of opinion. This study was conducted with the aim of examining the trends of public responses to an issue through a sentiment analysis approach based on the Support Vector Machine (SVM) algorithm. The research data sources come from one thousand YouTube user comments obtained online using scraping methods. The collected data then went through several processing stages, such as text cleaning and normalization, word weighting using the TF-IDF method, as well as balancing the data distribution with oversampling

Acknowledgment

techniques so that the classification process can run more optimally. The research results show that negative sentiment is the most dominant, with 581 comments or 58.1%, followed by neutral sentiment with 229 comments or 22.9%, and positive sentiment with 190 comments or 19.0%. Testing shows that SVM achieved an accuracy of 0.81 and is quite capable of recognizing positive and neutral sentiments, although it still has difficulty distinguishing between negative and neutral comments.

Key word: *Sentiment analysis, YouTube, SVM, text mining, social media*

©2026 Published by JIBEMA. Selection and/or peer-review under responsibility of JIBEMA

PENDAHULUAN

Perkembangan teknologi digital saat ini sudah membawa perubahan besar dalam kehidupan masyarakat, terutama dalam hal mendapatkan informasi dan menyampaikan pendapat. YouTube menjadi salah satu layanan digital yang memiliki jumlah pengguna cukup besar. Melalui *platform* ini, pengguna tidak hanya menikmati berbagai jenis konten, tetapi juga aktif memberikan tanggapan dan pandangan mereka pada bagian komentar. Banyaknya komentar yang muncul di setiap video membuat YouTube bisa dijadikan sebagai sumber data untuk melihat bagaimana pandangan masyarakat terhadap suatu isu.

Salah satu isu yang cukup sering dibahas adalah pembatasan akses media sosial bagi anak. Isu ini menjadi penting karena penggunaan media sosial pada anak memiliki dampak yang tidak selalu positif. Di satu sisi, media sosial bisa membantu anak dalam belajar, mencari informasi, dan mengembangkan kreativitas. Namun di sisi lain, ada juga risiko seperti terpapar konten yang tidak sesuai usia, *cyberbullying*, kecanduan, serta penyalahgunaan data pribadi. Hal inilah yang kemudian menimbulkan berbagai pendapat di masyarakat. Ada yang mendukung pembatasan, tetapi ada juga yang tidak setuju. Beragam tanggapan dari masyarakat dapat ditemukan pada kolom komentar di *platform* YouTube yang menunjukkan adanya perbedaan pandangan terhadap suatu isu.

Melihat banyaknya komentar yang muncul, tentu akan sulit jika dianalisis secara manual. Selain jumlahnya yang besar, bentuk data komentar juga tidak terstruktur. Karena jumlah data komentar cukup banyak, diperlukan metode pengolahan data secara otomatis agar proses analisis dapat dilakukan dengan lebih efisien. Salah satu teknik yang sering diterapkan dalam pengolahan data berbasis teks ialah *text mining*. Menurut (Mola et al., 2025, hal. 11) *text mining* adalah pola yang digunakan diambil dari dokumen atau teks dan relatif tidak terstruktur. Selain itu, analisis sentimen juga digunakan untuk mengetahui kecenderungan opini dari data teks tersebut. Menurut (Asrumi et al., 2023, hal. 1) Dalam

bidang pengolahan data teks, analisis sentimen digunakan untuk memahami kecenderungan pandangan seseorang terhadap suatu hal. Kajian ini mencakup identifikasi opini, *respons* emosional, sudut pandang, hingga bentuk penilaian masyarakat terhadap berbagai topik, seperti layanan, produk, tokoh, organisasi, maupun peristiwa tertentu. cenderung positif, cenderung negatif, maupun bersifat netral. Pendekatan yang diterapkan pada penelitian ini menggunakan metode *machine learning* yaitu *Support Vector Machine* (SVM), karena metode ini cukup sering digunakan dan mampu memberikan hasil yang baik dalam klasifikasi teks.

Beberapa penelitian sebelumnya juga sudah banyak membahas analisis sentimen dari komentar YouTube. Penelitian yang dilakukan oleh Susanto & Valentina, (2025) menunjukkan bahwa metode SVM berperan dalam menghasilkan tingkat ketepatan mencapai 85,37% dalam mengklasifikasikan sentimen, walaupun masih terdapat masalah pada data yang tidak seimbang. Penelitian oleh Saputra et al., (2026) juga menggunakan metode SVM dan memperoleh akurasi sebesar 85,42%, dengan hasil yang menunjukkan tanggapan dengan kecenderungan negatif memiliki jumlah lebih tinggi dibandingkan sentimen lain. Selanjutnya, studi Karimah et al., (2024) memanfaatkan metode *Naive Bayes* serta menunjukkan bahwa tingkat ketepatan model meningkat hingga 70,69% setelah menggunakan teknik SMOTE. Penelitian oleh Wicaksono & Nastiti, (2024) menunjukkan hasil yang sangat tinggi dengan metode LSTM, yaitu akurasi mencapai 99,61% dan sentimen negatif menjadi yang paling dominan. Kajian terdahulu dari Zhafira et al., (2021) membuktikan bahwa penerapan metode *Naive Bayes* dengan TF-IDF mampu mencapai akurasi hingga 97%. Widiyanto, (2025) menemukan bahwa sentimen netral paling dominan dengan nilai sebesar 96,79%. Hidayat et al., (2024) menunjukkan bahwa metode SVM menghasilkan akurasi sebesar 77% dalam analisis sentimen.

Umrona et al., (2025) menggunakan metode KNN dan memperoleh akurasi sebesar 70,08%, namun hasilnya masih kurang baik dalam mengenali sentimen negatif. Md et al., (2024) menunjukkan bahwa penggunaan teknik SMOTE mampu menghasilkan peningkatan hasil klasifikasi dari 76,5% ke 78,0%. Sementara itu, Thomas et al., (2021) menyimpulkan bahwa analisis sentimen pada YouTube sangat penting untuk memahami opini masyarakat, walaupun hasilnya sangat dipengaruhi oleh metode yang digunakan. Penelitian Adelia et al., (2023) memperlihatkan bahwa metode *Naive Bayes* menghasilkan tingkat keberhasilan klasifikasi sentimen sebesar 100% pada kategori positif serta 66,39% pada kategori negatif. Widia et al., (2024) dalam penelitiannya menemukan bahwa komentar YouTube terkait pemindahan ibu kota cenderung didominasi oleh sentimen negatif. Penelitian Muasaroh et al., (2025) juga menunjukkan bahwa penggunaan metode SVM memberikan hasil analisis yang lebih akurat dengan persentase ketepatan 97% dibandingkan *Random Forest* sebesar 95%, dengan hasil sentimen yang sebagian besar negatif. Selain itu, Zulqarnain et al., (2025) menemukan bahwa diskusi pada komentar YouTube menunjukkan adanya perbedaan pandangan yang cukup kuat disertai meningkatnya tingkat toksisitas, terutama pada isu yang sensitif.

Dari beberapa penelitian tersebut, dapat dilihat bahwa analisis sentimen sudah banyak dilakukan pada berbagai topik seperti politik, bencana, konflik, dan teknologi. Namun, penelitian yang membahas secara khusus tentang pembatasan akses media sosial bagi anak masih belum banyak ditemukan. Padahal isu ini cukup penting karena berkaitan langsung dengan penggunaan teknologi oleh anak di era sekarang. Berdasarkan hal tersebut, penelitian ini memiliki kebaruan karena fokus pada analisis sentimen masyarakat terhadap pembatasan akses media sosial bagi anak melalui media YouTube. Selain itu, penelitian ini juga menggunakan metode SVM yang dikombinasikan dengan teknik *oversampling* untuk mengatasi masalah ketidakseimbangan data, sehingga diharapkan hasil klasifikasi menjadi lebih baik. Sebagai solusi, penelitian ini menggunakan pendekatan *text mining* dengan tahapan *preprocessing*, pembobotan TF-IDF, dan klasifikasi menggunakan metode SVM. Dengan cara ini, informasi yang semula tidak tersusun rapi dapat diolah sehingga lebih jelas untuk dianalisis. Hasil kajian ini diharapkan dapat membantu menjelaskan respons masyarakat terhadap isu tersebut dan mendukung penelitian lain yang berkaitan di masa mendatang.

METODE PENELITIAN

Proses penelitian dilakukan secara online dengan mengambil data dari platform YouTube sebagai objek utama pengumpulan data. Objek yang diteliti berupa komentar pengguna YouTube yang membahas isu pembatasan akses media sosial bagi anak. Penelitian dilaksanakan pada periode Maret sampai Mei 2026, dimulai dari tahap pengumpulan data, pengolahan data, hingga analisis dan evaluasi model. Data yang digunakan dalam penelitian ini terdiri dari data primer dan data sekunder. Sumber data primer merupakan sekumpulan informasi tentang suatu peristiwa atau objek, di mana proses pengumpulan datanya melibatkan beberapa orang yang dijadikan sebagai sampel penelitian (Soesana et al., 2023, hal. 36). Perolehan data primer dilakukan secara langsung dari komentar pengguna YouTube melalui proses *scraping* dengan menerapkan bahasa pemrograman *Python* yang dijalankan pada *Google Colab*. Pengambilan data diterapkan dengan menggunakan kata kunci yang berkaitan dengan objek kajian agar komentar yang dikumpulkan memiliki hubungan langsung dengan topik penelitian. Sedangkan data sekunder merupakan sumber yang tidak langsung memberikan data kepada pengumpul data, misalnya lewat orang lain atau lewat dokumen (Soesana et al., 2023, hal. 38). Selain data utama, penelitian ini juga memanfaatkan data sekunder dari berbagai referensi ilmiah, seperti jurnal dan penelitian sebelumnya, guna menunjang proses analisis.

Populasi dalam penelitian ini adalah seluruh komentar pengguna YouTube yang membahas pembatasan akses media sosial bagi anak. Besarnya populasi tidak dapat dihitung secara pasti dikarenakan data pada YouTube terus bertambah, sehingga termasuk dalam kategori populasi tidak terbatas. Oleh karena itu, dalam proses pengambilan sampel, peneliti memilih data tertentu yang relevan dengan penelitian menggunakan teknik *purposive sampling*. Sebanyak 1.000 komentar dijadikan

sampel penelitian setelah melalui tahap seleksi berdasarkan beberapa ketentuan, seperti relevansi komentar terhadap topik, penggunaan bahasa Indonesia, serta adanya opini pada isi komentar. Langkah tersebut dilakukan agar hasil analisis lebih sesuai dengan tujuan penelitian.

Variabel pada penelitian ini berpusat pada respons masyarakat terhadap isu pembatasan media sosial untuk anak-anak. Respons tersebut kemudian diklasifikasikan menjadi tiga kategori sentimen, yaitu kategori positif, kategori negatif, dan kategori netral. Penentuan kategori dilakukan melalui proses pelabelan data secara manual dengan melibatkan ahli bahasa. Hal ini dilakukan agar interpretasi terhadap setiap komentar lebih tepat, terutama pada komentar yang memiliki makna tidak langsung seperti sindiran atau sarkasme.

Pengumpulan data dilakukan melalui observasi dan dokumentasi. Observasi dilakukan dengan melihat dan mengamati langsung komentar pada video YouTube yang berkaitan dengan topik penelitian. Sementara itu, dokumentasi dilakukan dengan mengumpulkan data komentar menggunakan teknik *scraping* secara otomatis. Data hasil pengumpulan selanjutnya disimpan ke dalam file penyimpanan agar dapat diproses pada tahap berikutnya. Analisis data dalam penelitian ini menggunakan pendekatan *text mining*. Proses analisis dilakukan secara bertahap. Tahap awal adalah *preprocessing* data, yang meliputi *case folding*, *cleaning*, *normalization*, *tokenization*, *stopword removal*, dan *stemming*. Tahapan ini bertujuan untuk membersihkan data dari elemen yang tidak diperlukan serta menyederhanakan teks agar lebih mudah diproses. *Preprocessing* data merupakan proses penting dalam analisis *data mining* yang bertujuan untuk membersihkan, mengubah format, dan mempersiapkan data agar lebih mudah dan akurat dalam proses analisis (Daniswara & Nuryana, 2023, hal. 97). Dengan proses ini, data yang awalnya tidak terstruktur dapat diubah menjadi lebih rapi dan siap dianalisis.

Tahap berikutnya adalah melakukan pemberian bobot pada kata menggunakan metode TF-IDF (*Term Frequency–Inverse Document Frequency*) untuk mengukur seberapa penting suatu kata dalam dokumen. Metode tf-idf adalah metode yang menjelaskan pentingnya sebuah kata bagi sebuah dokumen dalam suatu koleksi data, dan itulah mengapa tf-idf menggabungkan parameter lokal dan global karena mempertimbangkan tidak hanya istilah yang terisolasi tetapi juga istilah dalam koleksi dokumen (Afda, 2024, hal. 468). Metode ini berfungsi untuk mengalokasikan bobot pada tiap kata, sehingga kata yang memiliki peran lebih signifikan dalam sebuah dokumen dapat lebih mudah dikenali pada tahap analisis. Setelah tahap pembobotan, dilakukan proses klasifikasi menggunakan metode *Support Vector Machine* (SVM). Metode ini dipilih karena mampu menangani data teks dengan jumlah fitur yang cukup banyak dan sering digunakan dalam penelitian analisis sentimen. Metode *Support Vector Machine* (SVM) digunakan dalam penelitian ini untuk melakukan klasifikasi sentimen. SVM merupakan salah satu algoritma *machine learning* yang digunakan untuk klasifikasi dan bekerja dengan mencari hyperplane terbaik untuk memisahkan data ke dalam dua kelas yang berbeda (Eldo et al., 2024, hal. 1628). Untuk

mengatasi ketidakseimbangan jumlah data pada masing-masing kategori sentimen, digunakan teknik *oversampling*. Data kemudian dibagi menjadi data latih dan data uji dengan perbandingan 80:20, di mana data latih digunakan untuk membangun model, sedangkan data uji digunakan untuk mengevaluasi kinerja model.

Pengujian model dilakukan menggunakan *confusion matrix* dengan beberapa metrik evaluasi, yaitu akurasi, *presisi*, *recall*, dan *F1-score*. Penggunaan metrik ini bertujuan untuk mengetahui seberapa baik model dalam mengklasifikasikan sentimen serta melihat keseimbangan hasil klasifikasi pada setiap kategori. Seluruh proses pengolahan dan analisis data dilakukan menggunakan bahasa pemrograman Python dengan bantuan beberapa *library* seperti *pandas*, *numpy*, dan *scikit-learn*. Dengan rangkaian proses tersebut, data komentar YouTube yang awalnya tidak terstruktur dapat diolah menjadi informasi yang lebih jelas, sehingga dapat memberikan gambaran mengenai kecenderungan sentimen masyarakat terhadap isu yang diteliti.

HASIL DAN PEMBAHASAN

HASIL

Bagian ini berisi uraian hasil dari rangkaian proses penelitian yang sudah dijalankan sebelumnya, mulai dari proses pengumpulan data sampai ke evaluasi model klasifikasi. Data yang digunakan diperoleh dengan cara *scraping* komentar YouTube yang berkaitan dengan isu pembatasan akses media sosial untuk anak-anak.

Setelah data terkumpul, data tersebut kemudian melewati beberapa tahap pengolahan, seperti pelabelan data, *preprocessing*, pembobotan menggunakan metode TF-IDF, penyeimbangan data, pembagian data, sampai tahap pengelompokan data menggunakan metode *Support Vector Machine* (SVM). Hasil dari tiap tahapan tersebut disusun secara runtut agar dapat memberikan gambaran yang jelas tentang proses yang dilakukan dan bagaimana performa model dalam penelitian ini.

1. *Scraping Data*

Data penelitian dikumpulkan dengan cara *scraping* dari kolom komentar YouTube yang membahas tentang pembatasan akses media sosial pada anak. Dari hasil pengambilan tersebut, diperoleh 1000 komentar yang telah diseleski berdasarkan relevansi.

Tabel 1. Scraping Data

No	Komentar
1	Koruptor woouiiii ,!!!! Koruptor yg merugikan negara?? ㄿㄿ
2	Udah telat woi pembatasannya, anak anak mah udah tau semua
3	Kenapa sih buka HP HP kita yang minjemin orang aja ayah aku kok emang harus ya nggak boleh gitu hah kamu memang jadi emang kamu donatur emang kamu apa
4	Go terimo
....
1000	♥♥ komdigi love

2. Labeling Data

Setelah data terkumpul, dilakukan tahap pelabelan manual oleh ahli bahasa yang mengelompokkan data menjadi tiga jenis kategori ialah kategori positif, kategori negatif, dan kategori netral.

Tabel 2. Labeling Data

No	Full-text	Sentimen
1	Koruptor woouiiii ,!!!! Koruptor yg merugikan negara?? ㄿㄿ	negatif
2	Udah telat woi pembatasannya, anak anak mah udah tau semua	negatif
3	Kenapa sih buka HP HP kita yang minjemin orang aja ayah aku kok emang harus ya nggak boleh gitu hah kamu memang jadi emang kamu donatur emang kamu apa	negatif
4	Go terimo	positif
....
1000	♥♥ komdigi love	positif

3. Preprocessing

Tahap *preprocessing* data terdiri dari beberapa proses, yaitu mengubah huruf menjadi seragam (*case folding*), membersihkan data (*cleaning*), menormalkan kata, menghapus kata tidak penting (*stopword removal*), dan mengubah kata ke bentuk dasarnya (*stemming*). Hasil akhir dari proses ini ditampilkan pada tabel di bawah:

Tabel 3. Preprocessing

No	Komentar Asli	Hasil Akhir Preprocessing
1	Koruptor woouiiii ,!!!! Koruptor yg merugikan negara?? ㄿㄿ	koruptor woouiiii koruptor rugi negara
2	Udah telat woi pembatasannya, anak anak mah udah tau semua	telat woi batas anak anak mah tau semua
3	Kenapa sih buka HP HP kita yang minjem orang aja ayah aku kok emang harus ya nggak boleh gitu hah kamu memang jadi emang kamu donatur emang kamu apa	sih buka hp hp minjem orang ayah aku kok emang gitu hah kamu memang jadi emang kamu donatur emang kamu apa
4	Go terimo	go terimo
....
1000	♥♥ komdigi love	komdigi love

4. TF-IDF

Metode TF-IDF digunakan untuk mengubah data hasil *preprocessing* ke dalam representasi numerik.

Tabel 4. TF-IDF

No	aku	anak	apa	ayah	batas
1	0.0	0.0	0.0	0.0	0.0
2	0.0	0.393	0.0	0.0	0.287
3	0.119	0.0	0.122	0.244	0.0
4	0.0	0.0	0.0	0.0	0.0
....
1000	0.0	0.0	0.0	0.0	0.0

5. Oversampling

Berdasarkan proses pelabelan, diketahui bahwa distribusi data antar kategori sentimen tidak seimbang, dengan jumlah data pada sentimen negatif lebih dominan dibandingkan kategori lainnya. Ketimpangan ini dapat menyebabkan model memiliki kecenderungan bias terhadap kelas yang jumlahnya lebih besar. Sebagai solusi, dilakukan *oversampling* pada data pelatihan dengan menambah jumlah sampel pada kelas minoritas agar distribusi menjadi lebih merata. Melalui pendekatan ini, model diharapkan dapat belajar secara lebih proporsional dari setiap kategori sentimen dan

meningkatkan hasil klasifikasi.

6. *Split Data*

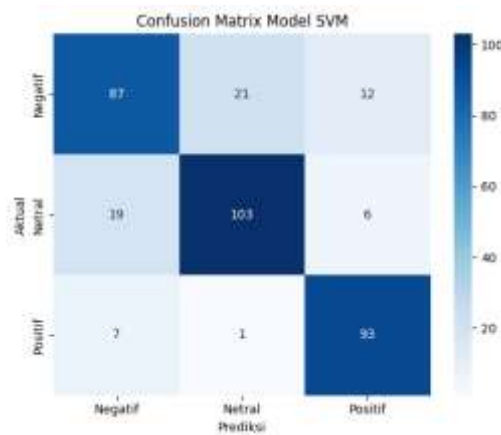
Dataset yang telah melalui proses penyeimbangan kemudian dipisahkan menjadi data latih dan data uji dengan komposisi 80% dan 20%. Pemisahan ini memungkinkan proses pelatihan dan pengujian dilakukan pada data yang tidak sama. Pada tahap ini, data latih digunakan dalam pembentukan model klasifikasi, sedangkan data uji berperan untuk mengevaluasi performa model pada data baru. Hal ini bertujuan untuk mengukur kemampuan generalisasi model secara objektif.

7. *Pelatihan Model Support Vector Machine*

Sebanyak 1394 data latih yang telah direpresentasikan dalam bentuk numerik menggunakan metode TF-IDF dengan 2502 fitur digunakan sebagai dasar dalam pelatihan model. Algoritma yang dipilih adalah *Support Vector Machine* (SVM) dengan *kernel linear*. Melalui proses ini, SVM mencari *hyperplane* terbaik untuk membedakan tiga kelas sentimen, yaitu negatif, netral, dan positif. Model yang dihasilkan selanjutnya diuji pada 349 data uji guna mengevaluasi kinerjanya.

8. *Confusion Matrix*

Confusion matrix memperlihatkan bagaimana hasil prediksi model dibandingkan dengan label sebenarnya. Dari hasil tersebut, sebagian besar data sudah berhasil dikenali dengan tepat, walaupun masih ada beberapa kesalahan klasifikasi. Ketepatan paling tinggi terlihat pada kategori positif, dengan 93 data berhasil diprediksi sesuai labelnya. Kesalahan pada kelas ini juga relatif sedikit, yaitu hanya 7 data yang bergeser ke negatif dan 1 data ke netral. Pada kategori netral, model mampu mengklasifikasikan 103 data dengan benar, namun masih terdapat kekeliruan, terutama yang cenderung bergeser ke kelas negatif (19 data) dan sebagian kecil ke positif (6 data). Berbeda dengan itu, pada kategori negatif, jumlah prediksi benar mencapai 87 data. Namun, kesalahan yang muncul cukup terlihat karena sebagian data berpindah ke netral (21 data) dan ke positif (12 data). Pola ini menunjukkan bahwa model masih mengalami kesulitan dalam membedakan data yang memiliki karakteristik serupa, khususnya antara kelas negatif dan netral. Secara umum, hasil *confusion matrix* ini menunjukkan bahwa performa model sudah cukup baik, tetapi masih terdapat keterbatasan dalam membedakan kelas dengan pola yang mirip. Hasil pengujian model SVM terhadap data uji ditampilkan melalui *confusion matrix* untuk mengetahui kesesuaian antara label sebenarnya dan label yang diprediksi oleh model



Gambar 1. Confusion Matrix

Sumber: data diolah peneliti (2026)

9. Evaluasi Model

Evaluasi model dilakukan menggunakan metrik *accuracy*, *precision*, *recall*, dan *f1-score* berdasarkan 349 data uji. Hasil pengujian menunjukkan bahwa model memperoleh nilai *accuracy* sebesar 0,81, yang menandakan bahwa sebagian besar data sudah dapat diklasifikasikan dengan cukup baik. Dilihat dari masing-masing kelas, performa terbaik terdapat pada kelas positif dengan nilai *precision* 0,84, *recall* 0,92, dan *f1-score* 0,88. Hal ini menunjukkan bahwa model cukup mampu mengenali data dengan sentimen positif. Pada kelas netral, hasil yang diperoleh juga tergolong baik, dengan nilai *precision* 0,82, *recall* 0,80, dan *f1-score* 0,81. Sementara itu, pada kelas negatif, performa model masih lebih rendah dibandingkan kelas lainnya, dengan nilai *precision* 0,77, *recall* 0,72, dan *f1-score* 0,75. Hal ini menunjukkan bahwa model masih mengalami kesulitan dalam membedakan data dengan sentimen negatif. Nilai *macro average* dan *weighted average* yang sama-sama sebesar 0,81 menunjukkan bahwa performa model secara umum sudah cukup merata pada setiap kelas, meskipun masih perlu ditingkatkan, terutama pada kelas negatif.

PEMBAHASAN

Berdasarkan pengolahan data yang dilakukan, tanggapan pengguna YouTube terhadap pembatasan akses media sosial bagi anak paling banyak berada pada kategori negatif. Dari 1000 komentar, terdapat 581 komentar negatif atau 58,1%. Selanjutnya, komentar netral berjumlah 229 atau 22,9%, sedangkan komentar positif berjumlah 190 atau 19,0%. Hasil ini memperlihatkan bahwa topik pembatasan media sosial bagi anak masih lebih banyak mendapat respon kritis dari masyarakat. Sebelum masuk ke tahap klasifikasi, komentar terlebih dahulu diproses melalui *preprocessing*. Tahap ini diperlukan karena data dari YouTube masih banyak mengandung simbol, kata tidak baku, singkatan, dan bentuk bahasa informal. Setelah dilakukan *case folding*, *cleaning*, *normalization*, *stopword removal*, dan *stemming*, data menjadi lebih bersih dan lebih mudah diolah pada tahap berikutnya.

Hasil pembobotan TF-IDF menghasilkan 2502 fitur kata dari 1000 komentar. Jumlah fitur tersebut menunjukkan bahwa komentar yang dianalisis memiliki variasi kata yang cukup banyak. Karena jumlah data pada tiap kelas belum seimbang, dilakukan *oversampling* agar setiap kategori sentimen memiliki jumlah data yang lebih proporsional. Setelah itu, data dibagi menjadi 1394 data latih dan 349 data uji. Proses klasifikasi kemudian dilakukan menggunakan metode SVM dengan kernel linear. Berdasarkan hasil evaluasi, model memperoleh nilai *accuracy* sebesar 0,81. Pada kelas positif, model menghasilkan *f1-score* sebesar 0,88. Pada kelas netral, nilai *f1-score* yang diperoleh sebesar 0,81. Sementara itu, kelas negatif memperoleh *f1-score* sebesar 0,75. Dari hasil tersebut dapat dilihat bahwa model sudah cukup baik dalam melakukan klasifikasi, tetapi hasil pada kelas negatif masih lebih rendah dibandingkan kelas lainnya. Rendahnya hasil pada kelas negatif kemungkinan disebabkan oleh bentuk komentar yang tidak selalu jelas menunjukkan penolakan. Beberapa komentar menggunakan sindiran, bahasa sehari-hari, atau kalimat yang maknanya dapat menyerupai sentimen netral maupun positif. Secara keseluruhan, metode SVM dengan pembobotan TF-IDF sudah dapat digunakan untuk menganalisis sentimen komentar YouTube pada topik ini, tetapi peningkatan masih diperlukan agar klasifikasi sentimen negatif dapat lebih tepat.

SIMPULAN

Dari 1000 komentar YouTube yang dianalisis, respon masyarakat terhadap pembatasan akses media sosial bagi anak paling banyak masuk ke sentimen negatif, yaitu 581 komentar atau 58,1%. Temuan ini sesuai dengan penelitian Widia et al., (2024) yang juga menyatakan bahwa isu sosial di YouTube sering didominasi oleh komentar negatif. Kesamaannya, kedua penelitian menunjukkan bahwa YouTube banyak digunakan masyarakat untuk menyampaikan kritik. Perbedaannya, penelitian ini membahas isu pembatasan akses media sosial bagi anak. Dari sisi metode, penelitian ini menggunakan *Support Vector Machine* (SVM) dan memperoleh akurasi 0,81. Hasil ini sejalan dengan Muasaroh et al., (2025) dan Hidayat et al., (2024) yang menyebutkan bahwa SVM dapat digunakan dengan baik untuk klasifikasi sentimen teks.

Sementara itu, sentimen netral berjumlah 229 komentar atau 22,9%, dan sentimen positif sebanyak 190 komentar atau 19,0%. Hasil penelitian juga menunjukkan bahwa metode *Support Vector Machine* (SVM), dengan proses *preprocessing*, pembobotan TF-IDF, penyeimbangan data, serta pembagian data latih dan data uji, dapat membantu mengelompokkan komentar ke dalam kategori positif, negatif, dan netral dengan hasil yang cukup baik, sehingga metode ini dapat digunakan untuk analisis sentimen pada data teks media sosial. Berdasarkan hasil penelitian yang telah dilakukan, disarankan agar penelitian selanjutnya memperbanyak jumlah data yang digunakan supaya hasil analisis lebih tepat dan mencerminkan kondisi yang lebih luas, menerapkan algoritma lain selain SVM seperti *Naive Bayes*, *Random Forest*, *K-Nearest Neighbor* (KNN), maupun pendekatan *Deep Learning* seperti

Long Short-Term Memory (LSTM) sebagai perbandingan untuk menentukan metode dengan kinerja terbaik, menggunakan teknik pengumpulan data yang lebih terstruktur seperti YouTube API agar data yang diperoleh lebih sistematis dan mudah diproses, serta meningkatkan proses *preprocessing* khususnya dalam menangani bahasa tidak baku, singkatan, dan berbagai variasi bahasa yang sering muncul pada media sosial

DAFTAR PUSTAKA

- Adelia, S., Kurniawan, F., Milanda, E., Santari, J., Kesuma, D. T., & Silvia, E. (2023). Analisis Sentimen Belajar Programming Pada Media Sosial Youtube Menggunakan Algoritma Klasifikasi Naive Bayes. *Journal of Information Technology Ampera*, 4(3), 254–264. <https://doi.org/10.51519/journalita.v4i3.430>
- Asrumi, Suharijadi, D., Setiari, agustina D., & Wulanda, D. P. (2023). *Analisis Sentimen dan Penggalan Opini*.
- Hidayat, 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. <https://doi.org/https://doi.org/10.33379/gtech.v8i3.4497>
- Karimah, A., Dwilestari, G., & Mulyawan. (2024). Analisis Sentimen Komentar Video Mobil Listrik Di Platform Youtube Dengan Metode Naive Bayes. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(1), 767–773. <https://doi.org/https://doi.org/10.36040/jati.v8i1.8373>
- Md, R., Restiyan, R. D., & Irsyad, H. (2024). Analisis Sentimen Masyarakat terhadap Perilaku Lawan Arah yang diunggah pada Media Sosial Youtube Menggunakan Naive Bayes. *BANDWIDTH: Journal of Informatics and Computer Engineering*, 02(02), 75–83. <https://doi.org/https://doi.org/10.53769/bandwidth.v2i2.706>
- Mola, S. adi S., Roma, R. V. . I. O., & Widiastuti, T. (2025). *Text Mining Analisis Sentimen dengan Lexicon*.
- Muasaroh, Y. I., Fatah, Z., & Baijuri, A. (2025). *Analisis Sentimen Komentar Youtube Terhadap Isu Ijazah Presiden Jokowi menggunakan Support Vector Machine dan Random Forest*. 1(2), 371–380. <https://ejurnal.sttdumai.ac.id/index.php/prosidingsemnas/article/view/1546>
- Saputra, A., Nurdiyani, I., Nurhidayah, U. S., Maesaroh, S., Informatika, T., Teknik, F., Bakti, U. M., & Barat, T. J. (2026). Analisis Sentimen Masyarakat Terhadap Penanganan Banjir Bandang di Pulau Sumatra Berdasarkan Komentar Youtube Menggunakan Metode Support Vector Machine (SVM). *Jurnal Ilmiah Multidisiplin*, 2(1), 2225–2239. <https://doi.org/https://doi.org/10.63822/8d1vka43>
- Soesana, A., Subakti, H., Karwanto, Fitri, anisa, Kuswandi, S., Sastri, L., Ilham, F., aswan, N., Hasibuan, F. artauli, & Lestari, H. (2023). *Metodologi Penelitian Kuantitatif*.
- Susanto, A., & Valentina, I. (2025). Analisis Sentimen Komentar YouTube terhadap Video “Purbaya Effect: Pertaruhan Ekonomi Indonesia” sebagai Cerminan Persepsi Publik Tahun 2025. *JUSIFOR : Jurnal Sistem Informasi dan Informatika*, 4(2), 185–191. <https://doi.org/10.70609/jusifor.v4i2.8262>
- Thomas, S., Yuliana, & Noviyanti. P. (2021). Study Analisis Metode Analisis Sentimen pada YouTube. *Journal of Information Technology*, 1(1), 1–7. <https://doi.org/10.46229/jifotech.v1i1.201>

- Umrona, R. D., Anwar, S. N., & Soelistijadi, R. (2025). Analisis Sentimen Komentar Youtube Terkait Kasus Pagar Laut Menggunakan Metode Knn (K-Nearest Neighbor). *Jurnal Informatika Teknologi dan Sains (JINTEKS)*, 7(3), 1537–1544. <https://doi.org/https://doi.org/10.51401/jinteks.v7i3.6251>
- Wicaksono, B., & Nastiti, V. R. S. (2024). Analisis Sentimen dalam Opini Publik di Chanel Youtube Indonesia Lawyers Club Tentang Isu Populer dengan Menggunakan Metode LSTM dan Bi-LSTM. *Jurnal Algoritma*, 22(2), 241–251. <https://doi.org/10.33364/algoritma/v.21-2.1696>
- Widia, Aqsalia, Z. Y., Sari, S., Khoirunisa, N. U., & Kurniawan, F. (2024). Optimasi Algoritma Naive Bayes Untuk Menganalisis Sentimen Pada Konten Pindahan Ibu Kota di Youtube. *Journal of Computer and Information Systems Ampera*, 5(2), 68–83. <https://doi.org/10.51519/journalcisa.v5i1.451>
- Widianto, F. (2025). Analisis Sentimen Komentar Youtube tentang Konflik Iran-Israel Menggunakan Orange Data Mining. *Sains Data Jurnal Studi Matematika dan Teknologi*, 3(2), 81–88. <https://doi.org/10.52620/sainsdata.v3i2.278>
- Zhafira, D. F., Rahayudi, B., & Indriati. (2021). Analisis Sentimen Kebijakan Kampus Merdeka Menggunakan Naïve Bayes dan Pembobotan TF-IDF Berdasarkan Komentar pada Youtube. *Jurnal Sistem Informasi, Teknologi Informasi, dan Edukasi Sistem Informasi (JUST-SI)*, 2(1), 55–63. <https://doi.org/10.25126/justsi.v2i1.24>
- Zulqarnain, Sultan, M. I., & Akbar, M. (2025). Analisis Sentimen Pemecatan Jokowi Pada Komentar Publik YouTube Tempo.co. *Jurnal Kopis: Kajian Penelitian dan Pemikiran Komunikasi Penyiaran Islam*, 07(02), 125–140. <https://doi.org/10.33367/kpi.v7i2.6888>