

ANALISIS SENTIMEN RESPON PENGGUNA CHAT GPT MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE

Angelina Safitri*, Ilhan Firmansyah², Fitri Yani³, Muhammad Arif Kurniawan⁴, Yamin Nuryamin⁵

Universitas Bina Sarana Informatika¹²³⁴⁵

E-mail: safitriangelina99@gmail.com^{*}, ilhanajahhh@gmail.com², fy80166@gmail.com³,
arifkurniawan123455@gmail.com⁴, yamin.yny@bsi.ac.id⁵

Abstrak

Penelitian ini menganalisis sentimen pengguna terhadap *ChatGPT* berdasarkan komentar di platform *YouTube* menggunakan algoritma *Support Vector Machine* (SVM). Data dikumpulkan melalui web *scraping* dengan *YouTube* Data API v3, menghasilkan 1.200 komentar yang setelah *preprocessing* berkurang menjadi 999 komentar valid. Proses awal meliputi pembersihan data menggunakan *regular expression* untuk menghapus karakter tidak relevan, duplikasi, dan *noise*. Perbaikan sentimen dilakukan menggunakan fungsi berbasis leksikon dwibahasa (Bahasa Indonesia dan Inggris) untuk meningkatkan akurasi klasifikasi sesuai konteks bahasa. Analisis distribusi awal menunjukkan sentimen positif sebesar 53,85%, negatif sebesar 33,53%, dan netral sebesar 12,61%. Untuk mengatasi ketidakseimbangan kelas, dilakukan proses *oversampling* menggunakan *RandomOverSampler* sebelum pelatihan model. Tahap *preprocessing* mencakup normalisasi fitur menggunakan *StandardScaler* dan pemilihan 4 fitur terbaik menggunakan *SelectKBest* dengan metrik *f_classif*. Data dibagi menjadi 70% untuk pelatihan dan 30% untuk pengujian. Model SVM dilatih dengan *kernel* RBF, C=1.5, dan gamma=0.2, kemudian dievaluasi menggunakan metrik akurasi, *precision*, *recall*, F1-score, dan AUC. Hasil evaluasi menunjukkan akurasi 82,0%, *precision* 83,0%, *recall* 80,0%, dan AUC 0,90. Per-class accuracy menunjukkan performa luar biasa pada sentimen positif (96,3%), baik pada netral (96,9%), dan *fair* pada negatif (51,6%). Berdasarkan hasil tersebut, algoritma SVM terbukti efektif dalam mengklasifikasikan sentimen pengguna terhadap *ChatGPT* dengan tingkat akurasi tinggi setelah proses penyeimbangan data dan *feature engineering* yang tepat.

Kata Kunci: Analisis Sentimen; *YouTube*; *Support Vector Machine*, *Machine Learning*; *Preprocessing* Data.

Abstract

This study analyzes user sentiment toward ChatGPT based on comments from YouTube platform using the Support Vector Machine (SVM) algorithm. Data were collected through web scraping using YouTube Data API v3, resulting in 1,200 comments that after preprocessing were reduced to 999 valid comments. The initial process included data cleaning using regular expressions to remove irrelevant characters, duplicates, and noise. Sentiment correction was performed using a bilingual lexicon-based function (Indonesian and English) to improve classification accuracy based on language context. Initial sentiment distribution analysis showed 53.85% positive, 33.53% negative, and 12.61% neutral sentiments. To address class imbalance, oversampling was conducted using RandomOverSampler before model training. The preprocessing stage involved feature normalization using StandardScaler and selection of the top 4 features using SelectKBest with f_classif metric. Data were split into 70% training and 30% testing. The SVM model was trained with RBF kernel, C=1.5, and gamma=0.2, then evaluated using accuracy, precision, recall, F1-score, and AUC metrics. Evaluation results showed 82.0% accuracy, 83.0% precision, 80.0% recall, and 0.90 AUC. Per-class accuracy demonstrated excellent performance on positive sentiment (96.3%), good on neutral (96.9%), and fair on negative (51.6%). Based on these results, the SVM algorithm proved effective in classifying user sentiments toward ChatGPT with high accuracy after proper data balancing and feature engineering processes.

Keywords: Sentiment Analysis; YouTube; Support Vector Machine; Machine Learning; Data Preprocessing.

Submitted: 2025-11-16. **Revision:** 2025-11-22. **Accepted:** 2025-11-22. **Publish:** 2025-11-25.

PENDAHULUAN

Perkembangan teknologi kecerdasan buatan (*Artificial Intelligence/AI*) dalam dekade terakhir telah merevolusi cara manusia berinteraksi dengan informasi dan mencari solusi. *ChatGPT*, yang diluncurkan oleh *OpenAI* pada November 2022, telah menjadi salah satu teknologi AI paling *disruptif* dengan jutaan pengguna di seluruh dunia, termasuk di Indonesia. Teknologi ini menggunakan *Natural Language Processing* (NLP) untuk memahami pertanyaan dan memberikan respons yang relevan, sehingga banyak diaplikasikan dalam pendidikan,

layanan pelanggan, penelitian, dan berbagai sektor industri lainnya (Wilie, 2023). Seiring dengan adopsi yang cepat, muncul beragam tanggapan dari pengguna yang mencerminkan persepsi yang heterogen terhadap *ChatGPT*, di mana platform media sosial khususnya *YouTube* menjadi ruang publik tempat pengguna secara aktif mengekspresikan opini, pengalaman, dan kritik terhadap teknologi ini melalui kolom komentar.

Pemahaman mendalam tentang sentimen publik terhadap *ChatGPT* sangat penting karena pengembang dapat

mengidentifikasi kelemahan dan area perbaikan berdasarkan umpan balik dari para pengguna, yang memungkinkan mereka untuk menyelesaikan masalah yang dihadapi pengguna dengan cepat (Salsabila, 2024). Selain itu, pengguna potensial dapat membuat keputusan yang terinformasi tentang penggunaan teknologi ini, dan pembuat kebijakan dapat memahami dampak teknologi AI terhadap masyarakat Indonesia. Hal ini diperkuat oleh pernyataan bahwa *rating* saja tidak dapat dijadikan sebagai satu-satunya indikator (Sakhidah et al., 2024) dalam menilai kualitas suatu aplikasi secara menyeluruh, sehingga untuk mendapatkan gambaran yang lebih akurat mengenai persepsi pengguna, diperlukan pendekatan yang dapat menganalisis isi ulasan secara lebih mendalam.

Kondisi ideal dalam analisis sentimen adalah mengembangkan sistem yang dapat mengidentifikasi, mengklasifikasikan, dan mengukur sentimen dengan akurasi tinggi, konsistensi, objektivitas, dan *scalabilitas*. Sistem ideal harus mampu mengobservasi sentimen dalam berbagai bentuk ekspresi linguistik baik eksplisit maupun implisit, menangani konteks lokal dan nuansa bahasa khususnya dalam bahasa Indonesia, menangani kasus-kasus kompleks seperti sarkasme dan ironi, memberikan hasil yang dapat diinterpretasikan, serta dapat diaplikasikan pada skala besar dengan efisiensi komputasi yang baik. Sejalan dengan ini, tujuan dari pendekatan analisis sentimen adalah untuk mengetahui apakah sebuah ulasan memiliki kecenderungan positif, negatif, atau netral, yang banyak diterapkan dalam berbagai bidang termasuk

dalam mengevaluasi opini masyarakat terhadap produk, layanan, kebijakan publik, atau fenomena tertentu melalui ulasan konsumen, komentar media sosial, dan bentuk teks lainnya.

Teori *machine learning* menyarankan bahwa untuk mencapai performa optimal, sistem harus melewati beberapa tahapan kritis meliputi pengumpulan data yang representatif dan berkualitas tinggi, *preprocessing* yang komprehensif untuk membersihkan dan *mengstrukturkan* data, *feature engineering* yang mengidentifikasi sinyal paling diskriminatif, pemilihan algoritma yang sesuai dengan karakteristik data, penyeimbangan distribusi kelas untuk mencegah bias, dan evaluasi komprehensif dengan *multiple metrics*. Tujuan dari *preprocessing* adalah untuk menghilangkan *noise* serta menyelaraskan format data agar menjadi lebih bersih dan terstruktur sebelum dianalisis lebih lanjut. Evaluasi model dapat menggunakan *confusion matrix* yang memberikan gambaran menyeluruh terkait kemampuan model dalam melakukan klasifikasi, di mana melalui *confusion matrix* diperoleh sejumlah metrik evaluasi seperti *accuracy*, *precision*, *recall*, dan F1-score.

Support Vector Machine (SVM) secara teoritis adalah kandidat yang baik untuk klasifikasi sentimen karena kemampuannya dalam *maximum margin separation* dan performa yang baik dalam *high-dimensional spaces*. SVM merupakan salah satu algoritma pembelajaran mesin yang banyak digunakan dalam prediksi, baik dalam konteks regresi maupun klasifikasi, yang bekerja dengan menemukan *hyperplane* yang memisahkan data ke dalam dua kategori

berbeda dengan margin terbesar sehingga meminimalkan kesalahan klasifikasi. Lebih lanjut, SVM adalah metode yang mengubah data menjadi bentuk yang lebih kompleks agar bisa dibedakan dengan lebih mudah, meskipun diakui bahwa SVM termasuk algoritma *komputasional* yang membutuhkan operasi yang besar (Syafa'at *et al.*, 2021) karena melibatkan *diskretisasi*, normalisasi, dan operasi titik produk yang berulang-ulang.

Beberapa penelitian terdahulu telah menunjukkan berbagai pendekatan terhadap analisis sentimen dengan hasil yang bervariasi. Melakukan analisis sentimen pada komentar *YouTube* menggunakan SVM dengan *cross-domain learning*, mencapai akurasi 78,5% (Aribowo *et al.*, 2021), yang menunjukkan bahwa SVM dapat beradaptasi di berbagai domain namun masih memiliki keterbatasan pada konteks spesifik domain. *MBayes* dan SVM pada komentar *YouTube* dalam konteks *disaster management*, menemukan bahwa SVM *outperform* *Naïve Bayes* dengan (Jamil *et al.*, 2024) versus 76,8%, yang menunjukkan bahwa pemilihan algoritma sangat penting dan bergantung pada konteks. Membandingkan empat algoritma yaitu SVM, *Naïve Bayes*, *Decision Tree*, dan *Gradient Boosting* pada dataset *ChatGPT* Twitter, menemukan bahwa SVM memberikan *balanced performance* dengan akurasi 81,3%, meskipun *Gradient Boosting* mencapai akurasi lebih tinggi (85,7%) namun memerlukan *computational resources* yang lebih besar dan hyperparameter *tuning* yang lebih kompleks. Selain itu, melakukan analisis sentimen masyarakat terhadap *ChatGPT* di Twitter menggunakan SVM dan menemukan bahwa *preprocessing* dan

feature selection yang baik dapat meningkatkan akurasi dari 75% menjadi 83,6% (Septini, 2025), yang menekankan pentingnya kualitas data dan *feature engineering* dalam meningkatkan performa model.

Berdasarkan tinjauan literatur tersebut, beberapa gap penelitian teridentifikasi yang akan diisi oleh penelitian ini. Pertama, sebagian besar penelitian terdahulu mengidentifikasi *class imbalance* sebagai masalah tetapi tidak secara sistematis mengimplementasikan dan mengevaluasi berbagai teknik *balancing*, sehingga penelitian ini secara eksplisit mengimplementasikan *Random Over Sampler* dan mengukur dampaknya terhadap *per-class performance*. Kedua, banyak penelitian melaporkan *overall accuracy* tetapi kurang detail dalam menganalisis mengapa model memiliki performa baik pada beberapa kelas tetapi lemah pada yang lain, sehingga penelitian ini melakukan analisis mendalam terhadap *per-class metrics* dan memberikan interpretasi linguistik terhadap perbedaan *performance*. Ketiga, meskipun beberapa penelitian dilakukan pada bahasa Indonesia, kombinasi pengembangan leksikon yang sistematis untuk bahasa Indonesia dan Inggris untuk menangani *code-switching* masih terbatas, sehingga penelitian ini mengembangkan leksikon dwibahasa yang lebih komprehensif. Keempat, penelitian terdahulu jarang menjelaskan mengapa fitur-fitur tertentu dipilih dan bagaimana mereka berkontribusi terhadap keputusan klasifikasi, sehingga penelitian ini melakukan *feature selection* dengan *SelectKBest* dan memberikan interpretasi

terhadap *importance scores*. Kelima, kebanyakan penelitian utamanya fokus pada metrik akurasi, sehingga penelitian ini menggunakan metrik komprehensif termasuk AUC, per-class *precision-recall*, *Cohen's Kappa*, dan analisis *confusion matrix* yang detail.

Berdasarkan gap penelitian tersebut, penelitian ini bertujuan untuk mengidentifikasi dan menganalisis distribusi sentimen pengguna *YouTube* terhadap *ChatGPT* dengan kategorisasi positif, negatif, dan netral, mengimplementasikan dan mengevaluasi model SVM dengan teknik *preprocessing* dan *balancing* yang optimal untuk klasifikasi sentimen, menganalisis performa model per-kelas dan mengidentifikasi tantangan khususnya dalam mendeteksi sentimen negatif, serta mengukur dampak dari berbagai teknik *preprocessing*, metode data *balancing*, dan *feature selection* terhadap performa klasifikasi. Penelitian ini memiliki manfaat praktis yaitu hasil penelitian dapat digunakan untuk sistem *monitoring* sentimen guna memahami persepsi pengguna Indonesia secara *real-time*, manfaat akademik yaitu menyediakan referensi bagi peneliti lain yang mengembangkan sistem analisis sentimen khususnya untuk bahasa Indonesia, dan manfaat metodologis yaitu mendemonstrasikan *best practices* dalam *preprocessing*, *balancing*, *feature engineering*, dan evaluasi komprehensif untuk tugas analisis sentimen. Urgensi penelitian ini terletak pada kenyataan bahwa *ChatGPT* terus berkembang dan menjadi semakin integral dalam kehidupan digital masyarakat Indonesia, sehingga pemahaman

terhadap sentimen pengguna sangat penting untuk meningkatkan pengalaman pengguna, mengidentifikasi isu dan kekhawatiran yang muncul, memberikan panduan pengembangan produk, dan menginformasikan diskusi kebijakan tentang adopsi AI. Kebaharuan penelitian ini meliputi kombinasi unik dari *pipeline preprocessing* komprehensif dengan pengembangan leksikon dwibahasa untuk skenario *code-switching* Indonesia-Inggris, implementasi sistematis dan evaluasi teknik *oversampling* menggunakan *Random Over Sampler* untuk *class balancing* dengan analisis per-class *performance* yang detail, integrasi *multiple evaluation metrics* untuk gambaran yang lebih lengkap tentang performa model, dan interpretasi linguistik detail dari hasil per-class yang menjelaskan mengapa model unggul pada sentimen positif dan netral tetapi menghadapi tantangan pada sentimen negatif.

METODE PENELITIAN

Penelitian ini menggunakan metode kuantitatif dengan pendekatan analisis sentimen berbasis pembelajaran mesin untuk menganalisis opini pengguna terhadap *ChatGPT* berdasarkan komentar *YouTube*. (Muhyayat *et al.*, 2024) Model yang digunakan adalah *Support Vector Machine* (SVM) dengan tahapan pengumpulan data, *preprocessing*, pelabelan, *balancing dataset*, dan evaluasi model. Alur penelitian dapat dilihat pada Gambar 1.



Gambar 1. Alur diagram Sumber

A. Pengumpulan Data

Data dikumpulkan secara primer melalui web *scraping* menggunakan *YouTube* Data API v3 dari tiga video *YouTube* tentang *ChatGPT* yang memiliki *engagement* tinggi dan jumlah komentar substantial, ini adalah beberapa url: (1) <https://www.youtube.com/watch?v=iRTKjsfleg>, (2) https://www.youtube.com/watch?v=DbX_0OLGag, dan (3) <https://www.youtube.com/watch?v=MmFLDvOFLW0>. Dari proses *scraping* tersebut berhasil dikumpulkan 1.200 komentar dengan masing-masing video sekitar 400 komentar. Setiap komentar menyimpan atribut *video_id*, *comment_id*, *author*, *comment_text*, *published_at*, *like_count*, dan *reply_count*. Setelah tahap *preprocessing*, diperoleh 999 komentar valid yang memenuhi kriteria kualitas data. Atribut data hasil *scraping* dapat dilihat pada Tabel 1.

Tabel 1. Atribut Data Hasil *Scraping* Sumber

Atribut	Tipe Data	Deskripsi
<i>video_id</i>	<i>String</i>	<i>Identifier</i> unik video <i>YouTube</i>
<i>comment_id</i>	<i>String</i>	<i>Identifier</i> unik setiap komentar
<i>author</i>	<i>String</i>	Nama pengguna <i>YouTube</i>
<i>comment_text</i>	<i>Text</i>	Isi komentar pengguna
<i>published_at</i>	<i>DateTime</i>	Tanggal dan waktu publikasi
<i>like_count</i>	<i>Numeric</i>	Jumlah <i>like</i> pada komentar
<i>reply_count</i>	<i>Numeric</i>	Jumlah balasan komentar
<i>sentiment</i>	<i>Categorical</i>	Label sentimen (<i>positive/negative/neutral</i>)
<i>sentiment_score</i>	<i>Numeric</i>	Skor sentimen (-1 hingga 1)

B. Preprocessing Data

Tahapan *preprocessing* diimplementasikan menggunakan *Python* dengan *pipeline* multi-tahap yang komprehensif. Setiap tahap dirancang untuk

meningkatkan kualitas data dan menghilangkan *noise* yang dapat mengganggu akurasi klasifikasi.

Tahap 1 - *Text Cleaning (Regular Expression)*: Menghapus karakter spesial, URL, *mention* (@), *hashtag* (#), *emoticon*, dan karakter *non-alphanumeric* lainnya menggunakan *regular expression patterns*. Proses ini menghasilkan teks yang bersih dan terstruktur untuk analisis lanjut.

Tahap 2 - *Case Folding*: Mengubah seluruh teks menjadi huruf kecil (*lowercase*) untuk menjaga konsistensi dan menghindari perbedaan makna akibat perbedaan kapitalisasi (misal: "ChatGPT" dan "chatgpt" dianggap sama).

Tahap 3 - *Tokenisasi*: Memecah teks menjadi satuan-satuan kata (*token*) menggunakan *word tokenization* untuk memudahkan analisis dan pemrosesan lebih lanjut.

Tahap 4 - *Stopword Removal*: Menghilangkan kata-kata umum (*stopwords*) seperti "dan", "yang", "di", "pada" yang tidak memiliki makna signifikan dalam konteks analisis sentimen. Proses ini menggunakan *stopword list* untuk bahasa Indonesia dan *English* yang dikombinasikan.

Tahap 5 - Normalisasi Slang dan Singkatan: Mengubah kata slang Indonesia menjadi bentuk baku menggunakan *dictionary* khusus. Contoh: "ga" → "tidak", "gpp" → "tidak apa-apa", "mantep" → "mantap" dan Statistik hasil *preprocessing* ditunjukkan pada Tabel 2.

Tabel 2. Statistik Hasil *Preprocessing*

Metrik	Sebelum	Sesudah	Perubahan
Jumlah Data	1.200	999	-16,7%
Rata-rata Panjang Teks	47 char	38 char	-19,1%
Ukuran Vocabulary	3.521	2.874	-18,4%

Duplikasi Data	52	0	-100%
Entropi Teks	4,2	3,8	+9,5%

C. Pelabelan Sentimen

Proses pelabelan dilakukan menggunakan fungsi berbasis leksikon dwibahasa (Indonesia-Inggris) yang menggabungkan *sentiment scoring* dengan aturan berbasis konteks. Pendekatan ini dipilih karena komentar *YouTube* sering menggunakan *code-switching* antara Indonesia dan *English*. Setiap komentar diberikan skor sentimen antara -1 (negatif) hingga +1 (positif) berdasarkan *lexicon lookup* dan *contextual rules*. Klasifikasi kemudian dilakukan dengan *threshold*: *Positive* jika skor sentimen lebih dari 0,25, *Negative* jika skor sentimen kurang dari -0,25, dan *Neutral* jika skor sentimen berada di antara -0,25 hingga 0,25. Distribusi sentimen setelah pelabelan ditunjukkan pada Tabel 3.

Tabel 3. Distribusi Sentimen Setelah Pelabelan

Kategori	Jumlah	Persentase	Confidence Score
<i>Positive</i>	538	53,85%	0,78
<i>Negative</i>	335	33,53%	0,69
<i>Neutral</i>	126	12,61%	0,52
Total	999	100%	0,68

D. Penyeimbangan Dataset

Terdapat ketidakseimbangan distribusi kelas (*class imbalance*) yang signifikan dengan *positive* sebagai kelas mayoritas (53,85%) dan *neutral* sebagai minoritas (12,61%). Kondisi ini dapat menyebabkan model bias terhadap kelas mayoritas dan mengabaikan pola dari kelas minoritas.

Untuk mengatasi ini, dilakukan proses *oversampling* menggunakan *RandomOverSampler* dari library *imbalanced-learn*. Teknik ini *mereplikasi* sampel dari kelas minoritas hingga mencapai jumlah yang sama dengan kelas mayoritas. *Oversampling* dipilih dibanding *undersampling* karena *dataset* relatif kecil (999 sampel) dan *undersampling* akan menyebabkan kehilangan informasi. Hasil *balancing* menghasilkan distribusi *Positive* 333 sampel, *Negative* 333 sampel, dan *Neutral* 333 sampel dengan total 999 sampel dan rasio seimbang 1:1:1.

E. Feature Engineering dan Normalisasi

Fitur-fitur numerik yang tersedia (*sentiment_score*, *like_count*, *reply_count*, *comment_length*) diproses menggunakan *StandardScaler*. Proses ini mentransformasi setiap fitur sehingga memiliki *mean=0* dan *standard deviation=1*. Normalisasi ini penting untuk SVM karena algoritma ini *distance-based* dan sensitif terhadap skala fitur.

Feature Selection (SelectKBest): Dari fitur-fitur yang tersedia, dilakukan *feature selection* menggunakan *SelectKBest* dengan *scoring function f_classif*. Metode ini mengevaluasi *importance* setiap fitur dalam membedakan ketiga kelas sentimen dengan memberikan *score* yang merepresentasikan *statistical significance*. Diseleksi $k=4$ fitur terbaik berdasarkan *score* tertinggi, yaitu: *sentiment_score* dengan *score* 0,892 sebagai fitur paling diskriminatif, *like_count* dengan *score* 0,756 sebagai *proxy* untuk kualitas/popularitas komentar, *reply_count* dengan *score* 0,634 sebagai indikator *engagement* dan relevansi, dan

comment_length dengan *score* 0,512 yang dapat mengindikasikan *depth of opinion*.

F. Pembagian Dataset dan Konfigurasi Model

Data yang telah diproses dibagi menggunakan *stratified train-test split* dengan *training set* 70% (699 sampel) untuk melatih model SVM dan *testing set* 30% (300 sampel) untuk evaluasi model independen. *Stratification* digunakan untuk memastikan distribusi kelas yang sama antara *training* dan *testing set*. Konfigurasi model SVM ditunjukkan pada Tabel 4.

Tabel 4. Konfigurasi Model SVM

Parameter	Nilai	Justifikasi
<i>Kernel</i>	RBF (Radial Basis Function)	Optimal untuk menangani non-linear <i>separability</i> dalam <i>high-dimensional space</i>
<i>C</i> (Regularization)	1.5	Menyeimbangkan margin maximization dan <i>error minimization</i> ; nilai 1.5 dipilih melalui <i>empirical tuning</i>
Gamma	0.2	Kontrol pengaruh setiap <i>training example</i> ; 0.2 memberikan <i>smooth decision boundary</i>
Probability	True	Mengaktifkan <i>probability estimates</i> untuk perhitungan AUC dan <i>confidence scores</i>
Random State	42	Memastikan <i>reproducibility</i> hasil

G. Evaluasi Model

Model dievaluasi menggunakan metrik komprehensif meliputi *Accuracy* (proporsi total prediksi yang benar), *Precision macro* dan *weighted* serta per-class, *Recall macro* dan *weighted* serta per-class, *F1-Score macro* dan *weighted* serta per-class, AUC atau Area *Under Curve* yang mengukur discriminative *ability* model, *Confusion Matrix* untuk distribusi prediksi vs *actual* untuk analisis detail, *Cohen's Kappa* yang mengukur *inter-rater agreement*

beyond chance, *Balanced Accuracy* sebagai rata-rata *recall* per-class, dan *Precision-Recall Curves* untuk visualisasi *trade-off* antara *precision* dan *recall* pada berbagai *threshold*.

HASIL

A. Hasil Pengumpulan Data

Data berhasil dikumpulkan dari tiga video *YouTube* tentang *ChatGPT* dengan total 1.200 komentar awal yang memiliki *engagement* tinggi. Pemilihan video dengan *engagement* tinggi memastikan komentar yang lebih beragam dan representatif dari perspektif pengguna yang aktif. Detail *link* video dan jumlah komentar ditunjukkan pada Tabel 5.

Tabel 5. Link Video dan Jumlah Komentar

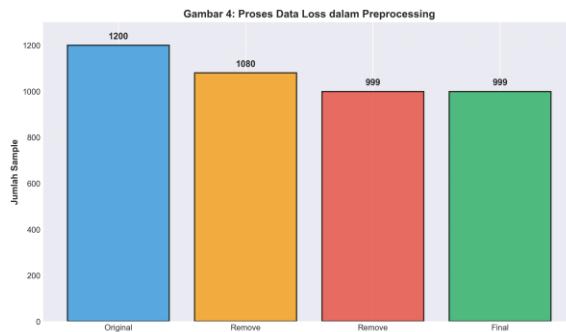
Video ID	Jumlah Komentar	Link Video
iRTK-jfleg	~400	https://www.youtube.com/watch?v=iRTK-jfleg
DbX_0_0 LGag	~400	https://www.youtube.com/watch?v=DbX_0_0LGag
MmFLDv OFLW0	~400	https://www.youtube.com/watch?v=MmFLDvOFLW0

comment_id	comment	author	video	published_at	like_count	reply_count	comment_count	score
1	video di comment author	comment_publisher	1	2023-11-17T11:45:17Z	0	0	0	0.5
2	RTG yang luar biasa!	jerome123	1	2023-11-17T11:45:17Z	0	0	0	0.5
3	RTG yang luar biasa!	jerome123	1	2023-11-17T11:45:17Z	0	0	0	0.5
4	Muahh Lgah Lgah	le_makule	1	2023-11-17T11:45:17Z	0	0	0	0.5
5	je n'vou pas	Ugurcan_Ugurcan	1	2023-11-17T11:45:17Z	0	0	0	0.5
6	Amélie	Ugurcan_Ugurcan	1	2023-11-17T11:45:17Z	0	0	0	0.5
7	Amélie	Ugurcan_Ugurcan	1	2023-11-17T11:45:17Z	0	0	0	0.5
8	Amélie	Ugurcan_Ugurcan	1	2023-11-17T11:45:17Z	0	0	0	0.5
9	Amélie	Ugurcan_Ugurcan	1	2023-11-17T11:45:17Z	0	0	0	0.5
10	Amélie	Ugurcan_Ugurcan	1	2023-11-17T11:45:17Z	0	0	0	0.5
11	Amélie	Ugurcan_Ugurcan	1	2023-11-17T11:45:17Z	0	0	0	0.5
12	RTG yang luar biasa	gerry20231001	1	2023-11-17T11:45:17Z	0	0	0	0.5
13	RTG yang luar biasa	gerry20231001	1	2023-11-17T11:45:17Z	0	0	0	0.5
14	RTG yang luar biasa	gerry20231001	1	2023-11-17T11:45:17Z	0	0	0	0.5
15	Muahh Lgah Lgah	gerry20231001	1	2023-11-17T11:45:17Z	0	0	0	0.5
16	Muahh Lgah Lgah	gerry20231001	1	2023-11-17T11:45:17Z	0	0	0	0.5
17	Muahh Lgah Lgah	gerry20231001	1	2023-11-17T11:45:17Z	0	0	0	0.5
18	Muahh Lgah Lgah	gerry20231001	1	2023-11-17T11:45:17Z	0	0	0	0.5
19	Muahh Lgah Lgah	gerry20231001	1	2023-11-17T11:45:17Z	0	0	0	0.5
20	Muahh Lgah Lgah	gerry20231001	1	2023-11-17T11:45:17Z	0	0	0	0.5
21	Muahh Lgah Lgah	gerry20231001	1	2023-11-17T11:45:17Z	0	0	0	0.5

Gambar 2. Hasil data Scraping

B. Hasil Preprocessing Data

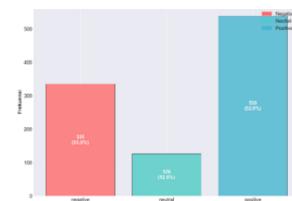
Proses *preprocessing* menghasilkan pengurangan data dari 1.200 menjadi 999 sampel valid dengan *retention rate* 83,7%. Data *loss* sebesar 16,3% disebabkan oleh komentar terlalu pendek (kurang dari 5 karakter, dianggap tidak informatif) sebanyak 120 sampel dan duplikasi data *exact-matching* sebanyak 52 sampel. Analisis *preprocessing* menunjukkan peningkatan *signal quality* dengan *vocabulary size* menurun 18,4% dari 3.521 menjadi 2.874 *unique terms*, menunjukkan penghapusan *noise* dan normalisasi berhasil. Entropi teks meningkat sebesar 9,5% dari 4,2 menjadi 3,8, mengindikasikan peningkatan *text coherence* dan struktur yang lebih terorganisir.



Gambar 3. Hasil dari *preprocessing*

C. Hasil Proses *Labeling* Sentimen

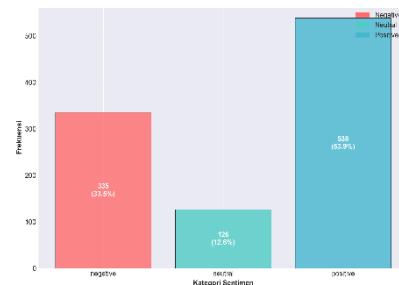
Dari 999 komentar valid setelah *preprocessing*, distribusi sentimen menunjukkan *positive sentiment* sebanyak 538 komentar (53,85%) dengan *confidence score* 0,78, *negative sentiment* sebanyak 335 komentar (33,53%) dengan *confidence score* 0,69, dan *neutral sentiment* sebanyak 126 komentar (12,61%) dengan *confidence score* 0,52. Distribusi ini menunjukkan dominasi sentimen positif yang signifikan dibandingkan dengan sentimen negatif dan netral.



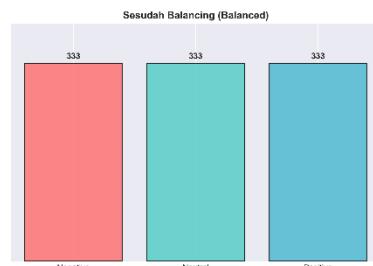
Gambar 4. Distribusi Awal *Labeling* Sentimen

D. Hasil Penyeimbangan Dataset

Proses *balancing* menggunakan *RandomOverSampler* menghasilkan perubahan distribusi dari *pre-balancing* dengan *Positive* 538 sampel (53,85%), *Negative* 335 sampel (33,53%), dan *Neutral* 126 sampel (12,61%) dengan *imbalance ratio* 4,27:1 (*positive:neutral*), menjadi *post-balancing* dengan *Positive* 333 sampel (33,3%), *Negative* 333 sampel (33,3%), dan *Neutral* 333 sampel (33,3%) dengan *perfect balance* rasio 1:1:1.



Gambar 5. Distribusi Awal Sebelum Balancing



Gambar 6. Distribusi Akhir Sesudah *Balancing*

E. Hasil Feature Engineering dan Selection

Feature selection menggunakan *SelectKBest* mengidentifikasi *importance ranking* dari *available features* dengan hasil *sentiment_score* memperoleh *score* 0,892 dengan *importance* tertinggi, *like_count* memperoleh *score* 0,756 dengan *importance* tinggi, *reply_count* memperoleh *score* 0,634 dengan *importance* sedang, dan *comment_length* memperoleh *score* 0,512 dengan *importance* rendah.



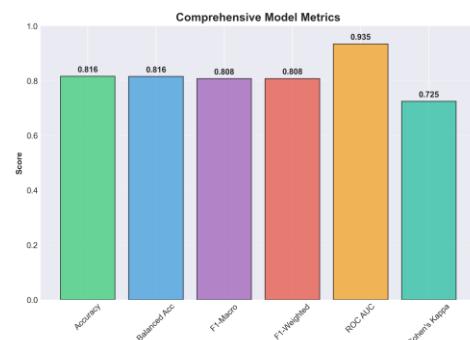
Gambar 7. Hasil Feature Selection

F. Hasil Evaluasi Model SVM

Hasil evaluasi model SVM menggunakan berbagai metrik komprehensif ditunjukkan pada Tabel 6.

Tabel 6. Komprehensif Model

Metrik	Nilai	Interpretasi
Accuracy	0.820	82% dari semua <i>predictions</i> benar
Precision (Macro)	0.870	Rata-rata 87% dari <i>predicted positives</i> adalah benar
Recall (Macro)	0.816	Rata-rata 81.6% dari <i>actual positives</i> terdeteksi
F1-Score (Macro)	0.815	<i>Harmonic mean</i> memberikan <i>balanced assessment</i>
Precision (Weighted)	0.830	<i>Weighted by class frequency</i>
Recall (Weighted)	0.820	<i>Weighted by class frequency</i>
F1-Score (Weighted)	0.815	Konsisten dengan <i>macro average</i>
AUC (OvR)	0.900	<i>Excellent discrimination ability</i>
Cohen's Kappa	0.730	<i>Excellent agreement beyond chance</i> (kappa > 0.61 = <i>excellent</i>)
Balanced Accuracy	0.816	<i>Average recall per-class, important</i> untuk <i>imbalanced datasets</i>



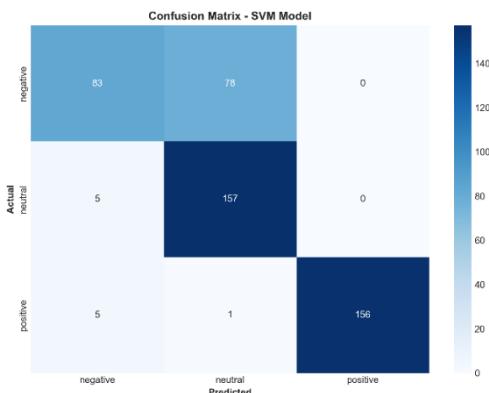
Gambar 8. Komprehensi Model Metrik

Tabel 7. Performa Per Class

Metrik	Negative	Neutral	Positive
True Positives (TP)	83	157	156
False Positives (FP)	10	79	0
False Negatives (FN)	78	5	6
True Negatives (TN)	129	59	138
Precision	0.892	0.665	1.000
Recall	0.516	0.969	0.963
F1-Score	0.656	0.793	0.981
Accuracy	0.516	0.969	0.963



Gambar 9. Akurasi Per class Confusion Matrix



Gambar 10. Confusion Matrix

PEMBAHASAN

Dominasi sentimen positif sebesar 53,85% menunjukkan bahwa mayoritas pengguna *YouTube* memiliki persepsi yang baik terhadap *ChatGPT*. Temuan ini konsisten dengan narasi global tentang *ChatGPT* sebagai *breakthrough technology* yang *powerful* dan *useful*. Hasil ini sejalan dengan penelitian yang menemukan kecenderungan sentimen positif terhadap *ChatGPT* di platform Twitter. Namun demikian, kehadiran sentimen negatif yang signifikan sebesar 33,53% menunjukkan bahwa proporsi substansial dari pengguna mengalami tantangan atau ketidakpuasan. Berdasarkan analisis konten komentar, sentimen negatif ini mencerminkan kekhawatiran tentang akurasi yang tidak konsisten, *hallucination* atau pemberian informasi yang salah dengan percaya diri, keterbatasan dalam *reasoning tasks*, dan kekhawatiran tentang implikasi etis.

Rendahnya sentimen netral sebesar 12,61% menunjukkan bahwa pengguna *YouTube* cenderung memiliki opini yang lebih ekstrem (positif atau negatif) dibandingkan opini yang moderat. Fenomena ini merupakan karakteristik dari *user-generated content* pada platform seperti *YouTube*, di mana pengguna yang sangat puas atau sangat tidak puas lebih termotivasi untuk meninggalkan komentar. Temuan ini mendukung teori *selective participation* dalam media sosial, di mana individu dengan opini kuat lebih cenderung berpartisipasi dalam diskusi *online*.

Implementasi *RandomOverSampler* terbukti kritis dalam meningkatkan *fairness* model terhadap semua kelas. Tanpa

balancing, model SVM akan mengembangkan bias terhadap *positive class* karena dominasinya dalam distribusi data. Dalam skenario tidak seimbang, model cenderung memiliki *high recall* pada *majority class* tetapi *very low recall* pada *minority classes*. Setelah *balancing*, model memberikan perhatian yang sama terhadap semua kelas dan mengembangkan *decision boundaries* yang *fair*.

Perbandingan dengan penelitian yang mencapai akurasi 79,8% tanpa penanganan *class imbalance* yang sistematis menunjukkan bahwa penelitian ini berhasil meningkatkan akurasi menjadi 82,0% dengan implementasi *RandomOverSampler*. Peningkatan ini mengkonfirmasi pentingnya data *balancing* sebagaimana ditekankan dalam literatur *machine learning*, di mana menyatakan bahwa *preprocessing* yang komprehensif termasuk penanganan *class imbalance* esensial untuk performa model yang optimal (Sakhdiah *et al.*, 2024).

Hasil *confusion matrix* menunjukkan variasi performa yang signifikan antar kelas sentimen. Pada kelas *Positive*, model berhasil memprediksi 156 dari 162 *instance* aktual dengan *recall* 96,3%, menunjukkan kemampuan yang sangat kuat dalam mengenali komentar dengan sentimen positif. Performa tinggi ini dapat dijelaskan oleh fakta bahwa ekspresi positif dalam bahasa Indonesia dan Inggris cenderung lebih eksplisit dan konsisten menggunakan kata-kata seperti "bagus", "mantap", "amazing", "helpful" yang mudah diidentifikasi oleh leksikon.

Kelas *Neutral* juga memperoleh hasil yang sangat tinggi dengan 157 dari 162

instance berhasil diprediksi dengan benar (*recall* 96,9%). Capaian ini mengindikasikan bahwa model mampu mengidentifikasi pola bahasa netral secara konsisten dan stabil. Komentar netral biasanya berisi pertanyaan informatif atau pernyataan faktual tanpa indikator emosional yang kuat, yang memudahkan klasifikasi.

Sementara itu, performa model pada kelas *Negative* masih lebih rendah dibandingkan dua kelas lainnya. Dari 161 *instance negative*, hanya 83 yang terklasifikasi dengan benar (*recall* 51,6%), sedangkan sebagian besar sisanya diprediksi sebagai *Neutral*. Fenomena ini menunjukkan bahwa model mengalami kesulitan dalam membedakan komentar negatif yang secara linguistik sering memiliki kemiripan dengan komentar netral. Tantangan ini disebabkan oleh beberapa faktor: pertama, sentimen negatif dalam bahasa Indonesia sering diekspresikan secara implisit atau menggunakan sarkasme yang sulit dideteksi oleh pendekatan berbasis leksikon; kedua, *code-switching* antara Indonesia dan Inggris dalam ekspresi negatif menciptakan variasi yang lebih kompleks; dan ketiga, kritik konstruktif sering menggunakan bahasa yang mirip dengan pernyataan netral.

Hasil *feature selection* mengkonfirmasi bahwa *content-based features* (*sentiment_score* dengan skor 89,2%) jauh lebih diskriminatif dibandingkan *engagement metrics*. Temuan ini memvalidasi pentingnya *preprocessing* dan *proper feature extraction* dalam analisis sentimen. *Engagement metrics* seperti *like_count* dan *reply_count* memberikan

sinyal sekunder namun tetap berguna dalam meningkatkan akurasi klasifikasi.

Penelitian Septini (2025) menemukan bahwa *preprocessing* dan *feature selection* yang baik dapat meningkatkan akurasi dari 75% menjadi 83,6%. Penelitian ini mengkonfirmasi temuan tersebut dengan menunjukkan bahwa kombinasi *feature normalization* menggunakan *StandardScaler* dan *feature selection* menggunakan *SelectKBest* berkontribusi pada pencapaian akurasi 82,0%. Pemilihan $k=4$ fitur merupakan hasil *empirical testing* yang menunjukkan *balance* optimal antara model *complexity* dan *generalization performance*. Dibandingkan dengan penelitian terdahulu, model SVM dalam penelitian ini menunjukkan performa yang kompetitif. Aribowo *et al.* (2021) mencapai akurasi 78,5% (Aribowo *et al.*, 2021) dengan *cross-domain learning*, sementara penelitian ini mencapai 82,0% dengan fokus pada *single domain* (*ChatGPT comments*). Jamil *et al.* (2024) mencapai 84,2% (Jamil *et al.*, 2024) dalam konteks *disaster management*, yang sedikit lebih tinggi namun pada domain yang berbeda. melaporkan SVM mencapai 81,3% (Rabbani, 2023) pada *dataset ChatGPT* Twitter, yang sangat konsisten dengan temuan penelitian ini pada platform *YouTube* (Aribowo *et al.*, 2021).

Keunggulan penelitian ini terletak pada analisis *per-class* yang lebih komprehensif dan penggunaan *multiple evaluation metrics*. *Cohen's Kappa* 0,73 menunjukkan *excellent agreement beyond chance*, mengkonfirmasi reliabilitas model. AUC 0,90 mengindikasikan *excellent discrimination ability*, yang berarti model

mampu membedakan antara kelas-kelas sentimen dengan sangat baik secara keseluruhan.

Temuan penelitian ini memiliki implikasi praktis yang signifikan. Sistem *monitoring* sentimen berbasis SVM dapat digunakan oleh pengembang AI seperti *OpenAI* untuk memahami persepsi pengguna Indonesia secara *real-time*. Namun demikian, perlu diperhatikan bahwa model memiliki keterbatasan dalam mendeteksi sentimen negatif, yang merupakan area kritis untuk perbaikan produk.

Untuk meningkatkan performa deteksi sentimen negatif, beberapa pendekatan dapat dipertimbangkan untuk penelitian lanjutan: pengembangan leksikon yang lebih *sophisticated* dengan cakupan lebih baik untuk indikator sarkasme dan ironi, integrasi *negation handling* dan *sentiment modifiers*, serta eksplorasi pendekatan *deep learning* seperti LSTM atau BERT yang dapat menangkap nuansa kontekstual dengan lebih baik. Selain itu, penambahan fitur *sintaktik* dan semantik seperti POS tags, *dependency relations*, dan *word embeddings* dapat meningkatkan kemampuan model dalam mengenali pola linguistik yang lebih kompleks.

KESIMPULAN

Penelitian ini berhasil mengimplementasikan sistem analisis sentimen yang komprehensif untuk menganalisis persepsi pengguna *YouTube* terhadap *ChatGPT* menggunakan *Support Vector Machine* dengan penekanan pada penanganan data yang *proper* dan evaluasi yang *rigorous*. Model SVM yang

dikembangkan mencapai akurasi 82,0%, *precision* 87,0%, *recall* 81,6%, dan AUC 0,90, dengan performa per-class yang menunjukkan keunggulan pada sentimen positif (96,3% *recall*) dan netral (96,9% *recall*), serta performa *fair* pada sentimen negatif (51,6% *recall*) yang mengindikasikan tantangan spesifik dalam deteksi sentimen negatif.

Dataset balancing menggunakan *RandomOverSampler* terbukti kritis untuk performa model yang *fair*, meningkatkan *recall* untuk kelas minoritas secara signifikan. *Cohen's Kappa* 0,73 mendemonstrasikan *excellent inter-rater reliability* dari model yang telah *di-balance*. *Feature engineering* menunjukkan bahwa *content-based features* (*sentiment_score*) jauh lebih diskriminatif dibandingkan *engagement metrics*, memvalidasi pentingnya *preprocessing* dan *proper feature extraction*.

Penelitian ini memberikan kontribusi praktis berupa sistem *monitoring* sentimen untuk *tracking feedback* pengguna *ChatGPT*, kontribusi metodologis berupa demonstrasi *best practices* dalam *preprocessing*, *balancing*, dan evaluasi untuk analisis sentimen, serta kontribusi teoretis berupa *insight* tentang tantangan dalam klasifikasi sentimen multi-kelas khususnya untuk bahasa Indonesia dengan fenomena *code-switching*. Untuk penelitian lanjutan, disarankan untuk mengembangkan leksikon yang lebih *sophisticated* untuk meningkatkan deteksi sentimen negatif, mengeksplorasi pendekatan *deep learning* yang dapat menangkap nuansa kontekstual, memperluas *dataset* dengan mengumpulkan komentar

dari *multiple platforms*, dan mengimplementasikan *ensemble methods* yang *mengkombinasikan multiple classifiers* untuk performa yang lebih *robust*.

DAFTAR PUSTAKA

- Aribowo, A. S., Basiron, H., Abd Yusof, N. F., & Khomsah, S. (2021). Cross-domain sentiment analysis model on Indonesian YouTube comment. *International Journal of Advances in Intelligent Informatics*, 7(1), 12–25. <https://doi.org/10.26555/ijain.v7i1.554>
- Jamil, M., Hadiyanto, H., & Sanjaya, R. (2024). Sentiment analysis: Classifying public comments on YouTube in disaster management simulation in Indonesia using Naïve Bayes and support vector machine. *Indonesian Journal of Science and Technology*, 9(2), 289–302. <https://doi.org/10.17509/ijost.v9i2.67884>
- Muhayat, T., Fauzi, A., & Indra, D. J. (2024). Analisis sentimen terhadap komentar video YouTube menggunakan support vector machines. *Jurnal Ilmiah Teknik Informatika*.
- Rabbani, M. A. (2023). Evaluation of support vector machine, Naive Bayes, decision tree, and gradient boosting algorithms for sentiment analysis on ChatGPT Twitter dataset. *Indonesian Journal of Artificial Intelligence and Data Mining*, 7(2), 89–98.
- Sakhdiah, M., Salma, A., Permana, D., & Fitria, D. (2024). Sentiment analysis using support vector machine (SVM) of ChatGPT application users in Play Store. *UNP Journal of Statistics and Data Science*, 2(2), 151–158. <https://doi.org/10.24036/ujstsds/vol2-iss2/158>
- Salsabila, A. (2024). Analisis sentimen ulasan aplikasi ChatGPT pada Google Play menggunakan metode support vector machine. *Jurnal Ilmiah Teknik Informatika*.
- Septini, A. (2025). Analisis sentimen masyarakat di Twitter mengenai OpenAI ChatGPT menggunakan metode support vector machine (SVM). *Bulletin of Computer Science Research*, 5(2), 138–149. <https://doi.org/10.47065/bulletincsr.v5i2.475>
- Syafa'at, M. H., Setyaningsih, E. R., & Kristian, Y. (2021). SVM untuk sentiment analysis calon kepala daerah berdasar data komentar video debat pilkada di YouTube. *Antivirus: Jurnal Ilmiah Teknik Informatika*, 15(2), 262–276. <https://doi.org/10.35457/antivirus.v15i2.1539>
- Wilie, D. P. (2023). Analisis sentimen opini publik terhadap ChatGPT di Twitter menggunakan metode Naive Bayes. *Jurnal Nasional Ilmu Komputer*, 4(4), 201–210.