

Sentiment Analysis of Youtube and Gotube Reviews on Google Play Using the Support Vector Machine (SVM) Method in Indonesia

Sri Raihan Putri ^{1*}, Asrianda ^{2*}, Lidya Rosnita ^{3*}

* Teknik Informatika, Universitas Malikussaleh, Indonesia

sri.200170009@mhs.unimal.ac.id ¹, asrianda@unimal.ac.id ², lidiyarosnita@unimal.ac.id ³

Article Info

Article history:

Received 2025-04-20

Revised 2025-06-20

Accepted 2025-06-26

Keyword:

*Sentiment Analysis,
Support Vector Machine (SVM),
Google Play Reviews,
YouTube,
GoTube.*

ABSTRACT

This research, titled Sentiment Analysis of YouTube and GoTube Reviews on Google Play Using the Support Vector Machine (SVM) Method in Indonesia, analyzes user perceptions of YouTube and GoTube based on Google Play reviews. The study is motivated by the growing popularity of video streaming apps in Indonesia and the limited sentiment analysis research on these platforms. The research collects 1,600 reviews (800 per app) from 2023-2024 using Python's Scrapy library. The data is split 70% for training and 30% for testing, undergoing text preprocessing (tokenization, stop word removal, stemming), TF-IDF weighting, and SVM classification with an RBF kernel. Evaluation metrics include accuracy, precision, recall, and F1-score, with PCA used for visualization. Results show 94.50% accuracy overall, 97.01% for YouTube, and 92.66% for GoTube. GoTube has higher positive sentiment (385 of 400 test reviews) than YouTube (345 of 400) but lower negative sentiment (15 vs. 55). However, the model exhibits a positive class bias due to data imbalance. The study concludes that SVM effectively detects positive sentiment, but balancing data and exploring non-linear methods could improve negative sentiment detection.



This is an open access article under the [CC-BY-SA](#) license.

I. PENDAHULUAN

Kemajuan teknologi digital telah membawa berbagai inovasi, termasuk dalam dunia aplikasi mobile seperti YouTube dan GoTube, yang menjadi salah satu platform streaming video populer di Indonesia[1]. YouTube, sebagai pelopor platform berbagi video, menawarkan berbagai konten hiburan, edukasi, hingga informasi terkini yang dapat diakses oleh pengguna di seluruh dunia. Sementara itu, GoTube hadir sebagai alternatif aplikasi yang menawarkan fitur-fitur serupa dengan kelebihan tertentu, seperti konsumsi data yang lebih hemat dan tampilan yang ramah pengguna[2]. Di Google Play, ulasan pengguna terhadap aplikasi ini menjadi sumber data penting yang mencerminkan pengalaman, kepuasan, atau keluhan mereka[3]. Ulasan ini sangat beragam, mulai dari yang positif, yang memuji fitur-fitur atau pengalaman penggunaan, hingga yang negatif, yang mengkritik kekurangan tertentu. Analisis ulasan ini memungkinkan pengembang untuk memahami kebutuhan dan preferensi

pengguna, serta meningkatkan kualitas layanan yang mereka tawarkan [4].

Analisis sentimen ulasan pengguna menjadi pendekatan yang relevan untuk menggali persepsi publik terhadap aplikasi YouTube dan GoTube. Sentimen yang teridentifikasi dapat membantu pengembang dalam mengambil langkah strategis untuk meningkatkan pengalaman pengguna. Di Indonesia, penelitian terkait analisis sentimen terhadap ulasan aplikasi seperti YouTube dan GoTube masih terbatas, meskipun aplikasi ini memiliki basis pengguna yang besar. Dengan pengguna internet yang terus meningkat dan keberagaman budaya serta bahasa di Indonesia, penting untuk memahami bagaimana ulasan pengguna mencerminkan persepsi terhadap aplikasi ini[5]. Penelitian ini bertujuan untuk menganalisis sentimen ulasan pengguna terhadap YouTube dan GoTube di Google Play menggunakan metode Support Vector Machine (SVM) [6].

Metode SVM dipilih karena kemampuannya dalam menangani data teks yang kompleks dan menghasilkan klasifikasi sentimen yang akurat. Algoritma ini bekerja

dengan mencari hyperplane optimal yang memisahkan data ke dalam kelas-kelas sentimen, seperti positif, negatif, dan netral[6]. Penelitian ini diharapkan dapat memberikan wawasan baru tentang bagaimana masyarakat Indonesia merespons aplikasi YouTube dan GoTube, serta memberikan rekomendasi untuk pengembangan lebih lanjut yang sesuai dengan kebutuhan pengguna lokal.

Penelitian sebelumnya tentang analisis sentimen ulasan aplikasi mola yang dilakukan oleh Muhammad Dikihendriyanto, Azhari Ali Ridha Dan Ultach Enri pada tahun 2022 dengan judul Analisis Sentimen Ulasan Aplikasi Mola Pada Google Play Store Menggunakan Algoritma Support Vector Machine (SVM). Knowledge Discovery in Database (KDD) adalah pendekatan yang digunakan dalam penelitian ini. Data yang digunakan terdiri dari 520 ulasan dari aplikasi MOLA, terdiri dari 312 ulasan positif dan 208 ulasan negatif. Dengan menggunakan kernel RBF (Function Basis Radial), hasil terbaik diperoleh pada skenario 1 (90:10). Kernel ini menghasilkan ketepatan 92,31%, ketepatan 96,3%, recall 89,66%, dan skor f1 - 92,86% [7].

Penelitian selanjutnya tentang analisis sentimen yang dilakukan oleh Vina Fitriyana, Lutfi Hakim, Dian Candra Rini Novitasari, Ahmad Hanif Asyhar pada tahun 2023 dengan judul Analisis Sentimen Ulasan Aplikasi Jamsostek Mobile Menggunakan Metode Support Vector Machine (SVM). Tujuan penelitian ini adalah untuk mengevaluasi kualitas aplikasi JMO dan menggunakan analisis sentimen untuk mengkategorikan ulasan atau opini ke dalam kategori positif, negatif, dan netral. Pada proses analisis ini, metode Support Vector Machine digunakan dengan pendekatan kernel linear untuk menentukan tingkat akurasi dari pengklasifikasian ulasan aplikasi JMO tersebut. Penelitian menunjukkan bahwa pengklasifikasian metode SVM terhadap analisis sentimen ulasan atau ulasan aplikasi JMO tersebut menghasilkan nilai akurasi terbaik dengan skor akurasi 96 persen, ketepatan 92 persen, recall 96 persen, dan skor f1 96 persen. Selain itu, hasil ulasan terbanyak adalah ulasan berbasis kernel [3].

Penelitian ini merancang sistem analisis sentimen ulasan YouTube dan GoTube di Google Play dengan metode SVM. Sistem ini mengklasifikasikan sentimen pengguna berdasarkan pola teks dalam ulasan untuk memahami persepsi mereka terhadap kedua platform.

II. METODE

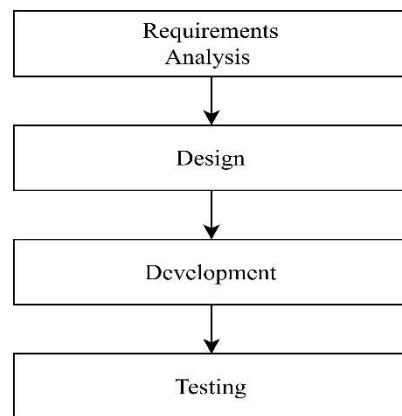
Pada tahap ini merupakan bagian menganalisa kebutuhan sistem yang akan dibangun, setelah memperolah data dari tahapan-tahapan sebelumnya maka data tersebut akan menjadi acuan dalam perancangan sistem. Dalam mendukung parancangan sistem implementasi analisis sentimen ulasan Youtube dan Go Tube berbahasa indonesia pada google play memakai algoritma Support vector machine (SVM)[8].

Penelitian ini memakai data yang berasal dari Ulasan di Google Play dengan melakukan websraping menggunakan library Python Scrapper. Data review yang akan diambil berupa ulasan yang terdapat pada situs Google Play. variabel

metadata data review yang diambil adalah username pengguna, isi dari ulasan dan jumlah bintang ulasan.

A. Alur Penelitian

Peneliti mengacu pada alur penelitian gambar 1 berikut.



Gambar 1. Alur Penelitian

1) *Requirements Analysis*: pada proses pengumpulan data peneliti mendapatkan data dari Google Play dengan cara web scraping pada aplikasi YouTube dan GoTube. Dalam dataset ini, terdapat 2000 ulasan untuk pembagian data set masing-masing data training 80% dan data testing 20%. Untuk Mengetahui Data Positif atau negatif tergantung dari sentimen dan rating, rating 1, 2, 3 sentimen negatif, 4, 5 sentimen positif yang diberikan oleh pengguna aplikasi youtube dan gotube. Pelabelan dilakukan secara otomatis dengan heuristik rating yang berikan pengguna.

TABEL 1
REVIEW REPRESENTATIF YANG DIANALISIS

Contoh Review	Klasifikasi	Alasan	Interpretasi
"GoTube keren banget, nonton video tanpa iklan!"	Positif	Fitur bebas iklan	Pengguna menghargai pengalaman menonton tanpa gangguan.
"Kenapa kadang GoTube tiba-tiba nutup sendiri ya?"	Negatif	Bug/ crash aplikasi	Masalah stabilitas aplikasi menjadi sumber ketidakpuasan.
"YouTube update terbaru malah bikin loading makin lama."	Negatif	Penurunan performa	Update aplikasi yang memperburuk performa memicu keluhan.
"Suka banget,	Positif	Penambahan fitur baru	Fitur tambahan meningkatkan kepuasan pengguna.
You Tube makin banyak fitur canggih kayak shorts."			
"Lumayan sih, kadang error kecil tapi masih oke dipakai."	Positif	Bug minor namun bisa ditoleransi	Pengguna tetap menerima aplikasi meski ada kekurangan kecil.

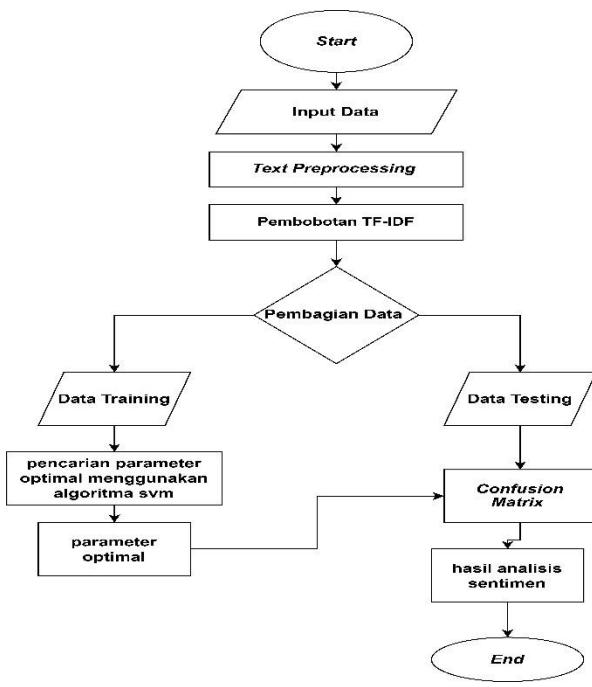
2. *Design*: Design atau Perancangan sistem mewakili sistem dalam alur kerja atau skema system.

3. *Development*: Tahapan ini adalah memasukkan metode yang dipilih ke dalam sistem memakai source code. Proses ini mencakup langkah-langkah seperti preprocessing teks TF-IDF, pemisahan data, pelatihan model Support Vector Machine (SVM), dan evaluasi. Pemahaman, penggunaan dan perbaikan di masa depan bergantung pada dokumentasi dan pemeliharaan kode.

4. *Testing*: Pengujian sistem dilakukan untuk memastikan apakah sistem berfungsi dengan baik dan memenuhi kebutuhan dan harapan dalam menjalankan fungsinya atau apakah ada ketidaksesuaian yang memerlukan perbaikan.

B. Skema Sistem

berikut skema sistem dari Analisis Sentimen Ulasan Youtube dan Go Tube Pada Google Play Menggunakan Metode Support Vector Machine (SVM) Di Indonesia.



Gambar 2. Skema Sistem

1. *Input Dataset*: Dataset berbentuk Comma-Separated Values (CSV) ini berisi teks ulasan, tanggal ulasan, rating, username.

2. *Text Preprocessing*: Melakukan pra-pemrosesan teks untuk membersihkan dan menyiapkan data ulasan. Menghapus tanda baca, angka, dan karakter khusus. Mengubah teks menjadi huruf kecil. Menghapus kata-kata umum yang tidak memberikan nilai analisis (stop words)[9]. Stemming atau lemmatization untuk mengubah kata ke bentuk dasarnya. Tokenisasi untuk memecah teks menjadi kata-kata atau token.

3. *Pembobotan TF-IDF*: Menggunakan teknik TF-IDF (Term Frequency-Inverse Document Frequency) untuk

mengubah teks ulasan menjadi fitur numerik[10]. Tahapan dalam pembobotan kata yaitu

a) *Term Frequency (TF)*: Term Frequency merupakan jumlah banyak kata yang muncul dalam suatu dokumen. Sedangkan wtf merupakan bobot dari tf yang dihitung dengan algoritma.

$$wtf_{td} = \begin{cases} 1 + \log_{10}(tf_{td}), & \text{if } tf_{td} > 0 \\ 0, & \text{if } tf_{td} = 0 \end{cases} \quad (1)$$

b) *Inverse Document Frequency (IDF)*: Jumlah dokumen yang mengandung suatu kata disebut dokumen frekuensi, juga disebut df. Sementara idf adalah bobot frekuensi atau jumlah dokumen. Semakin besar idf, semakin jarang kata muncul dalam dokumen

$$idf_t = \log_{10} \left(\frac{N}{df_t} \right) \quad (2)$$

c) *Term Frequency-Inverse Document Frequency (TF-IDF)*: TF-IDF merupakan pembobotan yang didapatkan dari hasil perkalian dari pembobotan Term Frequency (wtf) dengan Inverse Document Frequency (idf_t) dari suatu term.

$$w_{t,d} = w_{tf} \times idf_t \quad (3)$$

Keterangan:

$tf_{t,d}$: Jumlah kemunculan term (t) pada dokumen (d)

N : Jumlah dokumen teks

df_t : Jumlah dokumen yang mengandung term (t)

4. *Pembagian Data*: tahap ini dilakukan untuk membagi data menjadi data latih dan data uji[11].

5. *Data Training*: Pada tahap ini, pelatihan dilakukan dengan menggunakan 80% data latihan dari ulasan yang telah diproses dan diberi label (kelas) untuk membangun model.

6. *Data Testing*: Tahap ini dilakukan untuk mengetahui keakuratan model yang dibangun pada tahap training sebanyak 20%. Pada tahap ini, label (kelas) dari data diprediksi dengan menggunakan data uji.

7. *Klasifikasi data dengan Support Vector Machine (SVM)* : Penelitian ini menggunakan metode Support Vector Machine (SVM) dengan kernel Radial basis function (RBF)[12]. Rumus Radial basis function (RBF) :

$$K(x_i, x_d) = \exp(-Y \| X_i - X_d \|_2^2), Y > 0 \quad (4)$$

Kernel Radial basis function (RBF) dipilih karena kemampuan Menangani Data Nonlinear. Data ulasan di Google Play (termasuk sentimen positif, negatif, netral) biasanya tidak linear — artinya, hubungan antar fitur (kata-kata, frekuensi kata, dll) dengan label sentimen tidak bisa dipisahkan dengan garis lurus[13].

RBF dapat membuat batas keputusan yang melengkung, sehingga cocok untuk data seperti ini.

Berdasarkan banyak penelitian sebelumnya, SVM dengan RBF kernel sering memberikan akurasi yang tinggi pada tugas-tugas analisis sentimen[14].

8. *Confusion Matrix*: Pada tahap ini, mengukur akurasi, perhitungan, atau spesifikasi algoritma yang digunakan yaitu metode Support Vector Machine (SVM), dengan memastikan bahwa nilai dan standar yang digunakan sesuai. nilai akurasi dapat dihitung dengan menggunakan persamaan[15].

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \quad (5)$$

Nilai presisi menunjukkan jumlah data kategori positif yang diklasifikasikan dengan benar dibagi dengan total data positif yang diklasifikasikan. presisi dapat dihitung dengan menggunakan persamaan berikut

$$\text{Precision} = \frac{TP}{TP + FP} \times 100\% \quad (6)$$

Sementara itu, Recall menunjukkan bahwa sebagian besar data kategori positif telah terkласifikasi dengan benar oleh sistem.

$$\text{Recall} = \frac{TP}{TP + FN} \times 100\% \quad (7)$$

Kasus yang salah diidentifikasi dalam kumpulan data untuk menunjukkan tingkat kesalahan sistem yang digunakan dikenal sebagai error. berikut dapat digunakan untuk menghitung persentase error:

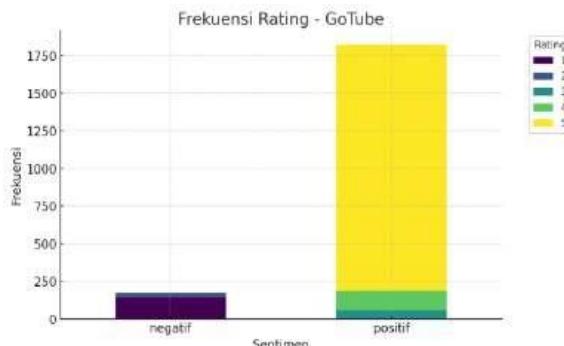
$$\text{Error} = \frac{FP}{TP} \times 100\% \quad (8)$$

III. HASIL DAN PEMBAHASAN

Penelitian ini menggunakan metode Support Vector Machine (SVM) untuk menganalisis sentimen pada ulasan YouTube dan GoTube di Google Play. Hasil penelitian ini akan dikembangkan menjadi sebuah website yang memungkinkan pengguna untuk melihat analisis sentimen. Website ini akan menampilkan klasifikasi sentimen (positif dan negatif), visualisasi data, serta perbandingan tingkat kepuasan pengguna antara YouTube dan GoTube.

Berikut adalah tabel frekuensi rating berdasarkan sentimen untuk data latih kedua aplikasi dimana masing masing menggunakan 2000 data sehingga totalnya ada 4000 data. Yang dijadikan data latih.

A. Frekuensi Rating dan Sentimen Aplikasi Gotube



Gambar 3. Frekuensi Sentimen Pada Data Latih Sentimen Gotube

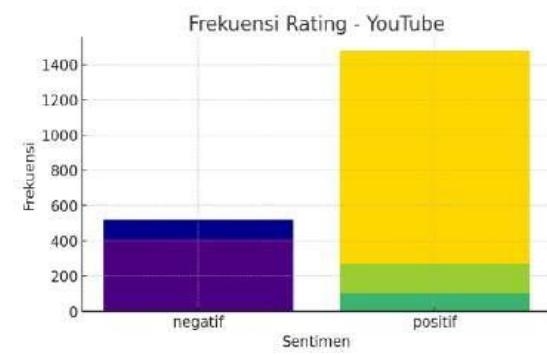
TABEL 1
FREKUENSI PADA DATA LATIH APLIKASI GOTUBE

Sentimen	Rating 1	Rating 2	Rating 3	Rating 4	Rating 5
Negatif	146	29	0	0	0
Positif	0	0	57	132	1636

B. Frekuensi Rating dan Sentimen Aplikasi Youtube

TABEL 2
FREKUENSI PADA DATA LATIH APLIKASI YOUTUBE

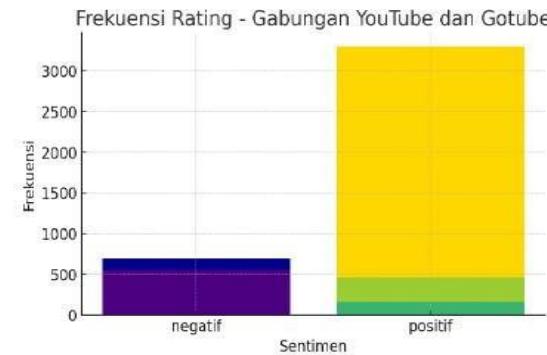
Sentimen	Rating 1	Rating 2	Rating 3	Rating 4	Rating 5
Negatif	412	109	0	0	0
Positif	0	0	103	168	1208



Gambar 4. Frekuensi Sentimen Pada Data Latih Sentimen Youtube

C. Data yang digunakan

Dimana dalam keseluruhan data aplikasi Gotube dan Youtube seperti berikut



Gambar 5. Frekuensi Sentimen Pada Data Latih Sentimen

TABEL 3
FREKUENSI SENTIMEN PADA DATA LATIH SENTIMEN

Sentimen	Rating 1	Rating 2	Rating 3	Rating 4	Rating 5
Negatif	412	109	0	0	0
Positif	0	0	103	168	1208

Dimana untuk seluruh data akan di tempatkan pada lampiran penelitian dikarenakan datanya yang hingga 4000 data latih.

1. Hasil Proses Pelatihan Model

Tahapan selanjutnya akan dilakukan proses pelatihan model diman akan digunakan dengan melakukan iterasi tahapan SVM pada kedua Analisa sentimen. Kemudian akan

dilakukan visualisasi hasil pelatihan model tadi dimana nantinya model yang sudah dilatih akan disimpan dalam bentuk file sehingga bisa digunakan kapanpun diperlukan tanpa harus melakukan pelatihan ulang. Dimana hasil evaluasi model nya sebagai berikut.

TABEL 4

EVALUASI MODEL UNTUK KOMBINASI DATA YOUTUBE DAN GOTUBE

Metric	Value
Accuracy	0.8388
Precision	0.8388
Recall	1.0000
F1 Score	0.9123

TABEL 5

EVALUASI MODEL UNTUK DATA YOUTUBE

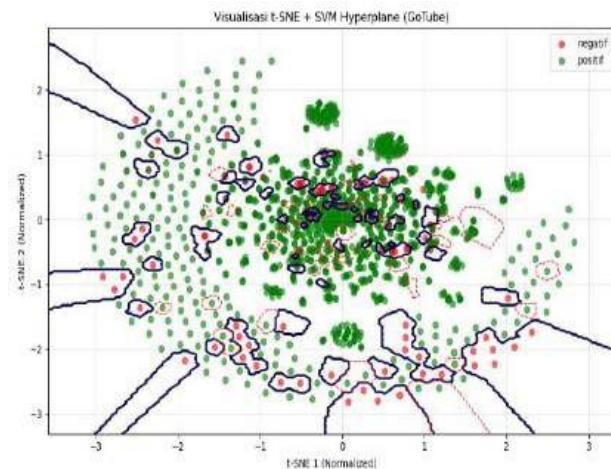
Metric	Value
Accuracy	0.7564
Precision	0.7564
Recall	1.0000
F1 Score	0.8613

TABEL 6

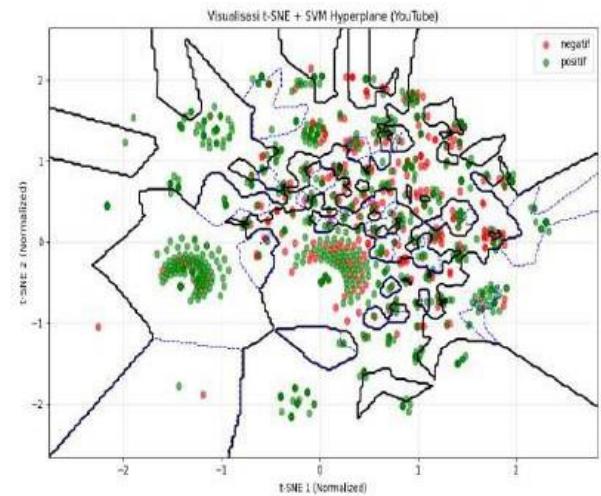
HASIL EVALUASI MODEL UNTUK DATA GOTUBE

Metric	Value
Accuracy	0.9190
Precision	0.9190
Recall	1.0000
F1 Score	0.9578

Visualiasi Model menggunakan kombinasi TNSE dan SVM Hyperplane.



Gambar 6. Visualisasi Model TNSE dan SVM Hyperplane Model untuk Data Gotube



Gambar 7. Visualisasi Model TNSE dan SVM Hyperplane Model untuk Data YouTube

2. Implementasi Model Pada Data Uji

Maka tahapan selanjutnya adalah menggunakan model yang sudah di simpan untuk prediksi data uji dimana dalam hal ini, akan menggunakan masing masing 400 data uji untuk masing masing GoTube dan YouTube. Sehingga totalnya 800 data untuk kedua data. Berikut adalah tabel frekuensi jumlah data berdasarkan rating untuk data Gotube :

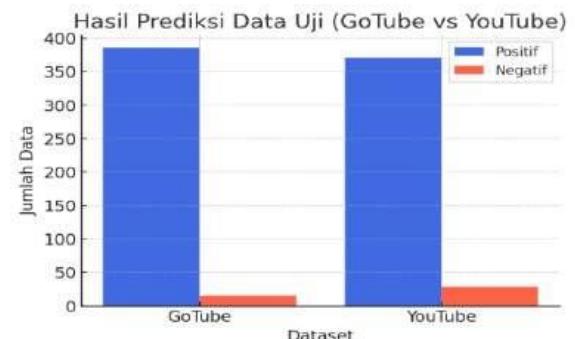
TABEL 7
DATA UJI UNTUK GOTUBE DAN YOUTUBE

Rating	Jumlah	Data GoTube
1	23	81
2	10	31
3	12	23
4	23	35
5	332	230

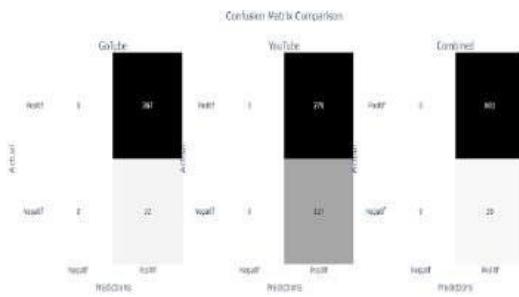
Dimana dengan menggunakan model yang sudah disimpan maka berikut adalah hasilnya:

TABEL 8
HASIL IMPLEMENTASI MODEL SVM UNTUK ANALISA SENTIMEN

Sumber Data	Positif	Negatif
GoTube	385	15
YouTube	371	29



Gambar 8. Grafik Perbandingan Analisa



Gambar 9. Hasil Confusion Matrix

1. *GoTube*: Pada GoTube, model memiliki performa yang sangat baik dalam memprediksi kelas "Positif" dengan 367 prediksi benar, tetapi tidak ada prediksi benar untuk kelas "Negatif". Hal ini menunjukkan bahwa model cenderung bias terhadap kelas "Positif" dan mungkin kurang mampu mengidentifikasi komentar "Negatif". Ini bisa disebabkan oleh ketidakseimbangan data latih atau model yang terlalu fokus pada fitur-fitur yang lebih umum pada kelas "Positif".

2. *YouTube*: Untuk YouTube, model juga menunjukkan bias yang kuat terhadap kelas "Positif" dengan 276 prediksi benar, sementara tidak ada prediksi benar untuk kelas "Negatif". Namun, jumlah false positive (124) lebih tinggi dibandingkan GoTube, yang menunjukkan bahwa model lebih sering salah mengklasifikasikan komentar "Negatif" sebagai "Positif". Ini mungkin mengindikasikan bahwa data YouTube memiliki karakteristik yang lebih kompleks atau beragam, sehingga model kesulitan membedakan kedua kelas.

3. *Combined*: Pada data gabungan (Combined), model kembali menunjukkan bias yang kuat terhadap kelas "Positif" dengan 601 prediksi benar, sementara tidak ada prediksi benar untuk kelas "Negatif". Meskipun jumlah prediksi benar untuk "Positif" meningkat, model tetap gagal mengidentifikasi kelas "Negatif". Ini menunjukkan bahwa masalah bias model terhadap kelas "Positif" konsisten di semua dataset, dan perlu dilakukan perbaikan seperti penyeimbangan data atau penyesuaian threshold klasifikasi untuk meningkatkan kemampuan model dalam mengidentifikasi kelas "Negatif".

3. Hasil Evaluasi Model pada data Uji

TABEL 8
HASIL EVALUASI MODEL SELURUH APLIKASI

Metric	Value
Accuracy	0.8388
Precision	0.8388
Recall	1.0000
F1 Score	0.9123

TABEL 9
HASIL EVALUASI MODEL APLIKASI YOUTUBE

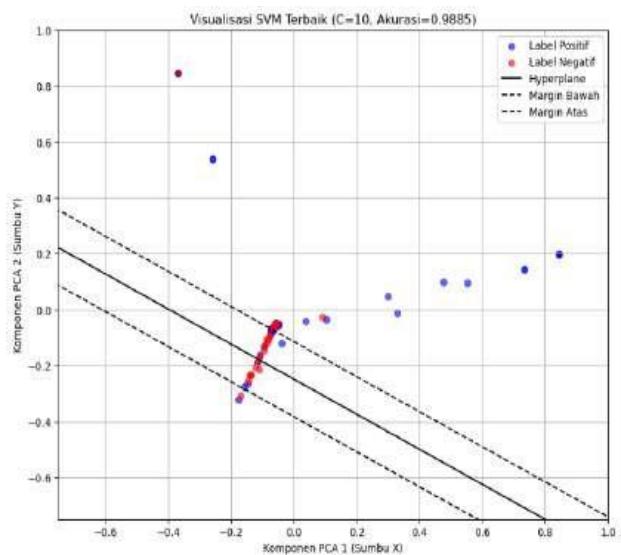
Metric	Value
Accuracy	0.7564
Precision	0.7564
Recall	1.0000
F1 Score	0.8613

TABEL 10
HASIL EVALUASI MODEL APLIKASI GOTUBE

Metric	Value
Accuracy	0.9190
Precision	0.9190
Recall	1.0000
F1 Score	0.9578

1. Hasil Visualisasi Model

Tahapan berikutnya adalah mengimplementasikan seluruh data yang sudah ada disebutkan sebelumnya. Dimana di visualisasikan modelnya dalam bentuk grafik Analisis Komponen Utama (PCA) dengan garis Hyperplane. visualisasi antara dua sumbu X dan Y dari nilai Support Vector Machine (SVM) yang Sudah di normalisasi agar dapat menggambarkan data secara keseluruhan dalam model Support Vector Machine (SVM) yang Sudah di bentuk. Dimana terdapat garis Hyperplane dan margin untuk memisahkan antara positif dan negative pada sentiment. Maka hasilnya sebagai berikut.



Gambar 10. Hasil Visualisasi SVM

Dimana dapat diambil kesimpulan sebagai berikut.

1. Distribusi Data:

a) *Sentimen Positif (biru)*: Titik-titik biru mewakili ulasan dengan sentimen positif. Sebagian besar titik biru terkonsentrasi di sisi kanan grafik (Komponen PCA 1 > 0) dan beberapa titik tersebar di sisi kiri (Komponen PCA 1 < 0).

b) *Sentimen Negatif (merah)*: Titik-titik merah mewakili ulasan dengan sentimen negatif. Titik merah lebih terkonsentrasi di sisi kiri grafik (Komponen PCA 1 < 0), tetapi ada beberapa titik merah yang berada di sisi kanan, tumpang tindih dengan titik biru.

c) *Tumpang Tindih*: Ada tumpang tindih yang cukup signifikan antara titik biru dan merah di sekitar area tengah grafik (Komponen PCA 1 sekitar -0.2 hingga 0.2). Hal ini menunjukkan bahwa pemisahan sentimen dalam ruang 2D

(setelah PCA) tidak sepenuhnya jelas, meskipun model SVM memiliki akurasi 98.85% dalam ruang asli.

2. *Hyperplane*: Hyperplane (garis hitam solid) membagi ruang 2D secara diagonal dari kiri bawah (Komponen PCA 1 ≈ -0.6 , Komponen PCA 2 ≈ -0.4) ke kanan atas

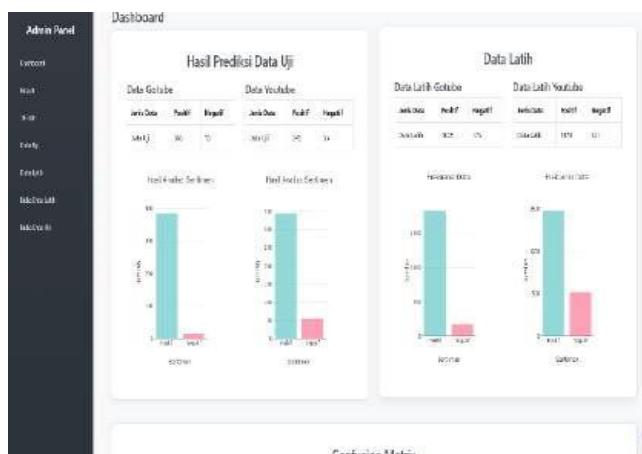
(Komponen PCA 1 ≈ 0.8 , Komponen PCA 2 ≈ 0.2). Hyperplane ini berusaha memisahkan titik biru (positif) di sisi kanan atas dan titik merah (negatif) di sisi kiri bawah.

3. *Margin*: Margin bawah dan atas (garis putus-putus hitam) berada di kedua sisi Hyperplane. Margin ini cukup sempit, yang menunjukkan bahwa model SVM dengan C=10 lebih fokus pada meminimalkan kesalahan klasifikasi daripada memaksimalkan margin (sesuai dengan nilai C yang besar, yang mengurangi toleransi terhadap kesalahan).

Dalam penelitian ini, PCA digunakan untuk mengurangi dimensi data fitur sebelum proses klasifikasi menggunakan SVM. Tujuannya adalah untuk menghilangkan redundansi informasi dan mempercepat proses training model. Kami memilih jumlah komponen utama yang mempertahankan 95% dari total variansi data, yang menghasilkan 50 komponen dari 300 fitur awal. Selain itu, PCA juga digunakan untuk visualisasi data pada ruang 2 dimensi, guna melihat separabilitas antar kelas sentimen (positif, netral, negatif). Untuk keperluan ini, kami memilih 2 komponen utama.

4. Hasil Implementasi Sistem

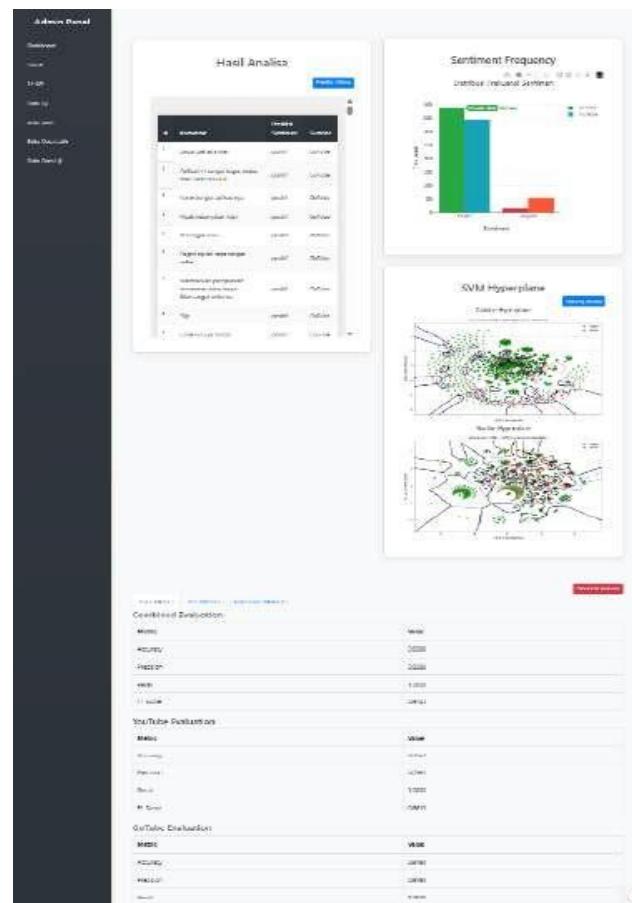
Berikut adalah hasil implementasi dalam bentuk website yang dibangun menggunakan Bahasa pemrograman python dan framework flask.



Gambar 11. Halaman Dashboard

Gambar 11 secara fungsional, halaman dashboard ini terbagi menjadi beberapa bagian utama yang menyajikan berbagai elemen visual dan data statistik yang relevan. Salah satu fitur utamanya adalah tabel dinamis yang menampilkan informasi terperinci dengan dukungan fitur filter dan pencarian, sehingga memudahkan pengguna dalam menavigasi dan mengelola data. Selain itu, halaman ini juga dilengkapi dengan grafik interaktif dalam berbagai format, seperti diagram batang, garis, dan lingkaran, yang

memberikan representasi visual terhadap tren atau pola yang terdapat dalam data.



Gambar 12. Halaman Hasil Analisa

Gambar 12 Halaman Hasil Analisa merupakan bagian dari Dashboard administrasi yang dirancang untuk menyajikan hasil analisis sentimen secara komprehensif. Fokus utama halaman ini mencakup prediksi sentimen, distribusi frekuensi sentimen, visualisasi model, serta metrik evaluasi. Secara fungsional, halaman ini menampilkan tabel prediksi sentimen, grafik distribusi frekuensi untuk membandingkan sentimen dari dua sumber data, visualisasi Hyperplane dari algoritma Support Vector Machine (SVM) dalam ruang t-SNE guna mempermudah pemahaman terhadap pemisahan kelas, dan tabel metrik evaluasi seperti akurasi, presisi, recall, dan F1-score dalam berbagai skenario pengujian.

Halaman ini juga dilengkapi dengan tombol interaktif untuk melatih model, mengevaluasi ulang performa, atau mengunggah data baru, sehingga memungkinkan pengguna untuk memantau kinerja model dan membuat keputusan berbasis data secara efisien.

No.	Komentar	Status
1	Sehr schönes Video!	
2	Super interessant!	
3	Ein sehr interessantes Video.	
4	Ein sehr interessantes Video.	
5	Ein sehr interessantes Video.	
6	Ein sehr interessantes Video.	
7	Ein sehr interessantes Video.	
8	Ein sehr interessantes Video.	
9	Ein sehr interessantes Video.	

Gambar 13. Halaman Data Uji

Gambar 13 Halaman Data Uji merupakan bagian dari Dashboard administrasi yang menyajikan tampilan terperinci dari data uji dalam format tabel, sehingga memungkinkan pengguna untuk meninjau informasi spesifik dari setiap entri data. Secara fungsional, halaman ini menyajikan kolom-kolom data yang relevan, dilengkapi dengan tombol untuk menyimpan perubahan atau hasil analisis. Selain itu, terdapat menu navigasi di sisi kiri yang memudahkan pengguna untuk berpindah ke fitur lain seperti Dashboard utama, halaman hasil analisis, maupun halaman pengelolaan data, sehingga mendukung proses verifikasi dan pengelolaan data uji secara efisien.

V. KESIMPULAN

Berdasarkan hasil dan pembahasan, dapat disimpulkan bahwa analisis sentimen terhadap ulasan aplikasi YouTube dan GoTube di Google Play Store Indonesia telah berhasil dilakukan menggunakan metode Support Vector Machine (SVM). Data ulasan dikumpulkan melalui teknik web scraping dan terdiri dari 1.600 ulasan, masing-masing 800 ulasan untuk aplikasi YouTube dan GoTube, dengan rentang waktu pengambilan data tahun 2023–2024. Data tersebut dibagi menjadi 80% untuk data latih dan 20% untuk data uji.

Model SVM yang dibangun mampu mencapai akurasi keseluruhan sebesar 94,50%, dengan akurasi per aplikasi sebesar 97,01% untuk YouTube dan 92,66% untuk GoTube. Evaluasi performa model menunjukkan bahwa YouTube memiliki nilai akurasi 75,64%, presisi 75,64%, recall 100,00%, dan F1-score 86,13%. Sementara itu, GoTube menunjukkan performa yang lebih tinggi dengan akurasi 91,90%, presisi 91,90%, recall 100,00%, dan F1-score 95,78%. Hal ini mengindikasikan bahwa ulasan terhadap GoTube cenderung lebih positif dibandingkan dengan YouTube.

Visualisasi menggunakan teknik t-SNE dan SVM hyperplane terhadap data YouTube (sebanyak 2.000 data setelah augmentasi) memperlihatkan distribusi sentimen positif (ditandai warna hijau) dan negatif (ditandai warna merah) yang tersebar dalam ruang dua dimensi yang telah

dinormalisasi, menunjukkan pemisahan kelas yang cukup jelas oleh model.

DAFTAR PUSTAKA

- [1] D. F. Zhafira, B. Rahayudi, and I. Indriati, “Analisis Sentimen Kebijakan Kampus Merdeka Menggunakan Naive Bayes dan Pembobotan TF-IDF Berdasarkan Komentar pada Youtube,” *Jurnal Sistem Informasi, Teknologi Informasi, dan Edukasi Sistem Informasi*, vol. 2, no. 1, pp. 55–63, Aug. 2021, doi: 10.25126/justsi.v2i1.24.
- [2] J. R. P. A. Yudha and S. Sundari, “Manfaat Media Pembelajaran YouTube terhadap Capaian Kompetensi Mahasiswa,” *Journal of Telenursing (JOTING)*, vol. 3, no. 2, pp. 538–545, Oct. 2021, doi:10.31539/joting.v3i2.2561.
- [3] V. Fitriyana, Lutfi Hakim, Dian Candra Rini Novitasari, and Ahmad Hanif Asyhar, “Analisis Sentimen Ulasan Aplikasi Jamsostek Mobile Menggunakan Metode Support Vector Machine,” *Jurnal Buana Informatika*, vol. 14, no. 01, pp. 40–49, Apr. 2023, doi: 10.24002/jbi.v14i01.6909.
- [4] L. Agustina, A. O. Fayardi, and I. Irwansyah, “Online Review Indikator Penilaian Kredibilitas Online dalam Platform E-commerce,” *Jurnal ILMU KOMUNIKASI*, vol. 15, no. 2, pp. 141–154, Nov.2018, doi: 10.24002/jik.v15i2.1320.
- [5] R. Rosita and D. Evalin, “Pengaruh Kualitas Konten Tik Tok Terhadap Customer Engagement Pada Customer Queensha,” *Jurnal Lentera Bisnis*, vol. 13, no. 2, pp. 1061–1071, May 2024, doi: 10.34127/jrlab.v13i2.1129.
- [6] D. Muallfah, Ramadhoni, R. Gunawan, and D. Mulyadipa Suratno, “Analisis Sentimen Komentar YouTube TvOne Tentang Ustadz Abdul Somad Dideportasi Dari Singapura Menggunakan Algoritma SVM,” *JURNAL FASILKOM*, vol. 13, no. 01, pp. 72–80, Jul. 2023, doi: 10.37859/jf.v13i01.4920.
- [7] M. D. Hendriyanto, A. A. Ridha, and U. Enri, “Analisis Sentimen Ulasan Aplikasi Mola Pada Google Play Store Menggunakan Algoritma Support Vector Machine,” *INTECOMS: Journal of Information Technology and Computer Science*, vol. 5, no. 1, pp. 1–7, Apr. 2022, doi: 10.31539/intecoms.v5i1.3708.
- [8] A. Z. Amrullah, A. Sofyan Anas, M. Adrian, and J. Hidayat, “Analisis Sentimen Movie Review Menggunakan Naive Bayes Classifier Dengan Seleksi Fitur Chi Square,” *Jurnal BITE*, vol. 2, no. 1, pp. 40–44, 2020, doi: 10.30812/bite.v2i1.804.
- [9] A. N. Indraini and I. Ernawati, “Analisis Sentimen Terhadap Pembelajaran Daring Di Indonesia Menggunakan Support Vector Machine (SVM),” *Jurnal Ilmiah FIFO*, vol. 14, no. 1, pp. 68–80, Jul. 2022, doi: 10.22441/fifo.2022.v14i1.007.
- [10] A. Deolika and E. Taufiq Luthfi, “Analisis Pembobotan Kata Pada Klasifikasi Text Mining,” *Jurnal Teknologi Informasi*, vol. 3, no. 2, 2019.
- [11] T. Tinaliah and T. Elizabeth, “Analisis Sentimen Ulasan Aplikasi PrimaKu Menggunakan Metode Support Vector Machine,” *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, vol. 9, no. 4, pp. 3436–3442, Dec. 2022, doi: 10.35957/jatisi.v9i4.3586.
- [12] N. Aula, M. Ula, and L. Rosnita, “Analisis Sentimen Review Customer Terhadap Perusahaan Ekspedisi Jne, & T Express Dan Pos Indonesia Menggunakan Metode Support Vector Machine (Svm),” *Journal Of Informatics And Computer Science*, vol. 9, no. 1, pp. 81–86, Apr. 2023, doi: 10.33143/jics.v9i1.2947.
- [13] S. B. Setiawan and R. Isnain, “Jurnal Media Informatika Budidarma Sentimen Analisis Masyarakat Terhadap Pembangunan IKN Menggunakan Algoritma Lexicon Based Approach dan Naïve Bayes,” *Jurnal Media Informatika*.
- [14] N. Aula, M. Ula, and L. Rosnita, “Analisis Sentimen Review Customer Terhadap Perusahaan Ekspedisi Jne, & T Express Dan Pos Indonesia Menggunakan Metode Support Vector Machine (Svm),” *Journal Of Informatics And Computer Science*, vol. 9, no. 1, pp. 81–86, Apr. 2023, doi: 10.33143/jics.v9i1.2947.
- [15] S. B. Setiawan and R. Isnain, “Jurnal Media Informatika Budidarma Sentimen Analisis Masyarakat Terhadap Pembangunan IKN Menggunakan Algoritma Lexicon Based Approach dan Naïve Bayes,” *Jurnal Media Informatika*.

- Menggunakan Algoritma Lexicon Based Approach dan Naïve Bayes,” Jurnal Media Informatika Budidarma, vol. 8, no. 2, pp. 1019–1030, 2024, doi: 10.30865/mib.v8i2.7605.
- [16] H. Harnelia, “Analisis Sentimen Review Skincare Skintific Dengan Algoritma Support Vector Machine (SVM),” Jurnal Informatika dan Teknik Elektro Terapan, vol. 12, no. 2, pp. 994–1002, Apr. 2024, doi: 10.23960/jitet.v12i2.4095.
- [17] D. Untuk, M. Persyaratan, M. Gelar, S. Komputer, D. Oleh, and A. Fatihin, “Analisis Sentimen Terhadap Ulasan Aplikasi Mobile Menggunakan Metode Support Vector Machine (Svm) Dan Pendekatan Lexicon Based.