

IMPLEMENTASI METODE SUPPORT VECTOR MACHINE (SVM) UNTUK ANALISIS SENTIMEN KOMENTAR YOUTUBE BERBASIS PYTHON

Oktaria Gisela, Sigit Auliana, Basuki Rakhim Setya Permana

Program Studi Ilmu Komputer, Fakultas Ilmu Komputer, Universitas Bina Bangsa
oktariagisela92@gmail.com

ABSTRAK

Penelitian ini bertujuan untuk mengimplementasikan metode Support Vector Machine (SVM) dalam melakukan analisis sentimen terhadap komentar pengguna pada platform YouTube berbasis Python. Analisis sentimen dilakukan untuk mengetahui kecenderungan opini publik terhadap suatu konten dalam bentuk sentimen positif, netral, dan negatif. Data yang digunakan berupa komentar YouTube yang diambil menggunakan API, kemudian diproses melalui tahapan text preprocessing yang meliputi cleaning, tokenizing, stopword removal, dan stemming. Selanjutnya, data teks diubah menjadi representasi numerik menggunakan metode Term Frequency-Inverse Document Frequency (TF-IDF). Proses klasifikasi dilakukan menggunakan algoritma SVM yang dikenal memiliki performa baik dalam menangani data berdimensi tinggi seperti teks. Sistem yang dikembangkan mampu melakukan analisis secara otomatis mulai dari pengambilan data hingga penyajian hasil dalam bentuk grafik, tabel, serta file unduhan CSV. Hasil pengujian menunjukkan bahwa sistem dapat berjalan dengan baik sesuai dengan fungsi yang dirancang. Berdasarkan pengujian White Box dan Black Box, seluruh fitur sistem berfungsi dengan benar. Selain itu, hasil User Acceptance Testing (UAT) menunjukkan tingkat kepuasan pengguna sebesar 90,25%, yang menunjukkan bahwa sistem mudah digunakan dan memberikan hasil yang informatif. Meskipun demikian, penelitian ini memiliki keterbatasan dalam menangani bahasa informal, sarkasme, serta ketidakseimbangan dataset. Oleh karena itu, pengembangan lebih lanjut diperlukan untuk meningkatkan akurasi dan performa sistem.

Kata Kunci: Analisis Sentimen, Support Vector Machine, YouTube, TF-IDF, Python

1. PENDAHULUAN

Perkembangan teknologi informasi dan komunikasi telah mendorong peningkatan penggunaan media sosial sebagai sarana berbagi informasi dan opini secara luas. Salah satu platform yang memiliki peran penting dalam hal ini adalah YouTube, yang memungkinkan pengguna untuk mengunggah video serta memberikan komentar sebagai bentuk interaksi. Komentar-komentar tersebut mengandung berbagai opini yang mencerminkan persepsi publik terhadap suatu konten, sehingga dapat dimanfaatkan sebagai sumber informasi yang bernilai untuk analisis lebih lanjut.

Seiring dengan meningkatnya jumlah data komentar yang dihasilkan setiap hari, analisis manual menjadi tidak efektif dan efisien. Oleh karena itu, diperlukan pendekatan berbasis teknologi seperti analisis sentimen untuk mengidentifikasi dan mengklasifikasikan opini pengguna ke dalam kategori tertentu, seperti positif, negatif, dan netral [1]. Analisis sentimen merupakan bagian dari text mining yang bertujuan untuk mengekstraksi informasi subjektif dari data teks dalam skala besar [2].

Dalam implementasinya, analisis sentimen memerlukan metode klasifikasi yang mampu menangani data teks dengan dimensi tinggi dan kompleksitas yang tinggi. Salah satu metode yang banyak digunakan adalah Support Vector Machine (SVM). Metode ini dikenal memiliki kemampuan yang baik dalam melakukan klasifikasi data dengan margin optimal serta mampu mengatasi permasalahan overfitting [3]. Selain itu, SVM juga efektif dalam

menangani data teks yang telah direpresentasikan dalam bentuk fitur numerik, seperti menggunakan metode Term Frequency-Inverse Document Frequency (TF-IDF) [4].

Beberapa penelitian sebelumnya telah menerapkan analisis sentimen pada media sosial menggunakan berbagai metode, seperti Naïve Bayes, K-Nearest Neighbor, dan SVM. Namun, sebagian penelitian masih menghadapi kendala dalam hal akurasi, penanganan bahasa informal, serta keterbatasan dalam visualisasi hasil analisis [5]. Selain itu, belum banyak penelitian yang mengintegrasikan proses pengambilan data secara langsung dari YouTube dengan sistem analisis yang terintegrasi dan mudah digunakan oleh pengguna.

Berdasarkan permasalahan tersebut, penelitian ini bertujuan untuk mengimplementasikan metode Support Vector Machine (SVM) dalam analisis sentimen komentar YouTube berbasis Python. Sistem yang dikembangkan tidak hanya mampu melakukan klasifikasi sentimen secara otomatis, tetapi juga menyajikan hasil analisis dalam bentuk visualisasi yang informatif, seperti grafik distribusi sentimen dan tabel hasil analisis. Dengan demikian, penelitian ini diharapkan dapat memberikan kontribusi dalam pengembangan sistem analisis sentimen yang lebih efektif, efisien, dan mudah digunakan.

2. TINJAUAN PUSTAKA

2.1. Analisis Sentimen

Analisis sentimen merupakan teknik dalam text mining yang digunakan untuk mengidentifikasi dan

mengklasifikasikan opini atau emosi yang terkandung dalam data teks menjadi kategori tertentu, seperti positif, negatif, dan netral. Teknik ini banyak digunakan untuk menganalisis opini pengguna terhadap produk, layanan, maupun konten digital [6]. Menurut Liu, analisis sentimen bertujuan untuk mengekstraksi informasi subjektif dari teks dalam jumlah besar secara otomatis [7]. Dengan meningkatnya penggunaan media sosial, analisis sentimen menjadi semakin penting untuk memahami persepsi publik secara cepat dan efisien.

2.2. Text Mining dan Natural Language Processing (NLP)

Text mining merupakan proses ekstraksi informasi yang berguna dari data teks yang tidak terstruktur menggunakan teknik statistik dan machine learning [8]. Dalam implementasinya, text mining tidak terlepas dari Natural Language Processing (NLP) yang berperan dalam memahami bahasa alami manusia. NLP memungkinkan sistem komputer untuk memproses teks melalui tahapan seperti tokenizing, stopword removal, dan stemming [9]. Kombinasi antara text mining dan NLP menjadi dasar utama dalam pengembangan sistem analisis sentimen berbasis data teks.

2.3. Representasi Fitur TF-IDF

Dalam analisis sentimen, data teks perlu diubah menjadi bentuk numerik agar dapat diproses oleh algoritma klasifikasi. Salah satu metode yang umum digunakan adalah Term Frequency-Inverse Document Frequency (TF-IDF). Metode ini memberikan bobot pada setiap kata berdasarkan frekuensi kemunculannya dalam dokumen serta tingkat keunikannya dalam keseluruhan dataset [10]. TF-IDF mampu meningkatkan kualitas fitur dengan menonjolkan kata-kata penting yang berpengaruh terhadap sentimen, sehingga banyak digunakan dalam penelitian analisis teks.

2.4. Support Vector Machine (SVM)

Support Vector Machine (SVM) merupakan salah satu algoritma machine learning yang digunakan untuk klasifikasi dan regresi. SVM bekerja dengan mencari hyperplane optimal yang mampu memisahkan data ke dalam kelas yang berbeda dengan margin maksimum [11]. Keunggulan SVM terletak pada kemampuannya dalam menangani data berdimensi tinggi, seperti data teks, serta memiliki tingkat generalisasi yang baik [12]. Oleh karena itu, SVM banyak digunakan dalam penelitian analisis sentimen karena mampu memberikan hasil klasifikasi yang akurat.

2.5. Penelitian Terkait

Beberapa penelitian sebelumnya telah menerapkan analisis sentimen pada media sosial dengan berbagai metode. Penelitian oleh Giachanou dan Crestani menunjukkan bahwa metode machine

learning seperti SVM dan Naïve Bayes memiliki performa yang baik dalam analisis sentimen pada data media sosial [13]. Penelitian lain juga menunjukkan bahwa penggunaan TF-IDF sebagai metode ekstraksi fitur dapat meningkatkan akurasi klasifikasi sentimen [10]. Namun, sebagian penelitian masih menghadapi kendala dalam menangani bahasa informal dan konteks kalimat, sehingga diperlukan pengembangan lebih lanjut dalam sistem analisis sentimen.

Penelitian yang dilakukan oleh S. Auliana, B. R. S. Permana, U. Kalsum, dan M. Mudofar dengan judul “Detecting Hoax News with Web-Based SVM Classifier: A Machine Learning Approach to Combat Misinformation Using Support Vector Machine (SVM) Algorithm” menerapkan metode Support Vector Machine (SVM) pada sistem berbasis web untuk mendeteksi berita hoaks. Hasil penelitian menunjukkan bahwa algoritma SVM mampu mengklasifikasikan berita hoaks secara efektif dan membantu proses identifikasi informasi yang tidak valid. Perbedaan dengan penelitian yang dilakukan saat ini terletak pada objek penelitian, dimana penelitian terdahulu berfokus pada deteksi berita hoaks, sedangkan penelitian ini menerapkan algoritma SVM untuk melakukan analisis sentimen terhadap komentar pengguna YouTube.

Penelitian yang dilakukan oleh B. R. Setya Permana berjudul “Penerapan Algoritma C4.5 untuk Klasifikasi Kepatuhan Konsumen” menggunakan algoritma C4.5 dalam proses klasifikasi data konsumen. Hasil penelitian menunjukkan bahwa algoritma tersebut mampu menghasilkan klasifikasi kepatuhan konsumen secara baik dan akurat. Perbedaan dengan penelitian ini terletak pada metode yang digunakan, dimana penelitian sebelumnya menggunakan algoritma C4.5, sedangkan penelitian ini menggunakan algoritma Support Vector Machine untuk klasifikasi sentimen.

Penelitian yang dilakukan oleh H. Hamdan, O. Romli, S. Auliana, dan B. R. S. Permana dengan judul “Utilization of Virtual Assistance (Chatbot) for an Integrated Information Portal as Part of a Marine Tourism Promotion Strategy in Banten” memanfaatkan teknologi chatbot sebagai media penyampaian informasi wisata bahari. Hasil penelitian menunjukkan bahwa chatbot mampu memberikan informasi secara otomatis dan meningkatkan efektivitas layanan informasi kepada pengguna. Adapun perbedaannya dengan penelitian ini adalah penelitian terdahulu berfokus pada pengembangan sistem chatbot, sedangkan penelitian ini lebih menitikberatkan pada text mining dan analisis sentimen terhadap komentar YouTube.

Penelitian yang dilakukan oleh B. R. S. Permana, S. Auliana, dan R. K. Octavian berjudul “DigiField: Real-Time Smart Farming Monitoring System Using React.js and Express.js for Agricultural Digitalization” mengembangkan sistem monitoring pertanian secara real-time menggunakan React.js dan

Express.js. Hasil penelitian menunjukkan bahwa sistem mampu memberikan informasi kondisi pertanian secara cepat dan akurat. Perbedaannya dengan penelitian ini terletak pada bidang kajian, dimana penelitian sebelumnya berfokus pada digitalisasi pertanian, sedangkan penelitian ini membahas klasifikasi sentimen komentar YouTube menggunakan metode SVM.

Penelitian yang dilakukan oleh B. R. S. Permana, G. D. P. Aryono, dan S. N. Aliffito dengan judul “Information System Audit of FIF Mobile Collection at PT Federal International Finance POS Baros Using the COBIT 5 Framework” menerapkan kerangka kerja COBIT 5 dalam audit sistem informasi. Hasil penelitian menunjukkan bahwa audit mampu mengevaluasi tata kelola teknologi informasi dan memberikan rekomendasi perbaikan. Perbedaan dengan penelitian ini adalah penelitian terdahulu berfokus pada audit sistem informasi, sedangkan penelitian ini menerapkan teknik machine learning dalam analisis sentimen.

Penelitian oleh H. R. A. Fauzan, S. Auliana, dan B. R. S. Permana yang berjudul “Perancangan Aplikasi Manajemen Hotspot Menggunakan Framework CodeIgniter di PT Jaya Sampurna Network” menggunakan Framework CodeIgniter untuk membangun sistem manajemen hotspot. Hasil penelitian menunjukkan bahwa sistem mampu membantu proses pengelolaan jaringan hotspot secara efektif. Perbedaannya dengan penelitian ini terletak pada objek dan metode penelitian, dimana penelitian sebelumnya berfokus pada sistem jaringan, sedangkan penelitian ini membahas analisis sentimen berbasis Python menggunakan algoritma SVM.

Penelitian yang dilakukan oleh H. Ramadhan, S. Auliana, dan B. R. S. Permana dengan judul “Penerapan Framework CodeIgniter 3 untuk Pengembangan Aplikasi Penjualan pada Toko Nazeera Food” menghasilkan sistem penjualan berbasis web yang mampu membantu proses transaksi secara digital. Penelitian tersebut menggunakan Framework CodeIgniter 3 sebagai teknologi utama. Adapun perbedaannya dengan penelitian ini adalah penelitian sebelumnya berfokus pada pengembangan sistem penjualan, sedangkan penelitian ini membahas klasifikasi komentar YouTube menggunakan metode Support Vector Machine.

Penelitian oleh M. K. Naufallinho, S. Auliana, dan B. R. S. Permana berjudul “Rancangan Aplikasi Augmented Reality Pengenalan Buah-Buahan Tropis pada SMPN 1 Kibin Menggunakan Unity 3D dan Blender” memanfaatkan teknologi Augmented Reality, Unity 3D, dan Blender untuk menciptakan media pembelajaran interaktif. Hasil penelitian menunjukkan bahwa penggunaan teknologi AR mampu meningkatkan minat belajar siswa. Perbedaan dengan penelitian ini terletak pada pendekatan yang digunakan, dimana penelitian sebelumnya menggunakan teknologi augmented reality,

sedangkan penelitian ini menggunakan machine learning untuk klasifikasi teks.

Penelitian yang dilakukan oleh R. N. Azizah, S. Auliana, dan B. R. Setya dengan judul “Perancangan Denah 3D Berbasis Augmented Reality Menggunakan Blender Terintegrasi Unity 3D pada Rumah Sakit Mata Achmad Wardi Kota Serang” menghasilkan sistem visualisasi denah rumah sakit berbasis augmented reality yang interaktif. Sistem tersebut mampu membantu pengguna dalam memahami tata letak ruangan rumah sakit. Perbedaannya dengan penelitian ini adalah penelitian terdahulu berfokus pada visualisasi berbasis AR, sedangkan penelitian ini membahas analisis sentimen komentar media sosial menggunakan algoritma SVM.

Penelitian yang dilakukan oleh T. I. R. Rahman, S. Auliana, dan B. R. S. Permana dengan judul “Pembuatan Game Edukasi Jenis-Jenis Hewan Air Menggunakan Aplikasi Construct 2” mengembangkan game edukasi menggunakan Construct 2 sebagai media pembelajaran. Hasil penelitian menunjukkan bahwa game edukasi mampu meningkatkan interaktivitas dan minat belajar pengguna. Adapun perbedaannya dengan penelitian ini terletak pada bidang kajian, dimana penelitian sebelumnya berfokus pada pengembangan game edukasi, sedangkan penelitian ini membahas analisis sentimen komentar YouTube berbasis Python dengan algoritma Support Vector Machine.

3. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan metode machine learning untuk melakukan analisis sentimen terhadap komentar pengguna pada platform YouTube. Metode yang digunakan adalah Support Vector Machine (SVM) dengan tahapan pengolahan data teks mulai dari pengambilan data hingga evaluasi model.

3.1. Sumber dan Pengumpulan Data

Data yang digunakan dalam penelitian ini berupa komentar pengguna YouTube yang diambil secara otomatis menggunakan YouTube Data API. Proses pengambilan data dilakukan berdasarkan Video ID yang dimasukkan oleh pengguna, dengan jumlah komentar yang dapat disesuaikan melalui parameter sistem. Data yang diperoleh berupa teks komentar yang bersifat tidak terstruktur.

Penggunaan API memungkinkan pengambilan data secara real-time dan efisien dibandingkan metode manual [14]. Data yang terkumpul kemudian disimpan dalam bentuk dataset untuk diproses lebih lanjut.

3.2. Tahapan Penelitian

Tahapan penelitian yang dilakukan dalam sistem analisis sentimen ini meliputi beberapa langkah utama sebagai berikut:

a. Data Collection

Mengambil data komentar YouTube berdasarkan

- Video ID menggunakan API.
- b. Text Preprocessing
Membersihkan dan menyiapkan data teks agar dapat diproses oleh sistem.
- c. Feature Extraction
Mengubah data teks menjadi representasi numerik menggunakan TF-IDF.
- d. Classification (SVM)
Mengklasifikasikan sentimen menggunakan algoritma SVM.
- e. Evaluation
Mengukur performa model menggunakan metrik evaluasi.

3.3. Text Preprocessing

Tahap preprocessing dilakukan untuk meningkatkan kualitas data sebelum dilakukan klasifikasi. Proses ini meliputi:

- a. Cleaning: Menghapus tanda baca, simbol, dan karakter tidak penting
- b. Case Folding: Mengubah huruf menjadi huruf kecil
- c. Tokenizing: Memecah kalimat menjadi kata
- d. Stopword Removal: Menghapus kata umum yang tidak bermakna
- e. Stemming: Mengubah kata menjadi bentuk dasar

Preprocessing sangat penting dalam analisis sentimen karena dapat mengurangi noise pada data teks [15].

3.4. Feature Extraction dengan TF-IDF

Setelah preprocessing, data teks diubah menjadi bentuk numerik menggunakan metode TF-IDF (Term Frequency-Inverse Document Frequency). TF-IDF memberikan bobot pada setiap kata berdasarkan tingkat kepentingannya dalam dokumen dan keseluruhan dataset.

Metode ini efektif dalam menonjolkan kata-kata penting yang berpengaruh terhadap sentimen, sehingga meningkatkan akurasi klasifikasi [10].

3.5. Klasifikasi Menggunakan SVM

Pada tahap ini, data yang telah direpresentasikan dalam bentuk numerik digunakan untuk melatih model klasifikasi menggunakan algoritma Support Vector Machine (SVM). SVM bekerja dengan mencari hyperplane optimal yang memisahkan data ke dalam kelas sentimen positif, netral, dan negatif [11].

Dalam penelitian ini digunakan kernel linear, karena lebih efektif untuk data teks yang memiliki dimensi tinggi. Model dilatih menggunakan data latih (training data) dan diuji menggunakan data uji (testing data).

3.6. Evaluasi Model

Evaluasi model dilakukan untuk mengetahui performa klasifikasi sentimen yang dihasilkan. Metrik yang digunakan meliputi:

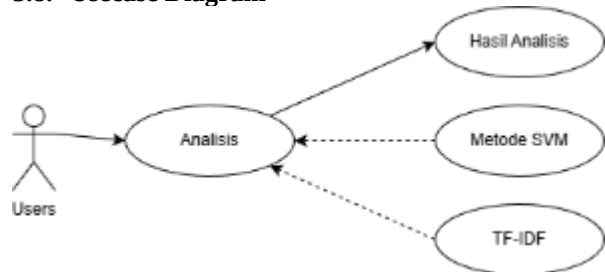
- a. Accuracy
- b. Precision
- c. Recall
- d. F1-Score

Metrik tersebut digunakan untuk mengukur tingkat ketepatan model dalam mengklasifikasikan data [16].

3.7. Alur Sistem Penelitian

Secara keseluruhan, alur sistem penelitian dimulai dari input Video ID oleh pengguna, pengambilan data komentar, preprocessing, ekstraksi fitur, klasifikasi menggunakan SVM, hingga penyajian hasil dalam bentuk visualisasi dan tabel. Sistem juga menyediakan fitur unduh hasil analisis dalam format CSV untuk memudahkan pengguna dalam pengolahan data lebih lanjut.

3.8. Usecase Diagram



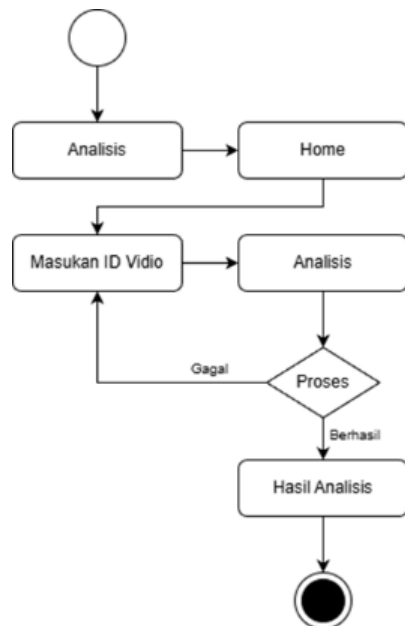
Gambar 1. Usecase Diagram

Use Case Diagram menggambarkan interaksi antara pengguna (user) dengan sistem aplikasi analisis sentimen komentar YouTube. Dalam sistem ini, pengguna berperan sebagai aktor utama yang melakukan beberapa aktivitas, yaitu memasukkan ID video YouTube, menjalankan proses pengambilan komentar, melakukan analisis sentimen, serta melihat hasil analisis. Sistem secara otomatis menangani proses pengambilan data melalui API, praproses teks, klasifikasi sentimen menggunakan Support Vector Machine, dan penyajian hasil. Diagram ini menunjukkan bahwa seluruh proses analisis terintegrasi dalam satu sistem berbasis web yang mudah diakses oleh pengguna tanpa memerlukan keahlian teknis pemrograman.

3.9. Activity Diagram

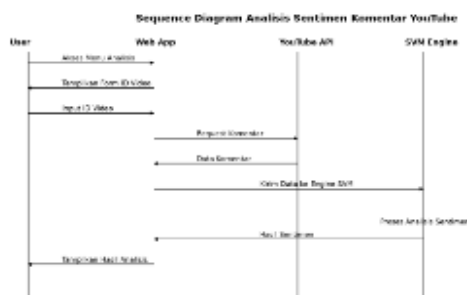
Activity Diagram menggambarkan alur aktivitas sistem secara berurutan mulai dari pengguna hingga sistem menghasilkan output. Proses dimulai ketika pengguna memasukkan ID video YouTube ke dalam aplikasi. Selanjutnya, sistem mengambil komentar menggunakan YouTube Data API, melakukan praproses teks, mengekstraksi fitur menggunakan TF-IDF, dan mengklasifikasikan sentimen menggunakan algoritma SVM. Setelah proses klasifikasi selesai, sistem menampilkan hasil analisis dalam bentuk persentase dan visualisasi. Diagram ini menunjukkan aliran proses yang sistematis dan berurutan, serta

menggambarkan bagaimana sistem bekerja secara otomatis dari awal hingga akhir.



Gambar 2. Activity Diagram

3.10. Sequence Diagram



Gambar 3. Sequence Diagram

Sequence Diagram pada Gambar 3 menggambarkan alur interaksi antara User, Web App, YouTube API, dan SVM Engine dalam proses analisis sentimen komentar YouTube. Proses dimulai ketika pengguna mengakses menu analisis pada aplikasi web, kemudian sistem menampilkan form untuk memasukkan ID video YouTube. Setelah pengguna menginput ID video, Web App mengirimkan permintaan pengambilan komentar ke YouTube API dan menerima data komentar sebagai respons. Data komentar tersebut selanjutnya dikirim ke SVM Engine untuk dilakukan proses analisis sentimen, yang mencakup pra-proses teks, ekstraksi fitur, dan klasifikasi sentimen menggunakan algoritma Support Vector Machine. Setelah proses analisis selesai, hasil sentimen dikirim kembali ke Web App dan ditampilkan kepada pengguna dalam bentuk informasi hasil analisis. Diagram ini menunjukkan bahwa seluruh proses berjalan secara terstruktur dan berurutan, sehingga mendukung

sistem analisis sentimen yang otomatis, efisien, dan mudah digunakan.

4. HASIL DAN PEMBAHASAN

4.1. Implementasi Sistem/Model

Implementasi sistem dalam penelitian ini bertujuan untuk merealisasikan metode Support Vector Machine (SVM) dalam melakukan analisis sentimen terhadap komentar pengguna pada platform YouTube. Sistem yang dikembangkan berbasis bahasa pemrograman Python dengan memanfaatkan beberapa library pendukung seperti pandas, scikit-learn, dan NLTK untuk mempermudah proses pengolahan data hingga klasifikasi sentimen.

Secara umum, implementasi sistem dilakukan melalui beberapa tahapan utama, yaitu pengumpulan data, preprocessing data, ekstraksi fitur, pelatihan model, serta proses klasifikasi. Setiap tahapan dirancang secara terstruktur agar sistem mampu menghasilkan analisis sentimen yang akurat dan efisien.

Pada tahap awal, dilakukan proses pengambilan data komentar dari YouTube menggunakan metode scraping atau pemanfaatan YouTube Data API. Data yang diperoleh berupa teks komentar yang kemudian disimpan dalam format dataset untuk diproses lebih lanjut. Dataset ini menjadi dasar dalam pelatihan model klasifikasi sentimen.

Selanjutnya, data yang telah dikumpulkan akan melalui proses data preprocessing. Tahap ini bertujuan untuk membersihkan dan menyiapkan data agar dapat diolah oleh sistem. Proses preprocessing meliputi pembersihan teks dari karakter yang tidak diperlukan, pengubahan huruf menjadi huruf kecil (case folding), pemisahan kata (tokenizing), penghapusan kata umum (stopword removal), serta pengubahan kata menjadi bentuk dasar (stemming). Tahapan ini sangat penting karena kualitas data akan sangat mempengaruhi hasil klasifikasi.

Setelah data diproses, dilakukan tahap feature extraction menggunakan metode Term Frequency-Inverse Document Frequency (TF-IDF). Metode ini digunakan untuk mengubah data teks menjadi representasi numerik yang dapat diproses oleh algoritma SVM. TF-IDF memberikan bobot lebih tinggi pada kata-kata yang dianggap penting dalam suatu dokumen, sehingga membantu model dalam mengenali pola sentimen.

Tahap berikutnya adalah model training, di mana algoritma SVM dilatih menggunakan data yang telah diekstraksi menjadi fitur numerik. Pada tahap ini, model akan mempelajari pola hubungan antara fitur dan label sentimen (positif, negatif, dan netral). Proses pelatihan dilakukan dengan membagi dataset menjadi data latih (training set) dan data uji (testing set).

Setelah model selesai dilatih, dilakukan proses klasifikasi untuk menentukan sentimen dari data komentar. Model yang telah terbentuk akan digunakan untuk memprediksi kelas sentimen dari

setiap komentar berdasarkan pola yang telah dipelajari sebelumnya.

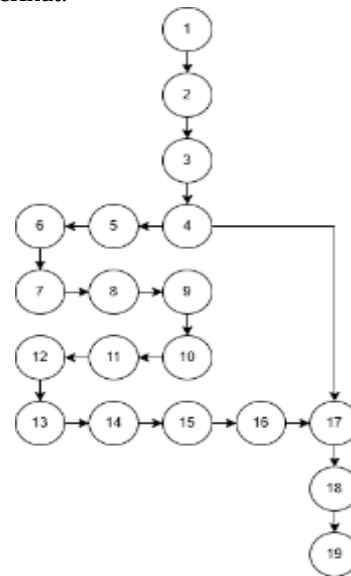
Untuk memastikan kinerja model, dilakukan tahap evaluasi menggunakan metrik seperti akurasi, precision, recall, dan f1-score. Hasil evaluasi ini digunakan untuk mengetahui seberapa baik model dalam mengklasifikasikan data serta untuk mengidentifikasi kekurangan yang perlu diperbaiki.

Dengan begitu implementasi sistem ini menghasilkan sebuah model analisis sentimen yang mampu mengolah data komentar YouTube secara otomatis dan mengelompokkan opini pengguna ke dalam kategori sentimen yang telah ditentukan. Sistem ini diharapkan dapat membantu dalam memahami persepsi publik terhadap suatu konten secara lebih cepat dan objektif.

4.2. Testing

Pengujian White Box Testing dilakukan untuk menguji struktur logika internal program pada fitur analisis sentimen. Fokus pengujian ini adalah memastikan bahwa setiap percabangan dan proses utama di dalam program telah dieksekusi dengan benar. Potongan program yang diuji merupakan bagian inti dari sistem, yaitu saat pengguna menekan tombol analisis untuk memulai proses pengambilan komentar dan klasifikasi sentimen.

Secara logika, alur flowgraph dapat gambarkan sebagai berikut:



Gambar 5. Flow Graph

Tabel 1. Deskripsi flowgraph

No	Node	Deskripsi
1	Node 1	Program mulai dijalankan pada fitur analisis sentimen
2	Node 2	Sistem memeriksa apakah tombol Analisis ditekan
3	Node 3	Sistem memvalidasi apakah Video ID telah diisi
4	Node 4	Jika input kosong, sistem menampilkan pesan peringatan
5	Node 5	Jika input lengkap, sistem mengambil data komentar dari YouTube
6	Node 6	Sistem membentuk data komentar ke dalam DataFrame
7	Node 7	Sistem melakukan preprocessing pada komentar
8	Node 8	Sistem mengubah hasil preprocessing menjadi fitur TF-IDF
9	Node 9	Sistem melakukan prediksi sentimen menggunakan model SVM
10	Node 10	Sistem menampilkan notifikasi bahwa analisis selesai
11	Node 11	Sistem menghitung persentase distribusi sentimen
12	Node 12	Sistem menampilkan metrik persentase positif, netral, dan negatif
13	Node 13	Sistem menampilkan grafik distribusi sentimen
14	Node 14	Sistem menampilkan tabel komentar dan hasil sentimen
15	Node 15	Sistem membuat file CSV dan menampilkan tombol unduh
16	Node 16	Proses analisis selesai
17	Node 17	Program mulai dijalankan pada fitur analisis sentimen
18	Node 18	Sistem memeriksa apakah tombol Analisis ditekan
19	Node 19	Sistem memvalidasi apakah Video ID telah diisi

[illegible]

Gambar 4. Listing Program

Kompleksitas siklomatis (*Cyclomatic Complexity*) digunakan untuk menentukan jumlah jalur independen pada program.

$$V(G) = E - N + 2$$

Keterangan:

E = jumlah edge

N = jumlah node

Berdasarkan flowgraph:

Jumlah Node (N) = 16

Jumlah Edge (E) = 16

Maka:

$$V(G) = 16 - 16 + 2 = 2$$

Namun, jika titik mulai dan selesai dianggap sebagai bagian alur linear utama, maka secara praktis fungsi ini memiliki 2 jalur independen utama, yaitu:

a. Jalur input tidak lengkap

b. Jalur input lengkap dan analisis berjalan penuh

Sehingga, dalam pengujian White Box, minimal diperlukan 2 test case utama untuk menguji seluruh logika percabangan yang ada.

Tabel 2. Jalur Eksekusi White Box

No	Jalur Eksekusi	Kondisi	Hasil yang Diharapkan
1	1 → 2 → 3 → 4 → 18 → 19	Input tidak valid (video_id kosong)	Sistem menampilkan pesan peringatan dan proses analisis tidak dijalankan
2	1 → 2 → 3 → 4 → 5 → 6 → 7 → 8 → 9 → 10 → 11 → 12 → 13 → 14 → 15 → 16 → 17 → 19	Input valid (video_id terisi)	Sistem mengambil komentar, melakukan preprocessing, klasifikasi SVM, menampilkan hasil (ringkasan, grafik, tabel), dan menyediakan file CSV

Tabel 3. Hasil White Box Testing

No	Jalur Uji	Skenario Pengujian	Hasil yang Diharapkan	Hasil Pengujian	Status
1	Jalur 1	Pengguna menekan tombol Analisis tanpa mengisi Video ID	Sistem menampilkan pesan peringatan dan tidak menjalankan proses analisis	Sistem menampilkan warning "Masukkan Video ID terlebih dahulu!"	Valid
2	Jalur 2	Pengguna mengisi Video ID dan menekan tombol Analisis	Sistem mengambil komentar dari YouTube	Sistem berhasil mengambil komentar sesuai jumlah yang ditentukan	Valid
3	Jalur 2	Data komentar berhasil diambil	Sistem membentuk DataFrame dari data komentar	Data berhasil dikonversi menjadi DataFrame	Valid
4	Jalur 2	Sistem melakukan preprocessing	Data komentar dibersihkan dan dinormalisasi	Data berhasil diproses tanpa error	Valid
5	Jalur 2	Sistem melakukan transformasi TF-IDF	Data teks diubah menjadi fitur numerik	Transformasi TF-IDF berhasil dilakukan	Valid
6	Jalur 2	Sistem melakukan klasifikasi SVM	Sistem menghasilkan label sentimen (positif, netral, negatif)	Sistem berhasil memprediksi sentimen setiap komentar	Valid
7	Jalur 2	Sistem menghitung distribusi sentimen	Sistem menampilkan persentase sentimen	Persentase berhasil dihitung dengan benar	Valid
8	Jalur 2	Sistem menampilkan ringkasan sentimen	Sistem menampilkan jumlah positif, netral, negatif	Ringkasan tampil dengan benar	Valid
9	Jalur 2	Sistem menampilkan Bar Chart	Grafik batang ditampilkan	Grafik tampil sesuai data	Valid
10	Jalur 2	Sistem menampilkan Pie Chart	Diagram lingkaran ditampilkan	Pie chart tampil dengan benar	Valid
11	Jalur 2	Sistem menampilkan Donut Chart	Donut chart ditampilkan	Donut chart tampil dengan benar	Valid
12	Jalur 2	Sistem menampilkan tabel komentar	Tabel komentar dan sentimen ditampilkan	Data tampil sesuai	Valid
13	Jalur 2	Sistem menampilkan contoh komentar	20 komentar pertama ditampilkan	Data tampil sesuai	Valid
14	Jalur 2	Sistem menyediakan fitur download CSV	File CSV dapat diunduh	File berhasil diunduh tanpa error	Valid

Berdasarkan hasil pengujian pada Tabel 3, seluruh jalur eksekusi program telah diuji dan menunjukkan hasil yang valid. Jalur pertama menunjukkan bahwa sistem mampu menangani kondisi input yang tidak lengkap dengan menampilkan pesan peringatan kepada pengguna. Hal ini penting untuk mencegah kesalahan proses sejak awal.

Pada jalur kedua, sistem berhasil menjalankan seluruh proses utama, mulai dari pengambilan data komentar, preprocessing, transformasi TF-IDF, klasifikasi menggunakan metode Support Vector Machine (SVM), hingga penyajian hasil dalam berbagai bentuk visualisasi dan file unduhan. Tidak ditemukan kesalahan pada setiap tahapan proses, yang menunjukkan bahwa logika program telah berjalan dengan baik.

Dengan demikian, dapat disimpulkan bahwa berdasarkan pengujian White Box Testing, struktur internal program telah sesuai dengan perancangan dan mampu menjalankan seluruh fungsi sistem dengan baik.

5. KESIMPULAN DAN SARAN

Penelitian ini berhasil mengimplementasikan metode Support Vector Machine (SVM) untuk analisis sentimen komentar pada platform YouTube berbasis Python. Sistem yang dikembangkan mampu melakukan proses secara terintegrasi mulai dari pengambilan data komentar, text preprocessing, ekstraksi fitur menggunakan TF-IDF, hingga klasifikasi sentimen ke dalam kategori positif, netral, dan negatif. Hasil pengujian menunjukkan bahwa sistem dapat berjalan dengan baik sesuai dengan fungsi yang dirancang, serta mampu menyajikan hasil analisis dalam bentuk visualisasi grafik dan tabel yang informatif. Dengan demikian, sistem ini dapat membantu dalam memahami opini publik secara lebih cepat dan efisien.

Meskipun demikian, masih terdapat beberapa keterbatasan yang dapat menjadi bahan pengembangan selanjutnya. Penelitian ini disarankan untuk dikembangkan dengan membandingkan metode klasifikasi lain guna meningkatkan akurasi hasil analisis. Selain itu, perlu dilakukan peningkatan pada tahap preprocessing, khususnya dalam menangani bahasa informal, sarkasme, dan data multibahasa. Pengembangan lebih lanjut juga dapat dilakukan dengan menambahkan fitur visualisasi yang lebih interaktif serta optimasi sistem agar mampu menangani dataset dalam jumlah besar secara lebih efisien.

DAFTAR PUSTAKA

- [1] B. Liu, *Sentiment Analysis and Opinion Mining*. Morgan & Claypool Publishers, 2012.
- [2] M. Medhat, A. Hassan, and H. Korashy, "Sentiment Analysis Algorithms and Applications: A Survey," *Ain Shams Engineering Journal*, vol. 5, no. 4, pp. 1093–1113, 2014.
- [3] C. Cortes and V. Vapnik, "Support-Vector Networks," *Machine Learning*, vol. 20, pp. 273–297, 1995.
- [4] G. Salton and C. Buckley, "Term-weighting approaches in automatic text retrieval," *Information Processing & Management*, 1988.
- [5] A. Giachanou and F. Crestani, "Like It or Not: A Survey of Twitter Sentiment Analysis Methods," *ACM Computing Surveys*, 2016.
- [6] M. Medhat, A. Hassan, and H. Korashy, "Sentiment Analysis Algorithms and Applications: A Survey," *Ain Shams Engineering Journal*, vol. 5, no. 4, pp. 1093–1113, 2014.
- [7] B. Liu, *Sentiment Analysis and Opinion Mining*. Morgan & Claypool Publishers, 2012.
- [8] C. C. Aggarwal and C. Zhai, *Mining Text Data*. Springer, 2012.
- [9] D. Jurafsky and J. H. Martin, *Speech and Language Processing*. 2023.
- [10] G. Salton and C. Buckley, "Term-weighting approaches in automatic text retrieval," *Information Processing & Management*, 1988.
- [11] C. Cortes and V. Vapnik, "Support-Vector Networks," *Machine Learning*, vol. 20, pp. 273–297, 1995.
- [12] N. Cristianini and J. Shawe-Taylor, *An Introduction to Support Vector Machines*. Cambridge University Press, 2000.
- [13] A. Giachanou and F. Crestani, "Like It or Not: A Survey of Twitter Sentiment Analysis Methods," *ACM Computing Surveys*, 2016.
- [14] Google Developers, "YouTube Data API Documentation," 2023.
- [15] S. Vijayarani, M. J. Ilamathi, and M. Nithya, "Preprocessing Techniques for Text Mining," *International Journal of Computer Science & Communication Networks*, 2015.
- [16] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, 3rd ed. Morgan Kaufmann, 2012.
- [17] S. Auliana, B. R. S. Permana, U. Kalsum, and M. Mudofar, "Detecting Hoax News with Web-Based SVM Classifier: A Machine Learning Approach to Combat Misinformation Using Support Vector Machine (SVM) Algorithm," *JINAV Journal of Information and Visualization*, vol. 6, no. 1, pp. 26–35, 2025.
- [18] B. R. Setya Permana, "Penerapan Algoritma C4.5 untuk Klasifikasi Kepatuhan Konsumen," *JSK*, vol. 4, no. 2, pp. 11–21, 2023.
- [19] H. Hamdan, O. Romli, S. Auliana, and B. R. S. Permana, "Utilization of Virtual Assistance (Chatbot) for an Integrated Information Portal as Part of a Marine Tourism Promotion Strategy in Banten," *JINAV Journal of Information and Visualization*, vol. 5, no. 2, pp. 206–219, 2024.
- [20] B. R. S. Permana, S. Auliana, and R. K. Octavian, "DigiField: Real-Time Smart Farming Monitoring System Using React.js and Express.js for Agricultural Digitalization," *JINAV Journal of Information and Visualization*, vol. 6, no. 1, pp. 113–122, 2025.
- [21] B. R. S. Permana, G. D. P. Aryono, and S. N. Aliffito, "Information System Audit of FIF Mobile Collection at PT Federal International Finance POS Baros Using the COBIT 5 Framework," *JINAV Journal of Information and Visualization*, vol. 6, no. 1, pp. 66–74, 2025.
- [22] H. R. A. Fauzan, S. Auliana, and B. R. S. Permana, "Perancangan aplikasi manajemen hotspot menggunakan framework CodeIgniter di PT Jaya Sampurna Network," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 9, no. 6, pp. 10239–10244, 2025.

- [24] H. Ramadhan, S. Auliana, and B. R. S. Permana, "Penerapan framework CodeIgniter 3 untuk pengembangan aplikasi penjualan pada Toko Nazeera Food," *Jurnal Sistem Informasi dan Teknologi (SINTEK)*, vol. 5, no. 2, pp. 208–216, 2025.
- [25] M. K. Naufallinho, S. Auliana, and B. R. S. Permana, "Rancangan aplikasi augmented reality pengenalan buah-buahan tropis pada SMPN 1 Kibin menggunakan Unity 3D dan Blender," *Jurnal Sosial Teknologi*, vol. 5, no. 8, pp. 3040–3059, 2025.
- [26] R. N. Azizah, S. Auliana, and B. R. Setya, "Perancangan Denah 3D Berbasis Augmented Reality Menggunakan Blender Terintegrasi Unity 3D pada Rumah Sakit Mata Achmad Wardi Kota Serang," *IRJE*, vol. 5, no. 4, pp. 1312–1320, 2025.
- [27] T. I. R. Rahman, S. Auliana, and B. R. S. Permana, "Pembuatan game edukasi jenis-jenis hewan air menggunakan aplikasi Construct 2," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 9, no. 6, pp. 10258–10264, 2025.