BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 3, April 2026 | Hal 886-898
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i3.995

Analisis Komentar Youtube Terhadap Kebijakan Bebas Impor Oleh Pemerintah Pusat Menggunakan Support Vector Machine

Ignasius Aditya Anggoro Putra*, Salmon, Kusnandar

Program Studi Sistem Informasi, STMIK Widya Cipta Dharma, Samarinda, Indonesia
Email: [1,*]2041901@wicida.ac.id, [2]salmon@wicida.ac.id, [3]kusnandar@wicida.ac.id
Email Penulis Korespondensi: 2041901@wicida.ac.id

Abstrak–YouTube sebagai media sosial berkembang menjadi sarana penting dalam mengekspresikan opini publik terhadap kebijakan pemerintah, termasuk kebijakan bebas impor. Penelitian ini bertujuan untuk menganalisis sentimen komentar pengguna YouTube terhadap kebijakan bebas impor menggunakan algoritma *Support Vector Machine* (SVM). Data diperoleh melalui *web scraping* menggunakan YouTube Data API v3 pada video kanal Kompas.com, dengan total 3.267 komentar mentah. Tahapan penelitian meliputi *preprocessing* teks, ekstraksi fitur menggunakan *Term Frequency–Inverse Document Frequency* (TF-IDF), pelabelan sentimen berbasis leksikon, serta klasifikasi menggunakan SVM. Untuk mengatasi ketidakseimbangan data, diterapkan metode *Synthetic Minority Oversampling Technique* (SMOTE). Evaluasi model dilakukan menggunakan *confusion matrix* dengan metrik akurasi, *precision, recall*, dan *F1-score*. Hasil penelitian menunjukkan bahwa model SVM tanpa *tuning* memperoleh akurasi 77,00%, sedangkan setelah *hyperparameter tuning* diperoleh akurasi 75,15% dengan performa yang lebih seimbang antar kelas. Penelitian ini membuktikan bahwa SVM efektif dalam mengklasifikasikan sentimen komentar YouTube secara objektif.

Kata kunci: Analisis Sentimen; Kebijakan Bebas Impor; *Support Vector Machine*; TF-IDF; YouTube

Abstract–YouTube has become an important platform for expressing public opinion on government policies, including the free import policy. This study aims to analyze the sentiment of YouTube user comments regarding the free import policy using the Support Vector Machine (SVM) algorithm. The data were collected through web scraping using the YouTube Data API v3 from a Kompas.com video, resulting in 3,267 raw comments. The research stages include text preprocessing, feature extraction using Term Frequency–Inverse Document Frequency (TF-IDF), lexicon-based sentiment labeling, and sentiment classification using SVM. To address data imbalance, the Synthetic Minority Oversampling Technique (SMOTE) was applied. Model performance was evaluated using a confusion matrix with accuracy, precision, recall, and F1-score metrics. The results show that the SVM model achieved an accuracy of 77.00% without tuning and 75.15% after hyperparameter optimization, with improved balance across sentiment classes. These findings indicate that SVM is effective for sentiment classification of YouTube comments.

Keywords: Sentiment Analysis; YouTube; Free Import Policy; Support Vector Machine; TF-IDF

1. PENDAHULUAN

Media sosial berkembang pesat dari abad akhir 20 hingga saat ini, membuat pengguna sangat terbantu dalam melakukan aktivitas setiap hari. *Youtube* merupakan platform media sosial yang menyediakan berbagai jenis konten video dengan durasi beragam, yang diproduksi oleh para kreator atau pembuat konten (YouTuber) melalui kreativitas masing-masing. Tingginya tingkat penggunaan *YouTube* di masyarakat menyebabkan salah satu platform ini semakin populer dan diminati oleh berbagai lapisan pengguna. Saat ini, *YouTube* menempati posisi paling populer sebagai salah satu platform media sosial di Indonesia. Sebagai layanan aliran video terbesar di dunia, *YouTube* menawarkan beragam konten, mulai dari hiburan hingga edukasi. Selain itu, *YouTube* turut dimanfaatkan sebagai media untuk menyampaikan informasi dan menyebarluaskan berita kepada masyarakat luas [1].

Komentar sering kali ditulis dalam bahasa informal, mengandung emotikon, singkatan, atau bahkan bahasa campuran yang dapat membingungkan [2]. Ragam bentuk sindiran dalam komunikasi memiliki variasi yang luas, seperti ironi, sinisme, satir, maupun sarkasme. Satir umumnya digunakan untuk menyampaikan kritik secara tajam, sedangkan sarkasme sering muncul sebagai ungkapan bernada kasar yang bertujuan merendahkan atau menghina. Komentar yang diberikan oleh masyarakat dapat menjadi sumber data yang kaya untuk memahami opini publik secara objektif dan satu waktu yang sama (*real-time*) baik dukungan maupun kritik terhadap kebijakan yang diterjadi. Respon, opini, pendapat dan reaksi terhadap video dapat dituangkan melalui fitur yang disediakan *Youtube* yaitu komentar yang merupakan tempat mengekspresikan suara bagi pengguna dalam menikmati sebuah video yang ditonton dan dalam komentar tesebut bersifat positif, negatif dan netral [3]. Analisis komentar teks sangat penting karena dapat memberikan gambaran objektif tentang persepsi publik terhadap masalah sensitif seperti ini. Namun, analisis manual menjadi kurang efektif dan rentan terhadap subjektivitas karena volume komentar yang besar. Untuk itu diperlukan metode machine learning seperti SVM dan untuk mengotomatisasi klasemen sentimen[4].

Dalam penelitian ini, terdapat sejumlah penelitian terdahulu yang relevan dan digunakan sebagai landasan serta acuan dalam penyusunan penelitian. Beberapa penelitian terdahulu yang mendukung penelitian ini antara lain sebagai berikut: Pertama, penelitian oleh Zulqarnain, Moehammad Iqbal Sultan, Muh. Akbar pada tahun 2025 dengan judul "*Analisis Sentimen Pemecatan Jokowi Pada Komentar Publik YouTube Tempo.co*". bertujuan untuk menganalisis sentimen publik terhadap isu pemecatan Presiden Joko Widodo (Jokowi) dan rencana pengembalian Pilkada melalui DPRD sebagaimana diangkat dalam komentar pada video YouTube Tempo.co. Hasil penelitian menunjukkan bahwa diskusi terkait isu ini memunculkan respons publik yang intens, dengan puncak aktivitas terjadi pada tanggal 21 Desember 2024. Kata-kata seperti "rakyat," "pilkada," dan "partai" serta emoji seperti dan mendominasi diskusi, mencerminkan

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 3, April 2026 | Hal 886-898
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i3.995

keterlibatan emosional audiens. Tingkat toksisitas juga meningkat signifikan pada hari yang sama, menegaskan adanya polarisasi dalam opini publik terhadap isu politik tersebut. Temuan ini memiliki signifikansi dalam konteks literasi digital, khususnya dalam memahami bagaimana media sosial berfungsi sebagai ruang publik untuk diskusi politik [5].

Kedua, penelitian oleh Suprayuandi Pratama, Majduddin, Medi Triawan pada tahun 2025 dengan judul "*Klasifikasi Sentimen Komentar pada Video 'Rendang Hilang di Palembang' oleh Willy Salim Menggunakan Algoritma Support*". Bertujuan untuk mengklasifikasikan sentimen komentar pengguna pada video "Rendang Hilang di Palembang" menggunakan algoritma SVM. Dengan mengkategorikan komentar berdasarkan sentimen, diharapkan hasil penelitian ini dapat memberikan gambaran umum tentang bagaimana publik merespons konten tersebut. Selain itu, hasil ini juga dapat memberikan masukan berharga bagi kreator konten dan pelaku media dalam memahami dinamika opini publik di platform digital [6].

Ketiga, penelitian oleh Adikara Alif Nurrahman, Muhammad Mauladi, Abdul Rahman pada tahun 2025 dengan judul "*Analisis Sentimen Masyarakat terhadap Kenaikan Harga Bahan Bakar Minyak Menggunakan Support Vector Machine dan SMOTE*". Untuk mengatasi ketidakseimbangan kelas dalam dataset, digunakan metode SVM sebagai model untuk klasifikasi sentimen terkait dengan kenaikan harga BBM. Untuk menangani ketidakseimbangan tersebut, SMOTE digunakan sebagai metode untuk menghasilkan data sintetis yang seimbang antara sentimen positif dan negatif, sehingga model SVM dapat lebih efektif dalam mengenali pola dari kedua kelas tersebut dan menganalisis sentimen masyarakat terkait dengan kenaikan harga BBM menggunakan metode SVM dan SMOTE, dengan fokus pada pengurangan ketidakseimbangan kelas dalam dataset[7].

Analisis sentimen, yang juga dikenal sebagai *opinion mining*, merupakan suatu bidang kajian yang berfokus pada pengidentifikasian, pengolahan, dan analisis opini, sikap, serta ekspresi emosi individu terhadap suatu peristiwa, isu, atau entitas tertentu[8]. Pemilihan dan identifikasi kata-kata emosional tersebut dapat dilakukan secara efektif dan efisien melalui penggunaan kamus atau leksikon sentimen sebagai acuan [9].

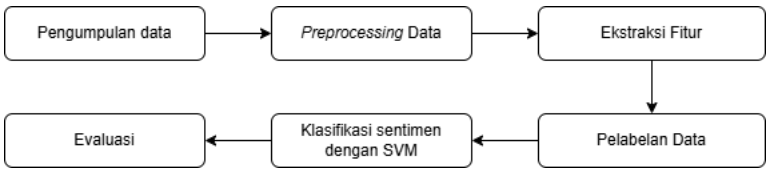
Support Vector Machine merupakan algoritma *machine learning* yang banyak digunakan untuk mengklasifikasi suatu topik karena kemampuan dalam mengelola data dengan dimensi banyak dan menghasilkan batas keputusan yang optimal. Dalam konteks pembobotan teks, teknik *Term Frequency-Inverse Document Frequency* (TF-IDF) dipilih untuk merepresentasikan data teks dengan memberikan bobot lebih tinggi pada kata-kata yang unik dan relevan dalam sebuah dokumen. Metode ini terbukti efektif dalam penelitian yang mengkaji ulasan produk di *platform* digital lainnya, dengan tingkat akurasi yang konsisten tinggi. Dengan menerapkan *Support Vector Machine* dalam analisis sentimen komentar pada *YouTube*, diharapkan dapat diperoleh model yang mampu secara akurat mengklasifikasikan opini masyarakat terhadap kebijakan pemerintah pusat.

2. METODOLOGI PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan metode analisis sentimen berbasis machine learning. Algoritma yang digunakan adalah Support Vector Machine untuk mengklasifikasikan sentimen komentar pengguna YouTube terhadap kebijakan bebas impor. Tahapan penelitian dilakukan secara terstruktur mulai dari pengumpulan data hingga evaluasi model.

2.1 Tahapan Penelitian

Berikut Gambar 1 merupakan tahapan penelitian.



Gambar 1. Tahapan Penelitian

Alur penelitian terdiri dari beberapa tahapan utama, yaitu:
a. Pengumpulan data
b. Preprocessing data
c. Ekstraksi fitur
d. Pelabelan data
e. Klasifikasi menggunakan SVM
f. Evaluasi model

2.2 Pengumpulan Data

Data yang digunakan dalam penelitian ini berupa komentar pengguna pada video YouTube yang membahas kebijakan bebas impor. Pengambilan data dilakukan menggunakan YouTube Data API v3 dengan memanfaatkan platform *Google Colaboratory* (*Google Colab*) untuk menjalankan proses *crawling* data secara otomatis[10]. Dataset yang diperoleh berjumlah 3.267 komentar dalam bentuk *raw data* yang belum memiliki label sentimen. Setiap data komentar dilengkapi

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 3, April 2026 | Hal 886-898
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i3.995

dengan atribut seperti ID komentar, nama pengguna, waktu publikasi, jumlah *like*, dan isi komentar. Data tersebut disimpan dalam format *.csv* untuk memudahkan proses pengolahan lebih lanjut. Proses pengambilan data (*scraping*) dilaksanakan dengan jarak waktu yang beragam antara bulan Oktober 2025 hingga bulan Desember 2025.

2.3 *Preprocessing* Data

Tahap *preprocessing* bertujuan untuk membersihkan dan menyiapkan data teks agar dapat diolah oleh algoritma *machine learning*. Tahapan *preprocessing* yang dilakukan meliputi:

a. Data Cleaning
   Menghapus karakter yang tidak relevan seperti tanda baca, angka, URL, *username*, hashtag, dan simbol lainnya.
b. Case Folding
   Mengubah seluruh huruf dalam teks menjadi huruf kecil (*lowercase*) untuk menyeragamkan penulisan.
c. Tokenization
   Memecah teks menjadi unit kata (token) berdasarkan spasi.
d. Stopword Removal
   Menghapus kata-kata umum yang tidak memiliki makna signifikan, seperti “dan”, “yang”, “di”.
e. Stemming
   Proses ini perlu dilakukan untuk menyeragamkan kata-kata yang ada pada data set agar menjadi kata dasar[11]. Mengubah kata menjadi bentuk dasar menggunakan library Sastrawi, misalnya “membeli” menjadi “beli”.

Hasil dari tahap ini adalah data teks yang lebih bersih, terstruktur, dan siap untuk tahap ekstraksi fitur.

2.4 Ekstraksi Fitur

Fitur ekstraksi yang digunakan dalam penelitian ini adalah menggunakan metode *Term Frequency–Inverse Document Frequency* (TF-IDF)[12]. Proses ini menggunakan parameter *max_features* sebesar 5000 serta *ngram_range* (1,2) yang mencakup unigram dan bigram. Pendekatan ini memungkinkan representasi teks menjadi lebih kaya karena mempertimbangkan kata tunggal dan kombinasi dua kata. Metode TF-IDF kemudian memberikan bobot pada setiap kata berdasarkan tingkat kepentingannya dalam dokumen. Dengan demikian, kata yang lebih relevan akan memiliki kontribusi lebih besar dalam meningkatkan akurasi dan efektivitas proses klasifikasi teks.

2.5 Pelabelan Data

Pelabelan data dilakukan dengan pendekatan *lexicon-based* menggunakan kamus sentimen Bahasa Indonesia atau *Indonesian Sentiment Lexicon*. Setiap komentar diberi skor berdasarkan penambahan nilai untuk kata positif dan pengurangan untuk kata negatif. Hasil skor kemudian menentukan klasifikasi, yaitu positif jika lebih dari nol, negatif jika kurang dari nol, dan netral jika sama dengan nol. Tahap ini menghasilkan dataset berlabel yang digunakan dalam proses pelatihan model klasifikasi sentimen.

2.6 Klasifikasi dengan Support Vector Machine

Model klasifikasi dibangun menggunakan algoritma Support Vector Machine dengan pendekatan LinearSVC. Proses dimulai dengan membagi data menjadi data latih dan data uji. Selanjutnya, dilakukan penyeimbangan data menggunakan metode SMOTE untuk mengatasi ketidakseimbangan kelas. Model kemudian dilatih menggunakan data latih dan diuji melalui proses prediksi pada data uji. Penggunaan SMOTE bertujuan mengurangi bias model terhadap kelas mayoritas sehingga meningkatkan kinerja klasifikasi secara keseluruhan.

2.7 Evaluasi Model

Dalam evaluasi model *Support Vector Machine* yang digunakan untuk analisis sentimen kebijakan bebas impor di *YouTube*, beberapa metrik penting diterapkan untuk memastikan kinerja model yang optimal[13]. Evaluasi model dilakukan menggunakan beberapa metrik, yaitu accuracy untuk mengukur tingkat ketepatan prediksi, precision untuk menilai ketepatan pada kelas tertentu, recall untuk mengukur kemampuan menemukan data relevan, serta F1-score sebagai rata-rata harmonis antara precision dan recall. Selain itu, dilakukan perbandingan performa model sebelum dan sesudah proses hyperparameter tuning menggunakan metode GridSearchCV guna memperoleh kombinasi parameter terbaik dan meningkatkan kinerja model klasifikasi.

3. HASIL DAN PEMBAHASAN

3.1 Pengumpulan Data

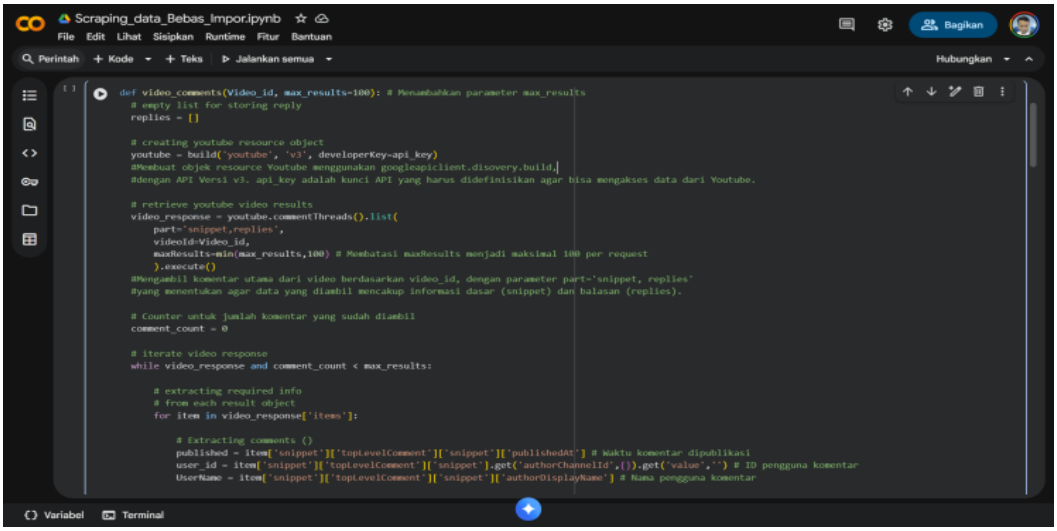
Proses pengambilan data dilakukan dengan menggunakan YouTube Data API v3 untuk mengakses komentar–komentar pada video. Hasil dari pengambilan data (*scraping*) ini dalam bentuk csv yang akan melewati tahap *preprocessing*. Data yang diambil merupakan data komentar pada video yang diunggah oleh kanal Kompas.com dengan judul “Prabowo: Siapa Saja Boleh Impor, Bebas, Enggak Usah Ada Kuota”. Video ini dipilih karena memperoleh perhatian yang cukup untuk membuka ruang diskusi mengenai kebijakan bebas impor yang di tetapkan oleh presiden ke-8 Republik Indonesia. Komentar-komentar pada video tersebut mencerminkan berbagai opini masyarakat, baik positif, negatif, maupun netral, sehingga sangat relevan untuk dianalisis menggunakan metode *sentiment analysis*. Selain itu, tingginya tingkat partisipasi

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 3, April 2026 | Hal 886-898
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i3.995

pengguna pada kolom komentar menjadikan data yang diperoleh cukup representatif untuk menggambarkan persepsi masyarakat terhadap isu kebijakan bebas impor. Dengan karakteristik data yang kaya dan beragam tersebut, penelitian ini memiliki peluang untuk menghasilkan pemahaman yang komprehensif mengenai kecenderungan sentimen masyarakat serta pola opini yang berkembang terkait kebijakan tersebut.

Gambar 1. Tangkapan layar pada kanal Youtube "Kompas.com"

Berdasarkan gambar 3, terlihat pada kanal Kompas.com dataset yang digunakan dalam penelitian ini merupakan hasil scraping komentar sebanyak 3.267 komentar dalam bentuk *raw* data. Setiap komentar disertai metadata seperti ID komentar, nama akun pengguna, tanggal unggahan, jumlah like, dan isi komentar. Data yang dikumpulkan masih bersifat tidak terstruktur (unstructured), komentar-komentar ini memiliki karakteristik bahasa yang sangat beragam, mencakup bahasa formal, informal, slang, serta variasi penulisan khas media sosial sehingga diperlukan serangkaian tahapan pemrosesan sebelum dapat dilakukan analisis lebih lanjut. Seluruh proses pengolahan data dilakukan menggunakan platform Google Colaboratory, yang memberikan fasilitas lingkungan komputasi berbasis cloud sehingga memudahkan dalam eksekusi program, pemanggilan API, serta pengolahan dataset dalam ukuran besar. YouTube Data API v3 digunakan sebagai media utama untuk pengambilan data, sedangkan Python beserta pustaka pendukung seperti Pandas, NumPy, dan Sastrawi digunakan untuk proses pengolahan dan pemodelan.

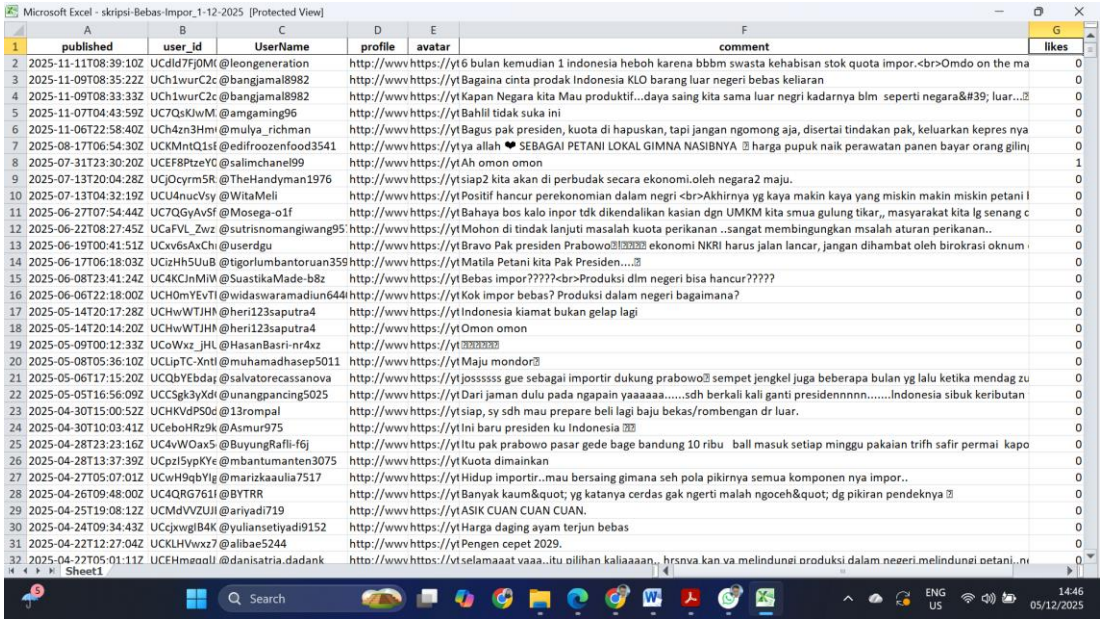


Gambar 2. Proses *scraping* data pada google colaboratory

Setelah proses *scraping* komentar selesai, pada Gambar 4 menunjukkan seluruh data yang berhasil diperoleh disimpan dalam format terstruktur, yaitu file berekstensi .csv dan .xlsx. Penyimpanan dalam format tersebut bertujuan untuk memastikan bahwa data dapat diproses, dibersihkan, dan dianalisis dengan lebih efisien pada tahap selanjutnya. Format .csv khususnya dipilih karena kompatibel dengan berbagai library analisis data pada Python, seperti *Pandas*, sehingga memudahkan proses manipulasi dan transformasi data.

Pada gambar 5 menunjukkan dataset yang terkumpul tidak hanya berisi teks komentar, tetapi juga mencakup sejumlah atribut pendukung yang sangat penting dalam penelitian analisis sentimen. Atribut tersebut meliputi identitas pengguna (user ID), waktu komentar dipublikasikan, jumlah interaksi (misalnya jumlah like), serta informasi tambahan lain yang dapat membantu memberikan konteks terhadap komentar. Teks komentar yang diperoleh merupakan data mentah yang mencerminkan reaksi spontan pengguna terhadap isu kebijakan bebas impor, sehingga menjadi sumber yang kaya untuk menggambarkan persepsi publik.

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 3, April 2026 | Hal 886-898
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i3.995



published	user_id	UserName	profile	avatar	comment	likes
2025-11-11T08:39:10Z	UCdld7Fj0M(	@leongeneration	http://www	https://yt	6 bulan kemudian 1 indonesia heboh karena bbbm swasta kehabisan stok quota impor. Omdo on the ma	0
2025-11-09T08:35:22Z	UCh1wurC2c	@bangjamal8982	http://www	https://yt	Bagaina cinta prodak Indonesia KLO barang luar negeri bebas keliaran	0
2025-11-09T08:33:33Z	UCh1wurC2c	@bangjamal8982	http://www	https://yt	Kapan Negara kita Mau produktif...daya saing kita sama luar negri kadarnya blm seperti negara' luar...	0
2025-11-07T04:43:59Z	UC7QsKJwM	@amgaming96	http://www	https://yt	Bahlil tidak suka ini	0
2025-11-06T22:58:40Z	UCh4zn3Hm	@mulya_richman	http://www	https://yt	Bagus pak presiden, kuota di hapuskan, tapi jangan ngomong aja, disertai tindakan pak, keluarkan kepres nya	0
2025-08-17T06:54:30Z	UCKMntQ1sE	@edifroozenfood3541	http://www	https://yt	ya allah ❤ SEBAGAI PETANI LOKAL GIMNA NASIBNYA harga pupuk naik perawatan panen bayar orang gilin	0
2025-07-31T23:30:20Z	UCEF8PtzeY(	@salimchanel99	http://www	https://yt	Ah omon omon	1
2025-07-13T20:04:28Z	UCjOcyrm5R	@TheHandyman1976	http://www	https://yt	siap2 kita akan di perbudak secara ekonomi.oleh negara2 maju.	0
2025-07-13T04:32:19Z	UCU4nucVsy	@WitaMeli	http://www	https://yt	Positif hancur perekonomian dalam negri Akhirnya yg kaya makin kaya yang miskin makin miskin petani l	0
2025-06-27T07:54:44Z	UC7QGyAvSf	@Mosega-o1f	http://www	https://yt	Bahaya bos kalo inpor tdk dikendalikan kasian dgn UMKM kita smua gulung tikar,, masyarakat kita lg senang c	0
2025-06-22T08:27:45Z	UCaFVL_Zwz	@sutrisnomangiwang95	http://www	https://yt	Mohon di tindak lanjuti masalah kuota perikanan ..sangat membingungkan msalah aturan perikanan..	0
2025-06-19T00:41:51Z	UCxv6sAxCh	@userdgu	http://www	https://yt	Bravo Pak presiden Prabowo ekonomi NKRI harus jalan lancar, jangan dihambat oleh birokrasi oknum	0
2025-06-17T06:18:03Z	UCizHh5UuB	@tigorlumbantoruan359	http://www	https://yt	Matila Petani kita Pak Presiden....	0
2025-06-08T23:41:24Z	UC4KCJnMiV	@SuastikaMade-b8z	http://www	https://yt	Bebas impor????? Produksi dlm negeri bisa hancur?????	0
2025-06-06T22:18:00Z	UCH0mYEvTI	@widaswaramadiun644	http://www	https://yt	Kok impor bebas? Produksi dalam negeri bagaimana?	0
2025-05-14T20:17:28Z	UCHwWTJHh	@heri123saputra4	http://www	https://yt	Indonesia kiamat bukan gelap lagi	0
2025-05-14T20:14:20Z	UCHwWTJHh	@heri123saputra4	http://www	https://yt	Omon omon	0
2025-05-09T00:12:33Z	UCoWxz_jHL	@HasanBasri-nr4xz	http://www	https://yt		0
2025-05-08T05:36:10Z	UCLipTC-Xn	tl @muhamadhasep5011	http://www	https://yt	Maju mondor	0
2025-05-06T17:15:20Z	UCQbYEbdaz	@salvatorecassanova	http://www	https://yt	jossssss gue sebagai importir dukung prabowo sempet jengkel juga beberapa bulan yg lalu ketika mendag zu	0
2025-05-05T16:56:09Z	UCCSgk3yXd	@unangpancing5025	http://www	https://yt	Dari jaman dulu pada ngapain yaaaaaa.....sdh berkali kali ganti presidennnnn.....Indonesia sibuk keributan	0
2025-04-30T15:00:52Z	UCHKVdPS0c	@13rompal	http://www	https://yt	siap, sy sdh mau prepare beli lagi baju bekas/rombengan dr luar.	0
2025-04-30T10:03:41Z	UCeboHRz9k	@Asmur975	http://www	https://yt	Ini baru presiden ku Indonesia	0
2025-04-28T23:23:16Z	UC4vWOax5-	@BuyungRafli-f6j	http://www	https://yt	Itu pak prabowo pasar gede bage bandung 10 ribu ball masuk setiap minggu pakaian trifh safir permai kapo	0
2025-04-28T13:37:39Z	UCpzI5ypKY	@mbantumanten3075	http://www	https://yt	Kuota dimainkan	0
2025-04-27T05:07:01Z	UCwH9qbYIg	@marizkaaulia7517	http://www	https://yt	Hidup importir..mau bersaing gimana seh pola pikirnya semua komponen nya impor..	0
2025-04-26T09:48:00Z	UC4QRG761I	@BYTRR	http://www	https://yt	Banyak kaum" yg katanya cerdas gak ngerti malah ngoceh" dg pikiran pendeknya	0
2025-04-25T19:08:12Z	UCMdVVZUJI	@ariyadi719	http://www	https://yt	ASIK CUAN CUAN CUAN.	0
2025-04-24T09:34:43Z	UCcjxwgIB4K	@yuliansetiyadi9152	http://www	https://yt	Harga daging ayam terjun bebas	0
2025-04-22T12:27:04Z	UCKLHVwxz7	@alibae5244	http://www	https://yt	Pengen cepet 2029.	0
2025-04-22T05:01:11Z	UCEHmggUI	@danisatria.dadank	http://www	https://yt	selamaaat yaaa..itu pilihan kaliaaaan.. hrsnya kan ya melindungi produksi dalam negeri melindungi petani..n	0

Gambar 3. Data *raw* hasil *scraping* pada komentar youtube

Hasil pengambilan data komentar melalui proses *scraping* disimpan dalam format berkas *.csv* dan selanjutnya diproses pada tahap *preprocessing*[4]. Dataset yang digunakan dalam penelitian ini tersedia dalam format *.csv* dan *.xlsx* dengan nama berkas *skripsi-bebas-impor_1-12-2025.xlsx*, yang memuat sebanyak 3.267 entri data. Setiap entri terdiri atas sejumlah atribut utama, meliputi waktu publikasi (*published*), *user id*, *username*, *profile URL*, *avatar URL*, isi komentar (*comment*), serta jumlah *likes*. Data tersebut dikumpulkan berdasarkan topik yang berkaitan dengan kebijakan bebas impor dan telah disusun secara sistematis dalam format terstruktur, sehingga siap untuk digunakan pada tahap *preprocessing* dan analisis sentimen selanjutnya.

3.2 *Text preprocessing*

Text preprocessing merupakan proses penyaringan dan penyederhanaan teks agar dapat diolah pada tahap selanjutnya. Data komentar yang diperoleh masih berupa *raw data* sehingga perlu melalui serangkaian tahapan *text preprocessing*.

a. Data *Cleaning*

Data *cleaning* merupakan langkah awal dalam rangkaian proses *text preprocessing* yang bertujuan untuk dokumen yang telah diperoleh selanjutnya dibersihkan dari berbagai karakter yang tidak relevan, seperti tag HTML, tanda pagar (*hashtag*), nama pengguna (*@username*), serta beragam tanda baca, antara lain titik, koma, tanda tanya, tanda seru, kurung, garis miring, dan simbol lainnya. Selain itu, seluruh karakter numerik serta simbol non-alfabet juga dihilangkan dari teks. Proses ini dilakukan untuk menghilangkan *noise* sehingga terbebas dari elemen-elemen yang tidak memiliki makna semantik dan teks menjadi lebih bersih dan siap dalam konteks pemodelan sentimen. Selain itu, langkah pembersihan juga mencakup penghapusan *ascii noise*, *mention*, spasi berlebih, serta baris kosong yang dapat mengganggu proses tokenisasi jumlah dataset awal adalah 3.267 komentar, dan setelah melalui tahap data *cleaning* jumlahnya tetap sama, yang menunjukkan bahwa komentar tidak memiliki banyak komponen yang harus dihapus secara keseluruhan, melainkan hanya dibersihkan dari karakter-karakter tertentu. Proses data *cleaning* ini memastikan bahwa teks berada dalam kondisi yang lebih terstruktur dan siap untuk diproses pada tahapan berikutnya seperti *case folding* dan *tokenizing*. Dengan menghapus elemen-elemen tersebut, teks yang dihasilkan menjadi lebih terstruktur dan fokus pada informasi linguistik yang bermakna. Tahapan ini juga membantu meningkatkan efektivitas proses tokenisasi dan ekstraksi fitur pada tahap selanjutnya. Pada akhirnya, data teks yang telah dibersihkan memiliki kualitas yang lebih baik untuk digunakan dalam pemodelan klasifikasi sentimen[11].

	comment	cleansing
0	6 bulan kemudian 1 indonesia heboh karena bbbm...	6 bulan kemudian 1 indonesia heboh karena bbbm...
1	Bagaina cinta prodak Indonesia KLO barang luar...	Bagaina cinta prodak Indonesia KLO barang luar...
2	Kapan Negara kita Mau produktif...daya saing k...	Kapan Negara kita Mau produktif daya saing kit...
3	Bahlil tidak suka ini	Bahlil tidak suka ini
4	Bagus pak presiden, kuota di hapuskan, tapi ja...	Bagus pak presiden kuota di hapuskan tapi jang...
...	...	...
3262	Lah kuma maneh we yg penting rakyat sejahtra d...	Lah kuma maneh we yg penting rakyat sejahtra d...
3263	@amni.z0r0 karena 1 gerakan dan 1 kata bro.<b...	z0r0 karena 1 gerakan dan 1 kata bro br br Jog...
3264	Syarat cinta tanah air adalah cinta produk dal...	Syarat cinta tanah air adalah cinta produk dal...
3265	Pretttt.. Kuota ngga ada ijin import dipersuli...	Pretttt Kuota ngga ada ijin import dipersulit ...
3266	Mentrinya spt jkw tukang boong	Mentrinya spt jkw tukang boong

3239 rows × 2 columns

Gambar 4. Hasil Cleaning

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 3, April 2026 | Hal 886-898
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i3.995

b. *Case folding*

Pada tahap ini, seluruh data teks dikonversi ke dalam bentuk huruf kecil (*lowercase*) dengan tujuan untuk menyeragamkan penulisan kata, sehingga memudahkan proses pengolahan dan pembacaan data oleh sistem komputer[14]. Langkah ini dilakukan untuk menyeragamkan bentuk penulisan sehingga mencegah adanya perbedaan pemaknaan terhadap kata yang sama akibat perbedaan penggunaan huruf kapital.

	comment	cleansing	case_folding
0	6 bulan kemudian 1 indonesia heboh karena bbbm...	6 bulan kemudian 1 indonesia heboh karena bbbm...	6 bulan kemudian 1 indonesia heboh karena bbbm...
1	Bagaina cinta prodak Indonesia KLO barang luar...	Bagaina cinta prodak Indonesia KLO barang luar...	bagaina cinta prodak indonesia klo barang luar...
2	Kapan Negara kita Mau produktif...daya saing k...	Kapan Negara kita Mau produktif daya saing kit...	kapan negara kita mau produktif daya saing kit...
3	Bahlil tidak suka ini	Bahlil tidak suka ini	bahlil tidak suka ini
4	Bagus pak presiden, kuota di hapuskan, tapi ja...	Bagus pak presiden kuota di hapuskan tapi jang...	bagus pak presiden kuota di hapuskan tapi jang...
...	...	...	...
3262	Lah kuma maneh we yg penting rakyat sejahtra d...	Lah kuma maneh we yg penting rakyat sejahtra d...	lah kuma maneh we yg penting rakyat sejahtra d...
3263	@amni.z0r0 karena 1 gerakan dan 1 kata bro.<b...	z0r0 karena 1 gerakan dan 1 kata bro br br Jog...	z0r0 karena 1 gerakan dan 1 kata bro br br jog...
3264	Syarat cinta tanah air adalah cinta produk dal...	Syarat cinta tanah air adalah cinta produk dal...	syarat cinta tanah air adalah cinta produk dal...
3265	Pretttt.. Kuota ngga ada ijin import dipersuli...	Pretttt Kuota ngga ada ijin import dipersulit ...	pretttt kuota ngga ada ijin import dipersulit ...
3266	Mentrinya spt jkw tukang boong	Mentrinya spt jkw tukang boong	mentrinya spt jkw tukang boong

3239 rows × 3 columns

Gambar 5. Hasil Case Folding

c. *Tokenization*

Proses tokenisasi dilakukan dengan cara memecah dokumen teks menjadi unit-unit kata (token) berdasarkan pemisah spasi, serta menghilangkan karakter tanda baca yang tidak diperlukan. Melalui tahapan ini, setiap kata dalam dokumen dapat diidentifikasi secara terpisah sehingga memudahkan proses analisis dan pengolahan data pada tahap selanjutnya[15].

	comment	cleansing	case_folding	tokenize
0	6 bulan kemudian 1 indonesia heboh karena bbbm...	6 bulan kemudian 1 indonesia heboh karena bbbm...	6 bulan kemudian 1 indonesia heboh karena bbbm...	[6, bulan, kemudian, 1, indonesia, heboh, kare...
1	Bagaina cinta prodak Indonesia KLO barang luar...	Bagaina cinta prodak Indonesia KLO barang luar...	bagaina cinta prodak indonesia klo barang luar...	[bagaina, cinta, prodak, indonesia, klo, baran...
2	Kapan Negara kita Mau produktif...daya saing k...	Kapan Negara kita Mau produktif daya saing kit...	kapan negara kita mau produktif daya saing kit...	[kapan, negara, kita, mau, produktif, daya, sa...
3	Bahlil tidak suka ini	Bahlil tidak suka ini	bahlil tidak suka ini	[bahlil, tidak, suka, ini]
4	Bagus pak presiden, kuota di hapuskan, tapi ja...	Bagus pak presiden kuota di hapuskan tapi jang...	bagus pak presiden kuota di hapuskan tapi jang...	[bagus, pak, presiden, kuota, di, hapuskan, ta...
...	...	...	...	...
3262	Lah kuma maneh we yg penting rakyat sejahtra d...	Lah kuma maneh we yg penting rakyat sejahtra d...	lah kuma maneh we yg penting rakyat sejahtra d...	[lah, kuma, maneh, we, yg, penting, rakyat, se...
3263	@amni.z0r0 karena 1 gerakan dan 1 kata bro.<b...	z0r0 karena 1 gerakan dan 1 kata bro br br Jog...	z0r0 karena 1 gerakan dan 1 kata bro br br jog...	[z0r0, karena, 1, gerakan, dan, 1, kata, bro, ...
3264	Syarat cinta tanah air adalah cinta produk dal...	Syarat cinta tanah air adalah cinta produk dal...	syarat cinta tanah air adalah cinta produk dal...	[syarat, cinta, tanah, air, adalah, cinta, pro...
3265	Pretttt.. Kuota ngga ada ijin import dipersuli...	Pretttt Kuota ngga ada ijin import dipersulit ...	pretttt kuota ngga ada ijin import dipersulit ...	[pretttt, kuota, ngga, ada, ijin, import, dipe...
3266	Mentrinya spt jkw tukang boong	Mentrinya spt jkw tukang boong	mentrinya spt jkw tukang boong	[mentrinya, spt, jkw, tukang, boong]

3239 rows × 4 columns

Gambar 6. Hasil Tokenisasi

d. Penghapusan *Stopword*

Proses *stopword removal* dilakukan untuk menghilangkan kata-kata yang memiliki nilai informatif rendah dan tidak berkontribusi signifikan terhadap konteks penelitian, seperti kata ganti dan kata umum lainnya, misalnya *saya*[16]. Proses tersebut bertujuan untuk mengurangi elemen teks yang tidak memberikan kontribusi terhadap analisis sentimen. Karakter-karakter yang tidak memiliki relevansi terhadap isi teks, seperti emoji dan berbagai simbol lainnya, dihilangkan dari data teks. Langkah ini dilakukan untuk mengurangi noise yang dapat mengganggu proses analisis dan pemodelan. Keberadaan karakter non-tekstual tersebut berpotensi menimbulkan kesalahan dalam proses ekstraksi fitur karena tidak memiliki makna linguistik yang jelas. Selain itu, pembersihan karakter ini membantu meningkatkan konsistensi struktur data teks[17].

	comment	cleansing	case_folding	tokenize	filtering/stopword removal
0	6 bulan kemudian 1 indonesia heboh karena bbbm...	6 bulan kemudian 1 indonesia heboh karena bbbm...	6 bulan kemudian 1 indonesia heboh karena bbbm...	[6, bulan, kemudian, 1, indonesia, heboh, kare...	[6, 1, indonesia, heboh, bbbm, swasta, kehabis...
1	Bagaina cinta prodak Indonesia KLO barang luar...	Bagaina cinta prodak Indonesia KLO barang luar...	bagaina cinta prodak indonesia klo barang luar...	[bagaina, cinta, prodak, indonesia, klo, baran...	[bagaina, cinta, prodak, indonesia, klo, baran...
2	Kapan Negara kita Mau produktif...daya saing k...	Kapan Negara kita Mau produktif daya saing kit...	kapan negara kita mau produktif daya saing kit...	[kapan, negara, kita, mau, produktif, daya, sa...	[negara, produktif, daya, saing, negri, kadarn...
3	Bahlil tidak suka ini	Bahlil tidak suka ini	bahlil tidak suka ini	[bahlil, tidak, suka, ini]	[bahlil, suka]
4	Bagus pak presiden, kuota di hapuskan, tapi ja...	Bagus pak presiden kuota di hapuskan tapi jang...	bagus pak presiden kuota di hapuskan tapi jang...	[bagus, pak, presiden, kuota, di, hapuskan, ta...	[bagus, presiden, kuota, hapuskan, ngomong, aj...
...	...	...	...	...	...
3262	Lah kuma maneh we yg penting rakyat sejahtra d...	Lah kuma maneh we yg penting rakyat sejahtra d...	lah kuma maneh we yg penting rakyat sejahtra d...	[lah, kuma, maneh, we, yg, penting, rakyat, se...	[kuma, maneh, we, yg, rakyat, sejahtra, impor,...
3263	@amni.z0r0 karena 1 gerakan dan 1 kata bro.<b...	z0r0 karena 1 gerakan dan 1 kata bro br br Jog...	z0r0 karena 1 gerakan dan 1 kata bro br br jog...	[z0r0, karena, 1, gerakan, dan, 1, kata, bro, ...	[z0r0, 1, gerakan, 1, bro, br, br, joget, gemo...
3264	Syarat cinta tanah air adalah cinta produk dal...	Syarat cinta tanah air adalah cinta produk dal...	syarat cinta tanah air adalah cinta produk dal...	[syarat, cinta, tanah, air, adalah, cinta, pro...	[syarat, cinta, tanah, air, cinta, produk, neg...
3265	Pretttt.. Kuota ngga ada ijin import dipersuli...	Pretttt Kuota ngga ada ijin import dipersulit ...	pretttt kuota ngga ada ijin import dipersulit ...	[pretttt, kuota, ngga, ada, ijin, import, dipe...	[pretttt, kuota, ngga, ijin, import, dipersuli...
3266	Mentrinya spt jkw tukang boong	Mentrinya spt jkw tukang boong	mentrinya spt jkw tukang boong	[mentrinya, spt, jkw, tukang, boong]	[mentrinya, spt, jkw, tukang, boong]

3239 rows × 5 columns

Gambar 7. Hasil penghapusan *Stopword*

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 3, April 2026 | Hal 886-898
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i3.995

e. *Stemming*

Stemming merupakan tahapan yang bertujuan mengembalikan setiap kata ke bentuk dasar (*root word*) dengan cara menghapus berbagai jenis imbuhan, baik awalan, akhiran, maupun bentuk morfologis lainnya yang melekat pada kata[3]. Menggunakan library Sastrawi, setiap kata pada komentar dikembalikan ke bentuk dasarnya (root word). Dengan demikian, kata-kata seperti "menyukai", dan "disukai" diubah menjadi "suka", sehingga mengurangi variasi kata yang tidak perlu. Tahapan ini merupakan bagian dari pemrosesan teks yang berfungsi menyederhanakan berbagai variasi kata menjadi bentuk dasar sesuai dengan kaidah morfologi bahasa. Dengan demikian, kata yang telah mengalami stemming tetap mempertahankan makna yang identik dengan bentuk dasarnya [18]. Hasil *preprocessing* secara keseluruhan menghasilkan data komentar yang lebih bersih, konsisten, dan terstruktur sehingga dapat digunakan sebagai masukan (*input*) pada tahap ekstraksi fitur menggunakan TF-IDF.

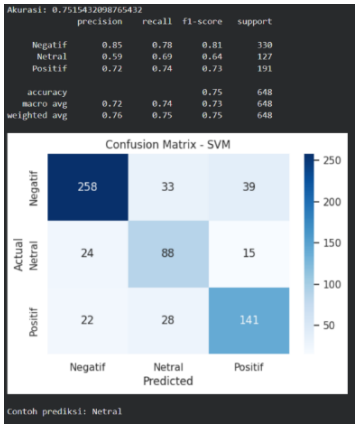
	comment	cleansing	case_folding	tokenize	filtering/stopword removal	stemming_data
0	6 bulan kemudian 1 indonesia heboh karena bbbm...	6 bulan kemudian 1 indonesia heboh karena bbbm...	6 bulan kemudian 1 indonesia heboh karena bbbm...	[6, bulan, kemudian, 1, indonesia, heboh, kare...	[6, 1, indonesia, heboh, bbbm, swasta, kehabis...	6 1 indonesia heboh bbbm swasta habis stok quo...
1	Bagaina cinta prodak Indonesia KLO barang luar...	Bagaina cinta prodak Indonesia KLO barang luar...	bagaina cinta prodak indonesia klo barang luar...	[bagaina, cinta, prodak, indonesia, klo, baran...	[bagaina, cinta, prodak, indonesia, klo, baran...	bagaina cinta prodak indonesia klo barang nege...
2	Kapan Negara kita Mau produktif...daya saing k...	Kapan Negara kita Mau produktif daya saing kit...	kapan negara kita mau produktif daya saing kit...	[kapan, negara, kita, mau, produktif, daya, sa...	[negara, produktif, daya, saing, negri, kadarn...	negara produktif daya saing negri kadar blm ne...
3	Bahlil tidak suka ini	Bahlil tidak suka ini	bahlil tidak suka ini	[bahlil, tidak, suka, ini]	[bahlil, suka]	bahlil suka
4	Bagus pak presiden, kuota di hapuskan, tapi ja...	Bagus pak presiden kuota di hapuskan tapi jang...	bagus pak presiden kuota di hapuskan tapi jang...	[bagus, pak, presiden, kuota, di, hapuskan, ta...	[bagus, presiden, kuota, hapuskan, ngomong, aj...	bagus presiden kuota hapus ngomong aja serta t...
5	ya allah ❤ SEBAGAI PETANI LOKAL GIMNA NASIBNYA...	ya allah SEBAGAI PETANI LOKAL GIMNA NASIBNYA h...	ya allah sebagai petani lokal gimna nasibnya h...	[ya, allah, sebagai, petani, lokal, gimna, nas...	[ya, allah, petani, lokal, gimna, nasibnya, ha...	ya allah tani lokal gimna nasib harga pupuk aw...
6	Ah omon omon	Ah omon omon	ah omon omon	[ah, omon, omon]	[ah, omon, omon]	ah omon omon
7	siap2 kita akan di perbudak secara ekonomi.ole...	siap2 kita akan di perbudak secara ekonomi ole...	siap2 kita akan di perbudak secara ekonomi ole...	[siap2, kita, akan, di, perbudak, secara, ekon...	[siap2, perbudak, ekonomi, negara2, maju]	siap2 budak ekonomi negara2 maju
8	Positif hancur perekonomian dalam negri Ak...	Positif hancur perekonomian dalam negri br Akh...	positif hancur perekonomian dalam negri br akh...	[positif, hancur, perekonomian, dalam, negri, ...	[positif, hancur, perekonomian, negri, br, yg,...	positif hancur ekonomi negri br yg kaya kaya m...
9	Bahaya bos kalo inpor tdk dikendalikan kasian ...	Bahaya bos kalo inpor tdk dikendalikan kasian ...	bahaya bos kalo inpor tdk dikendalikan kasian ...	[bahaya, bos, kalo, inpor, tdk, dikendalikan, ...	[bahaya, bos, kalo, inpor, tdk, dikendalikan, ...	bahaya bos kalo inpor tdk kendali kasi dgn umk...
10	Mohon di tindak lanjuti masalah kuota perikana...	Mohon di tindak lanjuti masalah kuota perikana...	mohon di tindak lanjuti masalah kuota perikana...	[mohon, di, tindak, lanjuti, masalah, kuota, p...	[mohon, tindak, lanjuti, kuota, perikanan, mem...	mohon tindak lanjut kuota ikan bingung msalah ...
11	Bravo Pak presiden Prabowo💗! 👌👍👍👍 ekonomi NKRI ...	Bravo Pak presiden Prabowo ekonomi NKRI harus ...	bravo pak presiden prabowo ekonomi nkri harus ...	[bravo, pak, presiden, prabowo, ekonomi, nkri,...	[bravo, presiden, prabowo, ekonomi, nkri, jala...	bravo presiden prabowo ekonomi nkri jalan lanc...
12	Matila Petani kita Pak Presiden....😭	Matila Petani kita Pak Presiden	matila petani kita pak presiden	[matila, petani, kita, pak, presiden]	[matila, petani, presiden]	matila tani presiden
13	Bebas impor????? Produksi dlm negeri bisa h...	Bebas impor br Produksi dlm negeri bisa hancur	bebas impor br produksi dlm negeri bisa hancur	[bebas, impor, br, produksi, dlm, negeri, bisa...	[bebas, impor, br, produksi, dlm, negeri, hancur]	bebas impor br produksi dlm negeri hancur
14	Kok impor bebas? Produksi dalam negeri bagaimana?	Kok impor bebas Produksi dalam negeri bagaimana	kok impor bebas produksi dalam negeri bagaimana	[kok, impor, bebas, produksi, dalam, negeri, b...	[impor, bebas, produksi, negeri]	impor bebas produksi negeri
15	Indonesia kiamat bukan gelap lagi	Indonesia kiamat bukan gelap lagi	indonesia kiamat bukan gelap lagi	[indonesia, kiamat, bukan, gelap, lagi]	[indonesia, kiamat, gelap]	indonesia kiamat gelap

Gambar 8. Hasil *Stemming*

3.3 Ektraksi Fitur

Setelah preprocessing, tahap berikutnya adalah melakukan ekstraksi fitur menggunakan metode Term Frequency–Inverse Document Frequency (TF-IDF). TF-IDF digunakan untuk mengubah data teks menjadi representasi numerik dalam bentuk vector space model. Metode ini memberikan nilai bobot pada setiap kata berdasarkan frekuensi kemunculannya dalam suatu dokumen (Term Frequency/TF) serta tingkat kesetaraannya pada keseluruhan kumpulan dokumen (Inverse Document Frequency/IDF). Dengan pendekatan tersebut, kata yang memiliki tingkat relevansi atau informasi tinggi akan memperoleh bobot yang lebih besar, sedangkan kata-kata yang bersifat umum dan sering muncul pada banyak dokumen akan diberikan bobot yang lebih rendah.

Pada penelitian ini, proses ekstraksi fitur menggunakan *TfidfVectorizer* diatur dengan parameter *max_features* sebesar 5.000 guna membatasi jumlah fitur pada kata-kata yang paling relevan. Selain itu, parameter *ngram_range* ditetapkan pada nilai (1,2) untuk mengakomodasi pembentukan fitur dalam bentuk unigram dan bigram. Penerapan bigram bertujuan untuk meningkatkan kemampuan model dalam memahami konteks frasa tertentu, seperti istilah "bebas impor", yang memiliki makna lebih spesifik dibandingkan kata tunggal. Proses pembobotan TF-IDF menghasilkan sebuah matriks *sparse* dengan ukuran (jumlah dokumen × jumlah fitur), yaitu (3.267 × 5.000), yang selanjutnya digunakan sebagai masukan utama dalam proses pemodelan menggunakan algoritma *Support Vector Machine*. Pemilihan pendekatan ini dinilai tepat karena SVM memiliki kinerja yang optimal dalam menangani data berdimensi tinggi dan bersifat *sparse*.

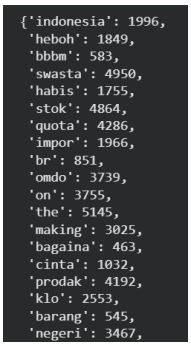


Gambar 9. Pemodelan *Support Vector Machine* (*Confusion Matrix*)

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 3, April 2026 | Hal 886-898
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i3.995

Dalam metode ini, komponen *Term Frequency* (TF) merepresentasikan jumlah kemunculan suatu kata dalam satu dokumen, sedangkan *Inverse Document Frequency* (IDF) menggambarkan tingkat kelangkaan kata tersebut pada keseluruhan dokumen. Pada penelitian ini, proses pembobotan dilakukan dengan memanfaatkan TfidfVectorizer, sehingga setiap kata dalam komentar dapat dikonversi menjadi nilai numerik sesuai tingkat kepentingannya untuk mendukung proses klasifikasi[19].

Berdasarkan gambar 11, pada tahap ini sebanyak 3.239 komentar yang telah melewati proses pra-pemrosesan digunakan sebagai data analisis. Proses pembobotan menghasilkan sebanyak 3.149 kata unik yang merepresentasikan berbagai istilah yang muncul di dalam keseluruhan dataset tersebut. Tahapan TF-IDF ini bertujuan mengonversi kumpulan dokumen teks menjadi representasi numerik dalam bentuk matriks istilah-dokumen. Melalui proses tersebut, setiap kata yang muncul dalam komentar diberikan bobot berdasarkan frekuensi kemunculannya pada suatu dokumen dibandingkan kemunculannya pada keseluruhan dokumen lainnya. Selain itu, setiap kata diberi indeks unik sehingga dapat diidentifikasi secara sistematis dalam vektor maupun matriks. Misalnya, kata "stok" memiliki indeks 4864, sedangkan kata "habis" memiliki indeks 1755, dan seterusnya. Tahap ini memiliki peranan yang krusial dalam mendukung proses analisis sentimen berbasis algoritma *Support Vector Machine*, karena memungkinkan model untuk mengidentifikasi serta membedakan tingkat kontribusi setiap kata dalam menentukan kategori sentimen dari suatu komentar.



Gambar 10. Daftar kata unik hasil tokenisasi

3.4 Pelabelan *Lexicon Based*

Setelah proses *pre-processing* menghasilkan data yang telah distandarisasi dalam variabel stemming_data, tahap berikutnya adalah melakukan pelabelan sentimen terhadap setiap komentar. Studi ini menerapkan pendekatan berbasis leksikon (*lexicon-based*), yaitu metode yang melakukan pencocokan setiap kata dalam komentar dengan himpunan kata positif dan negatif yang telah ditetapkan sebelumnya. Berdasarkan gambar 12, merupakan daftar kata unik hasil tokenisasi, pendekatan ini mencocokkan kata hasil stemming dengan kamus sentiment (*sentiment lexicon*) sehingga setiap komentar dapat diberikan label sesuai dengan kecenderungan emosional atau opini yang terkandung di dalamnya.

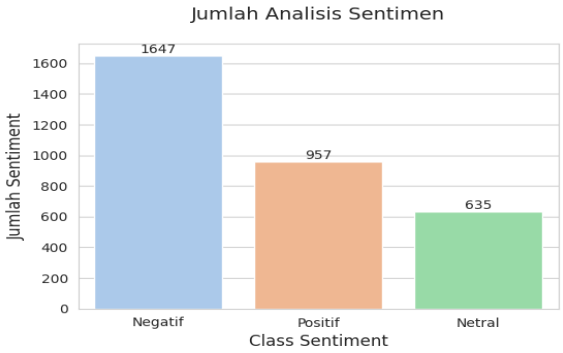
Gambar 11. Kode program untuk menghitung skor menggunakan *lexicon based*

Pada penelitian ini, daftar kata positif dan negatif diperoleh dari repositori GitHub yang menyediakan kumpulan lexicon Bahasa Indonesia dalam format.tsv. Kamus tersebut merupakan hasil kurasi dari proyek sentiment analysis yang banyak digunakan dalam penelitian-penelitian sebelumnya, sehingga memiliki tingkat validitas dan reliabilitas yang baik. Kamus yang digunakan pada penelitian ini adalah Indonesian Sentiment (InSet) *Lexicon*, yang berisi ratusan hingga ribuan kata berpolaritas positif dan negatif, lengkap dengan skor masing-masing kata yang menggambarkan tingkat kekuatan emosionalnya.

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 3, April 2026 | Hal 886-898
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i3.995

Setelah proses *pre processing* selesai, analisis sentimen dilakukan dengan menggunakan fungsi sentimen *analysis lexicon* Indonesia, yang menghitung skor sentimen secara sederhana berdasarkan kemunculan kata-kata dalam daftar kata positif dan negatif. Daftar kata tersebut diperoleh dari repositori *GitHub* yang menyediakan *lexicon* sentimen berbahasa Indonesia. Fungsi ini membandingkan setiap token dalam komentar dengan kata-kata yang ada di dalam *lexicon*. Setiap kemunculan kata positif akan menambah skor, sedangkan kata negatif akan menguranginya.

Hasil analisis sentimen terhadap komentar yang telah melalui tahapan pra-pemrosesan menunjukkan bahwa setiap komentar dapat dikelompokkan ke dalam tiga kategori utama, yakni sentimen negatif, positif, dan netral. Proses pengelompokan dilakukan menggunakan metode berbasis leksikon, yaitu dengan menghitung skor sentimen dari setiap komentar berdasarkan frekuensi kemunculan kata-kata yang tercantum dalam daftar kata positif dan negatif yang diperoleh dari repositori *GitHub*. Jika skor akhir dari sebuah komentar menunjukkan nilai positif, maka komentar tersebut diklasifikasikan sebagai sentimen positif; apabila skor bernilai negatif, maka dikategorikan sebagai sentimen negatif; dan apabila skor berada pada posisi netral atau nol, maka komentar dimasukkan ke dalam kategori sentimen netral.



Gambar 12. Jumlah Analisis Sentimen

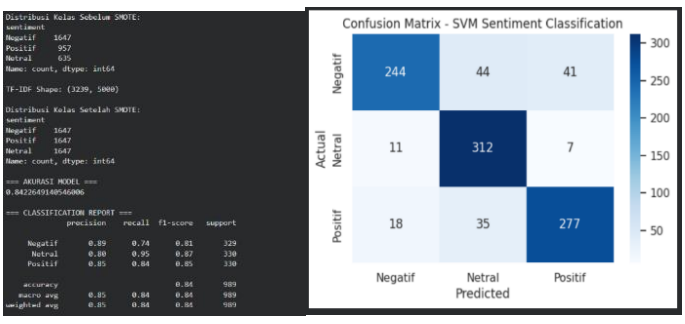
Pada Gambar 14 terlihat bahwa komentar dengan sentimen negatif mendominasi jumlah data, yaitu sebanyak 1.647 komentar, yang mencerminkan adanya kecenderungan mayoritas pengguna memberikan tanggapan yang kritis, kurang setuju, atau memiliki persepsi negatif terhadap kebijakan bebas impor tersebut.

Sementara itu, komentar dengan sentimen positif tercatat sebanyak 957 komentar, menunjukkan adanya warganet yang memberikan dukungan, pembelaan, atau pandangan yang lebih optimistis terhadap kebijakan yang dibahas. Adapun jumlah komentar dengan sentimen netral mencapai 635 komentar, yang mengindikasikan bahwa sebagian warganet memberikan opini secara objektif, informatif, atau tidak secara eksplisit mengekspresikan emosi tertentu terhadap topik tersebut.

3.5 Klasifikasi dengan *Support Vector Machine*

Support Vector Machine merupakan metode yang digunakan untuk melakukan prediksi, baik pada permasalahan klasifikasi maupun regresi. Secara fundamental, *Support Vector Machine* bekerja sebagai *linear classifier* yang memisahkan data ke dalam kelas berbeda menggunakan batas keputusan yang bersifat linier. Namun, perkembangan metode ini memungkinkan *Support Vector Machine* menangani permasalahan yang tidak dapat dipisahkan secara linier melalui penerapan kernel function, yang memetakan data ke ruang berdimensi lebih tinggi. Dengan demikian, pola non-linier dapat dipisahkan secara lebih efektif. Secara keseluruhan, *Support Vector Machine* menentukan garis atau bidang pemisah (*hyperplane*) dengan batas maksimum terhadap titik data terdekat dari masing-masing kelas agar diperoleh pemisahan yang optimal.

Setelah tahap pembobotan TF-IDF selesai, tahap selanjutnya dalam penelitian ini adalah melakukan penataan dengan menggunakan strategi *Support Vector Machine* (SVM). Berdasarkan gambar 15, model SVM akan memanfaatkan vektor TF-IDF yang dihasilkan sebagai elemen untuk mengisolasi kelas komentar positif, negatif, dan netral terhadap kebijakan bebas impor oleh pemerintah pusat. Proses persiapan model SVM dilakukan dengan informasi yang telah diproses sebelumnya dan diwakili oleh vektor TF-IDF [20].

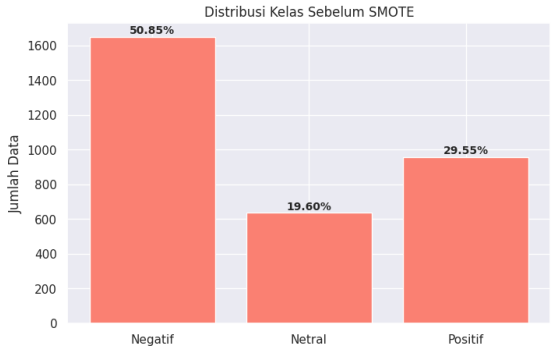


Gambar 13. Klasifikasi SVM dengan menggunakan metode SMOTE

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 3, April 2026 | Hal 886-898
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i3.995

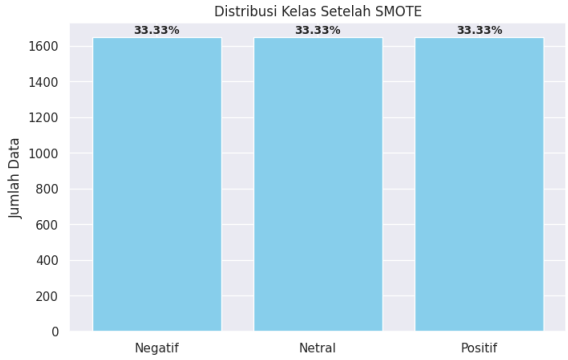
3.5.1 Penerapan SMOTE

Tahapan *Synthetic Minority Oversampling Technique* (SMOTE) berguna untuk melakukan penyeimbangan kata minoritas pada dataset yang akan berpengaruh terhadap hasil *accuracy*, *precision*, *recall* dan *F1 score* nya. Setelah dilakukannya ekstraksi fitur dan klasifikasi, dikarenakan terdapatnya ketidakseimbangan dataset yang digunakan, oleh karena itu untuk menyeimbangkan dataset, peneliti menggunakan metode *SMOTE* agar pelabelan antara sentimen positif, negatif, dan netral bisa diseimbangkan.



Gambar 14. Persentase Distribusi kelas sebelum SMOTE

Pembobotan menggunakan metode SMOTE diterapkan dengan tujuan untuk menyamakan jumlah data pada setiap kategori sentimen, yaitu positif, negatif, dan netral, sehingga tercipta distribusi data yang lebih seimbang. Pada kondisi awal, sebagaimana ditunjukkan pada Gambar 16, terdapat ketimpangan yang cukup signifikan, di mana jumlah komentar bersentimen negatif jauh lebih besar dibandingkan dua kelas lainnya, dengan rasio kurang lebih 3:1 antara sentimen negatif dan positif. Ketidakseimbangan ini berpotensi menimbulkan bias pada model klasifikasi, sebab algoritma cenderung lebih mudah mempelajari pola dari kelas mayoritas dan mengabaikan karakteristik kelas minoritas. Setelah proses SMOTE dilakukan, jumlah data pada setiap kelas meningkat hingga mencapai proporsi yang sama, yaitu rasio 1:1 antara sentimen negatif, positif, dan netral. Pada gambar 17, keseimbangan ini membuat proses pelatihan model menjadi lebih optimal, karena algoritma SVM mampu mempelajari representasi tiap kelas secara proporsional. Selain itu, penerapan SMOTE juga memberikan peningkatan terhadap performa evaluasi model, khususnya pada nilai recall dan f1-score untuk kelas positif dan netral yang sebelumnya kurang terwakili. Dengan demikian, teknik *oversampling* ini tidak hanya memperbaiki distribusi data, tetapi juga memberikan kontribusi penting terhadap konsistensi dan akurasi hasil klasifikasi sentimen secara keseluruhan.



Gambar 15. Persentase Distribusi kelas Setelah SMOTE

3.5.2 Bentuk Vektor TF-IDF

Dalam proses ini, peneliti menerapkan parameter *max_features* sebesar 5.000 untuk membatasi jumlah fitur pada kata-kata unik yang paling signifikan. Selain itu, digunakan *ngram_range* (1,2) sehingga model tidak hanya mempertimbangkan satu kata tunggal (unigram), tetapi juga pasangan dua kata berurutan (bigram) untuk menangkap konteks frasa yang lebih kompleks pada komentar YouTube. Penerapan *sublinear_tf=True* turut berfungsi menormalkan nilai frekuensi kata agar bobot TF lebih stabil dan tidak terlalu didominasi oleh kata yang memiliki kemunculan sangat tinggi. Setelah proses pelatihan vectorizer dilakukan, diperoleh matriks TF-IDF dengan ukuran sebagai berikut: *Shape TF-IDF train*: (2591, 5000) *Shape TF-IDF test*: (648, 5000).

Hasil tersebut menunjukkan bahwa setiap komentar pada data latih maupun data uji direpresentasikan dalam bentuk vektor berdimensi 5.000 fitur. Setiap elemen vektor tersebut merupakan nilai bobot TF-IDF yang merepresentasikan tingkat signifikansi sebuah kata atau frasa dalam keseluruhan dataset. Representasi numerik ini kemudian digunakan sebagai masukan bagi algoritma SVM dalam membangun model klasifikasi sentimen.

BULLETIN OF COMPUTER SCIENCE RESEARCH
ISSN 2774-3659 (Media Online)
Vol 6, No 3, April 2026 | Hal 886-898
https://hostjournals.com/bulletincsr
DOI: 10.47065/bulletincsr.v6i3.995

3.6 Evaluasi

Tahapan ini dilakukan untuk menguji performa model dari mesin yang telah dibangun, untuk menghitung akurasi dan mengidentifikasi klasifikasi algoritma *Support Vector Machine* menggunakan *confusion matrix* dengan menghitung *accuracy*, *precision*, *recall* dan *F1-Score*. Setelah proses training menggunakan algoritma *Linear Support Vector Classifier* (*LinearSV*C), diperoleh hasil evaluasi sebagai berikut.

1.6.1 Akurasi Model Sebelum Tuning

Tabel 1. Hasil Evaluasi Model SVM sebelum tuning

Kelas Sentimen	*Precision*	*Recall*	*F1-Score*	*Support*
Negatif	0,82	0,83	0,82	330
Netral	0,68	0,65	0,67	127
Positif	0,74	0,74	0,74	191
Accuracy	-	-	0,7700	648
Macro Avg	0,75	0,74	0,74	648
Weight Avg	0,77	0,77	0,77	648

Berdasarkan Tabel 1, hasil evaluasi model menunjukkan performa awal yang cukup baik dengan akurasi 77%. Kelas Negatif memiliki kinerja terbaik, sedangkan kelas Netral menjadi kelas yang paling sulit diklasifikasikan. Hal ini wajar mengingat komentar netral umumnya tidak memiliki indikator linguistik yang kuat. Pada tahap awal, model *Support Vector Machine* (SVM) dilatih menggunakan parameter *default* yang disediakan oleh algoritma *LinearSVC*. Model ini belum mengalami penyesuaian parameter (*hyperparameter tuning*), sehingga performa yang dihasilkan menggambarkan kemampuan dasar SVM dalam memproses data komentar yang telah direpresentasikan menggunakan TF-IDF. Nilai Akurasi: 0,7700 (77,00%). Akurasi ini berarti bahwa 77% dari total data uji berhasil diprediksi dengan benar oleh model, sedangkan 23% sisanya diklasifikasikan secara keliru. Model SVM memiliki kinerja dasar yang baik, namun masih menunjukkan kecenderungan bias terhadap kelas mayoritas, serta kesulitan dalam mengidentifikasi sentimen Netral.

2.6.1 Akurasi Model Setelah Tuning

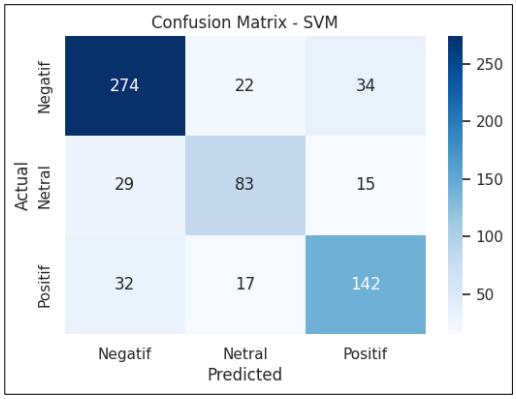
Tabel 2. Hasil Evaluasi Model SVM Setelah *Tuning*

Kelas Sentimen	Precision	Recall	F1-Score	Support
Negatif	0,85	0,78	0,81	330
Netral	0,59	0,69	0,64	127
Positif	0,72	0,74	0,73	191
Accuracy	-	-	0,7515	648
Macro Avg	0.72	0,74	0,73	648
Weighted Avg	0,76	0,75	0,75	648

Setelah dilakukan optimasi parameter menggunakan metode *GridSearchCV* dengan skema validasi silang *5-fold*, Tabel 2 menunjukkan hasil yang diperoleh kombinasi terbaik C = 1, class_weight = balanced. Model kemudian dilatih ulang menggunakan parameter tersebut dan diuji kembali. Nilai Akurasi: 0,7515 (75,15%). Akurasi model mengalami sedikit penurunan dari 77% menjadi 75,15%.

3.6.1 Analisis Confusion Matrix

Confusion matrix digunakan sebagai alat evaluasi untuk menilai kinerja model klasifikasi dengan membandingkan label aktual terhadap hasil prediksi yang dihasilkan oleh model. Dalam penelitian ini, *confusion matrix* terdiri atas tiga kategori sentimen, yaitu sentimen positif, negatif, dan netral.



Gambar 16. Confusion Matrix

Berdasarkan Gambar 18, hasil *confusion matrix* yang diperoleh, dapat diuraikan sebagai berikut.

a. Kelas sentimen negatif, dari total 330 komentar yang memiliki label sentimen negatif, sebanyak 274 komentar berhasil diklasifikasikan secara tepat sebagai sentimen negatif. Namun demikian, masih terdapat 22 komentar yang salah diklasifikasikan ke dalam kelas netral dan 34 komentar yang keliru diprediksi sebagai sentimen positif. Temuan ini menunjukkan bahwa model memiliki kemampuan yang cukup baik dalam mengidentifikasi sentimen negatif, meskipun masih terdapat sejumlah kesalahan klasifikasi ke kelas lainnya.
b. Kelas sentimen netral, pada kelas sentimen netral, dari 127 komentar, sebanyak 83 komentar berhasil diprediksi dengan benar. Sementara itu, 29 komentar diklasifikasikan secara keliru sebagai sentimen negatif dan 15 komentar sebagai sentimen positif. Hal ini mengindikasikan bahwa sentimen netral cenderung lebih sulit dikenali oleh model, karena karakteristik bahasanya sering kali berada di antara sentimen negatif dan positif.
c. Kelas sentimen positif, untuk kategori sentimen positif, dari total 191 komentar, sebanyak 142 komentar berhasil diklasifikasikan secara benar. Adapun 32 komentar salah diprediksi sebagai sentimen negatif dan 17 komentar sebagai sentimen netral. Hasil ini menunjukkan bahwa model cukup efektif dalam mengenali sentimen positif, meskipun masih terdapat tumpang tindih dengan kelas sentimen lainnya.

Secara keseluruhan, hasil *confusion matrix* menunjukkan bahwa model *Support Vector Machine* mampu melakukan klasifikasi sentimen dengan kinerja yang cukup baik, terutama pada kelas sentimen negatif dan positif yang memiliki pola linguistik yang lebih jelas. Kesalahan klasifikasi paling dominan terjadi pada kelas sentimen netral, yang secara umum bersifat ambigu dan tidak memiliki kecenderungan emosi yang kuat.

4. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilaksanakan, dapat disimpulkan bahwa penerapan algoritma *Support Vector Machine* (SVM) dalam analisis sentimen komentar pengguna YouTube terkait kebijakan bebas impor mampu menghasilkan performa klasifikasi yang baik dan konsisten. Pengumpulan data dalam penelitian ini dilakukan menggunakan teknik *web scraping* dengan memanfaatkan YouTube Data API v3, sehingga diperoleh sebanyak 3.267 komentar mentah dari video YouTube yang membahas kebijakan bebas impor. Seluruh data yang dikumpulkan merupakan komentar asli dari pengguna dan tidak mengalami pengurangan jumlah selama proses pengambilan data. Pada tahap preprocessing, khususnya pada proses *cleaning* atau *cleansing*, tidak terjadi perubahan pada jumlah data, melainkan hanya dilakukan pembersihan terhadap elemen–elemen yang tidak relevan, seperti simbol, tanda baca, angka, tautan, serta karakter non-teks lainnya. Dengan demikian, konsistensi dan integritas data tetap terjaga sejak tahap awal hingga tahap analisis, sehingga informasi utama yang terkandung dalam komentar tetap dapat dimanfaatkan secara optimal.

REFERENCES

[1] R. Firdaus, R. Al Hariri, and H. F. Amran, "Sentimen Analisis Masyarakat Tentang Penetapan Hari Raya Idul Adha Tahun 2023 Pada Video Youtube Menggunakan Algoritma Random Forest dan Support Vector Machine," *J. Fasilkom*, vol. 14, no. 1, pp. 278–285, 2024, doi: 10.37859/jf.v14i1.7012.
[2] R. T. Adek, Z. Fitri, and S. Chairani Siegar, "Analisis Sentimen Komentar Pada Saluran Youtube Beauty Vlogger Berbahasa Indonesia Menggunakan Metode Support Vector Machine," *J. Algoritm.*, vol. 5, no. 2, pp. 164–175, 2025, doi: 10.35957/algoritme.v5i2.9692.
[3] D. Marganingsih, H. Oktavianto, and G. Abdurrahman, "Analisis Sentimen Komentar Youtube Masterchef Indonesia Menggunakan Algoritma Support Vector Machine dan Gaussian Naïve Bayes," *J. Inform. dan Teknol. Pendidik.*, vol. 5, no. 1, pp. 16–26, 2025, doi: 10.59395/jitp.v5i1.117.
[4] Y. I. Muasaroh, Z. Fatah, and A. Baijuri, "Analisis Sentimen Komentar Youtube Terhadap Isu Ijazah Presiden Jokowi menggunakan Support Vector Machine dan Random Forest," *Pros. Semnas 2025 Sekol. Tinggi Teknol. Dumai*, vol. 1, no. 2, pp. 2581–267, 2025, [Online]. Available: https://ejurnal.sttdumai.ac.id/index.php/prosidingsemnas/article/view/1546/680
[5] M. I. Sultan and M. Akbar, "Analisis Sentimen Pemecatan Jokowi Pada Komentar Publik YouTube Tempo.co," vol. 07, no. 02, pp. 125–140, 2025, [Online]. Available: https://ejournal.uit-lirboyo.ac.id/index.php/kopis/article/view/6888/2187
[6] S. Pratama, Majduddin, and M. Triawan, "Klasifikasi Sentimen Komentar pada Video 'Rendang Hilang di Palembang' oleh Willy Salim Menggunakan Algoritma Support Vector Machine (SVM)," *SISKOMTI J. Sist. Inf. Komput. dan Teknol. Inf.*, vol. 7, no. 1, pp. 47–55, 2025, doi: 10.54342/625mqn95.
[7] A. A. Nurrahman, M. Mauladi, and A. Rahman, "Analisis Sentimen Masyarakat terhadap Kenaikan Harga Bahan Bakar Minyak Menggunakan Support Vector Machine dan SMOTE," *sudo J. Tek. Inform.*, vol. 4, no. 2, pp. 50–56, 2025, doi: 10.56211/sudo.v4i2.908.
[8] F. F. Abdulloh and I. R. Pambudi, "Analisis Sentimen Pengguna Youtube Terhadap Program Vaksin Covid-19," *CSRID (Computer Sci. Res. Its Dev. Journal)*, vol. 13, no. 3, p. 141, 2021, doi: 10.22303/csrid.13.3.2021.141-148.
[9] F. Caroline, R. G. S. Budi, and M. E. Al Rivan, "Analisis Sentimen Masyarakat terhadap Kasus Korupsi PT. Timah Menggunakan Metode Support Vector Machine," *J. Ilmu Komput. dan Inform.*, vol. 4, no. 1, pp. 43–50, 2024, doi: 10.54082/jiki.141.
[10] H. Hidayat, F. Santoso, and L. F. Lidimillah, "Analisis Sentimen Pengguna YouTube Tentang Rohingya Menggunakan Algoritma SVM (Support Vector Machine)," *G-Tech J. Teknol. Terap.*, vol. 8, no. 3, pp. 1729–1738, Jul. 2024, doi: 10.33379/gtech.v8i3.4497.
[11] A. M. Putra, Candra Saputra, Rahmaddeni, Safril Irsandi, and Vawana Muzaki, "Analisis Sentimen Masyarakat Terhadap Kasus Gas LPG 3 Kg Pada Youtube Kompas Menggunakan Metode Support Vector Machine," *Explore*, vol. 15, no. 2, pp. 163–171, 2025, doi: 10.35200/ex.v15i2.159.
[12] R. Asrianto and M. Herwinanda, "Analisis sentimen kenaikan harga kebutuhan pokok dimedia sosial youtube menggunakan algoritma support vector machine," *J. CoSciTech (Computer Sci. Inf. Technol.*, vol. 3, no. 3, pp. 431–440, Dec. 2022, doi:

10.37859/coscitech.v3i3.4368.

[13] N. A. Wahyuni, D. P. Ayu, and H. Irsyad, "Analisis Sentimen di Youtube Terhadap Kenaikan UKT Menggunakan Metode Support Vector Machine," *Arcitech J. Comput. Sci. Artif. Intell.*, vol. 4, no. 1, p. 57, Jun. 2024, doi: 10.29240/arcitech.v4i1.10829.

[14] T. Muhayat, A. Fauzi, and J. Indra, "Analisis Sentimen Terhadap Komentar Video Youtube Menggunakan Support Vector Machines," *Progresif J. Ilm. Komput.*, vol. 19, no. 1, p. 231, 2023, doi: 10.35889/progresif.v19i1.1060.

[15] A. N. Syafia, M. F. Hidayattullah, and W. Suteddy, "Studi Komparasi Algoritma SVM Dan Random Forest Pada Analisis Sentimen Komentar Youtube BTS," vol. 8, no. 3, pp. 207–212, 2023, [Online]. Available: https://pdfs.semanticscholar.org/7766/7b4bdc15a13c4644b1c2d17811ebf60d47d5.pdf

[16] A. Wijayanto and A. D. Defara, "Analisis Sentimen Komentar Youtube Mengenai Vaksin Covid-19 Menggunakan Support Vector Machine." [Online]. Available: http://pilar.unmermadiun.ac.id/index.php/pilarteknologi

[17] N. A. Laia and S. P. Barus, "Analisis Sentimen Pengguna Youtube Pada Video Berjudul '10 Tahun Jokowi Jadi Presiden,'" *JIKA (Jurnal Inform.*, vol. 9, no. 2, p. 169, 2025, doi: 10.31000/jika.v9i2.13470.

[18] L. Rofiqi and M. Akbar, "Analisis Sentimen Terkait RUU Perampasan Aset dengan Support Vector Machine," *JEKIN - J. Tek. Inform.*, vol. 4, no. 3, pp. 529–538, 2024, doi: 10.58794/jekin.v4i3.824.

[19] S. A. S. Mola, P. R. Lete, B. J. A. J. A. Pa, Triyanto, and T. Widiastuti, "Analisis Sentimen Menggunakan Metode Naive Bayes Dan Metode Support Vector Machine Pada Kasus Pelantikan Artis Sebagai Anggota Anggota Dpr Ri Tahun 2024," *HOAQ (High Educ. Organ. Arch. Qual. J. Teknol. Inf.*, vol. 15, no. 1, pp. 22–32, 2024, doi: 10.52972/hoaq.vol15no1.p22-32.

[20] A. A. Syam, G. Hardy M, A. Salim, D. F. Surianto, and M. Fajar B, "Analisis Teknik Preprocessing Pada Sentimen Masyarakat Terkait Konflik Israel-Palestina Menggunakan Support Vector Machine," *JIPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.*, vol. 9, no. 3, pp. 1464–1472, 2024, doi: 10.29100/jipi.v9i3.5527.