



Analisis Sentimen Terhadap Komentar Video IShowSpeed Tour Indonesia Pada YouTube Menggunakan Metode SVM

Daffa Ihza Kurniawan¹, Irwansyah^{2*}, Ahmad Taufik³

^{1,2}Teknik Informatika, Fakultas Teknologi Industri dan Informatika, Universitas Muhammadiyah Prof. Dr. HAMKA, Jakarta, Indonesia

³Sistem Informasi, Fakultas Teknologi, Institut Sosial dan Teknologi (ISTEK) Widuri, Jakarta, Indonesia

*✉ Corresponding Author: irwansyah@uhamka.ac.id

ARTICLE INFO

Vol. 1, No. 1, October 2025,
pp. 50-62

Article history:

Revised: 25 September 2025

Accepted: 5 October 2025

Published: 8 October 2025

Abstract

In September 2024, international influencer IShowSpeed's visit to Indonesia attracted public attention and triggered various reactions on social media, especially YouTube. The content in the form of tours and live broadcasts conducted by IShowSpeed generated various comments from users, ranging from positive, negative, to neutral sentiments. This study aims to analyze these sentiments using the Support Vector Machine (SVM) method with a linear kernel. A total of 43,778 comments were used in this study. The classification results showed an accuracy of 91.1%. For negative sentiment, precision 86%, recall 78%, and f1-score 82% were obtained. Neutral sentiment achieved 90% precision, 94% recall, and 92% f1-score. Meanwhile, positive sentiment obtained a precision, recall, and f1-score of 94% each. These findings show that the majority of user comments are positive, indicating that IShowSpeed and its content are well received by Indonesian audiences.

Keywords: Sentiment Analysis, Support Vector Machine, YouTube, IShowSpeed.

Abstrak

Pada September 2024, kedatangan *influencer* internasional IShowSpeed ke Indonesia menarik perhatian publik dan memicu berbagai reaksi di media sosial, khususnya *YouTube*. Konten berupa tur dan siaran langsung yang dilakukan IShowSpeed menimbulkan beragam komentar dari pengguna, mulai dari sentimen positif, negatif, hingga netral. Penelitian ini bertujuan untuk menganalisis sentimen tersebut menggunakan metode *Support Vector Machine* (SVM) dengan kernel linear. Sebanyak 43.778 komentar digunakan dalam penelitian ini. Hasil klasifikasi menunjukkan akurasi sebesar 91,1%. Untuk sentimen negatif diperoleh *precision* 86%, *recall* 78%, dan *f1-score* 82%. Sentimen netral mencapai *precision* 90%, *recall* 94%, dan *f1-score* 92%. Sementara itu, sentimen positif memperoleh *precision*, *recall*, dan *f1-score* masing-masing sebesar 94%. Temuan ini menunjukkan bahwa mayoritas komentar pengguna bersifat positif, menandakan bahwa kehadiran IShowSpeed dan kontennya diterima dengan baik oleh penonton Indonesia.

Kata Kunci: Analisis sentimen, *Support vector machine*, *YouTube*, IShowSpeed.



Copyright: © 2025 by the authors. Licensee DIGINTEL AI, Jakarta,

Indonesia. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License (CC BY-NC-SA 4.0)

<https://creativecommons.org/licenses/by-nc-sa/4.0/>

1. Pendahuluan

Di era digital saat ini, data telah menjadi aset penting bagi berbagai sektor industri. Pertumbuhan data yang masif menciptakan tantangan dalam pengelolaan dan analisisnya. Untuk mengatasi hal tersebut, teknik *data mining* hadir sebagai solusi dalam menggali informasi tersembunyi dari kumpulan data yang besar (Irwansyah et al., 2022). Salah satu teknik yang banyak digunakan adalah klasifikasi, yang berfungsi untuk mengelompokkan data ke dalam kategori tertentu. Dalam konteks ini, *Support Vector Machine* (SVM) menjadi salah satu algoritma yang efektif karena mampu membedakan data dengan akurasi tinggi melalui pencarian garis pemisah (*hyperplane*) optimal antar kelas data (Mualfah et al., 2023).

Di sisi lain, perkembangan platform digital seperti *YouTube* turut menyumbang peningkatan jumlah data, khususnya dalam bentuk komentar pengguna. *YouTube* kini tidak hanya menjadi media hiburan, tetapi juga ruang interaksi publik yang dinamis. Salah satu fenomena menarik dalam beberapa waktu terakhir adalah meningkatnya popularitas kreator konten internasional yang melakukan tur ke berbagai negara. Salah satu di antaranya adalah IShowSpeed, seorang *YouTuber* internasional yang melakukan tur ke Indonesia dan mendapat sorotan luas. Video-videonya dibanjiri komentar dari penonton, baik dari dalam maupun luar negeri, yang menunjukkan beragam respons berupa opini positif, negatif, maupun netral (Wijayanto, 2022).

Sayangnya, hingga saat ini masih sangat sedikit penelitian di Indonesia yang secara khusus menganalisis sentimen publik terhadap figur konten kreator internasional seperti IShowSpeed di media sosial. Mayoritas penelitian analisis sentimen di Indonesia masih berfokus pada isu politik, pelayanan publik, pariwisata, dan brand komersial. Hal ini menjadi celah penelitian yang penting untuk diisi, mengingat *influencer* seperti IShowSpeed memiliki basis penggemar yang besar dan dampak signifikan terhadap opini publik (Hermawan et al., 2023). Metode manual seperti ini jelas memiliki banyak keterbatasan, antara lain memakan waktu lama, membutuhkan tenaga yang besar, dan berpotensi menghasilkan hasil yang subjektif.

Namun, hingga saat ini, belum ditemukan penelitian di Indonesia yang secara spesifik menerapkan metode SVM untuk menganalisis sentimen komentar terhadap video IShowSpeed *Tour* Indonesia di *YouTube*. Padahal, kunjungan *influencer* internasional seperti IShowSpeed ke Indonesia memicu perhatian publik yang luas dan memunculkan berbagai tanggapan positif, negatif, maupun netral dari para penonton. Oleh karena itu, penelitian ini sangat penting dilakukan untuk mengisi celah penelitian yang ada sekaligus memberikan kontribusi ilmiah dalam bidang analisis sentimen di ranah konten hiburan dan *influencer* (Mualfah et al., 2023).

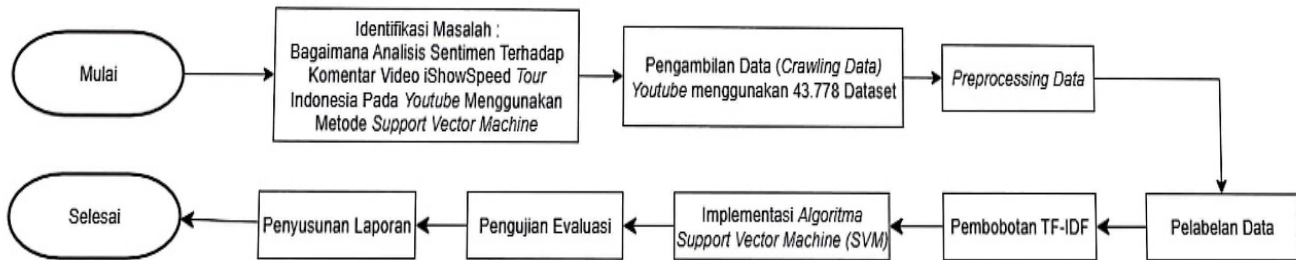
Analisis sentimen dapat dilakukan dengan memanfaatkan teknik *machine learning*, di mana metode *Support Vector Machine* (SVM) telah terbukti memberikan akurasi yang tinggi. Khususnya dalam klasifikasi teks, SVM mampu menghasilkan prediksi sentimen yang cukup tepat. Prinsip kerja SVM adalah dengan membuat *hyperplane* sebagai pembatas yang memisahkan data ke dalam kategori-kategori berbeda, sehingga sangat cocok digunakan dalam pengelompokan sentimen menjadi kategori seperti positif, negatif, dan netral (Husada & Paramita, 2021).

Penelitian ini dilakukan dengan tujuan untuk melaksanakan analisis sentimen dari komentar-komentar *YouTube* pada konten "*In Real Life IShowSpeed Tour Indonesia*" menggunakan metode *Support Vector Machine* (SVM) (Eldo et al., 2024). Melalui analisis ini, diharapkan mampu memberikan pemahaman yang lebih mendalam mengenai persepsi penonton terhadap tur tersebut, serta memperlihatkan bagaimana metode SVM dapat diterapkan dalam analisis sentimen pada platform media sosial seperti *YouTube* (Oktaviana et al., 2022).

Temuan dari penelitian ini dapat memberikan wawasan bagi kreator konten dalam memahami persepsi audiens, serta membantu peneliti dan praktisi dalam mengembangkan strategi komunikasi digital yang lebih efektif berdasarkan opini publik. Berdasarkan permasalahan, peneliti memiliki tujuan untuk menganalisis Sentimen Terhadap Komentar Video *IshowSpeed tour* Indonesia pada *YouTube* menggunakan metode *Support Vector Machine*.

2. Metodologi

Pada gambar 1 dibawah ini terdapat alur penelitian, akan menjelaskan langkah langkah yang akan dilakukan pada penelitian ini.



Gambar 1. Alur Penelitian

2.1. Pengambilan data (*Crawling*)

Dalam tahap pengumpulan data, peneliti melakukan proses '*crawling*' atau pengambilan data secara otomatis dari komentar pada video *YouTube* IShowSpeed selama kunjungannya ke Indonesia. Proses ini memanfaatkan *YouTube* Data API v3 untuk mengakses dan mengumpulkan data secara efisien (Irwansyah et al., 2024).

2.2. Preprocessing

Langkah pertama dalam pengolahan data adalah tahap pemrosesan. Pada fase ini, data dibersihkan dan diformat sedemikian rupa sehingga siap untuk tahap analisis berikutnya. Beberapa langkah yang dilakukan secara berurutan meliputi (Darwis et al., 2020) :

1. *Case Folding*: *Case folding* adalah teknik standarisasi teks dengan mengkonversi semua huruf kapital menjadi huruf kecil (a-z). Proses ini diperlukan karena dokumen teks seringkali mengandung campuran huruf besar dan kecil yang tidak konsisten. Dengan melakukan *case folding*, seluruh teks akan menjadi kecil sehingga memudahkan proses analisis lebih lanjut (Wiranata et al., 2023).
2. *Data Cleaning*: *Data cleaning* bertujuan memfilter dan menghapus berbagai *noise* dalam data seperti tag HTML, tautan, script, atau tanda baca yang tidak relevan (Bagaskara et al., 2024).
3. *Stopword Removal*: Dengan menggunakan daftar kata *stop* bahasa Indonesia, kata-kata umum seperti "dan", "yang", atau "di" yang sering muncul tapi tidak mengandung makna penting akan disaring. Hal ini dilakukan untuk memusatkan analisis pada kata-kata kunci yang benar-benar relevan (Muryono & Irwansyah, 2020).
4. *Tokenizing*: Tahap ini merupakan proses untuk memecah kalimat berdasarkan dari kata penyusunnya, kata ini dipecah berdasarkan spasi.
5. *Stemming*: *Tokenizing* adalah teknik pemecahan teks menjadi unit-unit kata individual (token) dengan menghilangkan spasi, angka, dan karakter pemisah lainnya.
6. *Normalized*: Normalisasi merupakan proses penyeragaman kata ke bentuk dasarnya. Contohnya, kata "berlari", "lari", dan "pelari" akan dinormalisasi menjadi "lari". Proses ini penting untuk menyederhanakan analisis teks dalam bahasa Indonesia.

2.3. Pelabelan data

Dalam tahap pelabelan data, setiap komentar dalam video *YouTube* IShowSpeed ke Indonesia dianalisis secara mendalam untuk menentukan sentimennya. Pelabelan mengategorikan komentar tersebut ke dalam tiga kelas: positif, negatif dan netral. Proses ini dilakukan dengan mengacu pada pedoman pelabelan yang telah

disusun sebelumnya, sehingga hasil klasifikasi menjadi lebih objektif dan konsisten (Anisah & Irwansyah, 2025).

2.4. Pembobotan TF-IDF

Setelah data dibersihkan dan dilabeli, dilakukan proses pembobotan kata menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF). Proses ini menghitung bobot setiap kata berdasarkan frekuensi kemunculannya dalam dokumen (TF) dan *inverse document frequency* (IDF) dalam keseluruhan dataset. Pembobotan TF-IDF menghasilkan representasi numerik dari teks komentar yang mencerminkan kepentingan relatif setiap kata dalam konteks analisis sentiment.

2.5. Pengujian Evaluasi

Untuk menilai kinerja model SVM, kita menggunakan *Confusion Matrix*. *Matrix* ini memberikan gambaran lengkap tentang hasil prediksi model, termasuk akurasi (tingkat ketepatan keseluruhan), presisi (ketepatan prediksi kelas positif), *recall* (kemampuan mengenali semua contoh kelas positif), dan *F1-score* (rata-rata harmonik antara presisi dan *recall*). Selain itu, analisis kesalahan dilakukan untuk mengidentifikasi pola-pola kesalahan klasifikasi yang dapat menjadi masukan untuk perbaikan model. Hasil evaluasi disajikan dalam bentuk visualisasi untuk memudahkan interpretasi (Muryono et al., 2021).

3. Hasil dan Pembahasan

3.1 Pengambilan Data

Dalam proses ini, pengambilan data menggunakan API dari *Google* yaitu *Youtube Data API v3*. Untuk mengambil data tersebut, menggunakan *Google Colabs* dan dengan menggunakan bahasa pemrograman *python* (Irwansyah et al., 2023). Berikut adalah gambar dari program pengambilan data. *API key* di sembunyikan untuk menjaga keamanan karena setiap pengguna memiliki kredensial API yang berbeda dari akun *Google* mereka. Code ditujukan pada gambar 2.

```
[ ] # isikan dengan api key Anda
    api_key = '#####'

    # Enter video id
    video_id = "9s6imDGjy08"

    # Call function
    comments = video_comments(video_id)

    comments
```

Gambar 2. Ilustrasi Pengambilan Data

Data yang didapatkan untuk penelitian ini diperoleh sebanyak 43.778 data. Berikut contoh data yang di ambil. Hasil *crawling* ditujukan pada tabel 1 berikut:

Tabel 1. Hasil *YouTube Crawling*

No	Date	Username	Comment
1	2024-09-21 08:53:14	@naylasung3386	saat dia bermain musik dia terlalu bersemangat kamu kelihatan senang sekali dengan beragam budaya yang kami punya terima kasih sudah

No	Date	Username	Comment
			datang di negara kami indonesia
2	2024-09-20 20:06:55	@haxim9020	@@justmif-w3jSEBENARNYA KALO MENURUT PENDAPAT KU KAMI GAK CLAIM CUMAN MENGAPRISIASI BUDAYA YANG DIBAWA TAPI SECARA BERLEBIHAN DAN MUNGKIN SESETENGAH WARGA KAMI FIKIR KALO INDONESIA AMA MALAYSIA INI KAYAK TEMEN BISA BERKONGSI BUDAYA TAPI KAMI SALAH SEHARUS
3	2024-09-20 01:28:47	@idahastuti2035	Malaysia jangan pakai nama Batik..batik itu nama yg dibuat indonesia..buat nama lain jangan ikut ikutan terus dan ngaku ngaku..malaysia ngga malu tah..bangga donk dengan budaya sendiri jangan iriii dan ngaku nagkuin budaya negara tetangga..

3.2 Preprocessing Data

Langkah ini merupakan tahap awal dalam mempersiapkan data sebelum dilakukan proses analisis. Tujuan utamanya merupakan agar membersihkan, menyaring, dan juga menyesuaikan data guna sesuai kebutuhan pengolahan menggunakan metode *Support Vector Machine*. Seterusnya, pada tahap ini dilakukan rangkaian normalisasi, yaitu mengubah kata-kata yang tidak baku atau singkatan berubah bentuk baku agar lebih konsisten dan mudah diproses. Tahap *preprocessing* data dilakukan untuk mendukung proses pelabelan dan klasifikasi. Dengan menghasilkan data yang lebih bersih dan terorganisir, kemungkinan terjadinya ketidakuratan yang dapat memengaruhi hasil pelabelan maupun klasifikasi dapat dikurangi.

3.2.1 Data Cleaning

Tahapan yang pertama dilakukan adalah Tahapan pertama dilakukan adalah proses *data cleaning*. Pada proses ini akan dilakukan penghapusan seperti emoticons, URL, *hashtag* dan karakter khusus . Hal ini bertujuan untuk mempermudah proses dan hasil dari klasifikasi. Pada tabel 2 merupakan contoh hasil data yang sebelum dilakukan *cleaning* dan sesudah dilakukan *cleaning*

Tabel 2. Hasil proses Data Cleaning

Sebelum <i>Cleaning</i>	Sesudah <i>Cleaning</i>
Setidak ya dia mengenalkan budaya Indonesia di mata dunia. Lebih bermanfaat meskipun dari negara lain. Daripada lu. Beban negara 🤔	Setidak ya dia mengenalkan budaya Indonesia di mata dunia Lebih bermanfaat meskipun dari negara lain Daripada lu Beban negara
Dateng ke indo bru sebentar sudah 16 juta view... luar biasa... Jika di malaysia mereka mengklaim batik akhirnya kamu tau dari mana batik itu berasal	Dateng ke indo bru sebentar sudah juta view luar biasa Jika di malaysia mereka mengklaim batik akhirnya kamu tau dari mana batik itu

Sebelum Cleaning	Sesudah Cleaning
yaitu dari negara dan budaya kami	berasal yaitu dari negara dan budaya kami

3.2.2 Case Folding

Tahapan setelah dilakukannya *data cleaning* dilanjutkan *case folding*. Merupakan langkah mengubah huruf data yang sudah dikumpulkan menjadi huruf kecil semua. Berikut contoh data sebelum dan sesudah dilakukan *case folding* dilihat pada tabel 3.

Tabel 3. Hasil Case Folding

Sebelum Case Folding	Sesudah Case Folding
Setidak ya dia mengenalkan budaya Indonesia di mata dunia Lebih bermanfaat meskipun dari negara lain Daripada lu Beban negara	setidak ya dia mengenalkan budaya Indonesia di mata dunia. lebih bermanfaat meskipun dari negara lain. dapada lu. beban negara
Dateng ke indo bru sebentar sudah juta view luar biasa Jika dimalaysia mereka mengklaim batik akhirnya kamu tau dari mana batik itu berasal yaitu dari negara dan budaya kami	dateng ke indo bru sebentar sudah juta view luar biasa jika dimalaysia mereka mengklaim batik akhirnya kamu tau dari mana batik itu berasal yaitu dari negara dan budaya kami

3.2.3 Stopwords Removal

Langkah berikutnya adalah melakukan proses *stopwords removal*, yakni langkah untuk menghilangkan kata-kata yang sering dipakai namun tidak ada pengaruh penting terhadap analisis. Dalam proses ini menerapkan pustaka `nltk.stopwords` dengan pilihan bahasa Indonesia. Proses ini dilakukan untuk menghapus kata kata penghubung yang ada pada dataset. Karena kata penghubung itu tidak dibutuhkan pada dataset. Berikut adalah contoh data sebelum dan sesudah dilakukan *stopwords removal* ditunjukkan pada tabel 4.

Tabel 4 Hasil Stopwords Removal

Sebelum Stopword Removal	Sesudah Stopwords Removal
setidak ya dia mengenalkan budaya Indonesia di mata dunia. lebih bermanfaat meskipun dari negara lain. dapada lu. beban negara	tidak ya kenal budaya indonesia mata dunia manfaat negara lu beban negara
dateng ke indo bru sebentar sudah juta view luar biasa jika dimalaysia mereka mengklaim batik akhirnya kamu tau dari mana batik itu berasal yaitu dari negara dan budaya kami	dateng indo bru bentar sudah juta view luar biasa jika Malaysia mereka klaim batik akhir kamu tau dari mana batik itu asal yaitu dari negara budaya kami

3.2.4 Tokenizing

Tahap ini termasuk salah satu tahapan penting dalam proses *preprocessing*, di mana teks kalimat dipecah menjadi potongan-potongan kata yang disebut token. Berikut contoh sebelum dan sesudah *tokenizing* dilakukan pada tabel 5.

Tabel 5. Hasil Tokenizing

Sebelum <i>Tokenizing</i>	Sesudah <i>Tokenizing</i>
tidak ya kenal budaya indonesia mata dunia manfaat negara lu beban negara	['tidak', 'ya', 'kenal', 'budaya', 'indonesia', 'mata', 'dunia', 'manfaat', 'negara', 'lu', 'beban', 'negara']
dateng indo bru bentar sudah juta view luar biasa jika malaysia mereka klaim batik akhir kamu tau dari mana batik itu asal yaitu dari negara buday kami	['dateng', 'indo', 'bru', 'bentar', 'sudah', 'juta', 'view', 'luar', 'biasa', 'jika', 'malaysia', 'mereka', 'klaim', 'batik', 'akhir', 'kamu', 'tau', 'dari', 'mana', 'batik', 'itu', 'asal', 'yaitu', 'dari', 'negara', 'buday', 'kami']

3.2.5 Normalized

Selanjut dilakukan proses *normalized*, proses terakhir yang ada pada *preprocessing*. Tahapan ini bertujuan guna mengonversi kata-kata tidak baku berubah bentuk baku dengan memanfaatkan beberapa sumber referensi, seperti kamus *colloquial-indonesian-lexicon*, kamus alay, KBBI daring, serta *library* Python kbbs. Berikut contoh hasil proses normalisasi dapat ditunjukkan pada tabel 6.

Tabel 6. Hasil Normalized

Sebelum <i>Normalized</i>	Sesudah <i>Normalized</i>
tidak ya kenal budaya indonesia mata dunia manfaat negara lu beban negara	tidak ya kenal budaya indonesia mata dunia manfaat negara lu beban negara
dateng indo bru bentar sudah juta view luar biasa jika malaysia mereka klaim batik akhir kamu tau dari mana batik itu asal yaitu dari negara buday kami	dateng indo baru bentar sudah juta view luar biasa jika malaysia mereka klaim batik akhir kamu tau dari mana batik itu asal yaitu dari negara buday kami

3.3 Pelabelan data

Setelah tahap *preprocessing*, proses dilanjutkan dengan pelabelan data yang bertujuan untuk mengelompokkan data ke dalam tiga kategori utama: positif, negatif, dan netral. Pelabelan dilakukan secara otomatis menggunakan kamus *lexicon-based*. Proses diawali dengan mengidentifikasi kata-kata yang paling sering muncul dalam dataset menggunakan kode Python. Setelah itu, nilai sentimen dihitung berdasarkan skor dari kamus dengan memeriksa setiap kata; jika kata ditemukan dalam kamus, nilainya ditambahkan, sedangkan jika tidak ditemukan, sistem mencoba menafsirkan artinya melalui konteks sekitarnya.

Jika terdapat kata baru yang belum ada dalam kamus, maka matriks akan diperluas; jika tidak, matriks yang ada diperbarui. Verifikasi kata dilakukan untuk menentukan nilai sentimen berdasarkan frekuensi kemunculan dan kekuatan asosiasinya terhadap kategori sentimen. Nilai sentimen akhir untuk setiap komentar dihitung dari total skor kata-kata dalam kalimat tersebut. Hasil akhir menunjukkan klasifikasi komentar ke dalam sentimen positif, negatif, atau netral, berikut hasil dari perhitungan matriksnya, di mana setiap kata yang telah diberi nilai berdasarkan kamus akan dihitung total nilai sentiment dalam satu kalimat, yang ditunjukkan pada gambar 3.

	dingin	awal	biadab	tenang	cepat	aneh	juta	pandang	suka	gila	...	nala	partisipasi	griya	prima	moral	konferensi	bidah	yukk	basi	sentiment	
0	1	1	2	1	1	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0	20
1	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	1	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0	4
4	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	1	1	0	0	...	0	0	0	0	0	0	0	0	0	0	3
6	0	0	0	0	0	0	1	0	1	1	...	0	0	0	0	0	0	0	0	0	0	8
7	0	0	0	0	0	0	1	0	0	1	...	0	0	0	0	0	0	0	0	0	0	8
8	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0	-5
9	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0	14

10 rows x 1781 columns

Gambar 3. Hasil Perhitungan Matriks

3.4 Pembobotan TF-IDF

Proses dimulai dengan menghitung *term frequency* (TF) berdasarkan jumlah kemunculan setiap kata. Hasil perhitungan TF kemudian melalui tahap tokenisasi untuk diproses lebih lanjut ke dalam perhitungan *inverse document frequency* (IDF). Setelah nilai IDF diperoleh, langkah berikutnya adalah mengalikan TF dengan IDF untuk mendapatkan nilai TF-IDF. Perhitungan bobot IDF dilakukan dengan menjumlahkan jumlah dokumen yang mengandung istilah tertentu, lalu digunakan untuk menghitung TF-IDF guna mengetahui bobot pentingnya suatu kata dalam keseluruhan dokumen. yang hasil prosesnya dapat dilihat pada gambar 4.

	term	TF	TF-IDF
0	0	{"kerumun": 0.07692307692307693, "dingin":...	{"kerumun": 0.40514325267274703, "dingin":...
1	1	{"tidak": 0.5, "hormat": 0.5}	{"tidak": 1.3904034348769188, "hormat": 2....
2	2	{"tidak": 0.3333333333333333, "sudam": 0.3...	{"tidak": 0.9269356232512791, "sudam": 3.0...
3	3	{"tshirt": 0.5, "aneh": 0.5}	{"tshirt": 4.225918278969115, "aneh": 3.24...
4	4	{"alir": 0.5, "sejarah": 0.5}	{"alir": 1.798602756159278, "sejarah": 2.8...
...
28097	28097	{"zoren": 0.3333333333333333, "simpan": 0....	{"zoren": 3.1834829488687797, "simpan": 2....
28098	28098	{"zoro": 0.3333333333333333, "kuat": 0.333...	{"zoro": 2.7658952927036573, "kuat": 1.702...
28099	28099	{"zpechojrbxtl": 0.5, "henti": 0.5}	{"zpechojrbxtl": 4.77522442330317, "henti":...
28100	28100	{"zulfazora": 0.14285714285714285, "lihat"...	{"zulfazora": 1.3643498352294772, "lihat":...
28101	28101	{"zw": 1.0}	{"zw": 9.55044884660634}

Gambar 4. Proses pembobotan TF-IDF

3.5 Implementasi Algoritma SVM

Pada proses ini digunakan algoritma *Support Vector Machine* (SVM) dengan kernel linear. Tahap awal dimulai dengan membagi dataset menjadi dua bagian, yaitu 80% untuk data latih (*training*) dan 20% untuk data uji (*testing*), menghasilkan 22.468 data latih dan 5.617 data uji.. Dapat dilihat pada gambar 5

```
len(X_test) #jumlah data test yg diuji/test
5617

len(X_train) #jumlah data latih
22468
```

Gambar 5. Pembagian Data Latih dan Data Uji

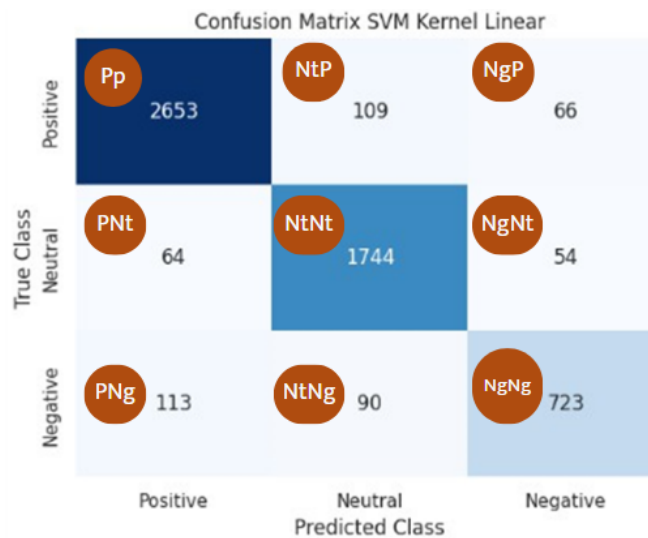
Selanjutnya, diterapkan *TF-IDF Vectorizer* untuk mengubah data teks menjadi representasi numerik, yang digunakan dalam proses klasifikasi. Algoritma SVM kemudian dijalankan dengan kernel linear, yang dipilih karena menghasilkan performa terbaik dibandingkan kernel lain seperti RBF. Prosesnya dapat dilihat pada gambar 6.

```
#TF-IDF vectorizer
vectorizer = TfidfVectorizer()
X_train_tfidf = vectorizer.fit_transform(X_train)
X_test_tfidf = vectorizer.transform(X_test)

# Klasifikasi SVM dengan kernel linear
svm = SVC(kernel='linear')
svm.fit(X_train_tfidf, y_train)
```

Gambar 6. TF-IDF Vectorizer dan Klasifikasi SVM dengan Kernel Linear

Hasil pengujian menunjukkan klasifikasi sentimen yang mencakup beberapa nilai penting: *Positive-Positive* (2653), *Neutral-Positive* (109), *Negative-Positive* (66), *Positive-Neutral* (64), *Neutral-Neutral* (1744), *Negative-Neutral* (54), *Positive-Negative* (113), *Neutral-Negative* (90), dan *Negative-Negative* (723). Hasil ini menggambarkan performa model dalam mengklasifikasikan komentar ke dalam kategori sentimen yang sesuai. Dapat dilihat pada gambar 7.



Gambar 7. Confusion Matrix

Tahap akhir dari implementasi algoritma adalah melakukan pengujian. Dari hasil pengujian tersebut, algoritma *Support Vector Machine* (SVM) menggunakan kernel linear berhasil memperoleh tingkat akurasi sebesar 91,1%. Dapat dilihat pada gambar 8.

```
accuracy = accuracy_score(y_test_cleaned, y_pred)
print("Akurasi:", accuracy)
```

Akurasi: 0.9116809116809117

Gambar 8. Hasil Akurasi Algoritma SVM

3.6 Pengujian Evaluasi

Bagian akhir dari penelitian ini dilakukan dengan menguji serta mengevaluasi data uji yang telah disiapkan. Proses ini bertujuan untuk memperoleh nilai *precision*, *recall*, dan *f1-score* dari klasifikasi sentimen, baik positif, netral maupun negatif.

	precision	recall	f1-score	support
-1.0	0.86	0.78	0.82	926
0.0	0.90	0.94	0.92	1862
1.0	0.94	0.94	0.94	2828
accuracy			0.91	5616
macro avg	0.90	0.89	0.89	5616
weighted avg	0.91	0.91	0.91	5616

Gambar 9. Hasil Evaluasi Algoritma SVM

Gambar 9. menampilkan hasil evaluasi performa model klasifikasi sentimen dengan menggunakan tiga metrik utama, yaitu *precision*, *recall*, dan *f1-score* untuk masing-masing kategori polaritas: negatif (-1.0), netral (0.0), dan positif (1.0). Berdasarkan hasil evaluasi, model memberikan performa terbaik pada kelas sentimen positif dan netral, yang masing-masing mencatatkan nilai *f1-score* sebesar 0.94 dan 0.92. Sebaliknya, performa model pada kelas sentimen negatif terlihat lebih rendah, dengan *f1-score* hanya mencapai 0.82, yang disebabkan oleh nilai *recall* yang relatif kecil, yaitu 0.78. Secara keseluruhan, akurasi model mencapai 91%. Artinya, sebanyak 91% dari total prediksi yang dihasilkan oleh model sesuai dengan label sebenarnya, menunjukkan bahwa model cukup andal dalam mengklasifikasikan data sentiment.

Pembahasan

Hasil penelitian ini menunjukkan bahwa metode *Support Vector Machine* (SVM) dengan kernel linear cukup andal dalam mengenali sentimen pada komentar YouTube. Model ini berhasil mencapai tingkat akurasi tinggi, yaitu 91,1%, yang berarti sebagian besar prediksi sistem sesuai dengan kenyataan. Temuan ini membuktikan bahwa SVM sangat cocok digunakan untuk menganalisis teks berbahasa Indonesia yang beragam, seperti komentar di media sosial yang sering menggunakan bahasa campuran dan gaya tidak formal.

Secara keseluruhan, hasil analisis memperlihatkan bahwa komentar positif lebih banyak dibandingkan komentar lainnya, dengan persentase sekitar 49% dari total data. Artinya, mayoritas masyarakat Indonesia memberikan respon positif terhadap kedatangan IShowSpeed ke Indonesia. Sebagian besar komentar positif berisi pujian dan rasa bangga karena IShowSpeed menunjukkan ketertarikan pada budaya Indonesia dan berinteraksi dengan masyarakat lokal. Hal ini menggambarkan bahwa penonton di Indonesia cenderung menyambut baik kreator asing yang menghargai budaya mereka.

Selanjutnya, komentar netral berada di posisi kedua, sekitar 30%. Komentar jenis ini biasanya hanya berupa tanggapan ringan, candaan, atau sekadar obrolan antar pengguna yang tidak menunjukkan emosi tertentu. Sedangkan komentar negatif, yang jumlahnya sekitar 21%, sebagian besar muncul karena isu sensitif, terutama terkait perdebatan budaya antara Indonesia dan Malaysia. Fenomena ini menunjukkan bahwa topik yang menyangkut identitas nasional masih bisa memicu perbedaan pendapat di ruang digital.

Dari sisi performa sistem, model bekerja sangat baik pada kategori positif dan netral, dengan nilai *precision* dan *recall* di atas 0,90. Artinya, sistem mampu mengenali dan mengklasifikasikan dua jenis komentar tersebut dengan akurat. Namun, untuk komentar negatif, nilai *recall*-nya sedikit lebih rendah, yaitu 0,78. Ini berarti masih ada beberapa komentar negatif yang tidak terdeteksi dengan benar. Salah satu penyebabnya adalah penggunaan bahasa yang sarkastik, bercampur bahasa Inggris, atau gaya bercanda khas media sosial, yang sulit dipahami oleh model berbasis kata. Hal ini juga ditemukan dalam penelitian sebelumnya oleh Mualfah et al. (2023) dan Wijayanto (2022), di mana bahasa tidak baku menjadi tantangan utama dalam analisis sentimen otomatis.

Selain itu, hasil penelitian ini menguatkan temuan-temuan sebelumnya bahwa kombinasi metode SVM dan pembobotan TF-IDF masih menjadi pilihan yang efektif untuk analisis teks pendek di media sosial. Metode ini mampu memberikan hasil stabil tanpa perlu penyetelan parameter yang rumit seperti pada model jaringan saraf (*neural network*).

Dari sisi sosial, penelitian ini memberikan gambaran bahwa interaksi lintas budaya di dunia digital bisa menghasilkan tanggapan positif apabila dilakukan dengan rasa hormat dan ketulusan. Respon publik Indonesia yang cenderung positif terhadap IShowSpeed mencerminkan sikap terbuka masyarakat terhadap figur internasional yang menghargai budaya lokal. Meski demikian, munculnya sebagian komentar negatif juga menjadi pengingat bahwa faktor budaya tetap sensitif, dan perlu diperhatikan oleh kreator global saat berinteraksi di ruang digital lintas negara.

4. Kesimpulan

Berdasarkan hasil penelitian, dari total 28.068 data yang diperoleh melalui proses *crawling YouTube*, ditemukan bahwa 13.846 data termasuk dalam sentimen positif, 8.944 sentimen negatif, dan 5.278 sentimen netral. Implementasi algoritma *Support Vector Machine* (SVM) dengan kernel linear menunjukkan akurasi sebesar 91,1%. Pada tahap evaluasi, performa untuk sentimen negatif menghasilkan *precision* 86%, *recall* 78%, dan *f1-score* 82%, sedangkan sentimen netral mencapai *precision* 90%, *recall* 94%, dan *f1-score* 92%. Untuk sentimen positif, *precision*, *recall*, dan *f1-score* masing-masing sebesar 94%. Dari hasil tersebut, dapat disimpulkan bahwa sentimen pengguna *YouTube* terhadap video IShowSpeed *Tour* Indonesia cenderung bersifat positif.

Penelitian ini masih memiliki beberapa keterbatasan, khususnya dalam menangani komentar dengan konteks bahasa campuran (*code-mixing*), penggunaan sarkasme, serta gaya bahasa informal khas media sosial. Oleh karena itu, untuk penelitian selanjutnya disarankan agar menggunakan metode *deep learning* seperti LSTM, Bi-LSTM, atau BERT yang mampu memahami konteks semantik lebih mendalam dibandingkan SVM. Menambah variasi data dari berbagai platform media sosial lain (misalnya X/Twitter, TikTok, atau Instagram) untuk memperoleh gambaran yang lebih luas tentang persepsi publik terhadap influencer internasional. Mengembangkan sistem analisis sentimen otomatis berbasis real-time, yang dapat digunakan oleh kreator konten dan lembaga riset untuk memantau opini publik secara langsung.

Daftar Pustaka

Anisah, S., & Irwansyah, I. (2025). Analisis Data Mining untuk Klasifikasi Kafe Populer di Jakarta Menggunakan Decision Tree dan Visualisasi dengan Tableau. *Jutisi : Jurnal Ilmiah Teknik Informatika Dan Sistem Informasi*, 14(2), 890. <https://doi.org/10.35889/jutisi.v14i2.2660>

- Bagaskara, W., Pusparini, N. N., & Irwansyah, I. (2024). Klasifikasi Penjadwalan Kerja Perawatan Air Conditioner (AC) Menggunakan Algoritma Decision Tree (C4.5) Pada PT Xyz. *Infotech: Journal of Technology Information*, 10(1), 11–20. <https://doi.org/10.37365/jti.v10i1.240>
- Darwis, D., Pratiwi, E. S., & Pasaribu, A. F. O. (2020). Penerapan Algoritma Svm Untuk Analisis Sentimen Pada Data Twitter Komisi Pemberantasan Korupsi Republik Indonesia. *Edutic - Scientific Journal of Informatics Education*, 7(1), 1–11. <https://doi.org/10.21107/edutic.v7i1.8779>
- Eldo, H., Ayuliana, A., Suryadi, D., Chrisnawati, G., & Judijanto, L. (2024). Penggunaan Algoritma Support Vector Machine (SVM) Untuk Deteksi Penipuan pada Transaksi Online. *Jurnal Minfo Polgan*, 13(2), 1627–1632. <https://doi.org/10.33395/jmp.v13i2.14186>
- Hermawan, A., Jowensen, I., Junaedi, J., & Edy. (2023). Implementasi Text-Mining untuk Analisis Sentimen pada Twitter dengan Algoritma Support Vector Machine. *JST (Jurnal Sains Dan Teknologi)*, 12(1), 129–137. <https://doi.org/10.23887/jstundiksha.v12i1.52358>
- Husada, H. C., & Paramita, A. S. (2021). Analisis Sentimen Pada Maskapai Penerbangan di Platform Twitter Menggunakan Algoritma Support Vector Machine (SVM). *Teknika*, 10(1), 18–26. <https://doi.org/10.34148/teknika.v10i1.311>
- Irwansyah, I., Dittyata, R., Rizal, R., & Wiyono, W. (2024). Optimalisasi Klasifikasi Uji Emisi Sepeda Motor Menggunakan Algoritma Naïve Bayes. *Infotech: Journal of Technology Information*, 10(2), 337–242. <https://doi.org/10.37365/jti.v10i2.327>
- Irwansyah, I., Wiranata, A. D., & M, T. T. (2023). Komparasi Algoritma Decision Tree, Naive Bayes Dan K-Nearest Neighbor Untuk Menentukan Kualitas Udara Di Provinsi Dki Jakarta. *Infotech: Journal of Technology Information*, 9(2), 193–198. <https://doi.org/10.37365/jti.v9i2.203>
- Irwansyah, I., Wiranata, A. D., Muryono, T. T., & Budiyantra, A. (2022). Sistem Pakar Deteksi Kerusakan Jaringan Local Area Network (Lan) Menggunakan Metode Beckward Chaining Berbasis Web. *Infotech: Journal of Technology Information*, 8(2), 135–142. <https://doi.org/10.37365/jti.v8i2.150>
- Mualfah, D., Ramadhoni, Gunawan, R., & Mulyadipa Suratno, D. (2023). Analisis Sentimen Komentar YouTube TvOne Tentang Ustadz Abdul Somad Dideportasi Dari Singapura Menggunakan Algoritma SVM. *Jurnal Fasilkom*, 13(01), 72–80. <https://doi.org/10.37859/jf.v13i01.4920>
- Muryono, T. T., & Irwansyah, I. (2020). Implementasi Data Mining Untuk Menentukan Kelayakan Pemberian Kredit Dengan Menggunakan Algoritma K-Nearest Neighbors (K-NN). *Infotech: Journal of Technology Information*, 6(1), 43–48. <https://doi.org/10.37365/jti.v6i1.78>
- Muryono, T. T., Taufik, A., & Irwansyah, I. (2021). Perbandingan Algoritma K-Nearest Neighbor, Decision Tree, Dan Naive Bayes Untuk Menentukan Kelayakan Pemberian Kredit. *Infotech: Journal of Technology Information*, 7(1), 35–40. <https://doi.org/10.37365/jti.v7i1.104>
- Oktaviana, N. E., Sari, Y. A., & Indriati. (2022). Analisis Sentimen Terhadap Kebijakan Kuliah Daring Selama Pandemi Menggunakan Pendekatan Lexicon Based Features Dan Support Vector Machine. *Jurnal Teknologi Informasi Dan Ilmu Komputer (JTIK)*, 9(2), 357–362. <https://doi.org/10.25126/jtiik.202295625>

- Wijayanto, A. (2022). Analisis Sentimen Komentar Youtube Mengenai Vaksin Covid-19 Menggunakan Support Vector Machine. *Jurnal Pilar Teknologi Jurnal Ilmiah Ilmu Ilmu Teknik*, 7(1), 24–31. <https://doi.org/10.33319/piltek.v7i1.118>
- Wiranata, A. D., Soleman, S., Irwansyah, I., Sudaryana, I. K., & Rizal, R. (2023). Klasifikasi Data Mining Untuk Menentukan Kualitas Udara Di Provinsi Dki Jakarta Menggunakan Algoritma K-Nearest Neighbors (K-NN). *Infotech: Journal of Technology Information*, 9(1), 95–100. <https://doi.org/10.37365/jti.v9i1.164>