



# Analisis Komentar Youtube Terhadap Polemik Ijazah Presiden Ke 7 Indonesia Menggunakan Support Vector Machine

Ignasius Aditya Anggoro Putra\*, Salmon, Kusnandar

Program Studi Sistem Informasi, STMIK Widya Cipta Dharma, Samarinda, Indonesia

Email: <sup>1</sup>\*2041901@wicida.ac.id, <sup>2</sup>salmon@wicida.ac.id, <sup>3</sup>kusnandar@wicida.ac.id

Email Penulis Korespondensi: 2041901@wicida.ac.id

**Abstrak**—Penelitian ini bertujuan menganalisis sentimen masyarakat terhadap polemik ijazah Presiden Joko Widodo melalui komentar pada platform YouTube. Sebanyak 20.294 komentar dikumpulkan dan diproses melalui tahapan pembersihan teks, normalisasi, tokenisasi, penghapusan stopword, serta stemming. Pelabelan sentimen dilakukan dengan pendekatan lexicon-based yang menghasilkan tiga kategori, yaitu positif, negatif, dan netral. Hasil pengujian menunjukkan bahwa kombinasi SVM, TF-IDF, dan SMOTE mampu menghasilkan performa klasifikasi yang baik dengan tingkat akurasi sebesar 86,87%. Model menunjukkan kemampuan yang lebih optimal dalam mengklasifikasikan sentimen negatif dan netral, meskipun sebagian sentimen positif masih cenderung diklasifikasikan sebagai netral. Secara keseluruhan, hasil penelitian ini menunjukkan bahwa analisis sentimen berbasis komentar YouTube dapat digunakan sebagai pendekatan yang efektif dalam memetakan opini publik terhadap isu sosial-politik secara cepat dan berskala luas. Ekstraksi fitur dilakukan menggunakan *Term Frequency–Inverse Document Frequency* (TF-IDF), sedangkan proses klasifikasi menggunakan algoritma Support Vector Machine (SVM). Hasil evaluasi menunjukkan akurasi sebesar 86,87% dan macro F1-score 0,87, yang mengindikasikan bahwa kombinasi TF-IDF, SMOTE, dan SVM efektif dalam mengklasifikasikan sentimen komentar YouTube terkait isu sosial-politik.

**Kata Kunci:** Analisis Sentimen; Joko Widodo; SVM; TF-IDF; YouTube

**Abstract**—This study aims to analyze public sentiment toward the controversy surrounding President Joko Widodo's academic credentials by examining user comments on YouTube. A total of 20,294 comments were collected and processed through text cleaning, normalization, tokenization, stopword removal, and stemming. Sentiment labels were assigned using a lexicon-based approach, producing positive, negative, and neutral categories. The experimental results indicate that the combination of SVM, TF-IDF, and SMOTE achieved strong classification performance, with an accuracy of 86.87%. The model demonstrated better performance in identifying negative and neutral sentiments, while some positive sentiments tended to be misclassified as neutral. Overall, this study shows that sentiment analysis based on YouTube comments can serve as an effective approach for mapping public opinion on socio-political issues in an automated and large-scale manner. Feature extraction utilized Term Frequency–Inverse Document Frequency (TF-IDF), and sentiment classification was performed using a Support Vector Machine (SVM). The model achieved an accuracy of 86.87% and a macro F1-score of 0.87, indicating that the integration of TF-IDF, SMOTE, and SVM is effective for large-scale sentiment classification of YouTube comments related to socio-political issues.

**Keywords:** Sentiment Analysis; Joko Widodo; SVM; TF-IDF; YouTube Comments

## 1. PENDAHULUAN

Tingginya minat masyarakat dalam menggunakan *YouTube* telah membuat *platform* ini semakin diminati oleh berbagai kalangan. *YouTube* menempati posisi teratas sebagai *platform* media sosial yang paling populer di Indonesia. Sebagai salah satu *platform streaming* video terkemuka di dunia, *youtube* menawarkan berbagai macam konten hiburan dan pendidikan. Selain itu, *youtube* juga dimanfaatkan sebagai media untuk menyebarkan berita dan informasi[1]. Dalam berkomunikasi media sosial sebagai sarana utama komunikasi publik telah mengubah cara mengekspresikan opini dan tanggapan terhadap isu-isu publik yang dinarasikan oleh berbagai macam *platform* media. Isu mengenai polemik ijazah Presiden ke-7 Indonesia menjadi salah satu topik yang sangat ramai diperbincangkan oleh banyak orang dan berbagai macam media dan *platform*, sehingga komentar–komentar yang diberikan dan komentar yang muncul mencerminkan opini publik yang sangat beragam.

Reaksi beragam yang dihadirkan dari masyarakat luas melalui media sosial dapat dengan mudah dijumpai dalam kolom komentar maupun di postingan. Kebebasan berpendapat menjadi salah satu aspek yang mendasari pengguna media sosial sebagai tempat untuk menurkan suatu opini. Tidak jarang ditemukan tuturan yang bersifat sindiran bahkan umpatan kasar di laman komentar maupun di dalam postingan. Hal ini menunjukkan bahwa komentar yang ditinggalkan oleh penonton memiliki peran penting dalam membentuk persepsi opini. Namun, tantangan dalam menganalisis komentar pengguna di *YouTube* sangatlah kompleks. Komentar sering kali ditulis dalam bahasa informal, mengandung 1 *Random Forest*, berdasarkan nilai akurasi dengan evaluasi menggunakan *Confusion Matrix*. Algoritma *Support Vector Machine* (SVM) memberikan performa klasifikasi yang lebih unggul dibandingkan *Random Forest*. SVM mencapai akurasi sebesar 97% dengan f1-score yang tinggi dan seimbang pada kedua kelas, yaitu 0.98 untuk kelas 0 dan 0.95 untuk kelas 1. Di sisi lain, *Random Forest* hanya mencapai akurasi 95%, dengan penurunan kinerja terutama pada kelas 1 yang memiliki f1-score 0.91, menunjukkan ketidakseimbangan dalam mendeteksi kelas minoritas. Dengan demikian, SVM terbukti lebih andal dan konsisten dalam menangani data sentimen, khususnya pada kasus dengan distribusi kelas yang tidak merata[3].

Kedua, penelitian oleh Zulqarnain, Mochammad 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 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[4].

Ketiga, penelitian oleh Dirgahayu Marganingsih, Hardian Oktavianto, Ginanjar Abdurrahman pada tahun 2025 dengan judul "Analisis Sentimen Komentar Youtube Masterchef Indonesia Menggunakan Algoritma *Support Vector Machine* Dan *Gaussian Naïve Bayes*". Penelitian ini membandingkan dua algoritma *machine learning*, yaitu *Support Vector Machine* dan *Gaussian Naïve Bayes*. Data yang digunakan sebanyak 2000 komentar Youtube dengan masing-masing komentar positif memperoleh 825 data yang berisi dukungan serta ucapan terhadap peserta dan komentar negatif sebanyak 1175 data yang berisi kekecewaan terhadap juri serta *channel* Masterchef Indonesia. Algoritma *Gaussian Naïve Bayes* menunjukkan performa terbaik berdasarkan data normal pada pengujian *K-fold* 10 langkah Uji 6 dengan akurasi tertinggi sebesar 65%, presisi sebesar 52% dan *recall* sebesar 81%. Kemudian dengan algoritma *Support Vector Machine* didapatkan nilai akurasi tertinggi sebesar 82% pada *K-fold* 10 Langkah Uji 10 dan akurasi terendah sebesar 71%, presisi sebesar 88% dan *recall* sebesar 72% pada *K-fold* 10 Langkah Uji 7. Algoritma *Gaussian Naïve Bayes* menunjukkan performa terbaik berdasarkan *Balancing Data Gaussian Naïve Bayes* pada pengujian *K-Fold* 10 langkah uji 8 dengan akurasi 72%, presisi 68%, *recall* 93%. Kemudian *Balancing Data Support Vector Machine* didapatkan nilai akurasi 85%, presisi 84% dan *recall* 89% pada *K-Fold* 10 langkah uji 6[5].

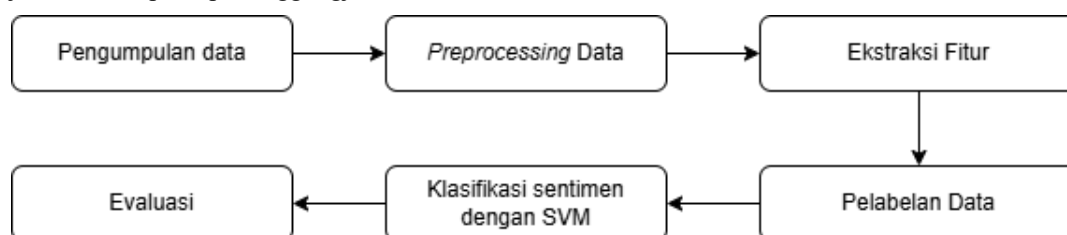
Analisis sentimen merupakan suatu teknik yang digunakan untuk mengekstraksi dan mengolah data opini secara otomatis dengan tujuan memahami makna serta kecenderungan sentimen yang terkandung dalam teks. Teknik ini berfokus pada identifikasi polaritas opini, yang umumnya diklasifikasikan ke dalam sentimen positif atau negatif. Dalam perkembangannya, analisis sentimen dapat dilakukan melalui berbagai pendekatan, di antaranya metode klasifikasi konvensional dan metode pembelajaran mesin, termasuk pendekatan berbasis aturan (*rule-based*). Pada teknik pembelajaran mesin, kata-kata yang mengandung muatan emosional dimanfaatkan sebagai fitur dalam proses klasifikasi. Pemilihan dan penentuan kata-kata emosional tersebut dapat dilakukan secara efektif dan efisien dengan memanfaatkan kamus sentimen, sehingga membantu meningkatkan akurasi dalam mengidentifikasi kecenderungan opini yang terdapat dalam data teks[6].

*Support Vector Machine* merupakan salah satu algoritma yang dikenal efektif dalam menangani berbagai permasalahan klasifikasi[7]. *Support Vector Machine* adalah salah satu algoritma *machine learning* yang banyak digunakan untuk mengklasifikasi suatu topik karena kemampuan dalam menangani data berdimensi tinggi dan menghasilkan margin 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 isu polemik ijazah.

## 2. METODOLOGI PENELITIAN

### 2.1 Tahapan Penelitian

Pada penelitian ini algoritma *Support Vector Machine* digunakan sebagai pengklasifikasian untuk mengukur serta menganalisis komentar pengguna *YouTube*. Metode penelitian analisis sentimen berupa komentar pada *Youtube* menggunakan algoritma *Support Vector Machine* merupakan serangkaian alur/tahapan kerja penelitian yang terstruktur secara garis besar yang menjadi pedoman penelitian secara sistematis agar penelitian tersebut tercapai. Tahapan penelitian meliputi proses pengumpulan data, pemrosesan awal data (*pre-processing*), ekstraksi fitur, pelabelan data, proses klasifikasi menggunakan algoritma SVM, serta evaluasi kinerja model. Setiap tahapan saling berkaitan dan berperan penting dalam menghasilkan analisis sentimen yang akurat. Proses pemrosesan dan ekstraksi fitur bertujuan untuk mengubah data teks menjadi representasi numerik yang sesuai dengan kebutuhan algoritma. Selanjutnya, tahap evaluasi digunakan untuk menilai tingkat akurasi dan efektivitas model dalam mengklasifikasikan sentimen komentar ke dalam kategori positif, negatif, dan netral. Rangkaian tahapan tersebut dirancang untuk menghasilkan analisis sentimen yang akurat, objektif, dan dapat dipertanggungjawabkan secara akademik.



Gambar 1. Tahap Penelitian



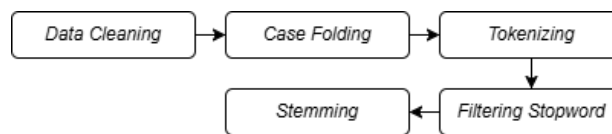
Data yang telah diolah dan diseimbangkan selanjutnya dipartisi menjadi data pelatihan dan data pengujian. Model SVM dilatihkan dengan data pelatihan untuk mengkategorikan sentimen menjadi positif, negatif, dan netral lalu dievaluasi dengan data pengujian menggunakan matrik presisi, *recall*, *F1-Score*, dan akurasi[8].

#### a. Pengumpulan Data

Data dikumpulkan dari komentar-komentar pada *link* atau alamat situs konten video di kanal media nasional *Youtube* Indonesia *Lawyers Club* menggunakan izin layanan dari *Youtube Data API V3*, untuk mengakses atribut data statistik dari video tersebut seperti tanggal komentar, nama pengguna akun, isi komentar, dan jumlah *like*. Proses pengumpulan data dilakukan menggunakan platform *Google Colab* untuk *crawling* data. Pada proses pelabelan, dilakukan pengkategorian data tersebut ke dalam sentimen positif, sentimen negatif dan sentimen netral[9]. Jumlah data komentar pada *link* yang telah didapat cukup banyak yaitu 20799 baris data yang belum memiliki label. Pengumpulan data dilakukan melalui proses *web scraping* secara mandiri menggunakan *YouTube Data API v3* sebagai media utama untuk pengambilan data, sedangkan Python beserta pustaka pendukung seperti *Pandas*, *NumPy*, dan *Sastrawi* digunakan untuk proses pengolahan dan pemodelan. Hasil dari pengambilan data (*scraping*) ini dalam bentuk *csv* yang akan melewati tahap *preprocessing*[3]. Dataset tersebut tersedia dalam format *.csv* dan *excel* dengan nama file *skripsi-ijazahPalsuPakJoko.csv/.xlsx* dan berisi total 20294 entri. Setiap entri terdiri dari beberapa atribut utama, seperti *published*, *user id*, *username*, *Profile URL*, *Avatar URL*, *comment* dan *likes*. Umumnya, dalam *web scraping* ada dua cara untuk proses pengambilan datanya, yaitu dengan cara manual dan otomatis[10].

#### b. Preprocessing Data

*Text preprocessing* adalah pembersihan, dan penyederhanaan teks agar dapat diproses lebih lanjut. Data komentar yang telah diperoleh merupakan *raw data* atau data mentah, yang perlu dilakukan tahapan *text preprocessing* terlebih dahulu. Tahapan *cleaning* menemukan dan menghapus objek karakter dalam teks yang tidak diperlukan seperti angka, simbol dan tanda baca, hashtag #, link URL, *emoticon* atau ekspresi wajah, dan karakter lainnya. Tahapan *case folding* merubah semua karakter yang terdeteksi terdapat huruf kapital akan menjadi huruf kecil. Tahapan *tokenization* bertugas memisahkan setiap kata dalam kalimat untuk masing-masing baris data komentar, kemudian menyatukannya kembali untuk diproses pada tahap *preprocessing* teks berikutnya. Tahapan *stopword* menemukan dan menghapus kata-kata yang tidak memiliki makna penting dalam sebuah kalimat, seperti kata penghubung, kata keterangan, kata ganti orang atau benda, dan kata depan. Tahapan *stemming* menghilangkan imbuhan pada kata yang memiliki awalan, akhiran, serta gabungan awalan dan akhiran.



Gambar 2. Tahap *preprocessing*

#### c. Ekstraksi Fitur

Setelah *preprocessing*, fitur-fitur penting dari teks komentar diekstraksi menggunakan metode *TF-IDF (Term Frequency-Inverse Document Frequency)* digunakan untuk mengukur seberapa penting kata dalam sebuah dokumen dibandingkan dengan kumpulan dokumen. Nilai tinggi *TF-IDF* menunjukkan bahwa kata tersebut memiliki signifikansi tinggi dalam dokumen[11].

#### d. Pelabelan Data

Proses ini dilakukan secara otomatis menggunakan *Google Colab Python* dengan menggunakan *coding* pada setiap data dan memberikan label sesuai dengan sentimennya, baik itu positif, negatif, atau netral. Pada penelitian ini, daftar kata positif dan negatif diperoleh dari repositori *GitHub* yang menyediakan kumpulan *lexicon* Bahasa Indonesia dalam format *.tsv*. 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 kekuatannya. Kumpulan kata positif dan negatif yang diambil dari *GitHub* (*Fajri Koto*) kamus *lexicon* dan digunakan sebagai acuan untuk menentukan nilai sentimen positif, negatif dan netral dalam data komentar. Langkah ini memiliki peran penting dalam pelatihan model nantinya karena membantu dalam mengklasifikasikan data berdasarkan sentimennya.

#### e. Klasifikasi dengan *Support Vector Machine*

Data yang telah diproses dan ekstraksi fiturnya akan diklasifikasikan menggunakan metode *Support Vector Machine*. Secara fundamental, *Support Vector Machine* beroperasi berdasarkan prinsip *linear classifier*, yaitu memisahkan kelas-kelas data yang dapat dipisahkan secara linier. Namun, seiring perkembangannya, *Support Vector Machine* telah diperluas sehingga mampu menangani permasalahan yang bersifat non-linier melalui penggunaan fungsi kernel pada ruang berdimensi tinggi. Problem *non-linier* dengan memasukkan konsep kernel pada ruang kerja berdimensi tinggi. Pada dasarnya, metode ini bekerja dengan cara mendefinisikan batas antara dua kelas dengan jarak maksimal dari data yang terdekat. Klasifikasi ini bertujuan untuk mencari keputusan yang terbaik ke dalam sentimen positif dan negatif, untuk selanjutnya data akan di training. Banyak penelitian telah menunjukkan bahwa *Support Vector Machine* adalah metode yang paling akurat untuk klasifikasi teks. Alur proses klasifikasi secara keseluruhan terdapat pada penelitian analisis sentimen ini diuraikan sebagai berikut.

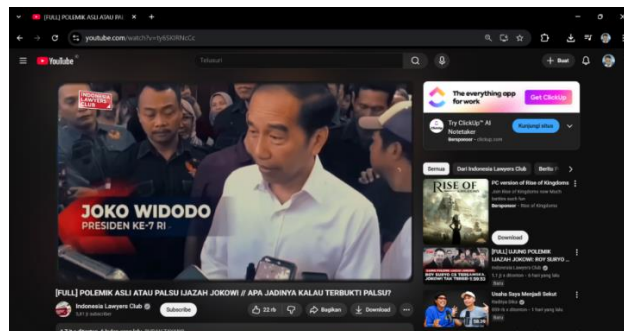
1. Mempersiapkan *raw* data komentar dengan nama file *skripsi-ijazahPalsuPakJoko.csv*.

2. Melakukan tahap *text preprocessing* pada *raw* data komentar dan menghasilkan *clean* dataset.
  3. Memberikan nilai bobot untuk setiap kata dengan fitur ekstraksi TF-IDF pada data input (isi komentar), kemudian mentransformasikan data setiap kata dalam komentar pada data latih dan data uji menjadi vektor-vektor bobot kata.
- f. Evaluasi
- Pengukuran performa hasil pengujian dari klasifikasi SVM terhadap analisis sentimen komentar pada data uji, menggunakan penghitungan evaluasi *confusion matrix*. *Confusion matrix* menampilkan tabel berdimensi 3x3 dikarenakan terdapat 3 jenis label yang diklasifikasikan yaitu komentar positif, komentar negatif, dan komentar netral. Evaluasi ini menampilkan banyaknya jumlah baris data komentar pada perbandingan label antara data aktual dengan data hasil prediksinya.

### 3. HASIL DAN PEMBAHASAN

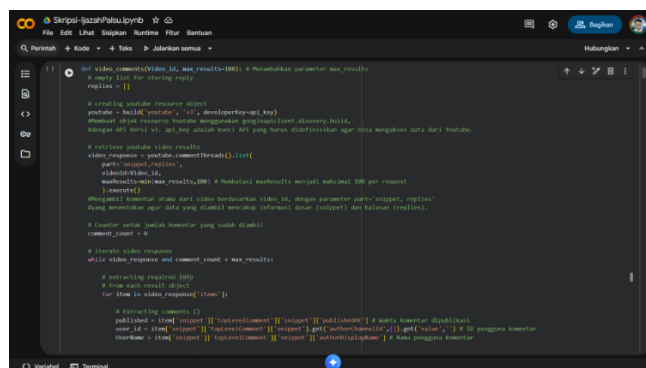
#### 3.1 Pengumpulan Data

Pengumpulan data pada penelitian ini dilakukan dengan memanfaatkan media sosial YouTube sebagai sumber data utama. Proses pengambilan data dilakukan menggunakan teknik *scraping* dengan bantuan YouTube Data API v3, yang memungkinkan peneliti untuk mengakses dan mengunduh data komentar secara sistematis dan terstruktur. Data yang diambil merupakan data komentar pada video yang diunggah oleh kanal Indonesia *Lawyers Club* dengan judul “[FULL] POLEMIK ASLI ATAU PALSU IJAZAH JOKOWI // APA JADINYA KALAU TERBUKTI PALSU?”. Video tersebut dipilih sebagai objek penelitian karena mengandung topik yang menimbulkan perdebatan publik dan menghasilkan beragam opini dari masyarakat. Data yang dikumpulkan berupa komentar-komentar pengguna YouTube yang berasal dari sebuah video dengan tingkat interaksi tinggi, sehingga diharapkan mampu merepresentasikan beragam opini masyarakat terhadap isu yang dibahas. Pemanfaatan API resmi YouTube dipilih karena memberikan akses data yang lebih akurat, legal, dan sesuai dengan ketentuan penggunaan platform. Komentar-komentar yang muncul pada video tersebut dianggap relevan untuk dianalisis karena mencerminkan persepsi, sikap, serta sentimen publik terhadap isu yang bersifat sensitif dan kontroversial dalam ranah politik dan hukum.



Gambar 3. Tangkapan Layar pada media social Youtube

Seluruh proses pengambilan data dilakukan secara otomatis dengan memanfaatkan YouTube Data API v3, sehingga memungkinkan peneliti untuk mengumpulkan komentar dalam jumlah besar secara efisien dan konsisten. Selanjutnya, proses pengolahan data dilakukan menggunakan platform Google Colaboratory, yang menyediakan lingkungan komputasi berbasis *cloud* serta mendukung penggunaan bahasa pemrograman Python dan berbagai pustaka pendukung analisis data. Penggunaan Google Colaboratory juga memudahkan peneliti dalam mengelola data, melakukan pra-pemrosesan teks, serta menerapkan metode analisis sentimen secara terintegrasi. Dengan demikian, data yang diperoleh dari media sosial YouTube dapat diolah secara optimal untuk menghasilkan informasi yang akurat dan dapat dipertanggungjawabkan dalam mendukung tujuan penelitian ini.



Gambar 4. Tampilan *crawling/scraping* data di *google colab*



Data yang diperoleh dan digunakan diolah langsung dari proses *crawling* mandiri. Data set tersebut dikumpulkan dan disimpan dalam bentuk format terstruktur (.csv) dan excel (.xlsx). Isi dari data set tersebut mencakup sejumlah informasi penting, antara lain komentar asli yang berkaitan dengan polemik ijazah, label sentimen yang menunjukkan kecenderungan opini (positif, negatif, dan netral), serta sejumlah meta data tambahan seperti ID pengguna, tanggal komentar dibuat, dan informasi akun pengguna. Adapun tampilan data mentah yang sudah di *crawling* dan diunduh dapat dilihat pada Gambar 5.

published	user_id	user	profile	avatar	comment	likes
2025-10-14T13:59:21	UCq0ND7b2oM	Bigwnt @RupertusRoland96	http://www	https://y	Pembodohan publik	0
2025-10-14T07:53:11	UCYvFCgAluP	eIRZLuA @HermanJamid	http://www	https://y	Klu ngurusin masalah Jokowi ..negara ini kacu	0
2025-10-08T04:17:11	UCB85x_LOGj	hwK43JC @SuryaAlmani-u5w	http://www	https://y	Pintar ngeles klu memngbada ijazah tunjukkan donk di depan masarak selesai knpa tunggu di pi	0
2025-10-07T12:08:3	UCBhqXGfTpkBnF	Cawd @UdSinegal	http://www	https://y	SeOrang Negarawan Patut Bangsa Kalao BeluOrnya MeNunjukkan Ijazah Nya Untuk Semua Mas	0
2025-09-28T12:14:11	UCkMzeZTeBzYg3Hh3J	@Bz21_Y	http://www	https://y	IJAZAH BISA DI PAKAI LAMAR KEJUAAN ITU AJA KALAU NDA SAMA DGN KEPRIADIAN PENUSHA	0
2025-09-28T12:10:0	UCkMzeZTeBzYg3Hh3J	@Bz21_Y	http://www	https://y	gmna negara HUKUM YG PEMEGANG HUKUM PENIAHAT SEMUA, PAKAI HUKUM YG GIMANA...	0
2025-09-27T07:21:0	UCmCm3D2_zuK6pD5pE	R @ross427	http://www	https://y	PA KARNI TDK BOSANKAH BAHAS IJASAH TERUS???? COBA BAHAS JOKOWI SEKARANG JADI PEN	1
2025-09-20T06:47:5	UCqit-decfuyTDNp	vMe @alagamob2860	http://www	https://y	itu Bung Andi cara berpikirnya gimana yah, kok di bulak-balik... Saya rakyat biasa dan sangat in	0
2025-09-17T03:19:3	UCmPqOOTAFyCk31rE	@bakritawang:7d	http://www	https://y	apaka kau sebelum masu pengacara tida di perissa ijasamu	0
2025-09-15T23:16:2	UC1ctWomhgQoGKZ1	@delimacantik2403	http://www	https://y	Sekelas Barack Obama saja menunjukkan identitas nya,tujuan nya agar masyarakat benar@#3;	0
2025-09-13T11:44:5	UCWc_uefKpWvRv3Y9A	@M.schurul	http://www	https://y	Ijazahnya itu palsu patut di acungi jempol Roy suryo cs. (Semangat banget Roy jangan kendhor	0
2025-09-12T00:29:2	UCCEGZCv05-ipuJ6lDne	@Rokingtag-yt	http://www	https://y	Ketika si tahi lalat bicara, gov skip -dr50k paling pintar. Cuehhhh	0
2025-09-12T00:08:1	UCCEGZCv05-ipuJ6lDne	@Rokingtag-yt	http://www	https://y	Siapa bisa nilai, mana yg otaknya kosong?	0
2025-09-12T00:03:4	UCCEGZCv05-ipuJ6lDne	@Rokingtag-yt	http://www	https://y	itu silver kemana sekarang kok gk keliatan lagi?	0
2025-09-29T06:47:5	UCqit-decfuyTDNp	vMe @alagamob2860	http://www	https://y	Orang curang mendidik	0
2025-08-24T13:07:3	UCDv5WUjyCo_oRZV5C	@dudfifanahidayat5929	http://www	https://y	Selama belum terbuka dri pihak UGM, maaf gua ga respect sebak apapun kampus nya itu	0
2025-08-24T13:01:0	UCDv5WUjyCo_oRZV5C	@dudfifanahidayat5929	http://www	https://y	Tersangka itu lavan sampean pak Roy	0
2025-08-24T12:38:4	UCDv5WUjyCo_oRZV5C	@dudfifanahidayat5929	http://www	https://y	Anak nya juga paltu	0
2025-08-19T06:24:1	UCv8NvU8d5iHQP4-Ljm	@endy4840	http://www	https://y	Org juga muak sama mbah AMIN...3	0
2025-08-19T02:53:0	UCV0MvHvY8q9z8EimstJ	@SamsulArifin-ff6jk	http://www	https://y	Si Andi usaha bngt palsu jd asli ..manik bngt di sogok gede eta.	0
2025-08-18T17:25:1	UCV0MvHvY8q9z8EimstJ	@SamsulArifin-ff6jk	http://www	https://y	Siapa yg mau ngatur pembobong ijasahnya gk jelas	0
2025-08-18T17:23:5	UCV0MvHvY8q9z8EimstJ	@SamsulArifin-ff6jk	http://www	https://y	Sulfester harus di hukum	0
2025-08-18T15:57:3	UCqgHhKqJcyZhvIRkm	@satrian09861	http://www	https://y	Saya juga yakin ijasahnya palsu	0
2025-08-18T07:44:0	UCX0LdH18hWcZlvvFPV	@DiasLoka-fhs	http://www	https://y	Setiap pblk publik ada konseksensiny tentang ijazah yg bersangkutan ....hal public keinginan ti	0
2025-08-17T13:35:2	UCG6Z2mUvNz8RBVzrW	@SugengHaryanto-c2g	http://www	https://y	peneliti kacampan	0
2025-08-17T13:33:3	UCG6Z2mUvNz8RBVzrW	@SugengHaryanto-c2g	http://www	https://y	si roy suruh aja jadi rekanan kpu biar dia uji dan teliti ijasah semua calon legialati dgn metoda i	0
2025-08-17T13:21:2	UCG6Z2mUvNz8RBVzrW	@SugengHaryanto-c2g	http://www	https://y	kebanakuan s noia ikinnya tdk sesuai	0

Gambar 5. Hasil *crawling* data dari komentar youtube

Hasil dari pengambilan data (*scraping*) ini dalam bentuk *csv* yang akan melewati tahap preprocessing[3]. Dataset tersebut tersedia dalam format *.csv* dan excel dengan nama file *skripsi-ijazahPalsuPakJoko.csv/xlsx* dan berisi total 20294 entri. Setiap entri terdiri dari beberapa atribut utama, seperti *published*, *user id*, *username*, *Profile URL*, *Avatar URL*, *comment* dan *likes*. Data ini dikumpulkan berdasarkan topik yang berkaitan dengan polemik ijazah, dan telah disusun secara terstruktur sehingga siap digunakan dalam tahap preprocessing dan analisis sentimen.

### 3.2 Text Processing

Pada tahap pra-pemrosesan data, dilakukan beberapa proses utama yang meliputi case folding, text cleansing, stemming, stopword removal, dan tokenization. Tahapan ini bertujuan untuk menyiapkan data teks agar lebih terstruktur dan mudah diolah pada tahap analisis selanjutnya. Proses awal yang dilakukan adalah case folding, yaitu mengonversi seluruh karakter dalam dokumen teks menjadi huruf kecil untuk menghindari perbedaan makna akibat variasi penggunaan huruf besar dan huruf kecil[12].

#### a. Data Cleaning

Tahap *data cleaning* merupakan proses awal dalam pengolahan data teks yang bertujuan untuk membersihkan dataset dari berbagai komponen yang tidak relevan dan berpotensi mengganggu proses analisis. Pada tahap ini dilakukan penghapusan karakter-karakter yang tidak memiliki nilai semantik dalam konteks analisis sentimen, seperti angka, tanda pagar (#), tanda mention (@), emoji, serta tautan (*link*) yang sering muncul pada komentar media sosial. Selain itu, pembersihan juga mencakup penghapusan karakter berbasis ASCII, simbol-simbol khusus, *mention* pengguna, dan baris kosong yang tidak mengandung informasi penting. Proses ini dilakukan untuk memastikan bahwa data teks yang digunakan benar-benar merepresentasikan opini pengguna secara murni dan dapat diproses secara optimal oleh algoritma klasifikasi.

Berdasarkan hasil pelaksanaan tahap *data cleaning*, jumlah dataset yang digunakan dalam penelitian ini tidak mengalami perubahan, yaitu tetap sebanyak 20.294 data komentar. Hal ini menunjukkan bahwa meskipun dilakukan proses pembersihan terhadap karakter dan komponen yang tidak relevan, setiap komentar masih memiliki konten teks yang bermakna dan layak untuk dianalisis lebih lanjut. Dengan tidak adanya data yang tereliminasi pada tahap ini, dapat disimpulkan bahwa dataset awal telah memenuhi kriteria kelengkapan data dan siap untuk dilanjutkan ke tahap pra-pemrosesan berikutnya, seperti tokenisasi, penghapusan stopword, dan stemming. Kondisi tersebut juga mengindikasikan bahwa proses *data cleaning* berhasil menjaga integritas data tanpa menghilangkan informasi penting yang dibutuhkan dalam analisis sentimen. Contoh *data cleaning* terdapat pada gambar 6.

comment	cleansing
0 Kalian salah kalau saling debat masalah palsu ...	Kalian salah kalau saling debat masalah palsu ...
1 Siapa yg berbohong (salah)Jokowi/(Prabowo)?	Siapa yg berbohong salah Jokowi Prabowo
2 Siapa yg salah,Prabowo atau siapa?	Siapa yg salah Prabowo atau siapa
3 Pembodohan publik	Pembodohan publik
4 Klu ngurusin masalah Jokowi ..negara ini kacu	Klu ngurusin masalah Jokowi negara ini kacu
...	...
20801 kejujuran itu lebih pnting dek	kejujuran itu lebih pnting dek
20802 @royroy7746 prucma klo ga sesuai dg fikiranmu ...	prucma klo ga sesuai dg fikiranmu bkai dibling...
20803 @OpokuuAk di pikiran buzzerp mngkin gk sesuai	di pikiran buzzerp mngkin gk sesuai
20804 Sudah jadi isu nasional dan dunia. Ada baiknya...	Sudah jadi isu nasional dan dunia Ada baiknya ...
20805 @royroy7746 kalo emg gak punya ijazah terus im...	kalo emg gak punya ijazah terus impactnya apa ...

Gambar 6. Hasil *Cleaning*



b. Case Folding

Case Folding merupakan mengubah data komentar menjadi *lower-case* atau penulisan dalam huruf kecil[13]. Mengolah data tidak terstruktur menjadi data terstruktur. Contoh *Case Folding* terdapat pada gambar 7.

	comment	cleansing	case_folding
0	Kalian salah kalau saling debat masalah palsu ...	Kalian salah kalau saling debat masalah palsu ...	kalian salah kalau saling debat masalah palsu ...
1	Siapa yg berbohong,(salah),Jokowi/(Prabowo)?	Siapa yg berbohong salah Jokowi Prabowo	siapa yg berbohong salah jokowi prabowo
2	Siapa yg salah,Prabowo atau siapa?	Siapa yg salah Prabowo atau siapa	siapa yg salah prabowo atau siapa
3	Pembodohan publik	Pembodohan publik	[pembodohan, publik]
4	Klu ngurusin masalah Jokowi ..negara ini kacu	Klu ngurusin masalah Jokowi negara ini kacu	[klu, ngurusin, masalah, jokowi, negara, ini, ...]
...	...	...	...
20801	kejujuran itu lebih prting dek	kejujuran itu lebih prting dek	[kejujuran, itu, lebih, prting, dek]
20802	@royroy7746 pruma klo ga sesuai dg fikiranmu ...	pruma klo ga sesuai dg fikiranmu bkai dibling...	[pruma, klo, ga, sesuai, dg, fikiranmu, bkai, ...]
20803	@OpokuuAk di pikiran buzzerp mngkin gk sesuai	di pikiran buzzerp mngkin gk sesuai	[di, pikiran, buzzerp, mngkin, gk, sesuai]
20804	Sudah jadi isu nasional dan dunia. Ada baiknya...	Sudah jadi isu nasional dan dunia. Ada baiknya ...	[isu, nasional, dunia, baiknya, ...]
20805	@royroy7746 kalo emg gak punya ijazah terus im...	kalo emg gak punya ijazah terus impactnya apa ...	[kalo, emg, gak, punya, ijazah, terus, impactnya, ...]

Gambar 7. Hasil Case Folding

c. Tokenizing

Tokenizing merupakan Proses tokenisasi merupakan tahapan yang bertujuan untuk membagi dokumen teks menjadi unit-unit kata atau token. Token-token yang dihasilkan selanjutnya digunakan untuk mengidentifikasi kata kunci berdasarkan kategori *Parts of Speech* (PoS), sehingga setiap kata dapat dikenali fungsi dan perannya dalam struktur kalimat[14]. Hal ini dilakukan untuk mengetahui kemunculan kata. Contoh *Tokenizing* terdapat pada gambar 8.

	comment	cleansing	case_folding	tokenize
0	Kalian salah kalau saling debat masalah palsu ...	Kalian salah kalau saling debat masalah palsu ...	kalian salah kalau saling debat masalah palsu ...	[kalian, salah, kalau, saling, debat, masalah, ...]
1	Siapa yg berbohong,(salah),Jokowi/(Prabowo)?	Siapa yg berbohong salah Jokowi Prabowo	siapa yg berbohong salah jokowi prabowo	[siapa, yg, berbohong, salah, jokowi, prabowo]
2	Siapa yg salah,Prabowo atau siapa?	Siapa yg salah Prabowo atau siapa	siapa yg salah prabowo atau siapa	[siapa, yg, salah, prabowo, atau, siapa]
3	Pembodohan publik	Pembodohan publik	[pembodohan, publik]	[pembodohan, publik]
4	Klu ngurusin masalah Jokowi ..negara ini kacu	Klu ngurusin masalah Jokowi negara ini kacu	[klu, ngurusin, masalah, jokowi, negara, ini, ...]	[klu, ngurusin, masalah, jokowi, negara, ini, ...]
...	...	...	...	...
20801	kejujuran itu lebih prting dek	kejujuran itu lebih prting dek	kejujuran itu lebih prting dek	[kejujuran, itu, lebih, prting, dek]
20802	@royroy7746 pruma klo ga sesuai dg fikiranmu ...	pruma klo ga sesuai dg fikiranmu bkai dibling...	pruma klo ga sesuai dg fikiranmu bkai dibling...	[pruma, klo, ga, sesuai, dg, fikiranmu, bkai, ...]
20803	@OpokuuAk di pikiran buzzerp mngkin gk sesuai	di pikiran buzzerp mngkin gk sesuai	di pikiran buzzerp mngkin gk sesuai	[di, pikiran, buzzerp, mngkin, gk, sesuai]
20804	Sudah jadi isu nasional dan dunia. Ada baiknya...	Sudah jadi isu nasional dan dunia. Ada baiknya ...	sudah jadi isu nasional dan dunia ada baiknya ...	[sudah, jadi, isu, nasional, dan, dunia, ada, ...]
20805	@royroy7746 kalo emg gak punya ijazah terus im...	kalo emg gak punya ijazah terus impactnya apa ...	kalo emg gak punya ijazah terus impactnya apa ...	[kalo, emg, gak, punya, ijazah, terus, impactn, ...]

Gambar 8. Hasil Tokenizing

d. Filtering Stopword/Stopword Removal

Filtering Stopword/Stopword Removal merupakan proses menghapus kata yang dianggap tidak penting pada kalimat. untuk menghapus kata-kata umum yang banyak digunakan tapi tidak memberikan pengaruh sentimen pada suatu kalimat. Proses *stopword* yang digunakan pada penelitian ini adalah dengan memanfaatkan *library corpus stopwords* bahasa Indonesia[15]. Untuk mempermudah proses *remove stopwords*, digunakan library NLTK berbahasa Indonesia.

	comment	cleansing	case_folding	tokenize	filtering/stopword removal
0	Kalian salah kalau saling debat masalah palsu ...	Kalian salah kalau saling debat masalah palsu ...	kalian salah kalau saling debat masalah palsu ...	[kalian, salah, kalau, saling, debat, masalah, ...]	[salah, debat, palsu, nya, kerna, memang, asi, ...]
1	Siapa yg berbohong,(salah),Jokowi/(Prabowo)?	Siapa yg berbohong salah Jokowi Prabowo	siapa yg berbohong salah jokowi prabowo	[siapa, yg, berbohong, salah, jokowi, prabowo]	[yg, berbohong, salah, jokowi, prabowo]
2	Siapa yg salah,Prabowo atau siapa?	Siapa yg salah Prabowo atau siapa	siapa yg salah prabowo atau siapa	[siapa, yg, salah, prabowo, atau, siapa]	[yg, salah, prabowo]
3	Pembodohan publik	Pembodohan publik	[pembodohan, publik]	[pembodohan, publik]	[pembodohan, publik]
4	Klu ngurusin masalah Jokowi ..negara ini kacu	Klu ngurusin masalah Jokowi negara ini kacu	[klu, ngurusin, masalah, jokowi, negara, ini, ...]	[klu, ngurusin, masalah, jokowi, negara, ini, ...]	[klu, ngurusin, jokowi, negara, kacu]
...	...	...	...	...	...
20801	kejujuran itu lebih prting dek	kejujuran itu lebih prting dek	kejujuran itu lebih prting dek	[kejujuran, itu, lebih, prting, dek]	[kejujuran, prting, dek]
20802	@royroy7746 pruma klo ga sesuai dg fikiranmu ...	pruma klo ga sesuai dg fikiranmu bkai dibling...	pruma klo ga sesuai dg fikiranmu bkai dibling...	[pruma, klo, ga, sesuai, dg, fikiranmu, bkai, ...]	[pruma, klo, ga, sesuai, dg, fikiranmu, bkai, ...]
20803	@OpokuuAk di pikiran buzzerp mngkin gk sesuai	di pikiran buzzerp mngkin gk sesuai	di pikiran buzzerp mngkin gk sesuai	[di, pikiran, buzzerp, mngkin, gk, sesuai]	[pikiran, buzzerp, mngkin, gk, sesuai]
20804	Sudah jadi isu nasional dan dunia. Ada baiknya...	Sudah jadi isu nasional dan dunia. Ada baiknya ...	sudah jadi isu nasional dan dunia ada baiknya ...	[isu, nasional, dunia, baiknya, ...]	[isu, nasional, dunia, baiknya, dibuka, sidang, ...]
20805	@royroy7746 kalo emg gak punya ijazah terus im...	kalo emg gak punya ijazah terus impactnya apa ...	kalo emg gak punya ijazah terus impactnya apa ...	[kalo, emg, gak, punya, ijazah, terus, impactn, ...]	[kalo, emg, gak, ijazah, impactnya, dianulir, ...]

Gambar 9. Hasil Stopword Removal

e. Stemming

Stemming merupakan proses mengembalikan kata menjadi kata dasar dengan membuang awalan, imbuhan dan akhiran. *Stemming* adalah langkah dalam pengolahan teks yang mengubah kata-kata dari bentuk variasi menjadi bentuk dasarnya sesuai struktur morfologi. Ini berarti kata-kata yang diubah tetap memiliki arti dan makna yang sama dengan kata dasarnya[16]. Proses ini perlu dilakukan untuk menyeragamkan kata-kata yang ada pada data set agar menjadi kata dasar. *Stemming* pada data set berbahasa Indonesia dilakukan dengan memanfaatkan *library* Sastrawi. Penerapan stemming memiliki peranan penting dalam analisis sentimen, khususnya pada tahap klasifikasi teks. Kata-



kata yang telah distem cenderung memiliki makna yang lebih konsisten, sehingga membantu algoritma klasifikasi dalam mengenali pola sentimen secara lebih akurat. Selain itu, proses stemming dapat meningkatkan keandalan hasil analisis dengan mengurangi ambiguitas makna serta meningkatkan efisiensi pemrosesan data. Oleh karena itu, stemming menjadi salah satu tahapan penting dalam pra-pemrosesan teks untuk menghasilkan model analisis sentimen yang lebih optimal dan dapat dipertanggungjawabkan secara ilmiah[17].

	comment	cleansing	case_folding	tokenize	filtering/stopword removal	stemming data
0	Kalian salah kalau saling debat masalah palsu ...	Kalian salah kalau saling debat masalah palsu ...	kalian salah kalau saling debat masalah palsu ...	[kalian, salah, kalau, saling, debat, masalah, ...]	[salah, debat, palsu, nya, kerna, memng, asli, ...]	salah debat palsu nya kerna memng asli memng l...
1	Siapa yg berbohong (salah)Jokowi/Prabowo)?	Siapa yg berbohong salah Jokowi Prabowo	siapa yg berbohong salah jokowi prabowo	[siapa, yg, berbohong, salah, jokowi, prabowo]	[yg, berbohong, salah, prabowo]	yg bohong salah jokowi prabowo
2	Siapa yg salah,Prabowo atau siapa?	Siapa yg salah,Prabowo atau siapa	siapa yg salah prabowo atau siapa	[siapa, yg, salah, prabowo, atau, siapa]	[yg, salah, prabowo]	yg salah prabowo
3	Pembodohan publik	Pembodohan publik	pembodohan publik	[pembodohan, publik]	[pembodohan, publik]	bodoh publik
4	Klu ngurusin masalah Jokowi negara ini kacu	Klu ngurusin masalah Jokowi negara ini kacu	klu ngurusin masalah jokowi negara ini kacu	[klu, ngurusin, masalah, jokowi, negara, ini, ...]	[klu, ngurusin, jokowi, negara, kacu]	klu ngurusin jokowi negara kacu
5	Pintar ngeles klu memngbada ijazah tunjukkan d...	Pintar ngeles klu memngbada ijazah tunjukkan do...	pintar ngeles klu memngbada ijazah tunjukkan do...	[pintar, ngeles, klu, memngbada, ijazah, tunjuk...	[pintar, ngeles, klu, memngbada, ijazah, tunjuk...	pintar ngeles klu memngbada ijazah tunjuk donk ...
6	SeOrang Negarawan Patut Bangsa Kalao BelOnya ...	SeOrang Negarawan Patut Bangsa Kalao BelOnya ...	seorang negarawan patut bangsa kalao belonya ...	[seorang, negarawan, patut, bangsa, kalao, bel...	[negarawan, patut, bangsa, kalao, belonya, me...	negarawan patut bangsa kalao belonya tunjuk l...
7	UJAZAH BISA DI PAKAI LAMAR KERJAAN ITU AJA KAL...	UJAZAH BISA DI PAKAI LAMAR KERJAAN ITU AJA KAL...	ijazah bisa di pakai lamar kerjaan itu aja kal...	[ijazah, bisa, di, pakai, lamar, kerjaan, itu, ...]	[ijazah, pakai, lamar, kerjaan, aja, nda, dgn, ...]	ijazah pakai lamar kerja aja nda dgn pribadi u...
8	gmna negara HUKUM YG PEMEGANG HUKUM PENJAHAT S...	gmna negara HUKUM YG PEMEGANG HUKUM PENJAHAT S...	gmna negara hukum yg pemegang hukum penjahat s...	[gmna, negara, hukum, yg, pemegang, hukum, pen...	[gmna, negara, hukum, yg, pemegang, hukum, pen...	gmna negara hukum yg pegang hukum jahat pakai ...
9	PA KARNI TDK BOSANKAH BAHAS UJAZAH TERUS?????	PA KARNI TDK BOSANKAH BAHAS UJAZAH TERUS COBA ...	pa kami tdk bosankah bahas ijazah terus coba ...	[pa, kami, tdk, bosankah, bahas, ijazah, terus...	[pa, kami, tdk, bosankah, bahas, ijazah, coba...	pa kami tdk bosan bahas ijazah coba bahas jok...
10	Itu Bung Andi cara berpikirnya gimana yah, kok...	Itu Bung Andi cara berpikirnya gimana yah kok ...	itu bung andi cara berpikirnya gimana yah kok ...	[itu, bung, andi, cara, berpikirnya, gimana, y...	[andi, berpikirnya, gimana, yah, bulak, rakyat, rakyat...	andi pikir gimana yah bulak rakyat beabsahan l...
11	apaka kau sebelum masu pengacara tida di peris...	apaka kau sebelum masu pengacara tida di peris...	apaka kau sebelum masu pengacara tida di peris...	[apaka, kau, sebelum, masu, pengacara, tida, di, peris...	[apaka, kau, masu, pengacara, tida, perissa, i...	apaka kau masu acara tida perissa ijassamu
12	Sekelas Barack Obama saja menunjukan identitas...	Sekelas Barack Obama saja menunjukan identitas...	sekelas barack obama saja menunjukan identitas...	[sekelas, barack, obama, saja, menunjukan, iden...	[sekelas, barack, obama, menunjukan, identitas...	kelas barack obama tunjuk identitas nya tju n...
13	Ijazahnya itu palsu patut di acungi jempol Roy...	Ijazahnya itu palsu patut di acungi jempol Roy...	ijazahnya itu palsu patut di acungi jempol roy...	[ijazahnya, itu, palsu, patut, di, acungi, jem...	[ijazahnya, palsu, patut, acungi, jempol, roy...	ijazah palsu patut acungi jempol roy suryo ca s...
14	Ketika si tahi lalat bicara, gw skip   Sok pal...	Ketika si tahi lalat bicara gw skip   Sok pal...	ketika si tahi lalat bicara gw skip br sok pal...	[ketka, si, tahi, lalat, bicara, gw, skip, br...	[si, tahi, lalat, bicara, gw, skip, br...	si tahi lalat bicara gw skip br sok pintar cue...
15	Siapa bisa nilai, mana yg otaknya	Siapa bisa nilai mana yg otaknya	siapa bisa nilai mana yg	[siapa, bisa, nilai, mana, yg,	[nilai, yg, otaknya, kosong]	nilai yg otak kosong

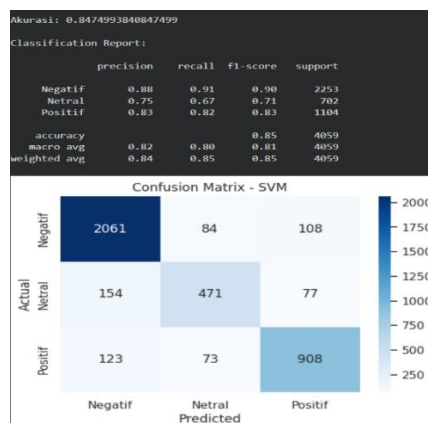
Gambar 10. Hasil Stemming

### 3.3 Ekstraksi Fitur

Ekstraksi fitur pada penelitian ini menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF). TF-IDF digunakan untuk mengubah komentar *youtube* menjadi representasi numerik yang dapat diproses oleh algoritma *Support Vector Machine*. Metode TF-IDF memberikan bobot pada setiap kata berdasarkan tingkat kemunculannya dalam suatu dokumen (TF) dan kelangkaannya pada keseluruhan dokumen (IDF). Kata yang penting akan memiliki bobot lebih besar, sedangkan kata umum memiliki bobot rendah.

Dalam penelitian ini, *TfidfVectorizer* dikonfigurasi dengan `max_features=5000` untuk menghasilkan 5000 fitur paling relevan, serta `Ngram_range=(1,2)` untuk menangkap unigram dan bigram. Penggunaan bigram meningkatkan kemampuan model mengidentifikasi konteks frasa seperti "ijazah palsu". Hasil TF-IDF berupa matriks sparse berdimensi (`n_documents × 5000`), yang kemudian digunakan sebagai input utama pemodelan SVM. Pemilihan metode ini efektif karena SVM bekerja optimal pada data berdimensi tinggi dan sparse.

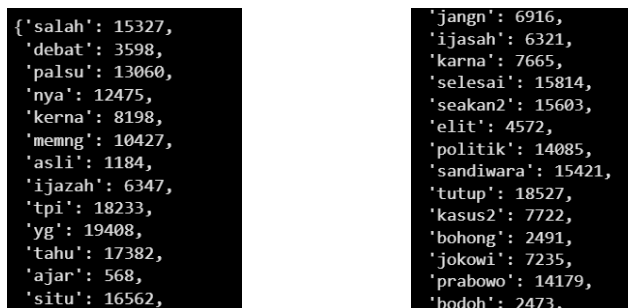
Hasil dari proses ekstraksi fitur menggunakan TF-IDF berupa matriks berdimensi (`n_documents × 5000`) dengan format *sparse matrix*, di mana setiap baris merepresentasikan satu komentar dan setiap kolom merepresentasikan fitur kata atau frasa terpilih. Matriks ini selanjutnya digunakan sebagai input utama dalam proses pemodelan menggunakan algoritma *Support Vector Machine*. Pemilihan TF-IDF sebagai metode ekstraksi fitur dinilai tepat karena mampu menghasilkan representasi data teks yang informatif dan efisien. Selain itu, karakteristik data TF-IDF yang berdimensi tinggi dan bersifat *sparse* sangat sesuai dengan kemampuan SVM yang dikenal efektif dalam menangani data dengan jumlah fitur yang besar, sehingga diharapkan dapat meningkatkan kinerja klasifikasi sentimen secara optimal.



Gambar 11. Pemodelan *Support Vector Machine* (Confusion Matrix)

Perhitungan ini menentukan seberapa relevan sebuah kata di dalam sebuah dokumen. Setelah proses ini, dapat dilanjutkan untuk menggunakan vektor fitur TF-IDF sebagai input pada model *Support Vector Machine* (SVM). Algoritma TF-IDF menentukan nilai bobot untuk setiap kata dalam dataset. TF mengukur seberapa sering sebuah kata muncul dalam dokumen, sedangkan IDF merujuk pada jumlah dokumen (DF) yang memiliki kata tersebut. Implementasi *TfidfVectorizer*[18].

Pada tahap ini, digunakan sebanyak 20294 komentar yang telah melewati tahapan *pre processing*. Proses pembobotan menghasilkan total 19824 kata unik yang mewakili berbagai istilah yang muncul dalam kumpulan data tersebut. Tahapan *TF-IDF* ini bertujuan untuk mengubah kumpulan dokumen teks menjadi representasi numerik dalam bentuk matriks istilah-dokumen. Proses ini memungkinkan setiap kata yang muncul dalam tweet diberi bobot berdasarkan seberapa sering kata tersebut muncul di satu dokumen dibandingkan kemunculannya di seluruh dokumen lainnya. Dalam proses ini, setiap kata diberikan indeks unik untuk mengidentifikasinya dalam vektor atau matriks. Misalnya, kata "salah" memiliki indeks 15327, sedangkan kata "palsu" memiliki indeks 13060, dan seterusnya. Tahapan ini penting untuk mendukung proses analisis sentimen menggunakan metode *Support Vector Machine* karena memungkinkan algoritma untuk mengenali dan membedakan bobot pengaruh dari masing-masing kata dalam menentukan sentimen dari sebuah komentar.



```
{'salah': 15327,
'debat': 3598,
'palsu': 13060,
'nya': 12475,
'kerna': 8198,
'memng': 10427,
'asli': 1184,
'ijazah': 6347,
'tpi': 18233,
'yg': 19408,
'tahu': 17382,
'ajar': 568,
'situ': 16562,
'jangan': 6916,
'ijazah': 6321,
'karna': 7665,
'selesai': 15814,
'seakan2': 15603,
'elit': 4572,
'politik': 14085,
'sandiwara': 15421,
'tutup': 18527,
'kasus2': 7722,
'bohong': 2491,
'jokowi': 7235,
'prabowo': 14179,
'bodoh': 2473,
```


Gambar 12. Daftar kata unik hasil tokenisasi

### 3.4 Pelabelan *Lexicon Based*

Setelah proses *pre-processing* menghasilkan data dalam variabel `stemming_data`, dilakukan tahapan pelabelan sentimen terhadap data tersebut. Proses pelabelan ini menggunakan pendekatan *lexicon-based*, yaitu dengan membandingkan kata-kata dalam setiap komentar dengan daftar kata positif dan negatif yang diperoleh dari repositori *GitHub*. Daftar kata tersebut merupakan hasil kurasi dari proyek analisis sentimen berbahasa Indonesia dan telah tersedia dalam format `.tsv` atau `.txt`. Setiap komentar yang telah melalui proses *pre-processing* diubah menjadi daftar token, lalu dianalisis berdasarkan jumlah kemunculan kata positif dan negatif sesuai dengan isi *lexicon*.

Meskipun proses pelabelan data dalam penelitian ini dilakukan secara otomatis, tetap disadari bahwa validasi manual memiliki peranan penting untuk memastikan tingkat reliabilitas hasil pelabelan. Oleh karena itu, sebagian data, yaitu sebesar 10% dari keseluruhan dataset, dipilih secara acak untuk dilakukan pengecekan manual. Pengecekan ini bertujuan untuk membandingkan hasil pelabelan otomatis dengan interpretasi peneliti terhadap sentimen yang terkandung dalam data tersebut.

Hasil dari proses validasi manual tersebut digunakan sebagai dasar untuk menilai keandalan model pelabelan yang diterapkan. Langkah ini menjadi penting mengingat analisis sentimen berbasis *machine learning* memiliki potensi kerentanan terhadap bias algoritmik serta ambiguitas bahasa yang dapat memengaruhi akurasi hasil klasifikasi[19]. Dapat dilihat pada gambar 13.



```
import csv
import csv
lexicon_positive = dict()
with open('./positive.csv', 'r') as csvfile:
    reader = csv.reader(csvfile, delimiter=',')
    next(reader)
    for row in reader:
        lexicon_positive[row[0]] = int(row[1])

lexicon_negative = dict()
with open('./negative.csv', 'r') as csvfile:
    reader = csv.reader(csvfile, delimiter=',')
    next(reader)
    for row in reader:
        lexicon_negative[row[0]] = int(row[1])

def sentiment_analysis_lexicon_indonesia(text):
    score = 0
    for word in text:
        if (word in lexicon_positive): score += lexicon_positive[word]
    for word in text:
        if (word in lexicon_negative): score += lexicon_negative[word]
    polarity=""
    if (score > 0):
        sentiment='Positif'
    elif (score < 0):
        sentiment='Negatif'
    else:
        sentiment='Netral'
    return score, sentiment
```

Gambar 13. Kode program untuk menghitung skor menggunakan *lexicon based*



Pada Gambar 14 berisi kumpulan kata-kata positif yang diambil dari (Fajri Koto) kamus *lexicon* dan digunakan sebagai acuan untuk menentukan nilai sentimen positif dalam data komentar. Kata-kata ini berfungsi sebagai indikator untuk mengidentifikasi opini atau perasaan positif yang terkandung dalam teks, sehingga membantu dalam proses analisis sentimen secara lebih akurat dan terstruktur.

	word	weight
0	hai	3
1	merekam	2
2	ekstensif	3
3	paripurna	1
4	detail	2

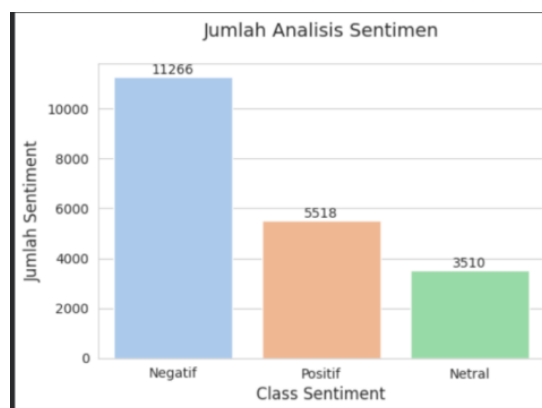
**Gambar 14.** Kumpulan kata positif Kamus *Lexicon*

Pada Gambar 15 berisi kumpulan kata-kata negatif yang diambil dari (Fajri Koto) kamus *lexicon* dan digunakan sebagai acuan untuk menentukan nilai sentimen negatif dalam data komentar.

	word	weight
0	putus tali gantung	-2
1	gelebah	-2
2	gobar hati	-2
3	tersentuh (perasaan)	-1
4	isak	-5

**Gambar 15.** Kumpulan kata negatif Kamus *Lexicon*

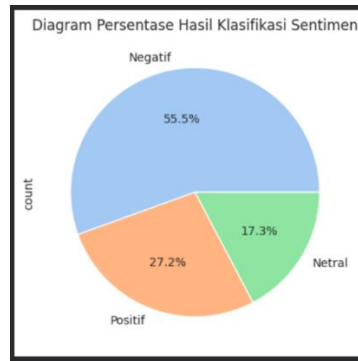
Setelah proses *pre processing* selesai, analisis sentimen dilakukan dengan menggunakan fungsi sentimen *analysis lexicon* Indonesia, yang menghitung skor sentimen secara sederhana berdasarkan kemunculan kata-kata dalam daftar kata positif dan negatif. Daftar kata tersebut diperoleh dari repositori *GitHub* 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 mengurangnya. Hasil akhir dari perhitungan skor ini digunakan untuk mengklasifikasikan setiap komentar ke dalam kategori positif, negatif, atau netral.



**Gambar 16.** Jumlah Analisis Sentimen

Hasil analisis sentimen terhadap data komentar yang telah melalui proses *pre processing* menunjukkan bahwa setiap komentar dapat diklasifikasikan ke dalam tiga kategori utama, yaitu sentimen negatif, positif, dan netral. Proses klasifikasi ini dilakukan dengan metode perhitungan sederhana, yakni menghitung skor sentimen berdasarkan jumlah kemunculan kata-kata yang terdapat dalam daftar kata positif dan negatif yang diperoleh dari repositori *GitHub*.

Skor Sentimen dari setiap komentar kemudian disimpan dalam struktur berupa *DataFrame* untuk mempermudah proses analisis lanjutan dan visualisasi hasil. Setelah seluruh data dianalisis dan diklasifikasikan, hasilnya divisualisasikan dalam bentuk diagram untuk menggambarkan distribusi sentimen secara keseluruhan. Hal ini mencerminkan bahwa sebagian besar adanya kelompok pengguna yang memberikan tanggapan yang kritis, kurang setuju atau bersifat kebencian terhadap isu polemik ijazah. Komentar dengan sentimen positif tercatat sebanyak 5518 komentar, menunjukkan pengguna yang memberikan respon positif dan memberikan dukungan pembelaan terhadap isu polemik ijazah. Sementara itu, jumlah komentar dengan sentimen netral mencapai 3510 komentar, yang mengindikasikan bahwa sebagian pengguna menyampaikan opini secara objektif atau tidak secara eksplisit mengekspresikan emosi terhadap topik tersebut.



**Gambar 17.** Hasil Lexicon Based

Berdasarkan hasil presentasi klasifikasi sentimen pada gambar 15, terlihat bahwa sentimen negatif mendominasi dengan persentase sebesar 55,5%, diikuti oleh sentimen positif sebesar 27,2% dan netral sebesar 17,3%. Hal ini menunjukkan bahwa sebagian besar komentar yang dianalisis cenderung memberikan respons yang bernada negatif terhadap isu polemik ijazah pada video *youtube Indonesia Lawyers Club*. Dominasi sentimen negatif ini mengindikasikan bahwa sentimen publik di media sosial lebih banyak diwarnai oleh kritik, ketidakpuasan, ketidaksukaan, dan kebencian dibandingkan dengan dukungan dan pandangan optimis atau sikap netral. Dengan kata lain, opini yang bersifat negatif memiliki pengaruh paling besar dalam membentuk persepsi publik di berbagai *platform* salah satunya *youtube* terkait isu tersebut.

### 3.5 Klasifikasi dengan *Support Vector Machine*

Support Vector Machine (SVM) adalah suatu teknik untuk melakukan prediksi, baik dalam kasus klasifikasi maupun regresi. SVM memiliki prinsip dasar linear classifier yaitu kasus klasifikasi yang secara linier dapat dipisahkan, namun SVM telah dikembangkan agar dapat bekerja pada problem non-linier dengan memasukkan konsep kernel pada ruang kerja berdimensi tinggi. Pada dasarnya, metode ini bekerja dengan cara mendefinisikan batas antara dua kelas dengan jarak maksimal dari data yang terdekat.

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

```
Distribusi Kelas Sebelum SMOTE:
sentiment
Negatif    11266
Positif     5518
Netral      3510
Name: count, dtype: int64

TF-IDF Shape: (20294, 5000)

Distribusi Kelas Setelah SMOTE:
sentiment
Negatif    11266
Netral     11266
Positif     11266
Name: count, dtype: int64

=== AKURASI MODEL ===
0.8687869822485207

=== CLASSIFICATION REPORT ===
              precision    recall  f1-score   support

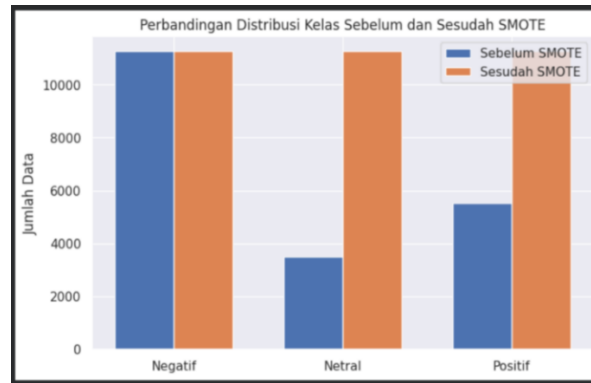
   Negatif         0.93     0.88     0.90     2253
    Netral         0.81     0.91     0.85     2253
    Positif         0.89     0.82     0.85     2254

 accuracy                   0.87     6760
 macro avg              0.87     0.87     0.87     6760
 weighted avg           0.87     0.87     0.87     6760
```

**Gambar 18.** Klasifikasi SVM dengan menggunakan metode SMOTE

#### 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 19.** Penerapan *SMOTE*

Pembobotan *SMOTE* ini dilakukan agar antara sentimen positif, negatif, dan netral bisa disetarakan. Seperti pada Gambar 19 yang dimana sebelum menggunakan *SMOTE*, perbandingan data antara sentimen positif dan negatif mencapai 3:1, setelah dilakukannya *SMOTE*, rasio data menapai 1:1 antara sentimen negatif dan sentimen positif.

### 3.5.2 Bentuk Vektor TF-IDF

Sebelum data dimodelkan menggunakan *SVM*, dilakukan ekstraksi fitur menggunakan TF-IDF (Term Frequency–Inverse Document Frequency). Teknik ini mengonversi teks menjadi representasi numerik berdasarkan frekuensi kata dan tingkat kepentingannya dalam keseluruhan dokumen. Sebelum model dilatih, teks komentar diubah menjadi representasi numerik menggunakan TF-IDF Vectorizer. Adapun hasil dari proses TF-IDF adalah TF-IDF Shape: (20.294, 5.000).

Hasil transformasi TF-IDF ditunjukkan oleh output TF-IDF Shape artinya terdapat 20.294 komentar YouTube. Setiap komentar direpresentasikan sebagai vektor 5.000 dimensi. Pemilihan 5.000 fitur bertujuan menjaga keseimbangan antara: representasi informasi, kompleksitas model, kemampuan komputasi. TF-IDF terbukti efektif untuk analisis teks karena menurunkan bobot kata-kata umum dan meningkatkan bobot kata yang lebih informatif. Setiap komentar direpresentasikan sebagai vektor dengan 5.000 fitur (kata unik). Fitur sebanyak 5.000 dipilih sebagai batas maksimum agar model tetap akurat namun efisien secara komputasi.

### 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 (*LinearSVC*), diperoleh hasil evaluasi sebagai berikut.

#### 3.6.1 Akurasi Model

Akurasi: 0,8687 (86,87%), artinya, Nilai akurasi ini tergolong tinggi mengingat: dataset berasal dari komentar media sosial yang bersifat *noisy*, jumlah kelas mencapai tiga (*multi-class classification*), banyaknya variasi bahasa informal. Hal ini menunjukkan bahwa kombinasi TF-IDF + *SMOTE* + *SVM* merupakan konfigurasi yang efektif untuk analisis sentimen komentar YouTube. *SVM* mampu memberikan prediksi sentimen yang benar sebesar 86,87% dari seluruh data uji. Nilai akurasi ini tergolong sangat baik, terutama untuk kasus *multi-class* (3 kelas) dan dataset sosial media yang memiliki karakteristik *noisy*.

#### 3.6.2 Classification Report

**Tabel 1.** *Classification report*

Sentimen	Precision	Recall	F1-Score	Support
Negatif	0,93	0,88	0,90	2253
Netral	0,81	0,91	0,85	2253
Positif	0,89	0,82	0,85	2254

##### a. Kelas Negatif

*Precision* tertinggi: 0.93 → model sangat jarang salah memprediksi kelas negatif. Prediksi kelas negatif sangat akurat.  
*Recall*: 0.88 → sebagian kecil komentar negatif yang salah prediksi/terklasifikasi. Model sangat kuat dalam mendeteksi komentar negatif.

##### b. Kelas Netral

*Recall* tertinggi: 0.91 → model sangat baik mendeteksi komentar netral. *F1-score*: 0.85 → performa stabil/seimbang antara akurasi dan kemampuan deteksi (*precision & recall*).

##### c. Kelas Positif

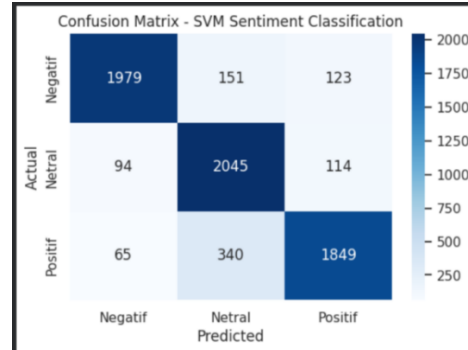
*Precision*: 0.89 → prediksi kelas Positif cukup baik/akurat. *Recall*: 0.82 → beberapa komentar positif dianggap netral.  
Secara keseluruhan:



Macro average  $F1$ -score = 0.87, menunjukkan performa model memiliki performa merata di seluruh kelas. Penerapan SMOTE terbukti meningkatkan performa kelas positif dan netral yang sebelumnya minoritas.

### 3.6.3 Analisis confusion matrix

Confusion matrix memberikan gambaran detail mengenai prediksi model terhadap label sebenarnya.



Gambar 20. Analisis Confusion Matrix

Confusion matrix adalah alat yang digunakan untuk mengevaluasi performa model klasifikasi dalam machine learning, dengan memungkinkan pengukuran hasil prediksi dalam dua atau lebih kelas. Pada gambar 20 confusion matrix menampilkan persebaran prediksi model dibandingkan label sebenarnya.

Tabel 2. Confusion Matrix

Actual \ Predicted	Negatif	Netral	Positif
Negatif	1979	151	123
Netral	94	2045	114
Positif	65	340	1849

- Kelas Negatif**  
Terdapat 1979 komentar negatif diprediksi benar. Salah diprediksi sebagai Netral (151) dan Positif (123). Model paling akurat pada kelas ini → sesuai dengan precision tinggi (0.93).
- Kelas Netral**  
Terdapat 2045 komentar netral diprediksi benar. Salah diprediksi sebagai Negatif (94) dan Positif (114). Nilai recall tertinggi (0.91), artinya model sangat baik mengenali komentar netral.
- Kelas Positif**  
Terdapat 1849 komentar positif diprediksi benar. Salah diprediksi sebagai Netral (340) → menunjukkan beberapa komentar positif dianggap moderat atau tengah oleh model. Nilai recall 0.82 cukup baik, tetapi masih bisa ditingkatkan.

Kesimpulan dari Confusion Matrix, model *Support Vector Machine* sangat kuat dalam mengenali komentar negatif dan netral. Kesalahan terbesar terjadi antara kelas Positif → Netral, karena konteks kedua kelas ini seringkali mirip dalam komentar publik. Secara keseluruhan, model memberikan performa prediksi yang stabil dan seimbang setelah penerapan SMOTE.

## 4. KESIMPULAN

Berdasarkan hasil pengujian yang telah dilakukan, dapat disimpulkan bahwa algoritma *Support Vector Machine* (SVM) berhasil diterapkan dalam analisis sentimen komentar YouTube terkait polemik ijazah Presiden ke-7 Indonesia. Data penelitian diperoleh melalui proses crawling menggunakan YouTube API v3 dan selanjutnya diproses melalui tahapan pra-pemrosesan teks yang meliputi cleaning, case folding, tokenisasi, stopword removal, dan stemming. Kombinasi metode SVM dengan teknik ekstraksi fitur TF-IDF serta penerapan SMOTE terbukti efektif dalam meningkatkan kinerja model, khususnya dalam menangani ketidakseimbangan data. Model yang dibangun mampu mencapai tingkat akurasi sebesar 86,87%, yang menunjukkan kemampuan klasifikasi opini publik yang baik. Hasil evaluasi juga menunjukkan bahwa model lebih optimal dalam mengidentifikasi sentimen negatif dan netral, sementara sebagian sentimen positif masih cenderung diklasifikasikan sebagai netral. Secara keseluruhan, penelitian ini membuktikan bahwa analisis sentimen berbasis media sosial dapat digunakan sebagai alat yang efektif untuk memetakan persepsi publik terhadap isu sosial dan politik secara otomatis, cepat, dan dalam skala yang luas.

## 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/algoritm.v5i2.9692.
- [3] 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.
- [4] M. I. Sultan and M. Akbar, "Analisis Sentimen Pemecatan Jokowi Pada Komentar Publik YouTube Tempo.co," vol. 07, no. 02, pp. 125–140, 2025.
- [5] 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.
- [6] 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.
- [7] 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.
- [8] 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.
- [9] 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.
- [10] 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.
- [11] 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.
- [12] 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.
- [13] 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>
- [14] 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.
- [15] A. M. Putra, Candra Saputra, Rahmadden, Safiril 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.
- [16] 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.
- [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] 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.
- [19] K. E. V Sihombing, M. A. I. Pakereng, and U. Kristen, "Analisis Sentimen Komentar Youtube Terhadap Wawancara Presiden Prabowo Menggunakan Machine Learning Dan Orange Data Mining," *STORAGE: Jurnal Ilmiah Teknik dan Ilmu Komputer*, vol. 4, no. 4, pp. 327–336, 2025. doi: 10.55123/storage.v4i4.6478
- [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," *JUPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.)*, vol. 9, no. 3, pp. 1464–1472, 2024, doi: 10.29100/jupi.v9i3.5527.