

Klasifikasi Sentimen Komentar Youtube Demo DPR RI Menggunakan Support Vector Machine

Siti Aulia Rahmadhani¹, Lia Dwi Rusanti², Harun Al Rosyid³

^{1,2,3}Program Studi Pendidikan Teknologi Informasi, Universitas Negeri Surabaya, Surabaya, Indonesia

Email : siti.22038@mhs.unesa.ac.id, liadwi.22010@mhs.unesa.ac.id, harunrosyid@unesa.ac.id

Article Information

Article history

Received 5 October 2025

Revised 9 November 2025

Accepted 30 November 2025

Available 30 December 2025

Keywords

Classification
Sentiment Analysis
Youtube Comments
Support Vector Machine
DPR RI

Corresponding Author:

Name of corresponding author,
Affiliation,
Email :
siti.22038@mhs.unesa.ac.id

Abstract

Demonstrations against the Indonesian House of Representatives (DPR RI) have triggered extensive public opinion flows on social media; however, sentiment mapping of Indonesian-language comments on YouTube live broadcasts of political issues still requires more structured methodological reporting and evaluation. This study aims to classify public sentiment from 1,493 YouTube comments related to DPR RI demonstrations using the Support Vector Machine (SVM) algorithm. Data were collected via the YouTube Data API and subsequently processed through text cleaning, case folding, normalization, tokenization, stopword removal, and stemming. Sentiment labeling was performed using an Indonesian lexicon-based approach to generate three sentiment classes (positive, negative, and neutral), with neutral sentiment being dominant. Feature representation was constructed using CountVectorizer, and the SVM model was trained using an 80:20 split for training and testing data. Model performance was evaluated using accuracy, precision, recall, and F1-score metrics, achieving an accuracy of 92.4% (weighted performance of 0.924). Word frequency analysis was also employed to identify dominant terms within each sentiment class. These findings demonstrate the effectiveness of SVM in mapping digital public perceptions on political issues and highlight its potential to support data-driven policy evaluation.

Keywords : *Classification, Sentiment Analysis, YouTube Comments, Support Vector Machine, DPR RI*

Abstrak

Demonstrasi terhadap Dewan Perwakilan Rakyat Republik Indonesia (DPR RI) memicu arus opini publik di media sosial, namun pemetaan sentimen pada komentar YouTube siaran langsung isu politik berbahasa Indonesia masih memerlukan pelaporan metode dan evaluasi yang lebih terstruktur. Penelitian ini bertujuan mengklasifikasikan sentimen publik pada 1.493 komentar YouTube terkait demonstrasi DPR RI menggunakan Support Vector Machine (SVM). Data dikumpulkan melalui YouTube Data API, kemudian diproses melalui cleaning, case folding, normalisasi, tokenisasi, stopword removal, dan stemming. Pelabelan sentimen dilakukan menggunakan metode leksikon Bahasa Indonesia untuk menghasilkan tiga kelas (positif, negatif, netral), dengan dominansi sentimen netral. Representasi fitur dibentuk menggunakan CountVectorizer, lalu model SVM dilatih dengan pembagian data latih dan uji 80:20. Kinerja model dievaluasi menggunakan akurasi, presisi, recall, dan F1-score, dengan akurasi mencapai 92,4% (weighted performance 0,924). Analisis frekuensi kata turut digunakan untuk mengidentifikasi kata-kata dominan pada tiap kelas sentimen. Temuan ini menegaskan efektivitas SVM untuk pemetaan persepsi publik digital pada isu politik dan berpotensi mendukung evaluasi kebijakan berbasis data.

Kata Kunci : *Klasifikasi, Analisis Sentimen, Komentar Youtube, Support Vector Machine, DPR RI*

Copyright@2025 Siti Aulia Rahmadhani, Lia Dwi Rusanti, Harun Al Rosyid

This is an open access article under the [CC-BY-NC-SA](https://creativecommons.org/licenses/by-nc-sa/4.0/) license.



1. Pendahuluan

Negara Kesatuan Republik Indonesia (NKRI) mendasarkan sistem demokrasi pada prinsip kedaulatan rakyat sehingga secara kontekstual kepentingan masyarakat harus disuarakan dan di akomodasi melalui jalur legislatif, yang diwakili oleh Dewan Perwakilan Rakyat (DPR) dari tingkat daerah hingga tingkat pemerintahan pusat (Asbanu et al., 2025). DPR sebagai Lembaga legislatif memiliki peran pengambilan keputusan, pembentukan kebijakan, dan pengawasan terhadap pemerintah (Jelly & Prihana, 2025). Setiap tindakan serta kebijakan yang dihasilkan sering menjadi sorotan publik dan memicu berbagai reaksi masyarakat. Dalam praktiknya, kebijakan dan tindakan lembaga publik kerap memunculkan respons pro–kontra; pada era digital, respons tersebut semakin terekam sebagai jejak opini publik di platform media sosial dan berpotensi diukur secara komputasional untuk memahami dinamika sikap masyarakat (O'Connor et al., 2010).

Interaksi manusia diperluas dengan kemajuan teknologi sehingga tidak hanya terbatas pada komunikasi langsung, tetapi juga melalui media sosial (Arsini et al., 2023). Media sosial membuka peluang bagi setiap penggunanya untuk saling bertukar informasi, dan memberikan kebebasan bagi mereka untuk memilih, merumuskan, serta menyampaikan pesan, informasi, atau pendapat sesuai dengan keinginannya (Jamilah & Wahyuni, 2020). Seiring dengan banyaknya jejaring sosial seperti Facebook, Instagram, Twitter, Youtube, dan banyak lagi membawa keuntungan bagi masyarakat dalam aspek kemudahan mengakses informasi, kehadiran YouTube memberikan dampak terhadap cepatnya laju penyebaran informasi yang disajikan dalam bentuk audio visual, dimanfaatkan oleh banyak pihak sebagai wadah untuk branding dan publikasi (Zhafira et al., 2021). Di dalam komentar YouTube, pengguna dapat menyampaikan pendapat, kritik, atau dukungan terhadap isu yang sedang populer. Hal ini juga diperkuat dengan sistem demokrasi Indonesia dimana kebebasan berpendapat merupakan hak yang diakui oleh negara tercantum dalam pasal 28 ayat (3) UUD 1945. Komentar YouTube pada konten berita terkait DPR RI berisi informasi yang mencerminkan persepsi, kritikan, atau dukungan masyarakat terhadap tindakan dan kebijakan DPR sehingga jika dianalisis dapat dijadikan evaluasi pemerintah.

Pada peristiwa demonstrasi di depan Gedung DPR RI, percakapan publik di media sosial YouTube berlangsung masif dan cepat. Dilihat pada komentar *live* YouTube Liputan6 berjudul “Situasi Terkini Demo di Depan Gedung DPR RI” menunjukkan bahwa satu video menghasilkan 1.453 komentar berbahasa Indonesia yang mempresentasikan opini spontan masyarakat. Jumlah komentar tersebut mencerminkan tingginya perhatian dan keterlibatan masyarakat sekaligus menunjukkan bahwa opini publik dalam ruang digital berkembang dalam skala besar, beragam dan bersifat tidak terstruktur. Hal ini didukung oleh data platform global yang dilaporkan oleh (DataReportal, 2025) menunjukkan bahwa YouTube mencapai jangkauan lebih dari

2,53 miliar pengguna secara global, artinya platform ini merupakan salah satu sumber data terbesar dengan pengguna aktif berjumlah milyaran orang di seluruh dunia.

Dari perspektif data, media sosial menjadi sumber data yang kaya, menyediakan data berkapasitas besar dan *real-time* (Hou et al., 2021). Karakteristik data yang dihasilkan media sosial yaitu volume data yang besar dan tidak terstruktur (Octiva et al., 2024). Analisis opini secara manual tidak efisien maka untuk menganalisis kompleksitas data tersebut, *Machine learning* menjadi salah satu teknologi dalam mendukung analisis sentimen, teknologi ini memungkinkan sistem belajar dari data dan menghasilkan prediksi atau klasifikasi secara otomatis (Siswoyo & Utomo, 2025). Secara konseptual, analisis sentimen (*sentiment analysis*) mengelompokkan opini berbasis teks menjadi polaritas tertentu (mis. positif, netral, negatif) untuk menghasilkan ringkasan kuantitatif yang dapat ditindaklanjuti (Pang & Lee, 2008; Junaedi et al., 2024). Kemampuan analisis sentimen telah digunakan di berbagai sektor seperti prediksi pasar keuangan, isu kesehatan, pemasaran, politik, dan manajemen keadaan darurat karena kemampuannya mengukur opini publik (Cas´anez-Ventura & Cuenca-Jim´enez, 2023).

Analisis sentimen dipilih dalam penelitian ini karena kemampuannya dalam menyajikan insight kuantitatif terhadap opini, dengan cara mengklasifikasikan opini yang diekspresikan dalam teks menjadi kategori positif, netral, atau negatif (Junaedi et al., 2024). Kemampuan analisis sentimen telah digunakan di berbagai sektor seperti prediksi pasar keuangan, isu kesehatan, pemasaran, politik, dan manajemen keadaan darurat karena kemampuannya mengukur opini publik (Cas´anez-Ventura & Cuenca-Jim´enez, 2023).

Peneliti memfokuskan penggunaan algoritma *Support Vector Machine* sebab dikenal efektif menangani data dengan dimensi tinggi pada komentar YouTube, mampu mencapai akurasi hingga 86,33% dalam studi terdahulu (Asrianto & Herwinanda, 2022). Keberhasilan penerapan SVM dalam menganalisis sentimen masyarakat terhadap isu-isu publik di komentar Youtube juga diperkuat oleh berbagai penelitian lain (Riwanto et al., 2025) (Hidayat et al., 2024). Penelitian lain juga menunjukkan bahwa penggunaan SVM kombinasi dengan TF-IDF menghasilkan performa yang sangat baik dalam klasifikasi komentar media sosial dengan F1-score mencapai 0,92, sehingga memperkuat bukti bahwa SVM konsisten efektif dalam menganalisis opini masyarakat pada isu-isu publik (Wahyuni et al., 2024). Berdasarkan hasil penelitian terdahulu tersebut, Penelitian ini mengusulkan klasifikasi sentimen oleh model SVM (positif, netral, negatif). SVM digunakan karena mampu mengolah data teks yang berdimensi tinggi tanpa mengalami penurunan performa (Anadas et al., 2025).

Ruang lingkup penelitian mencakup implementasi algoritma SVM dan ekstraksi fitur *CountVectorizer* dengan fokus utama menganalisis polaritas sentimen terhadap isu kebijakan DPR RI dengan sumber data yang dikumpulkan dari komentar video siaran langsung pada kanal YouTube Liputan6 yang berjudul “Situasi Terkini Demo di Depan

Gedung DPR RI”. Hasil klasifikasi akan dievaluasi menggunakan matriks akurasi, presisi, *recall*, dan *F1-score*, untuk menganalisis opini publik yang dapat dijadikan sebagai bahan evaluasi kebijakan DPR RI. Rumusan masalah dalam penelitian ini adalah bagaimana hasil pemetaan sentimen opini publik terhadap isu demonstrasi di depan Gedung DPR RI yang terekam dalam komentar YouTube secara objektif. Serta sejauh mana algoritma *Support Vector Machine (SVM)* dengan ekstraksi fitur *CountVectorizer* efektif mengklasifikasikan komentar berbahasa Indonesia ke dalam kategori sentimen positif, netral, dan negatif.

Nilai kebaruan penelitian ini terletak melalui kombinasi algoritma SVM dengan ekstraksi fitur *CountVectorizer* pada data teks berbahasa Indonesia yang berdimensi tinggi dan informal yaitu komentar YouTube. Kombinasi metode ini bertujuan untuk memodelkan dan mengklasifikasikan opini publik yang sensitif dan tidak terstruktur. Penelitian ini memberikan manfaat karena hasil analisis sentimen diharapkan dapat berfungsi sebagai model dan bahan evaluasi objektif berbasis data bagi lembaga legislatif ataupun pemerintah dalam memahami persepsi antara kebijakan dengan aspirasi masyarakat.

2. Kajian Terdahulu

Analisis sentimen pada media sosial telah banyak dimanfaatkan secara global untuk memahami dinamika opini publik, khususnya dalam hal tanggapan masyarakat terhadap kebijakan dan isu sosial negara. Media sosial dipandang sebagai sarana penting bagi ekspresi publik terhadap isu-isu kebijakan karena menyediakan ruang bagi masyarakat untuk menyampaikan pendapat dan sikap secara terbuka, sekaligus memberikan gambaran persepsi warga kepada kebijakan pemerintah (Rizky & Alrasyid, 2025). Di antara platform media sosial, YouTube menjadi salah satu media penting yang memungkinkan masyarakat mengekspresikan pandangan mereka secara terbuka dan langsung terhadap berbagai inisiatif pemerintah (Septiana et al., 2025).

Penelitian terkait analisis sentimen komentar YouTube menggunakan *Support Vector Machine (SVM)* telah banyak dilakukan untuk memahami opini publik terhadap berbagai isu, mulai dari kebijakan pendidikan, program sosial, hingga respon terhadap konten digital. (Wahyuni et al., 2024) menganalisis komentar YouTube terkait kenaikan Uang Kuliah Tunggal (UKT) dan mengklasifikasikan komentar menjadi positif, netral, dan negatif. Penelitian ini menunjukkan bahwa SVM mampu mencapai akurasi 0,88 serta presisi 0,83, *recall* 0,90, dan *F1-score* 0,86, membuktikan bahwa SVM efektif untuk mengklasifikasikan opini publik pada isu kebijakan pendidikan. Penelitian lain oleh (Santoso et al., 2025) menggunakan dataset 15.257 komentar YouTube pada kanal politik, dengan tahapan *preprocessing* seperti *case folding*, normalisasi slang, tokenisasi, *stopword removal*, dan *stemming*. Hasil menunjukkan bahwa SVM tetap unggul dibanding Naïve Bayes dan Random Forest, dengan akurasi hingga 84,14%.

Chairani meneliti komentar dari kanal YouTube “*beauty vlogger*” menerapkan algoritma SVM dipadukan TF-IDF untuk ekstraksi fitur (Chairani Siregar et al., 2024), hasil analisis diperoleh bahwa SVM efektif mengklasifikasikan komentar informal dan *user-generated*. Sedangkan (Hidayat et al., 2024) meneliti komentar YouTube terkait isu pengungsi Rohingya dan menemukan akurasi 77%, *recall* 100%, dan presisi 77%, yang membuktikan kemampuan SVM dalam menangani data komentar sosial-politik yang sensitif. Penelitian lain yang menganalisis 13.692 komentar YouTube pada video promosi bisnis kuliner menunjukkan bahwa SVM lebih unggul dibanding Naïve Bayes dengan akurasi 86,55%, memperlihatkan konsistensi performa SVM pada dataset besar dan heterogen (Ahfaz et al., 2025).

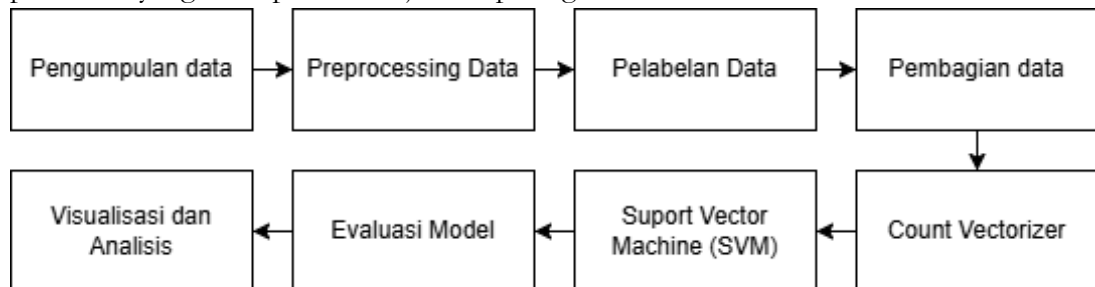
Meskipun berbagai penelitian tersebut menunjukkan efektivitas SVM dalam klasifikasi komentar YouTube, sebagian besar masih berfokus pada pendidikan, hiburan, atau konten komersial, sedangkan penelitian yang secara spesifik menilai opini publik terhadap lembaga legislatif dan kebijakan pemerintah, seperti DPR RI, masih sangat terbatas. Penelitian ini memanfaatkan SVM dan *CountVectorizer* untuk menganalisis komentar YouTube terkait DPR RI dengan tujuan melengkapi penelitian terdahulu sekaligus memberikan gambaran yang lebih spesifik mengenai persepsi masyarakat terhadap kebijakan legislatif. Pendekatan ini diharapkan tidak hanya menghasilkan klasifikasi yang akurat, tetapi juga dapat menjadi dasar evaluasi kebijakan berbasis data opini masyarakat.

Temuan-temuan tersebut konsisten dengan penelitian lain yang menekankan bahwa normalisasi slang/variasi bahasa informal dan perancangan preprocessing yang tepat dapat meningkatkan kinerja SVM pada komentar YouTube (Saputra et al., 2025), serta bahwa penerapan SVM juga sering digunakan pada skenario ulasan berbasis platform digital di Indonesia dengan variasi sumber data dan skema evaluasi (Muhayat et al., 2023; Nursalim & Novita, 2023; Putri et al., 2025). Meskipun demikian, sebagian studi masih berfokus pada topik non-kebijakan atau non-institusional, sementara studi bertopik kebijakan/isu publik sering beralih ke model transformer untuk menangkap nuansa bahasa yang lebih kompleks (Yulita et al., 2023).

Berdasarkan pemetaan tersebut, dapat disimpulkan bahwa penelitian analisis sentimen komentar YouTube dengan SVM sudah banyak dilakukan pada beragam domain, tetapi kajian yang secara spesifik mengkaji sentimen komentar YouTube terkait Demonstrasi DPR RI masih jarang ditemukan dalam rujukan yang ditelaah. Oleh sebab itu, penelitian ini diposisikan untuk mengisi gap domain tersebut dengan memanfaatkan SVM sebagai baseline terukur, sekaligus menekankan perlunya pelaporan yang kuat (desain validasi, metrik per kelas, dan analisis kesalahan) agar kontribusinya relevan secara ilmiah dan dapat dibandingkan dengan studi sebelumnya (Mao et al., 2024; Rodríguez-Ibáñez et al., 2023).

3. Metodologi Penelitian

Dalam penelitian ini, komentar Youtube terhadap video siaran langsung Liputan6 dianalisis dengan menerapkan algoritma *Support Vector Machine (SVM)*. Alur penelitian yang diterapkan ditunjukkan pada gambar 1.



Gambar 1. Alur Penelitian

Pada tahap pengumpulan data, data diperoleh dari kolom komentar pada video siaran langsung yang diunggah oleh kanal YouTube Liputan6 yang berjudul “Situasi Terkini Demonstrasi di Depan Gedung DPR RI” dan ditayangkan secara *live* pada tanggal 30 Agustus 2025. Pengumpulan data dilakukan melalui teknik *Web Scraping*, yaitu proses mengambil data secara otomatis dari laman daring menggunakan *software* (Khder, 2021). YouTube Data API v3 dari *Google Cloud Console* dimanfaatkan untuk mendukung proses pengumpulan data dalam penelitian ini. Dilakukan untuk mengakses *endpoint* komentar, mengambil data secara terstruktur, dan menghasilkan dataset dalam format CSV.

Pada tahap preprocessing data, tahapan *preprocessing* berfungsi untuk meningkatkan kualitas teks sebelum dianalisis. Data yang telah diperoleh kemudian diolah untuk membuang data yang tidak berguna. Tahapan ini terdiri dari proses (1) *Cleaning* untuk menghapus angka, emoji, simbol, tanda baca, dan karakter tidak relevan. (2) *Case folding*, teks dikonversi ke bentuk huruf kecil. (3) Normalisasi kata tidak baku yang banyak muncul di komentar YouTube agar sesuai dengan bentuk formal bahasa Indonesia. (4) *Tokenizing* yaitu setiap kalimat diuraikan menjadi komponen kata sehingga menghasilkan kumpulan token. (5) *Stopword Removal* untuk pembuangan kata-kata umum (*stopword*) seperti konjungsi. (6) *Stemming* untuk mengembalikan kata ke bentuk dasarnya sehingga memudahkan proses pelabelan dan klasifikasi.

Pada tahap pelabelan data, setelah data teks dibersihkan setiap komentar diberikan label sentimen menggunakan metode leksikon Bahasa Indonesia. Pada tahap ini digunakan daftar kata bermakna positif dan negatif untuk mengidentifikasi kecenderungan sentimen pada setiap komentar. Proses pelabelan dilakukan dengan menghitung kemunculan kata positif dan negatif pada setiap dokumen sehingga menghasilkan tiga klasifikasi meliputi sentimen positif, negatif, dan netral. Proses pelabelan data sentimen dilakukan menggunakan pendekatan *lexicon-based*, dengan

memanfaatkan daftar kata positif (L_{pos}) dan negatif (L_{neg}) yang disusun secara manual. Setiap komentar dianalisis dengan menghitung jumlah kemunculan kata positif dan kata negatif dalam teks. Setiap komentar direpresentasikan sebagai himpunan kata $W = \{w_1, w_2, \dots, w_n\}$. Jumlah kata positif (Pos) dan kata negatif (Neg) dihitung menggunakan fungsi indikator sebagai berikut:

$$Pos = \sum_{i=1}^n I(w_i \in L_{pos})$$

$$Neg = \sum_{i=1}^n I(w_i \in L_{neg})$$

I merupakan fungsi indikator yang bernilai 1 jika kondisi terpenuhi dan 0 jika tidak. Penentuan label sentimen dilakukan berdasarkan aturan berikut:

$$Sentimen = \begin{cases} Positif, & \text{jika } Pos > Neg \\ Negatif, & \text{jika } Neg > Pos \\ Netral, & \text{jika } Pos = Neg \end{cases}$$

Hasil perbandingan jumlah kata tersebut digunakan untuk menentukan label sentimen, tanpa melibatkan bobot numerik atau nilai ambang (*threshold*).

```

positif = [
    'dukung','mendukung','setuju','sepakat','baik','bagus','mantap','mantul',
    'hebat','keren','top','luarbiasa','terbaik','sukses','maju','positif',
    'cocok','salut','hormat','kagum','berkah','adil','jujur','benar','bijak',
    'sabar','istiqomah','semangat','lanjutkan','gaskeun','ayo','semoga',
    'berhasil','doa','sukses'
]

negatif = [
    'tolak','bakar','lenyapkan','bubar','bubarkan','marah','kesal','jengkel','benci','muak',
    'jijik','kecewa','sakit','sakit hati','goblok','tolol','bodoh','goblok',
    'bego','brengsek','bangsat','najis','kampret','korupsi','curang','maling',
    'mafia','busuk','rusak','bobrok','hancur','kotor','buruk','kacau','chaos',
    'ricuh','rusuh','anarkis','berbahaya','ancam','tembak','gas','pukul',
    'kekerasan','brutal','gagal','fatal','jelek','masalah','lemah','hina',
    'rendah'
]

def get_sentiment(text):
    words = text.split()
    pos = sum(1 for w in words if w in positif)
    neg = sum(1 for w in words if w in negatif)

    if pos > neg:
        return "Positif"
    elif neg > pos:
        return "Negatif"
    else:
        return "Netral"

```

Gambar 2. Pseudocode Labelling dengan *lexicon-based*

Gambar 2 diatas menunjukkan implementasi langsung dari rumus dan pseudocode yang digunakan dalam penelitian ini. Daftar leksikon positif dan negatif disusun secara manual berdasarkan konteks komentar YouTube. Fungsi `get_sentiment()` menghitung jumlah kata positif dan negatif, kemudian menentukan label sesuai aturan *lexicon-based*.

Pada tahap pembagian data, dataset yang sudah melalui proses pelabelan kemudian dibagi ke dalam dua bagian, yaitu 80% untuk data latih dan 20% untuk data uji. Pemisahan data dilakukan dengan memastikan setiap kelas memiliki proporsi yang seimbang untuk menjaga keseimbangan proporsi kelas pada masing-masing subset data. Pembagian dataset ini bertujuan menilai kemampuan generalisasi model dalam mengklasifikasikan komentar yang belum pernah dilihat sebelumnya.

Pada tahap *CountVectorizer*, mengkonversi data teks menjadi bentuk fitur numerik menggunakan *CountVectorizer*. Proses ini menghasilkan matriks fitur berdimensi tinggi berdasarkan frekuensi kata di tiap dokumen. Hasil ekstraksi fitur inilah kemudian digunakan sebagai masukan bagi algoritma *Support Vector Machine* dalam proses pelatihan dan prediksi.

Pada tahap *Support Vector Machine*, setelah data direpresentasikan dalam bentuk fitur, dilakukan pelatihan model *Support Vector Machine* (SVM) kernel linear. SVM adalah algoritma klasifikasi berfungsi membagi data menjadi kelas berbeda dengan cara menentukan hyperplane terbaik, dikenal lebih konsisten dalam menangani data berdimensi tinggi dan cenderung menghasilkan akurasi lebih baik (Permana et al., 2025). Model dilatih menggunakan data latih untuk mempelajari pola relasi antara kata dan sentimen. Selanjutnya dilakukan prediksi pada data uji untuk mengukur performa model.

Pada tahap evaluasi model, evaluasi model dilakukan melalui pengukuran akurasi, perhitungan classification report, serta visualisasi confusion matrix untuk melihat tingkat kesalahan prediksi pada masing-masing kelas. *Confusion matrix* secara visual menunjukkan bahwa model memiliki tingkat keberhasilan klasifikasi paling tinggi pada kategori Netral, diikuti kategori Negatif dan Positif.

Pada tahap visualisasi dan analisis, visualisasi model dilakukan melalui *confusion matrix* dan *Classification Report* untuk melihat tingkat kesalahan prediksi pada masing-masing kelas.

4. Hasil dan Pembahasan

Sumber data pada penelitian ini berasal dari platform YouTube dengan menggunakan video ID Demonstrasi DPR RI. Proses pengumpulan data dilakukan melalui YouTube Data API yang diakses menggunakan API key dari *Google Cloud Console*. Tahap pengambilan data menghasilkan total 1.453 komentar dari pengguna YouTube yang kemudian disimpan ke dalam file berekstensi .csv. Seluruh data mentah selanjutnya menjalani tahap *preprocessing* untuk mengatasi masalah kualitas data seperti kata tidak lengkap, inkonsistensi, *noise*, redundansi, dan duplikasi, jika tidak diselesaikan informasi yang akan diekstraksi dari data tidak akurat (ÇETİN & YILDIZ, 2022). Tahapan ini terdiri dari beberapa proses yaitu proses *Cleaning*, penghapusan karakter tidak diperlukan dengan menghapus angka, punctuation, emoji serta simbol-simbol tertentu. Kemudian proses *Case Folding* guna menyeragamkan teks ke format huruf kecil (*lowercase*) supaya menghindari perbedaan arti antara kata yang sama namun memiliki kapitalisasi berbeda. Dilanjutkan proses Normalisasi yaitu mengubah penggunaan bahasa informal atau singkatan, tahap ini memiliki representasi standar sesuai kamus bahasa Indonesia. Selanjutnya pada proses *Tokenizing*, pemecahan kalimat menjadi kata. Kemudian pada proses *Stopword Removal* menggunakan pustaka Sastrawi untuk menghilangkan kata-kata

penghubung atau preposisi yang tidak memberikan nilai polaritas signifikan. Proses terakhir *Stemming*, mengembalikan kata ke bentuk dasar dari setiap token. Hasil tahap *preprocessing* data dicantumkan pada tabel 1.

Tabel 1. Hasil Tahap Preprocessing

Tahapan	Hasil
Data awal	Tuntaskn K, zolimn. Hcrkan, Kbatiln, Tuk Org2 Yg Pninds Rkyt, Wahai Pjuang Rkyt Yg Mulia,
Cleaning	Tuntaskn K zolimn Hcrkan Kbatiln Tuk Org Yg Pninds Rkyt Wahai Pjuang Rkyt Yg Mulia
Case folding	tuntaskn k zolimn hcrkan kbatiln tuk org yg pninds rkyt wahai pjuang rkyt yg mulia
Normalisasi	tuntaskn ke zolimn hcrkan kbatiln untuk orang yang pninds rkyt wahai pjuang rkyt yang mulia
Tokenizing	tuntaskn,ke,zolimn,hcrkan,kbatiln,untuk,orang,yang,pninds,rkyt,wahai,pjuang,rkyt,yang,mulia
Stopword Removal	tuntaskn,zolimn,hcrkan,kbatiln,orang,pninds,rkyt,pjuang,rkyt,mulia

Catatan: Hasil data tahapan *preprocessing* pada tabel hanya sampel dari satu komentar Youtube

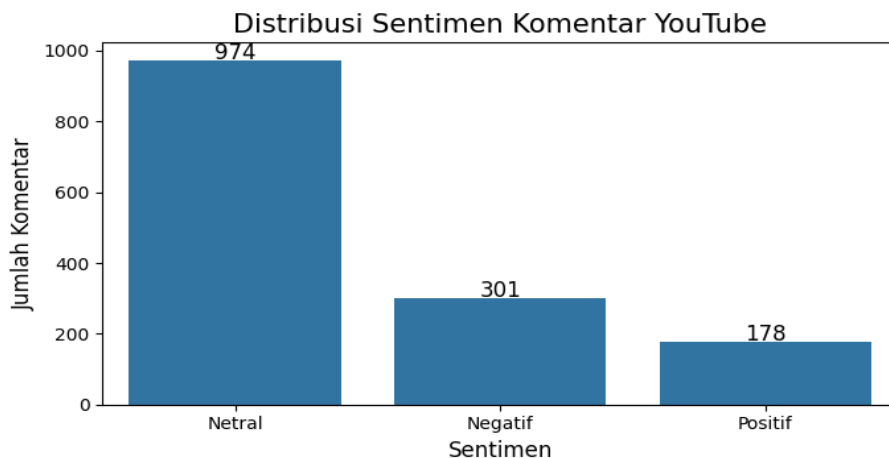
Setelah tahap *preprocessing* selesai, data kemudian diberi label menggunakan pendekatan *lexicon-based* Leksikon sentimen yang digunakan dalam penelitian ini disusun secara manual oleh peneliti berdasarkan daftar kata bermuatan positif dan negatif dalam bahasa Indonesia. Proses pelabelan dilakukan dengan menghitung jumlah kemunculan kata positif dan kata negatif pada setiap komentar. Apabila jumlah kata positif lebih dominan, komentar diberi label positif; jika kata negatif lebih dominan, diberi label negatif; dan apabila tidak ditemukan dominasi keduanya, komentar diklasifikasikan sebagai netral. Pendekatan ini tidak menggunakan pembobotan numerik maupun nilai ambang tertentu (*threshold*), sehingga penentuan label sepenuhnya bergantung pada aturan leksikal yang telah ditetapkan. Hasil akhirnya komentar terbagi atas tiga jenis sentimen, yaitu positif, negatif, dan netral. Tabel 2 menunjukkan hasil pembagian sentimen setiap kalimat yang kemudian akan dianalisis distribusinya:

Tabel 2. Hasil Pelabelan Data

	author	comment	publishedAt	sentiment
0	@murniatiagustina8556	Tuntaskn K, zolimn. Hcrkan, Kbatiln, Tuk Org2 Yg Pninds Rkyt, Wahai Pjuang Rkyt Yg Mulia,	2025-09-14T19:45:36Z	Netral

	author	comment	publishedAt	sentiment
1	@murniatiagustina8556	Wahai Saudara2ku Pejuang K, adiln Alloh, Sllu Mlindungn Stiap Langkh Gerak Kki Klian. Amien, Stiap Tets Kringt Jerh Payh Klian Berbuah Mans, Aml Phala Klian Slmt Dunia And Akherat,	2025-09-14T19:31:50Z	Netral
2	@murniatiagustina8556	Ass, mualaikm Wr Wb, Bissmillah, Irohman, nirohim, Wahai Saudara2ku Smga Klian Slmt, Dri Sgla Musbh, Dmi Berjuang Tuk K, bebrn	2025-09-14T19:25:56Z	Netral
3	@agusdemon659	segera bubarkan dpr yang sekarang ,, ganti dengan gen z	2025-09-14T15:35:11Z	Negatif
4	@Agoengriendyanto	Semua saudaraku dr pihak manapun sesama anak CC nb.adam jg dr baik", minimalisirkan cedera dn mg TDK terjadi adanya korban jw. Amin y robal alamin...	2025-09-14T12:52:33Z	Netral
...	...	...	...	...
1448	@wahyuhhasbi6916	rakyat skrng di bungkam dgn tdk ada nya udd informasi dgn di hilang kan live di tiktok sehingga membuat masyarakat tambah marah	2025-08-30T14:55:43Z	Negatif
1449	@Lilaintana	Semoga selamat semua rakyat jelata. Hidup rakyatt	2025-08-30T14:55:41Z	Netral
1450	@IkikChanel2	Rekaman ini	2025-08-30T14:55:24Z	Netral
1451	@FikriFikri-s8b	Tiktok error	2025-08-30T14:55:14Z	Netral
1452	@septiyanawulan3402	Ayooo bakar	2025-08-30T14:54:57Z	Netral

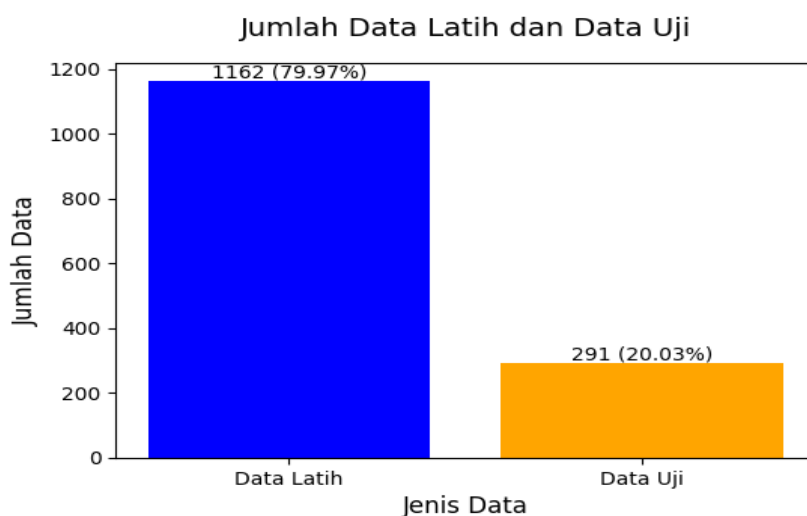
Fungsi pelabelan data setiap sentimen tabel 2 supaya mendukung upaya penelitian untuk menafsirkan bagaimana penilaian publik melalui komentar YouTube demonstrasi DPR RI.



Gambar 3. Hasil Distribusi Jumlah Sentimen

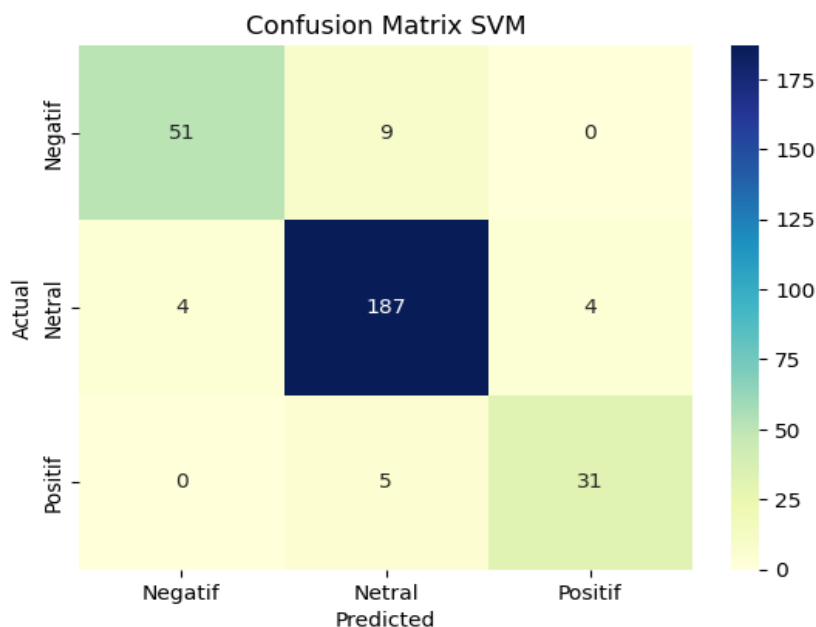
Berdasarkan hasil pelabelan, distribusi sentimen menunjukkan dominansi pada sentimen netral. Hal ini dapat dijelaskan oleh karakteristik pengguna yang memberikan tanggapan dalam bentuk pengamatan situasi, informasi, doa, maupun pernyataan yang minim ekspresi emosional yang eksplisit. Selain itu, penggunaan pendekatan leksikon menyebabkan komentar yang tidak mengandung kata bermuatan sentimen yang jelas cenderung diklasifikasikan sebagai netral.

Sebelum melakukan pemodelan, terlebih dahulu memisahkan data menjadi dua yakni, latih dan uji. Persentase pembagian yaitu data latih 80% dan data uji 20%. Dari total 1.453 data, 1.162 data ditetapkan sebagai data latih, sementara 291 sisanya sebagai data uji. Data latih berfungsi membangun dan mengajarkan SVM agar mampu mengenali pola sentimen, sementara data uji dipakai untuk menilai dan mengukur sejauh mana model SVM yang telah dilatih mampu melakukan klasifikasi dengan baik.



Gambar 4. Hasil Jumlah Data Latih dan Data Uji

Setelah model dilatih kemudian dievaluasi menggunakan *Confusion Matrix* dan *Classification Report*. Gambar 5 menunjukkan *Confusion Matrix* sedangkan gambar 6 menunjukkan *Classification Report*. *Confusion Matrix* efektif untuk menganalisis klasifikasi multi-kelas untuk menunjukkan distribusi prediksi benar dan salah pada setiap kelas sentiment (Heydarian et al., 2022). SVM menunjukkan kemampuan klasifikasi yang kuat terhadap kelas netral dengan 187 prediksi benar dari total 291 data uji kelas tersebut. Pada kelas positif, SVM mampu mengklasifikasikan 31 data dengan benar tetapi masih terjadi kesalahan prediksi ke kelas netral sebanyak 5 data. Sehingga tantangan utama SVM, terletak pada pembedaan antara sentimen positif dan netral.



Gambar 5. Hasil Confusion Matrix

Evaluasi selanjutnya menggunakan *Classification Report* yang memuat matriks *Accuracy*, *Precision*, *Recall*, dan *F1-Score*, matriks tersebut penting sebagai indikator dalam menilai kualitas model dalam tugas klasifikasi (Kusuma & Cahyono, 2023). Dari evaluasi diperoleh, SVM berhasil mencapai performa sangat baik dalam mengklasifikasikan sentimen dengan tingkat akurasi sebesar 92.4% dari total 291 sampel data uji. Analisis per kelas menunjukkan SVM sangat baik melakukan klasifikasi sentimen Netral (*F1-Score* 0.944) dan Negatif (*F1-Score* 0.887). Kelas negatif memiliki presisi tertinggi (0.927), yang menunjukkan bahwa hampir semua prediksi negatif oleh SVM adalah benar. Sementara itu, SVM menunjukkan tantangan terbesar pada klasifikasi kelas positif, yang memiliki nilai *Recall* terendah (0.861) dan *F1-Score* terendah (0.873). *Recall* rendah ini mengindikasikan bahwa SVM masih gagal mengidentifikasi sebagian komentar yang seharusnya berlabel positif. Kesenjangan kinerja ini diperkirakan terjadi akibat ketidakseimbangan data (*imbalanced data*), di mana jumlah sampel positif jauh lebih kecil

dibandingkan sampel netral serta keterbatasan fitur berbasis frekuensi kata yang digunakan.

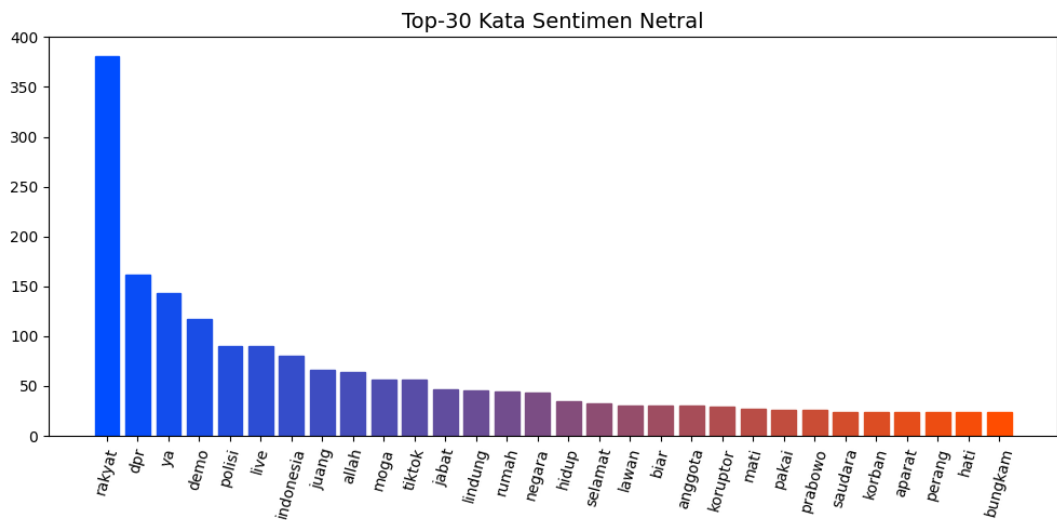
Keterbatasan pendekatan *lexicon-based* juga berkontribusi terhadap kesalahan klasifikasi. Model kurang mampu menangkap variasi bahasa kompleks seperti sarkasme, negasi, bahasa slang, dan *code-mixing* yang umum ditemukan dalam komentar YouTube. Komentar seperti “Wahai Saudara2ku Pejuang K, adiln Alloh, Sllu Mlindungan...” cenderung diklasifikasikan sebagai netral karena tidak mengandung kata positif eksplisit, meskipun sebenarnya menunjukkan dukungan. Komentar provokasi seperti “Ayooo bakar” diklasifikasikan sebagai netral, padahal mengandung makna negatif. Variasi bahasa non-standar, singkatan, serta campuran bahasa Indonesia dengan bahasa daerah atau bahasa asing juga menyebabkan sebagian makna emosional serta penggunaan bahasa tidak baku, singkatan, dan ejaan yang tidak sesuai kaidah, seperti pada komentar “Tuntaskan K, zolimn. Herkan, Kbatiln”, menyebabkan beberapa kata bermuatan sentimen tidak teridentifikasi dalam kamus leksikon.

Performa SVM yang diukur melalui *weighted average* (0.924) menunjukkan kapabilitas klasifikasi yang tinggi dan dapat diandalkan. Akurasi tinggi belum tentu mencerminkan kebenaran sentimen, karena model belajar dari label yang dihasilkan oleh aturan leksikon.

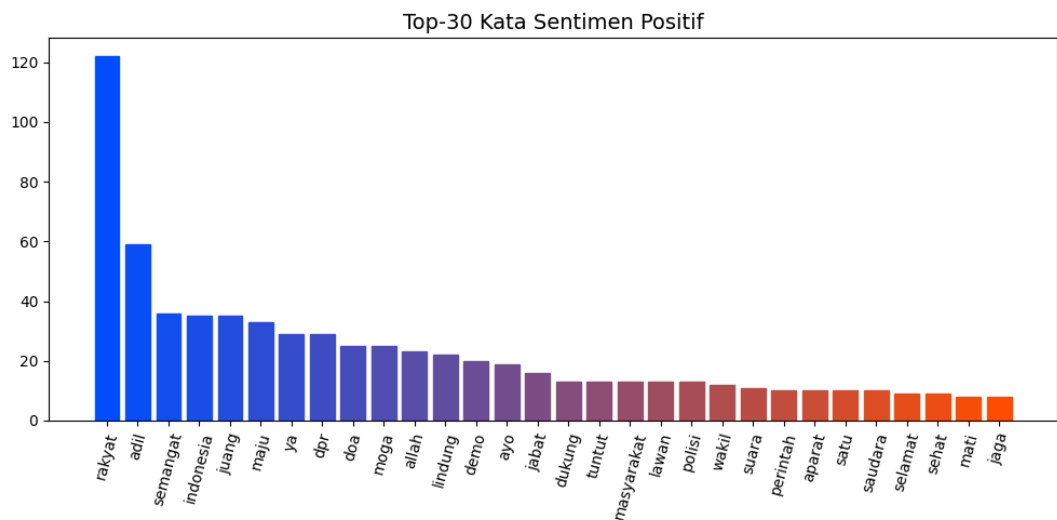
	precision	recall	f1-score	support
Negatif	0.927	0.850	0.887	60.000
Netral	0.930	0.959	0.944	195.000
Positif	0.886	0.861	0.873	36.000
accuracy	0.924	0.924	0.924	0.924
macro avg	0.914	0.890	0.902	291.000
weighted avg	0.924	0.924	0.924	291.000

Gambar 6. Hasil Classification Report

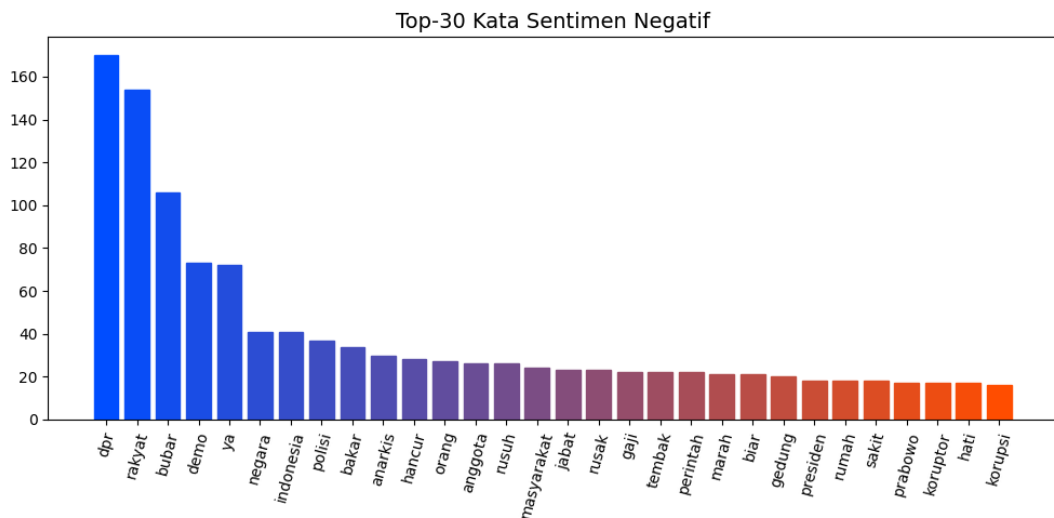
Analisis tambahan dilakukan dengan menghitung frekuensi kemunculan kata pada setiap kategori sentimen. Diagram batang menampilkan 30 kata dengan frekuensi tertinggi pada masing-masing kelas sentimen, sehingga memberikan gambaran kata-kata dominan yang merepresentasikan opini publik terhadap isu demonstrasi DPR RIi. Hasil ini digunakan sebagai pendukung interpretasi pola sentimen yang dihasilkan oleh model klasifikasi. Visualisasi frekuensi kata disajikan pada gambar 7 hingga gambar 9.



Gambar 7. Top 30 Kata Sentimen Netral



Gambar 8. Top 30 Kata Sentimen Positif



Gambar 9. Top 30 Kata Sentimen Negatif

Adanya perbedaan kata yang menonjol pada setiap kelompok sentimen memberikan gambaran mengenai pola bahasa pengguna. Pada sentimen netral, kata-kata seperti “rakyat” dan “dpr” sering muncul karena sebagian besar komentar berisi penyebutan aktor atau objek isu tanpa disertai ekspresi emosional yang eksplisit sehingga secara leksikal tidak memenuhi kriteria kata bermuatan sentimen positif maupun negatif. Pada sentimen positif, kemunculan kata seperti “rakyat” dan “adil” menunjukkan adanya dukungan atau harapan normatif terhadap keadilan dan kepentingan masyarakat. Namun, kata-kata tersebut sering digunakan dalam konteks implisit dalam bentuk doa, seruan moral, atau ungkapan religius. Sehingga menyebabkan sebagian komentar positif berpotensi salah diklasifikasikan sebagai netral ketika tidak disertai kata positif yang kuat menurut leksikon mengakibatkan turunnya kemampuan model dalam membedakan sentimen positif dan netral. Untuk sentimen negatif didominasi oleh kata-kata seperti “dpr” dan “rakyat” yang muncul dalam konteks kritik, ketidakpuasan, atau protes.

Penggunaan kata yang sama pada ketiga kelas sentimen menunjukkan bahwa kata-kata tersebut bersifat kontekstual dan tidak selalu merepresentasikan polaritas sentimen tertentu. Ketergantungan model pada frekuensi kata tanpa mempertimbangkan arti kata berdasarkan situasi menyebabkan kesalahan klasifikasi, terutama ketika kata yang sama digunakan dalam nada sarkastik, mengandung negasi, atau dikombinasikan dengan bahasa slang dan *code-mixing*. Kata yang sama muncul di berbagai kelas sentimen, model kesulitan membedakan sentimen hanya dari frekuensi kata sehingga komentar positif sering tertarik ke kelas netral.

5. Kesimpulan

Penelitian ini menganalisis sentimen komentar YouTube terkait demonstrasi di Gedung DPR RI Jakarta menggunakan pendekatan klasifikasi berbasis Support Vector Machine (SVM). Hasil pelabelan sentimen menunjukkan dominasi kelas netral, yaitu 974 komentar, diikuti sentimen negatif 301 komentar dan sentimen positif 178 komentar. Evaluasi model SVM pada skenario pembagian data latih–uji 80:20 menunjukkan kinerja klasifikasi yang tinggi dengan akurasi 92,4%, sehingga SVM dinilai efektif sebagai baseline untuk pemetaan opini publik digital pada teks berbahasa Indonesia. Selain itu, analisis frekuensi kata membantu mengidentifikasi kosakata dominan pada tiap polaritas sentimen, sehingga memberikan gambaran konteks isu yang paling menonjol dalam diskursus komentar.

Secara praktis, temuan ini dapat dimanfaatkan untuk pemantauan ringkas opini publik berbasis data pada isu kebijakan/aksi massa melalui platform media sosial, misalnya dengan memantau proporsi sentimen per periode dan kata kunci dominan yang muncul. Namun, penelitian ini memiliki keterbatasan, terutama pada ketergantungan pelabelan berbasis leksikon, cakupan data yang bersumber dari satu konteks video/kanal tertentu, serta ketidakseimbangan distribusi kelas yang dapat mempengaruhi interpretasi kinerja per kelas. Penelitian selanjutnya disarankan melakukan validasi label berbasis anotasi manusia, membandingkan representasi fitur seperti misalnya TF-IDF n-gram dan strategi penyeimbangan kelas, menerapkan validasi yang lebih robust (misalnya k-fold cross-validation), serta mengeksplorasi model berbasis transformer untuk meningkatkan ketahanan terhadap bahasa informal, sarkasme, dan variasi konteks komentar.

6. Ucapan Terima Kasih

Ucapan terima kasih disampaikan kepada seluruh pihak atas dukungan serta kontribusi secara langsung maupun tidak langsung, selama penelitian berlangsung. Rasa terima kasih ditunjukkan kepada Universitas Negeri Surabaya khususnya teruntuk Program Studi Pendidikan Teknologi Informasi yang telah memberikan bimbingan akademik.

7. Pernyataan Penulis

Kami sebagai penulis, menyatakan bahwa penelitian ini dilakukan secara cermat dan menyeluruh untuk menganalisis sentimen komentar masyarakat terkait Demonstrasi DPR RI melalui platform YouTube. Seluruh data dikumpulkan dan diproses melalui tahap pra-pemrosesan yang komprehensif untuk memastikan keakuratan dan konsistensi analisis.

Penulis menyampaikan bahwa tidak terdapat konflik kepentingan dalam publikasi artikel ini. Selain itu, seluruh data dan isi makalah dijamin bebas dari plagiarisme, dan penulis bertanggung jawab sepenuhnya atas keaslian karya ini.

Bibliografi

- Ahfaz, A., Ramdani, R., Pratama, Y. B., Pramudyantoro, A., Altiarika, E., & Wahyuzi, Z. (2025). Perbandingan sentimen komentar YouTube pada video promosi bisnis kuliner di Bangka Belitung menggunakan algoritma machine learning. *Journal of Information Technology and Society (JITS)*, 3(1). <https://jits.unmuhbabel.ac.id/>
- Anadas, S., Suryono, R. R., Studi, P., Informasi, S., Teknik, F., Teknokrat, U., Lampung, B., Studi, P., Ilmu, M., Teknik, F., Teknokrat, U., & Lampung, B. (2025). Analisis sentimen publik terhadap program Lapor Mas Wapres periode 2024 di media sosial X menggunakan metode Naive Bayes dan Support Vector Machine (SVM). *Jurnal Pendidikan dan Teknologi Indonesia (JPTI)*, 5(8), 2183-2192. <https://doi.org/10.52436/1.jpti.934>
- Arsini, Y., Azzahra, H., Tarigan, K. S., & Azhari, I. (2023). Pengaruh media sosial terhadap kesehatan mental pada remaja. *Journal Research and Education Studies*, 3(2), 32-38. <https://doi.org/10.61994/educate.v1i1.114>
- Asbanu, N. R., Belo, C. M. F., & Saingo, Y. A. (2025). Analisis peran DPR dalam menyuarakan kepentingan rakyat berdasarkan Sila Keempat Pancasila. *Jurnal Ilmiah Multidisiplin Indonesia*, 3(1), 120-134. <https://doi.org/10.61404/jimi.v3i1.364>
- Asrianto, R., & Herwinanda, M. (2022). Analisis sentimen kenaikan harga kebutuhan pokok di media sosial YouTube menggunakan algoritma support vector machine. *Jurnal Computer Science and Information Technology*, 3(3), 431-440. <https://doi.org/10.37859/coscitech.v3i3.4368>
- Cetin, V., & Yildiz, O. (2022). A comprehensive review on data preprocessing techniques in data analysis. *Pamukkale Universitesi Muhendislik Bilimleri Dergisi*, 28(2), 299-312. <https://doi.org/10.5505/pajes.2021.62687>
- Chairani Siregar, S., Adek, R. T., & Fitri, Z. (2024). Sentiment analysis of comments on YouTube channel beauty vlogger in Indonesian language using support vector machine method. *Proceedings of International Conference on Multidisciplinary Engineering (ICOMDEN)*, 2, 1-6. <https://doi.org/10.29103/icomden.v2.xxxx>
- DataReportal. (2025, December). *Essential YouTube Statistics*. <https://datareportal.com/essential-youtube-stats>

- Heydarian, M., Doyle, T. E., & Samavi, R. (2022). MLCM: Multi-label confusion matrix. *IEEE Access*, 10, 19083-19095.
- Hidayat, Santoso, F., & Lidimillah, L. F. (2024b). Analisis sentimen pengguna YouTube tentang Rohingya menggunakan algoritma SVM (Support Vector Machine). *Jurnal Teknologi Terapan*, 8(3), 1729-1738. <https://doi.org/10.33379/gtech.v8i3.4497>
- Hou, Q., Han, M., Qu, F., & He, J. S. (2021). Understanding social media beyond text: A reliable practice on Twitter. *Computational Social Networks*, 8(4). <https://doi.org/10.1186/s40649-021-00088-x>
- Jamilah, F., & Wahyuni, P. (2020). Ujaran kebencian dalam kolom komentar YouTube pada tahun politik pemilihan presiden 2019. *Jurnal Penelitian Pendidikan Bahasa Indonesia, Daerah, dan Asing*, 3(2), 325-341. <https://doi.org/10.31540/silamparibisa.v3i2.1109>
- Jelly, G. Y., & Prihana, E. (2025). Peran dan tantangan DPR-RI dalam mewujudkan demokrasi di Indonesia. *Jurnal Kajian Pemerintah & Kebijakan Publik*, 1(2), 53-63.
- Junaedi, Gunawan, A. H., Kuswanto, V., & Jonathan. (2024). Eksplorasi algoritma support vector machine untuk analisis sentimen destinasi wisata di Indonesia. *Bit-Tech (Binary Digital - Technology)*, 7(2). <https://doi.org/10.32877/bt.v7i2.1810>
- Khder, M. A. (2021). Web scraping or web crawling: State of art, techniques, approaches and application. *International Journal of Advances in Soft Computing & Its Applications*, 13(3). <https://doi.org/10.15849/IJASCA.211128.11>
- Kusuma, I. H., & Cahyono, N. (2023). Analisis sentimen masyarakat terhadap penggunaan e-commerce menggunakan algoritma k-nearest neighbor. *Jurnal Informatika: Jurnal Pengembangan IT*, 8(3), 302-307. <https://doi.org/10.30591/jpit.v8i3.5734>
- Mao, Y., Liu, Q., & Zhang, Y. (2024). Sentiment analysis methods, applications, and challenges: A systematic literature review. *Journal of King Saud University - Computer and Information Sciences*, 36(4), 102048. <https://doi.org/10.1016/j.jksuci.2024.102048>
- Muhayat, T., Fauzi, A., & Indra, J. (2023). Analisis sentimen terhadap komentar video YouTube menggunakan support vector machines. *Progresif: Jurnal Ilmiah Komputer*, 19(1). <https://doi.org/10.35889/progresif.v19i1.1060>
- Nursalim, A., & Novita, R. (2023). Sentiment analysis of comments on Google Play Store, Twitter and YouTube to the MyPertamina application with support vector machine. *Jurnal Teknik Informatika (JUTIF)*, 4(6), 1305-1312. <https://doi.org/10.52436/1.jutif.2023.4.6.1059>

- O'Connor, B., Balasubramanyan, R., Routledge, B. R., & Smith, N. A. (2010). From tweets to polls: Linking text sentiment to public opinion time series. *Proceedings of the International AAAI Conference on Web and Social Media*.
- Octiva, C. S., Fajri, T. I., Sulistiarini, E. B., Suharjo, & Nuryanto, U. W. (2024). Penggunaan teknik data mining untuk analisis perilaku pengguna pada media sosial. *Jurnal Minfo Polgan*, 13, 1074-1078. <https://doi.org/10.33395/jmp.v13i1.13936>
- Pang, B., & Lee, L. (2008). Opinion mining and sentiment analysis. *Foundations and Trends in Information Retrieval*, 2(1-2), 1-135. <https://doi.org/10.1561/15000000011>
- Permana, T. D., Pratama, Y. B., Wahyuzi, Z., Altiarika, E., & Pramudyantoro, A. (2025). Perbandingan performa algoritma Naive Bayes dan terhadap industri esports di Indonesia. *Jurnal Ilmiah Nusantara*, 2(6), 1391-1399. <https://doi.org/10.61722/jinu.v2i6.6753>
- Putri, S. R., Asrianda, A., & Rosnita, L. (2025). Sentiment analysis of YouTube and GoTube reviews on Google Play using the support vector machine (SVM) method in Indonesia. *Journal of Applied Informatics and Computing (JAIC)*, 9(3). <https://doi.org/10.30871/jaic.v9i3.9461>
- Riwanto, M. H., Ardhiyansyah, P., Alfiansyah, R. A. A. A., Waek, G., & Fahlapi6, R. (2025). Penerapan program makan bergizi gratis adalah langkah. *Jurnal Multidisiplin Saintek*, 6(12), 1-14.
- Rizky, G. P., & Alrasyid, W. (2025). AI-Based Sentiment Analysis of Social Media to Detect Public Opinion on Government Policies. In *Journal Basic Science and Technology* (Vol. 14, Issue 2).
- Rodriguez-Ibanez, M., Casanez-Ventura, A., Castejon-Mateos, F., & Cuenca-Jimenez, P.-M. (2023). A review on sentiment analysis from social media platforms. *Expert Systems with Applications*, 223, 119862. <https://doi.org/10.1016/j.eswa.2023.119862>
- Santoso, B. A., Nugroho, I., & Asyfiya, D. U. (2025). Perbandingan algoritma Naive Bayes, support vector machine, dan random forest untuk analisis sentimen komentar politik YouTube. *TIN: Terapan Informatika Nusantara*, 6(4), 391-400. <https://doi.org/10.47065/tin.v6i4.8326>
- Saputra, A. N. A., Saputro, R. E., & Saputra, D. I. S. (2025). Enhancing sentiment analysis accuracy using SVM and slang word normalization on YouTube comments. *Sinkron: Jurnal dan Penelitian Teknik Informatika*, 9(2), 687-699. <https://doi.org/10.33395/sinkron.v9i2.14613>

- Septiana, L., Yasin, V., & Sianipar, A. Z. (2025). *Analyzing public sentiment on youtube comments regarding the free lunch policy using the Support Vector Machine (SVM) algorithm*. 14(1), 129–138. www.ejournal.isha.or.id/index.php/Mandiri
- Siswoyo, B., & Utomo, N. A. P. (2025). Pemanfaatan machine learning untuk klasifikasi sentimen pelanggan pada media sosial. *Jurnal Teknologi Informasi Digital*, 1(1), 29-34.
- Wahyuni, N. A., Ayu, D. P., & Irsyad, H. (2024a). Analisis sentimen di YouTube terhadap kenaikan UKT menggunakan metode support vector machine. *Journal of Computer Science and Artificial Intelligence*, 4(1), 57-71. <https://doi.org/10.29240/arcitech.v4i1.10829>
- Yulita, I. N., Wijaya, V., Rosadi, R., Sarathan, I., Djuyandi, Y., & Prabuwo, A. S. (2023). Analysis of government policy sentiment regarding vacation during the COVID-19 pandemic using the bidirectional encoder representation from transformers (BERT). *Data*, 8(3), 46. <https://doi.org/10.3390/data8030046>
- Zhafira, D. F., Rahayudi, B., & Indriati, I. (2021). Analisis sentimen kebijakan Kampus Merdeka menggunakan naive bayes dan pembobotan TF-IDF berdasarkan komentar pada YouTube. *Jurnal Sistem Informasi, Teknologi Informasi, dan Edukasi Sistem Informasi*, 2(1), 55-63. <https://doi.org/10.25126/justsi.v2i1.24>