

Analisis Sentimen Komentar YouTube terhadap Trailer Kedua Grand Theft Auto VI Menggunakan TF-IDF dan Support Vector Machine

Hafiz Pratama^{1*}, Nanda Begti Rizal Nugroho², Agustina Purwatiningsih³

¹Teknik Informatika, Fakultas Ilmu Komputer

Universitas Duta Bangsa Surakarta

1*240103057@mhs.udb.ac.id

²Teknik Informatika, Fakultas Ilmu Komputer

Universitas Duta Bangsa Surakarta

240103072@mhs.udb.ac.id

³Fakultas Ilmu Komputer

Universitas Duta Bangsa Surakarta

agustinapurwa@udb.ac.id

Abstrak— Perilisan trailer kedua Grand Theft Auto VI oleh Rockstar Games menuai gelombang antusiasme publik yang masif di YouTube. Studi ini berupaya memilah polaritas komentar penonton atas trailer tersebut menjadi tiga golongan: positif, netral, serta negatif. Melalui YouTube Data API v3, data dihimpun lalu dilewatkan pada rangkaian penyaringan teks, pembuangan entri ganda, dan normalisasi, sehingga tersisa 6.445 komentar bersih. Penandaan polaritas ditempuh secara otomatis lewat pendekatan berbasis leksikon VADER, sementara ciri tekstualnya diekstraksi memakai Term Frequency-Inverse Document Frequency (TF-IDF) dan digolongkan dengan algoritma Support Vector Machine (SVM) berkernel linear. Ketimpangan kelas ditangani melalui pembobotan kelas seimbang berikut penalaan parameter via GridSearch. Berdasarkan evaluasi, model membukukan akurasi 73,55%, F1-score weighted 0,7245, dan F1-score macro 0,6715, disertai skor validasi silang lima lipatan $0,7054 \pm 0,0100$ yang menandakan model cukup stabil. Sebaran sentimen paling banyak ditempati kelas netral (53,6%), disusul positif (28,8%) dan negatif (17,5%). Hasil ini mengisyaratkan bahwa sebagian besar penonton merespons trailer dengan nada netral sampai antusias, sekaligus menyoroti kelemahan pelabelan leksikon saat berhadapan dengan sarkasme dan komentar negatif, yang berujung pada rendahnya performa kelas negatif.

Kata kunci— analisis sentimen, komentar YouTube, Grand Theft Auto VI, TF-IDF, Support Vector Machine

Abstract— Rockstar Games' launch of the second Grand Theft Auto VI trailer sparked an enormous surge of public reaction across YouTube. The present work sets out to categorize viewers' comment sentiment regarding the trailer into three groups: positive, neutral, and negative. Comments were gathered via the YouTube Data API v3, then run through text cleaning, duplicate removal, and normalization, leaving 6,445 clean comments for analysis. Sentiment labels were assigned automatically through the VADER lexicon-based approach, whereas the textual features were drawn out using Term Frequency-Inverse Document Frequency (TF-IDF) and categorized by means of a linear-kernel Support Vector Machine (SVM). Class imbalance was mitigated by applying balanced class weighting alongside parameter tuning through GridSearch. Evaluation results record an accuracy of 73.55%, a weighted F1-score of 0.7245, and a macro F1-score of 0.6715, together with a 5-fold cross-validation score of 0.7054 ± 0.0100 that points to model stability. Within the sentiment distribution, the neutral class was most prevalent at 53.6%, trailed by positive at 28.8% and negative at 17.5%. Such findings suggest that the majority of viewers greeted the trailer with neutral to enthusiastic reactions, while simultaneously exposing the shortcomings of lexicon-based labeling in detecting sarcasm and negative remarks, which in turn produced comparatively weak performance on the negative class.

Keywords— sentiment analysis, YouTube comments, Grand Theft Auto VI, TF-IDF, Support Vector Machine

I. PENDAHULUAN

Kini industri gim video menjelma sebagai salah satu pilar terbesar dalam ranah hiburan digital dunia, dan peluncuran sebuah judul besar kerap menjadi peristiwa budaya yang menarik perhatian jutaan orang. Grand Theft Auto VI, seri terbaru yang dikembangkan Rockstar Games, merupakan salah satu gim paling dinanti dalam satu dekade terakhir. Ketika trailer kedua gim ini dirilis pada 6 Mei 2025, video tersebut dengan cepat menembus ratusan juta penayangan lintas platform sekaligus menuai ratusan ribu komentar di YouTube yang mencerminkan tingginya

antusiasme sekaligus beragamnya opini publik [1].

Komentar dalam jumlah besar tersebut merupakan sumber data berharga untuk memahami persepsi audiens, namun volume yang sangat besar membuat analisis manual menjadi tidak praktis. Analisis sentimen, yang tergolong sebagai salah satu cabang pemrosesan bahasa alami, menghadirkan cara otomatis untuk memilah opini menjadi tiga golongan: positif, netral, maupun negatif. Sehingga pola persepsi publik dapat dipetakan secara sistematis [2]. Pendekatan ini telah banyak dimanfaatkan untuk memetakan opini masyarakat Indonesia pada

berbagai isu, mulai dari layanan publik hingga peristiwa nasional [3].

Pada ranah media sosial berbahasa Indonesia, algoritma pembelajaran mesin konvensional semisal Support Vector Machine (SVM) kerap dimanfaatkan berkat kinerjanya yang konsisten dalam menangani klasifikasi teks berdimensi tinggi. SVM lazimnya disandingkan dengan pembobotan fitur *Term Frequency-Inverse Document Frequency* (TF-IDF), yang berperan mempertegas kata-kata bersifat diskriminatif di dalam korpus [4]. Selain itu, tahap prapemrosesan teks terbukti memberi pengaruh yang cukup besar terhadap mutu hasil klasifikasi sentimen pada data media sosial [5].

Sejumlah penelitian terdahulu telah menerapkan kombinasi TF-IDF dan SVM secara khusus pada data komentar YouTube. Penelitian terhadap komentar pada kanal e-sport Team RRQ mencatat akurasi sebesar 75%, dengan capaian paling menonjol justru berada di kelas sentimen negatif, sehingga membuktikan kemampuan kombinasi tersebut pada data berbahasa informal yang sarat slang [6]. Penelitian terhadap opini publik pada kanal Mata Najwa juga membuktikan bahwa SVM mampu mengklasifikasikan komentar dengan baik [7], sementara studi komparasi pada komentar penggemar BTS menunjukkan SVM unggul dibandingkan Random Forest [8].

Selain pada YouTube, efektivitas SVM dengan TF-IDF teruji pada beragam domain. Penelitian pada opini film horor Indonesia [9] serta klasifikasi quick count Pemilihan Presiden 2024 di platform X [10] memperkuat bukti bahwa kombinasi ini kompetitif. Pemilihan kernel turut memengaruhi performa, sebagaimana ditunjukkan pada studi perbandingan kernel SVM untuk sentimen vaksinasi COVID-19 yang menemukan kernel linear memberikan keseimbangan kinerja yang baik [11].

Persoalan mendasar dalam analisis sentimen adalah penyediaan label. Pelabelan manual menghasilkan kualitas tinggi tetapi memerlukan waktu dan tenaga besar, sehingga metode pelabelan otomatis yang bertumpu pada leksikon kian banyak dilirik. Salah satunya adalah

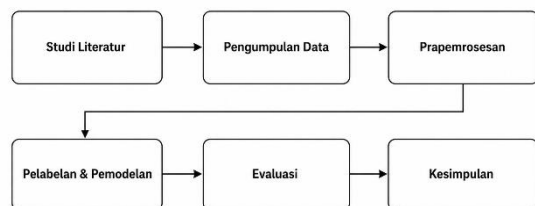
Reasoner (VADER), sebuah model leksikon yang dirancang khusus untuk teks informal media sosial [12]. VADER terbukti efisien sebagai alternatif pelabelan manual karena mampu memangkas waktu anotasi menjadi hitungan detik [13], dan telah berhasil diterapkan untuk pelabelan otomatis data ulasan maupun cuitan berbahasa Indonesia yang kemudian diklasifikasikan dengan berbagai algoritma [14].

Meskipun demikian, pendekatan leksikon memiliki keterbatasan dalam menangkap konteks, sarkasme, dan ironi yang lazim ditemukan pada komentar daring [15]. Tantangan lain yang umum dihadapi adalah ketidakseimbangan kelas (*class imbalance*), yaitu kondisi salah satu kelas sentimen mendominasi data sehingga model cenderung memihak kelas mayoritas. Beberapa penelitian mengatasinya dengan teknik penyeimbangan data seperti SMOTE untuk meningkatkan performa pada kelas minoritas [16].

Berdasarkan tinjauan tersebut, terlihat bahwa penerapan TF-IDF dan SVM dengan pelabelan VADER pada komentar berbahasa Inggris di domain gim video, khususnya pada momen perilis trailer Grand Theft Auto VI, masih belum banyak dieksplorasi. Padahal momentum tersebut menghasilkan volume komentar yang sangat besar dengan karakteristik bahasa yang khas. Kekosongan itulah yang hendak dijawab dalam penelitian ini.

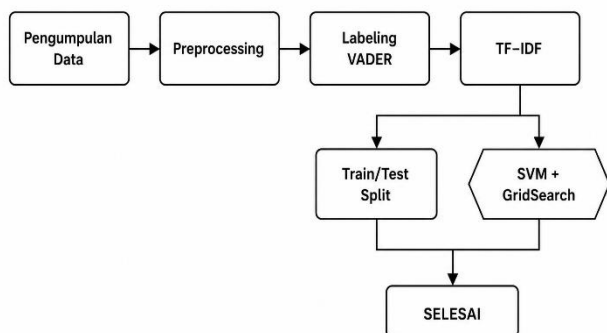
Studi ini diarahkan untuk mengupas sentimen komentar penonton YouTube atas trailer kedua Grand Theft Auto VI melalui kombinasi TF-IDF dan SVM yang dipadukan pelabelan otomatis VADER. Tak hanya itu, penelitian ini turut mencermati karakteristik sebaran sentimen sekaligus keterbatasan metode yang dipakai, sehingga temuannya diharapkan bisa menjadi acuan bagi kajian analisis sentimen di ranah gim video maupun konten hiburan digital.

II. METODOLOGI PENELITIAN



Gambar 1. Diagram Alir Penelitian Analisis Sentimen Komentar YouTube terhadap Trailer Kedua Grand Theft Auto VI

Pendekatan kuantitatif diterapkan dalam penelitian ini melalui alur kerja berjenjang yang mencakup lima tahapan pokok: pengumpulan data, prapemrosesan teks, pelabelan sentimen, kemudian ekstraksi fitur, klasifikasi serta ditutup dengan evaluasi model. Keseluruhan tahapan tersebut dijalankan dengan bahasa pemrograman Python di lingkungan Google Colaboratory, yang ditopang pustaka Scikit-learn, NLTK, serta vaderSentiment.



Gambar 2. Flowchart Proses Analisis Sentimen Komentar YouTube terhadap Trailer Kedua Grand Theft Auto VI

Pengumpulan Data. Sumber data berupa komentar diambil dari video trailer kedua Grand Theft Auto VI yang tayang di kanal resmi Rockstar Games sejak 6 Mei 2025. Penghimpunan dilakukan pada 24 Juni 2026 dengan memanfaatkan mekanisme *commentThreads* pada YouTube Data API v3, dan berhasil menjaring sebanyak 9.000 komentar mentah. Selain komentar utama, proses *scraping* turut merekam metadata pelengkap, yaitu nama pengguna, jumlah suka, serta tanggal komentar. Dari keseluruhan data mentah tersebut, sebanyak 2.555 komentar (28,4%) tereliminasi pada tahap prapemrosesan karena terdeteksi sebagai duplikat maupun entri kosong, sehingga tersisa 6.445

komentar bersih yang dipakai sebagai basis analisis.

Prapemrosesan Teks. Tahap ini bertujuan membersihkan data mentah agar layak dianalisis. Langkah-langkah yang dilakukan meliputi penghapusan komentar duplikat dan kosong, konversi huruf menjadi huruf kecil (*case folding*), penghapusan URL, mention, tagar, emoji, angka, serta tanda baca. Selanjutnya dilakukan penyaringan kata henti (*stopword removal*) berbahasa Inggris yang dilengkapi daftar kata henti khusus untuk istilah dominan namun tidak diskriminatif seperti nama gim dan pengembang. Setelah prapemrosesan, diperoleh 6.445 komentar bersih yang siap dianalisis.

Pelabelan Sentimen. Untuk menandai polaritas komentar, penelitian ini mengandalkan VADER, yakni model leksikon yang memang dirancang bagi teks informal, secara otomatis. Keluaran VADER berupa nilai *compound* yang bergerak pada rentang minus satu sampai satu. Komentar bernilai *compound* $\geq 0,10$ ditetapkan sebagai positif, nilai $\leq -0,10$ sebagai negatif, sedangkan nilai di antara keduanya tergolong netral. Pemilihan ambang 0,10 didasari pertimbangan agar batas antarkelas menjadi lebih tegas ketimbang memakai ambang baku 0,05.

Ekstraksi Fitur dan Klasifikasi. Representasi numerik dari teks bersih dibentuk memakai TF-IDF dengan konfigurasi rentang *n-gram* satu hingga dua kata, kuota fitur maksimum 20.000, ambang minimum kemunculan dokumen sebesar dua, serta batas atas frekuensi dokumen 0,9 guna meredam kata yang terlampau umum. Bobot TF-IDF diperoleh dari perkalian *term frequency* dan *inverse document frequency*, yang dirumuskan:

$$w_{t,d} = tf_{t,d} \times \log \log \left(\frac{N}{df_t} \right)$$

dengan $tf_{t,d}$ adalah frekuensi term t pada dokumen d , N jumlah total dokumen, dan df_t jumlah dokumen yang memuat term t .

Selanjutnya data dipisah secara terstratifikasi ke dalam porsi 80% pelatihan dan 20% pengujian. Tahap klasifikasi mengandalkan SVM

berkernel linear (LinearSVC) dengan pembobotan kelas seimbang. SVM mencari *hyperplane* pemisah optimal $f(x) = w^T x + b$ yang memaksimalkan margin antarkelas melalui optimasi:

$$\frac{1}{2} \|w\|^2 + C \sum_{i=1}^n \xi_i$$

dengan w vektor bobot, b bias, C parameter regularisasi, dan ξ_i *slack variable*. Parameter C dioptimasi melalui GridSearch berbasis validasi silang lima lipatan, dengan nilai C terbaik sebesar 0,1.

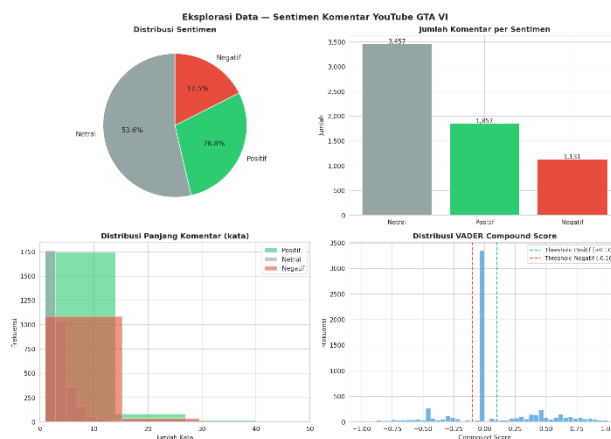
Evaluasi Model. Pengukuran kinerja model bertumpu pada sejumlah metrik, meliputi akurasi, presisi, recall, hingga F1-score, yang masing-masing dihitung dalam skema *weighted* maupun *macro*. Sebagai pelengkap, validasi silang lima lipatan dijalankan guna memastikan kestabilan model, sementara *confusion matrix* ditelaah untuk memetakan pola kekeliruan klasifikasi di tiap kelas sentimen.

III. HASIL DAN PEMBAHASAN

Bagian ini memaparkan hasil dari setiap tahapan penelitian, mulai dari hasil prapemrosesan dan pelabelan data, distribusi sentimen, hingga kinerja model klasifikasi, yang kemudian diikuti dengan pembahasan terhadap temuan yang diperoleh.

a. Distribusi Sentimen Hasil Pelabelan

Dari 6.445 komentar bersih yang berhasil dilabeli menggunakan VADER, diperoleh distribusi sentimen yang tidak seimbang. Kelas netral mendominasi dengan 3.457 komentar atau 53,6%, diikuti kelas positif sebanyak 1.857 komentar (28,8%), dan kelas negatif sebanyak 1.131 komentar (17,5%). Dominasi kelas netral ini sejalan dengan karakteristik komentar pada video trailer, yang banyak berisi ungkapan antisipasi, pertanyaan seputar tanggal rilis, maupun komentar penanda kehadiran yang secara leksikal tidak memuat bobot emosional yang kentara.



Gambar 3. Eksplorasi data dan distribusi sentimen komentar.

Gambar 3 menyajikan eksplorasi data yang mencakup distribusi sentimen, jumlah komentar per kelas, distribusi panjang komentar, serta sebaran nilai *compound VADER*. Terlihat bahwa mayoritas komentar memiliki panjang yang pendek, yaitu di bawah lima belas kata, dan sebaran nilai *compound* menumpuk di sekitar nol. Penumpukan nilai *compound* di sekitar nol inilah yang menyebabkan besarnya proporsi kelas netral.

b. Analisis Kata Dominan

Untuk memahami karakteristik tiap kelas sentimen, dilakukan visualisasi word cloud terhadap kata-kata yang paling sering muncul. Hasilnya ditunjukkan pada Gambar 4.



Gambar 4. Word cloud kata dominan pada masing-masing kelas sentimen.

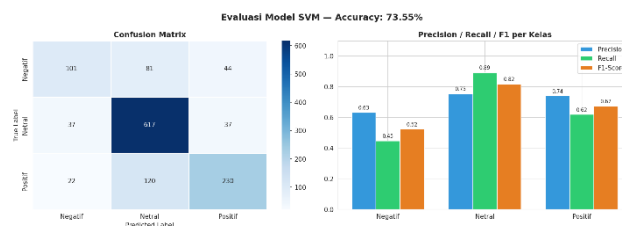
Hasil word cloud memperlihatkan fenomena yang menarik sekaligus menjadi tantangan utama. Kata-kata seperti trailer, may, today, dan day muncul secara dominan pada ketiga kelas sentimen sekaligus. Tumpang tindih kosakata antar kelas ini menyulitkan model dalam membedakan sentimen, karena fitur yang sama hadir pada kelas yang berbeda. Pada kelas positif muncul kata bermuatan apresiatif seperti love, better, dan please, sementara pada kelas negatif muncul kata seperti delayed, flops, dan crazy yang mencerminkan kekecewaan.

c. Kinerja Model Klasifikasi

Model SVM berkernell linear dilatih menggunakan fitur TF-IDF dengan parameter regularisasi C optimal sebesar 0,1 hasil GridSearch. Evaluasi pada data uji sebanyak 1.289 komentar menghasilkan akurasi sebesar 73,55%, F1-score *weighted* 0,7245, dan F1-score *macro* 0,6715. Hasil validasi silang lima lipatan memberikan nilai F1 sebesar 0,7054 dengan simpangan baku 0,0100. Nilai simpangan baku yang sangat kecil ini menunjukkan bahwa model bersifat stabil dan konsisten antar-lipatan. Rincian kinerja per kelas disajikan pada Tabel 1.

Tabel 1. Hasil Evaluasi Model per Kelas Sentimen

Kelas	Presisi	Recall	F1
Negatif	0,63	0,45	0,52
Netral	0,75	0,89	0,82
Positif	0,74	0,62	0,67
Akurasi	0,7355		



Gambar 5. Confusion matrix dan perbandingan presisi, recall, serta F1 per kelas.

Tabel 1 dan Gambar 5 menunjukkan bahwa kinerja model bervariasi antar kelas. Kelas netral memperoleh F1-score tertinggi sebesar 0,82 dengan recall 0,89, yang wajar mengingat kelas ini paling dominan dalam data. Sebaliknya, kelas negatif memperoleh kinerja terendah dengan F1-score 0,52 dan recall hanya 0,45. Berdasarkan *confusion matrix*, dari 226 komentar negatif pada data uji, sebanyak 81 komentar salah diklasifikasikan sebagai netral dan 44 sebagai positif.

d. Pembahasan

Rendahnya kinerja pada kelas negatif merupakan temuan yang perlu dicermati. Terdapat tiga faktor utama yang saling berkaitan. Pertama, ketimpangan distribusi kelas yang cukup mencolok mendorong model untuk lebih condong menebak kelas mayoritas, yakni netral, kendati pembobotan kelas seimbang sudah diterapkan. Kedua, tumpang tindih kosakata

antar kelas sebagaimana terlihat pada word cloud menyebabkan fitur TF-IDF kesulitan memisahkan kelas negatif dari netral. Ketiga, keterbatasan VADER selaku pendekatan yang bertumpu pada leksikon saat harus mengidentifikasi sarkasme dan ironi, yang kerap muncul pada komentar negatif bernuansa gaming.

Sebagai ilustrasi, komentar yang mengandung sarkasme atau ungkapan antusias dengan diksi bernada literal negatif sering kali tidak terdeteksi dengan tepat oleh pendekatan leksikon. Hal ini menegaskan bahwa batas atas kinerja model sangat dipengaruhi oleh kualitas pelabelan otomatis, karena model SVM pada dasarnya mempelajari pola keputusan yang dihasilkan oleh VADER.

Secara umum, perpaduan TF-IDF dan SVM terbukti sanggup memilah sentimen komentar dengan tingkat akurasi yang terbilang memadai untuk data komentar media sosial yang berkarakter informal serta kaya ragam bahasa. Angka akurasi 73,55% yang diraih ini selaras dengan rentang temuan pada penelitian serupa di ranah analisis sentimen media sosial. Dominasi sentimen netral dan positif yang mencapai lebih dari 82% secara kumulatif mengindikasikan bahwa trailer kedua Grand Theft Auto VI secara umum diterima dengan respons yang netral hingga antusias oleh penonton.

IV. KESIMPULAN

Penelitian ini berhasil menerapkan kombinasi TF-IDF dan Support Vector Machine untuk menganalisis sentimen 6.445 komentar penonton YouTube terhadap trailer kedua Grand Theft Auto VI. Model yang dibangun mencapai akurasi 73,55% dengan F1-score *weighted* 0,7245 dan F1-score *macro* 0,6715. Nilai validasi silang lima lipatan sebesar 0,7054 dengan simpangan baku 0,0100 membuktikan bahwa model bersifat stabil dan konsisten antar-lipatan, sehingga tujuan penelitian untuk mengklasifikasikan sentimen komentar dapat dikatakan tercapai.

Analisis yang dilakukan memperlihatkan bahwa sentimen penonton paling banyak ditempati kategori netral sebesar 53,6%, diikuti positif 28,8% dan negatif 17,5%. Dominasi

sentimen netral dan positif yang secara kumulatif mencapai lebih dari 82% mengindikasikan bahwa trailer tersebut diterima dengan respons yang netral hingga antusias. Di sisi lain, penelitian ini menemukan bahwa kinerja klasifikasi pada kelas negatif relatif rendah akibat ketidakseimbangan kelas, tumpang tindih kosakata antarkelas, berikut keterbatasan pelabelan berbasis leksikon saat mengidentifikasi sarkasme dan ironi.

Penelitian selanjutnya disarankan untuk mengeksplorasi metode pelabelan berbasis model transformer seperti RoBERTa guna meningkatkan kualitas label, menerapkan teknik penyeimbangan data, serta membandingkan beberapa algoritma klasifikasi agar performa pada kelas minoritas dapat ditingkatkan.

UCAPAN TERIMA KASIH

Rasa syukur yang mendalam penulis panjatkan kepada Tuhan Yang Maha Esa, yang atas rahmat serta kemudahan-Nya penelitian ini dapat rampung dengan lancar. Penghargaan tulus juga penulis tujukan bagi kedua orang tua, yang doa dan sokongannya tak pernah surut, kepada teman-teman yang senantiasa memberikan semangat, serta kepada teman-teman seperjuangan di Laboratorium 6 atas kebersamaan dan dukungannya selama proses penyusunan penelitian ini.

REFERENSI

- [1] J. Connatser, "New GTA 6 trailer fails to break the 24-hour view world record, here's why," TweakTown, 8 Mei 2025. [Online]. Tersedia: <https://www.tweaktown.com/news/105118/new-gta-6-trailer-fails-to-break-the-24-hour-view-world-record-heres-why/index.html>. [Diakses: 14 Juli 2026].
- [2] C. J. Hutto and E. Gilbert, *VADER: A parsimonious rule-based model for sentiment analysis of social media text*, in Proc. 8th Int. AAAI Conf. Weblogs Soc. Media (ICWSM), Ann Arbor, MI, 2014, pp. 216-225. <https://doi.org/10.1609/icwsml.v8il.14550>
- [3] A. Andhini, F. N. Handayani, I. Diasih, and Nurmawati, *Analisis sentimen opini publik pada channel YouTube Mata Najwa menggunakan metode SVM*, J. Teknik Inform. Teknol. Inf., vol. 5, no. 2, pp. 139-154, 2025. <https://doi.org/10.55606/jutiti.v5i2.5426>
- [4] D. Septiani and I. Isabela, *Analisis term frequency inverse document frequency (TF-IDF) dalam temu kembali informasi pada dokumen teks*, Sist. Teknol. Inf. Indones. (SINTESIA), vol. 1, no. 2, pp. 81-88, 2022.
- [5] S. Khairunnisa, Adiwijaya, and S. Al Faraby, *Pengaruh text preprocessing terhadap analisis sentimen komentar masyarakat pada media sosial Twitter (studi kasus pandemi COVID-19)*, J. Media Inform. Budidarma, vol. 5, no. 2, pp. 406-414, 2021. <https://doi.org/10.30865/mib.v5i2.2835>
- [6] K. Sakun and C. Rudianto, *Analisis sentimen komentar pada kanal YouTube Team RRQ menggunakan algoritma Support Vector Machine*, J. Teknol. Manaj. Ind. Terap. (JTMIT), vol. 5, no. 2, pp. 1321-1329, 2026. <https://doi.org/10.55826/jtmit.v5i2.1702>
- [7] T. Muhyat, A. Fauzi, and D. J. Indra, *Analisis sentimen terhadap komentar video YouTube menggunakan Support Vector Machine*, Progresif J. Ilm. Komput., vol. 19, no. 1, pp. 231-240, 2023. <http://dx.doi.org/10.35889/progresif.v19i1.1060>
- [8] A. N. Syafia, M. F. Hidayatullah, and W. Sutedy, *Studi komparasi algoritma SVM dan Random Forest pada analisis sentimen komentar YouTube BTS*, J. Inform. J. Pengemb. IT, vol. 8, no. 3, pp. 207-212, 2023. <https://doi.org/10.30591/jpit.v8i3.5064>
- [9] J. E. B. Sinulingga and H. C. K. Sitorus, *Analisis sentimen masyarakat terhadap film horor Indonesia menggunakan metode SVM dan TF-IDF*, J. Manaj. Inform. (JAMIKA), vol. 14, no. 1, pp. 42-53, 2024. <https://doi.org/10.34010/jamika.v14i1.11946>
- [10] R. A. Pranata, Rudiman, and N. A. Verdikha, *Metode pembobotan TF-IDF untuk klasifikasi teks quick count pemilihan Wakil Presiden Indonesia 2024 pada X Twitter dengan metode SVM*, J. Teknol. Inf. (JTI), vol. 18, no. 2, pp. 126-137, 2024.
- [11] T. M. Permata Aulia, N. Arifin, and R. Mayasari, *Perbandingan Kernel Support Vector Machine (Svm) Dalam Penerapan Analisis Sentimen Vaksinasi Covid-19*, SINTECH Sci. Inf. Technol. J., vol. 4, no. 2, pp. 139-145, 2021. <https://doi.org/10.31598/sintechjournal.v4i2.762>
- [12] D. Abimanyu, E. Budianita, E. P. Cynthia, F. Yanto, and Yusra, *Analisis sentimen akun Twitter Apex Legends menggunakan VADER*, J. Nas. Komputasi Teknol. Inf. (JNKTI), vol. 5, no. 3, pp. 423-431, 2022. <https://doi.org/10.32672/jnkti.v5i3.4382>
- [13] A. S. Rizkia, W. Wufron, and F. F. Roji, *Analisis sentimen Coretax: perbandingan pelabelan data manual, transformers-based, dan lexicon-based pada performa IndoBERT*, MALCOM: Indones. J. Mach. Learn. Comput. Sci., vol. 5, no. 3, pp. 1037-1048, 2025. <https://doi.org/10.57152/malcom.v5i3.2151>
- [14] Y. Asri, W. N. Suliyanti, D. Kuswardani, and M. Fajri, *Pelabelan otomatis lexicon VADER dan klasifikasi Naive Bayes dalam menganalisis sentimen data ulasan PLN Mobile*, PETIR, vol. 15, no. 2, pp. 264-275, 2022. <https://doi.org/10.33322/petir.v15i2.1733>
- [15] A. D. H. Setiawan and W. Maharani, *Understanding public sentiments on the 2024 presidential election through BERT-powered analysis*, Edumatic J. Pendidik. Inform., vol. 9, no. 1, pp. 89-98, 2025. <https://doi.org/10.29408/edumatic.v9i1.29267>
- [16] T. A. Dewi and E. Mailoa, *Perbandingan implementasi metode SMOTE pada algoritma Support Vector Machine (SVM) dalam analisis sentimen opini masyarakat tentang Mixue*, J. Indones. Manaj. Inform. Komun. (JIMIK), vol. 4, no. 3, pp. 849-855, 2023. <https://doi.org/10.35870/jimik.v4i3.289>