



Analisis Sentimen Komentar YouTube terhadap Video “Purbaya Effect: Pertaruhan Ekonomi Indonesia” sebagai Cerminan Persepsi Publik Tahun 2025

Adi Susanto^{1*}, Imelda Valentina Octavia²

¹ Teknologi Informasi, Fakultas Sains dan Teknologi, Universitas Ibrahimy, Indonesia

² Sistem Informasi, Fakultas Sains dan Teknologi, Universitas Ibrahimy, Indonesia

Info Artikel

Riwayat Artikel:

Diterima : **14 Oktober 2025**

Direvisi : **17 Desember 2025**

Diterbitkan : **23 Desember 2025**

Kata Kunci:

Analisis Sentimen,
Support Vector Machine (SVM),
YouTube,
Kebijakan Ekonomi,
TF-IDF.

ABSTRAK

Media sosial khususnya YouTube telah menjadi platform penting dalam membentuk opini publik terhadap isu ekonomi nasional. Penelitian ini bertujuan menganalisis sentimen publik terhadap kebijakan ekonomi Indonesia melalui komentar YouTube menggunakan pendekatan *machine learning*. Dataset terdiri dari 1.637 komentar yang dibagi menjadi data *training* (1.309) dan *testing* (328) dengan empat kategori sentimen: Positive, Negative, Neutral, dan Mixed Sentiment. Algoritma *Support Vector Machine* (SVM) LinearSVC diimplementasikan dengan ekstraksi fitur TF-IDF. Hasil penelitian menunjukkan model SVM mencapai akurasi 85,37% dengan waktu *training* 0,04 detik. Distribusi sentimen didominasi Neutral (75,32%), diikuti Positive (14,42%), Negative (7,15%), dan Mixed Sentiment (3,12%). Performa terbaik dicapai pada kategori Neutral dengan precision 0,87 dan recall 0,99 (F1-score 0,93). Namun, model gagal total mengklasifikasikan *Mixed Sentiment* (F1-score 0.00) dan memiliki kinerja rendah pada *Negative* (Recall 0.25) yang merupakan dampak dari severe class imbalance. Mayoritas publik menunjukkan sikap *wait-and-see* terhadap kebijakan ekonomi, mengindikasikan kematangan literasi ekonomi masyarakat Indonesia dalam merespons isu nasional.

Keywords:

Sentiment Analysis,
Support Vector Machine (SVM),
YouTube,
Economic Policy,
TF-IDF

ABSTRACT

Social media, particularly YouTube, has become a crucial platform in shaping public opinion regarding national economic issues. This study aims to analyze public sentiment towards Indonesia's economic policies through YouTube comments using a machine learning approach. The dataset consists of 1,637 comments, divided into 1,309 training data and 328 testing data, with four sentiment categories: Positive, Negative, Neutral, and Mixed Sentiment. The Support Vector Machine (SVM) LinearSVC algorithm was implemented alongside TF-IDF feature extraction. The results indicate that the SVM model achieved an accuracy of 85.37% with an efficient training time of 0.04 seconds. The sentiment distribution was dominated by Neutral (75.32%), followed by Positive (14.42%), Negative (7.15%), and Mixed Sentiment (3.12%). The best performance was achieved in the Neutral category with a precision of 0.87 and a recall of 0.99 (F1-score 0.93). However, the model demonstrated significant weaknesses due to severe class imbalance: the model completely failed to classify Mixed Sentiment (F1-score 0.00) and showed low performance on Negative (Recall 0.25). The majority of the public exhibited a wait-and-see attitude toward economic policies, indicating the maturity of economic literacy among the Indonesian society in responding to national issues.

Penulis Korespondensi:

Adi Susanto,
Teknologi Informasi,
Universitas Ibrahimy
Email: adisusanto@ibrahimyy.ac.id

This is an open access article under the [CC BY-SA](#) license



1. PENDAHULUAN

Era digital telah mengubah cara masyarakat mengekspresikan pendapat terhadap berbagai isu, termasuk kebijakan ekonomi nasional. YouTube sebagai salah satu platform media sosial terbesar di dunia menjadi ruang publik virtual dimana jutaan pengguna berbagi pandangan melalui kolom komentar [1]. Video berjudul

"Purbaya Effect: Pertaruhan Ekonomi Indonesia" yang diunggah oleh Raymond Chin pada 5 Oktober tahun 2025 telah menarik perhatian publik dengan ribuan komentar yang mencerminkan beragam persepsi masyarakat terhadap kondisi ekonomi Indonesia.

Analisis sentimen merupakan cabang dari text mining yang bertujuan mengidentifikasi dan mengekstrak informasi subjektif dari teks [2]. Dalam konteks media sosial, analisis sentimen dapat memberikan wawasan berharga tentang opini publik terhadap isu tertentu. Pendekatan *machine learning*, khususnya *Support Vector Machine* (SVM), telah terbukti efektif dalam klasifikasi sentimen karena kemampuannya menangani data berdimensi tinggi dan memberikan hasil yang robust [3]. Penelitian terbaru menunjukkan bahwa integrasi algoritma pembelajaran mesin dengan analisis teks media sosial memungkinkan ekstraksi insight yang lebih mendalam tentang dinamika opini publik [4].

Support Vector Machine (SVM) dikenal memiliki performa yang baik dalam klasifikasi teks. Dalam analisis sentimen terhadap kebijakan Tapera menggunakan data komentar publik dari Twitter, SVM menghasilkan akurasi sebesar 68,42%. Meskipun akurasi tersebut berada di bawah Naïve Bayes yang mencapai 69,17%, SVM tetap menunjukkan efektivitas yang layak dalam menangani data teks kompleks dari media social [5]. Pengembangan lebih lanjut dalam optimasi parameter SVM telah membuktikan peningkatan signifikan dalam mengenali pola sentimen kompleks pada dataset skala besar [6].

Meskipun telah banyak penelitian terkait analisis sentimen pada media sosial, masih terdapat celah penelitian yang belum banyak dieksplorasi, yaitu: (1) terbatasnya studi analisis sentimen komentar YouTube berbahasa Indonesia dengan fokus pada isu ekonomi, (2) pendekatan klasifikasi multi-kelas (empat kategori) belum umum digunakan, dan (3) kebutuhan akan dataset tahun 2025 yang lebih relevan dengan kondisi ekonomi terkini. Oleh karena itu, penelitian ini bertujuan untuk mengisi gap tersebut dengan menganalisis sentimen publik terhadap kebijakan ekonomi Indonesia melalui komentar YouTube menggunakan algoritma *Support Vector Machine* (SVM) [7]. Studi komparatif mengenai efektivitas berbagai algoritma *machine learning* dalam konteks bahasa Indonesia menunjukkan potensi signifikan untuk penelitian lebih lanjut dalam domain analisis sentimen [8].

Penelitian ini memberikan kontribusi penting dalam bidang analisis sentimen berbahasa Indonesia, khususnya pada isu ekonomi nasional. Salah satu hasil utama dari penelitian ini adalah tersusunnya dataset komentar YouTube yang berfokus pada topik ekonomi, yang dapat dimanfaatkan untuk studi lanjutan. Metode klasifikasi sentimen multi-kelas diterapkan menggunakan algoritma *Support Vector Machine* (SVM), dengan tujuan mengidentifikasi persepsi publik secara lebih terstruktur. Selain itu, penelitian ini juga mengeksplorasi tantangan dalam menangani data yang tidak seimbang (*imbalanced dataset*), serta mengevaluasi performa model secara menyeluruh. Temuan dari penelitian ini diharapkan mampu memberikan wawasan bagi para pemangku kebijakan mengenai opini publik terhadap kebijakan ekonomi, sekaligus memperkaya pendekatan metodologis dalam studi analisis sentimen di Indonesia.

2. METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan metode *machine learning*. Desain penelitian bersifat deskriptif-analitik untuk mengidentifikasi pola sentimen dan evaluatif untuk menguji performa model klasifikasi. Metode penelitian terdiri dari beberapa tahapan sebagai berikut.

2.1. Pengumpulan Data

Data penelitian ini diperoleh dari komentar YouTube yang dikumpulkan melalui video berjudul "*Purbaya Effect: Pertaruhan Ekonomi Indonesia*" yang diunggah oleh kanal reymondchin. Pengumpulan data dilakukan pada tahun 2025 menggunakan YouTube Data API v3, dengan total sebanyak 1.637 komentar yang berhasil dihimpun. Proses seleksi data dilakukan secara ketat untuk memastikan kualitas dan relevansi komentar terhadap topik penelitian. Adapun kriteria yang digunakan meliputi: komentar harus menggunakan bahasa Indonesia, tidak termasuk spam atau komentar duplikat, memiliki konten substansial dengan minimal tiga kata, serta dinilai relevan dengan isu ekonomi Indonesia yang menjadi fokus utama dalam video tersebut. Dataset ini diharapkan mampu memberikan representasi opini publik yang bermakna dalam analisis sentimen dan pemetaan persepsi masyarakat terhadap dinamika ekonomi nasional.

Protokol pengumpulan data melalui API telah menjadi standar dalam penelitian media sosial dan memberikan akses terhadap data yang terstruktur dan dapat diverifikasi [9].

2.2. Preprocessing Data

Tahapan *preprocessing* dilakukan untuk membersihkan dan menyiapkan data sebelum proses klasifikasi [10]. Tahapan yang dilakukan meliputi:

1. *Cleaning*: Menghapus URL, mention, hashtag, karakter spesial, angka, dan whitespace berlebih
2. *Case Folding*: Mengkonversi semua teks ke lowercase untuk standarisasi
3. *Tokenization*: Memecah teks menjadi token individual
4. *Stopword Removal*: Menghapus kata-kata umum yang tidak informatif menggunakan library Sastrawi
5. *Normalisasi*: Memperbaiki kata-kata slang dan singkatan ke bentuk baku

Optimasi tahapan *preprocessing* melalui penggunaan *library* khusus pemrosesan teks berbahasa Indonesia telah meningkatkan kualitas fitur yang dihasilkan [11]

2.3. Pelabelan Data

Pelabelan dilakukan secara manual oleh peneliti dengan empat kategori sentimen [12]:

1. *Positive*: Komentar menyatakan dukungan, optimisme, atau penilaian positif
2. *Negative*: Komentar menyatakan kritik, pesimisme, atau penilaian negatif
3. *Neutral*: Komentar informatif tanpa sentimen jelas atau pertanyaan
4. *Mixed Sentiment*: Komentar mengandung sentimen positif dan negatif sekaligus

Dataset dibagi menggunakan *stratified split* untuk mempertahankan proporsi kelas dengan rasio 80:20, yaitu training set sebanyak 1.309 komentar dan testing set sebanyak 328 komentar

2.4. Feature Extraction

Ekstraksi fitur menggunakan TF-IDF (*Term Frequency-Inverse Document Frequency*) yang merupakan metode representasi teks numerik [13]. Pada tahap ekstraksi fitur, metode TF-IDF digunakan untuk merepresentasikan teks dalam bentuk numerik. Parameter yang diterapkan meliputi *max_features* sebesar 5000, *ngram_range* (1, 2) yang mencakup unigram dan bigram, serta *min_df* sebesar 2 untuk mengeliminasi kata-kata yang terlalu jarang muncul.

Pemilihan TF-IDF didasarkan pada kemampuannya dalam memberikan bobot lebih tinggi terhadap kata-kata yang penting namun jarang muncul dalam korpus, sehingga menghasilkan fitur yang lebih informatif dan relevan untuk proses klasifikasi [14].

2.5. Model Klasifikasi SVM

Support Vector Machine (LinearSVC) dipilih sebagai model klasifikasi dengan parameter $C=1.0$, $\text{max_iter}=1000$, dan $\text{random_state}=42$. SVM bekerja dengan menemukan hyperplane optimal yang memaksimalkan margin antara kelas-kelas data. LinearSVC dipilih karena efektif untuk data berdimensi tinggi seperti TF-IDF vectors, robust terhadap overfitting, dan memiliki waktu training yang efisien.

2.6. Evaluasi Model

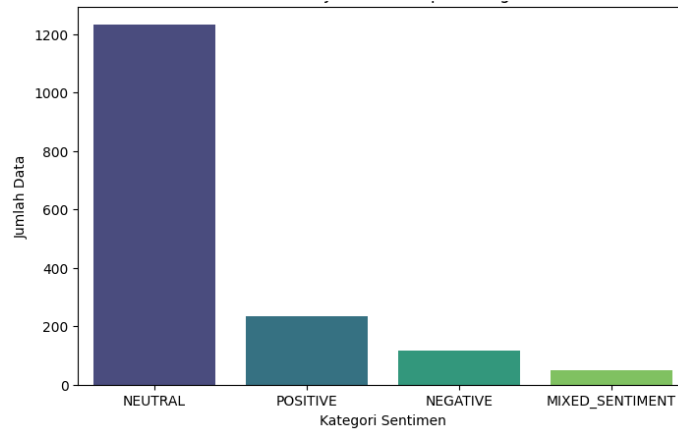
Metrik evaluasi yang digunakan meliputi *accuracy*, *precision*, *recall*, *F1-score*, dan *confusion matrix*. *Accuracy* mengukur proporsi prediksi benar dari total prediksi. *Precision* mengukur proporsi prediksi positif yang benar. *Recall* mengukur proporsi data positif yang terdeteksi dengan benar. *F1-score* adalah *harmonic mean* dari *precision* dan *recall*. *Confusion matrix* digunakan untuk visualisasi performa model per kelas [15].

Implementasi menggunakan Python 3.13.3 dengan library pandas untuk manipulasi data, scikit-learn untuk *machine learning* (SVM, TF-IDF, metrics), matplotlib dan seaborn untuk visualisasi data, serta Sastrawi untuk text preprocessing bahasa Indonesia.

3. HASIL DAN ANALISIS

3.1. Distribusi Sentimen

Hasil pelabelan manual terhadap 1.637 komentar menunjukkan distribusi sentimen sebagaimana terlihat pada Tabel 1 dan Gambar 1.



Gambar 1. Distribusi Jumlah Data per Kategori Sentimen

Tabel 1. Distribusi Sentimen Komentar YouTube

Kategori Sentimen	Jumlah	Persentase
Neutral	1.233	75.32%
Positive	236	14.42%
Negative	117	7.15%
Mixed Sentiment	51	3.12%
Total	1.637	100.01%

Berdasarkan Tabel 1, kategori Neutral mendominasi dengan 1.233 komentar atau 75,32% dari total dataset. Kategori Positive menempati posisi kedua dengan 236 komentar (14,42%), diikuti Negative sebanyak 117 komentar (7,15%), dan Mixed Sentiment hanya 51 komentar (3,12%). Distribusi ini menunjukkan severe class imbalance yang dapat mempengaruhi performa model klasifikasi.

Gambar 1 menampilkan visualisasi distribusi jumlah data per kategori sentimen dalam bentuk bar chart. Terlihat jelas dominasi kategori Neutral yang jauh lebih tinggi dibanding kategori lainnya. Rasio sentimen Positive terhadap Negative sebesar 2.02:1 menunjukkan optimisme moderat publik terhadap kebijakan ekonomi Indonesia.

3.2. Performa Model SVM

Model SVM yang dilatih dengan 1.309 data training menunjukkan performa sebagaimana tercantum pada Tabel 2.

Tabel 2. Metrik Performa Model SVM

Metrik	Nilai
Overall Accuracy	85.37%
Macro Average Precision	0.55
Macro Average Recall	0.46
Macro Average F1-Score	0.49
Weighted Average Precision	0.80
Weighted Average Recall	0.84
Weighted Average F1-Score	0.81

Model SVM mencapai akurasi keseluruhan 85.37% dan Weighted average metrics lebih tinggi dibanding macro average karena didominasi performa kelas Neutral yang memiliki jumlah data terbanyak. Macro average

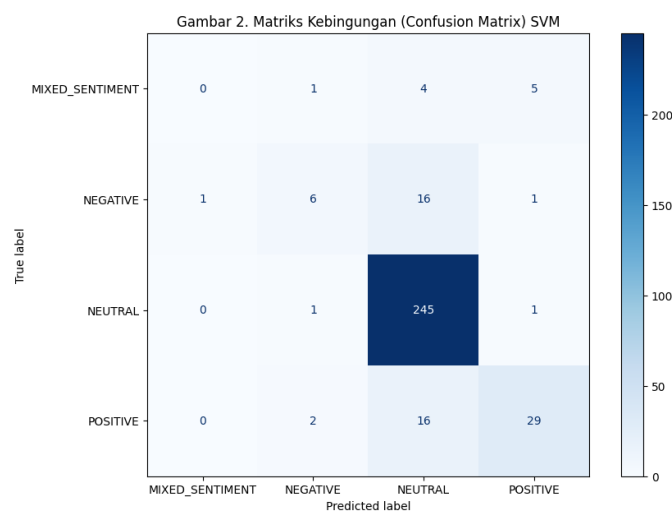
yang lebih rendah menunjukkan performa kurang merata antar kelas. Tabel 3 menampilkan performa model per kategori sentimen yang menunjukkan perbedaan signifikan antar kelas.

Tabel 3. Performa Model per Kategori Sentimen

Kelas	Precision	Recall	F1-Score	Support
Mixed Sentiment	0.00	0.00	0.00	10
Negative	0.60	0.25	0.35	24
Neutral	0.87	0.99	0.93	247
Positive	0.81	0.62	0.70	47

Berdasarkan Tabel 3, kategori Neutral menunjukkan performa terbaik dengan precision 0.87, recall 0.99, dan F1-score 0.93. Recall 0.99 menunjukkan hampir semua data neutral (245 dari 247) berhasil terdeteksi dengan benar. Kategori Positive memiliki precision yang baik (0.81) namun recall yang masih moderat (0.62), mengindikasikan bahwa 16 data *Positive* terklasifikasi salah, sebagian besar sebagai *Neutral*. Kategori Negative menunjukkan performa yang lemah dengan Recall 0.25 dan F1-score 0.35, di mana 18 dari 24 data *Negative* gagal dideteksi, sebagian besar salah diklasifikasikan sebagai *Neutral* (lihat Gambar 2). Kategori Mixed Sentiment mengalami kegagalan total dengan semua metrik bernilai 0.00 karena dataset sangat kecil dan karakteristik linguistik yang ambigu.

3.3. Analisis Confusion Matrix



Gambar 2. Confusion Matrix

Dari confusion matrix terlihat bahwa 245 dari 247 data *Neutral* berhasil diprediksi dengan benar, menunjukkan konsistensi model pada kelas mayoritas. Untuk kategori Positive, dari 47 data actual, hanya 29 yang diprediksi benar, sementara 16 diprediksi sebagai *Neutral* dan 2 sebagai *Negative*. Hal ini menunjukkan model konservatif dalam memberikan label *Positive* dan cenderung mengklasifikasikan komentar dengan sentimen ringan sebagai *Neutral*.

Kategori Negative mengalami *confussion* yang tinggi dengan hanya 6 dari 24 data yang diprediksi benar. Sebanyak 16 data *Negative* diprediksi sebagai *Neutral*, 1 sebagai *Mixed Sentiment*, dan 1 sebagai *Positive*. Kesalahan ini kemungkinan disebabkan penggunaan sarkasme atau kritik tersirat yang sulit dideteksi oleh model, diperparah oleh bias model terhadap kelas *Neutral*.

Kategori *Mixed Sentiment* mengalami kegagalan total dengan semua 10 data salah diklasifikasikan (True Positive = 0). Data tersebut tersebar menjadi 1 sebagai *Negative*, 4 sebagai *Neutral*, dan 5 sebagai *Positive*. Total 10 data *Mixed Sentiment* (*support*) diuji, namun model tidak berhasil memprediksi satupun dengan benar. Model cenderung mengklasifikasikan *Mixed Sentiment* berdasarkan sentimen dominan yang muncul dalam teks, bukan mengenali karakteristik campuran dari sentimen tersebut.

3.4. Analisis Dataset Imbalance

Dataset mengalami *severe imbalance* dengan rasio *Neutral: Positive: Negative: Mixed* sebesar 75.32:14.42:7.15:3.12. Imbalance ini berdampak signifikan terhadap performa model, dimana model bias terhadap kelas mayoritas (*Neutral*) dan performa buruk pada kelas minoritas (*Mixed Sentiment* dan *Negative*). Perbedaan antara Weighted F1-score (0.81) dan Macro F1-score (0.49) menjadi bukti kuat, di mana *Weighted metrics* lebih tinggi karena didominasi performa kelas *Neutral* yang memiliki jumlah data terbanyak. Untuk mengatasi masalah ini, beberapa solusi dapat diterapkan seperti class weighting pada SVM untuk memberikan bobot lebih tinggi pada kelas *minority*, *oversampling minority class* menggunakan teknik SMOTE (*Synthetic Minority Over-sampling Technique*), *undersampling majority class*, atau menggunakan ensemble methods yang lebih robust terhadap imbalance.

3.5. Interpretasi Persepsi Publik

Dominasi sentimen neutral (75,32%) mencerminkan sikap *wait-and-see* masyarakat terhadap kebijakan ekonomi Indonesia. Hal ini mengindikasikan masyarakat menunggu bukti konkret implementasi kebijakan sebelum memberikan penilaian definitif, awareness tinggi tentang kompleksitas isu ekonomi, dan kematangan literasi ekonomi publik yang meningkat. Rasio *positive: negative* sebesar 2.02:1 menunjukkan optimisme moderat publik terhadap arah kebijakan ekonomi. Kepercayaan publik terhadap kebijakan lebih tinggi dibanding skeptisisme, dengan harapan terhadap perbaikan kondisi ekonomi nasional. Meskipun sentimen *negative* hanya 7,15%, hal ini mengindikasikan publik tidak serta-merta menerima semua kebijakan dan menunjukkan adanya *critical thinking* dalam mengevaluasi policy. *Mixed sentiment* yang sangat rendah (3,12%) menunjukkan publik cenderung memiliki stance yang jelas tanpa ambivalensi signifikan. Hal ini mencerminkan discourse yang lebih matang dibanding platform media sosial lainnya seperti Twitter yang cenderung lebih emosional dan polarisasi.

4. KESIMPULAN

Performa terbaik dicapai pada kategori *Neutral* dengan F1-score 0,93 dan Recall 0,99, sementara kategori *Mixed Sentiment* mengalami kegagalan total klasifikasi (F1-score 0,00) dan *Negative* memiliki kinerja sangat rendah (Recall 0.25) karena *severe dataset imbalance* dan kompleksitas linguistik. Mayoritas publik menunjukkan sikap *wait-and-see* terhadap kebijakan ekonomi dengan rasio optimisme moderat (*Positive:Negative* $\approx$ 2.02:1).

Penelitian ini memberikan kontribusi metodologis implementasi SVM untuk klasifikasi sentimen multi-kelas pada teks berbahasa Indonesia dan memberikan insight tentang persepsi publik terhadap kebijakan ekonomi nasional. Untuk penelitian lanjutan disarankan mengatasi dataset imbalance menggunakan teknik *balancing*, eksplorasi *deep learning models* seperti BERT atau IndoBERT, implementasi *aspect-based sentiment analysis*, dan analisis temporal untuk tracking evolusi sentimen dari waktu ke waktu.

REFERENSI

- [1] R. A. Kaloka, "Pemanfaatan Youtube Sebagai Community Engagement Kegiatan Hubungan Masyarakat Skala Mikro," *Interak. J. Ilmu Komun.*, vol. 12, no. 2, pp. 321–331, 2023, doi: 10.14710/interaksi.12.2.321-331.
- [2] Y. C. Hua, P. Denny, J. Wicker, and K. Taskova, "A systematic review of aspect-based sentiment analysis: domains, methods, and trends," *Artif. Intell. Rev.*, vol. 57, no. 11, p. 296, 2024, doi: 10.1007/s10462-024-10906-z.
- [3] N. Lovely, N. L. R. Rahma, D. Almeida, and N. A. Rakhmawati, "Analisis Sentimen Opini Publik pada Platform X terhadap Pembangunan IKN Menggunakan Support Vector Machine (SVM)," *Etika Teknol. Inf.*, vol. 1, no. 1, pp. 1–7, 2024, doi: 10.5281/zenodo.14238234.
- [4] A. R. Padri, "Klasifikasi Kemacetan Lalu Lintas Di Indonesia Menggunakan Metode Naive Bayes Classification," *Simetris J. Tek. Mesin, Elektro dan Ilmu Komput.*, vol. 14, no. 2, pp. 297–310, 2023, doi: 10.24176/simet.v14i2.10102.

- [5] E. Sihombing, M. Halmi Dar, and F. Aini Nasution, "Comparison of Machine Learning Algorithms in Public Sentiment Analysis of TAPERA Policy," *Int. J. Sci. Technol. Manag.*, vol. 5, no. 5, pp. 1089–1098, 2024, doi: 10.46729/ijstm.v5i5.1164.
- [6] I. P. D. W. Darmawan, G. A. Pradnyana, and I. B. N. Pascima, "Optimasi Parameter Support Vector Machine Dengan Algoritma Genetika Untuk Analisis Sentimen Pada Media Sosial Instagram," *SINTECH (Science Inf. Technol. J.)*, vol. 6, no. 1, pp. 58–67, 2023, doi: 10.31598/sintechjournal.v6i1.1245.
- [7] A. A. Ningtyas, A. Solichin, and R. Pradana, "Analisis Sentimen Komentar Youtube Tentang Prediksi Resesi Ekonomi Tahun 2023 Menggunakan Algoritme Naïve Bayes," *Bit (Fakultas Teknol. Inf. Univ. Budi Luhur)*, vol. 20, no. 1, p. 9, 2023, doi: 10.36080/bit.v20i1.2317.
- [8] M. H. Wicaksono, M. D. Purbolaksono, and S. Al Faraby, "Perbandingan Algoritma Machine Learning untuk Analisis Sentimen Berbasis Aspek pada Review Female Daily," *eProceedings Eng.*, vol. 10, no. 3, pp. 3591–3600, 2023.
- [9] Chando and N. Batubara, "Sistem Akusisi Data untuk Polling berbasis Media Sosial dan Autentikasi User menggunakan reCaptcha Tugas Akhir diajukan untuk memenuhi salah satu syarat memperoleh gelar sarjana dari Program Studi Sarjana Teknik Informatika Chando Anggara Natanael Batubar," vol. 6, no. 2, pp. 9528–9536, 2019.
- [10] F. S. R. Siregar and D. E. Ratnawati, "Klasifikasi Data Paten Berdasarkan Judul dan Abstrak Menggunakan Support Vector Machine," vol. 9, no. 9, pp. 1–7, 2025.
- [11] A. F. Aufar, Mochamad Alfian Rosid, A. Eviyanti, and I. R. I. Astutik, "Optimizing Text Preprocessing for Accurate Sentiment Analysis on E-Wallet Reviews," *JICTE (Journal Inf. Comput. Technol. Educ.)*, vol. 7, no. 2, pp. 42–50, 2023, doi: 10.21070/jicte.v7i2.1650.
- [12] W. B. Zulfikar, A. R. Atmadja, and S. F. Pratama, "Sentiment Analysis on Social Media Against Public Policy Using Multinomial Naive Bayes," *Sci. J. Informatics*, vol. 10, no. 1, pp. 25–34, 2023, doi: 10.15294/sji.v10i1.39952.
- [13] P. Sulistiyawati *et al.*, "Prediksi Kata Kasar Berbahasa Indonesia Menggunakan Machine Learning Berbasis Mobile Infrastructure," *Transmisi*, vol. 24, no. 2, pp. 55–61, 2022, doi: 10.14710/transmisi.24.2.55-61.
- [14] A. Firizkiansah, A. Muhammad, and I. R. Maulana, "Optimasi Klasifikasi Data Teks Menggunakan Algoritma Logistic Regression dengan TF-IDF dan SMOTE," *JIKOMTI J. Ilm. Ilmu Komput. dan Teknol. Inf.*, vol. 2, no. 1, pp. 29–36, 2025, [Online]. Available: <https://ojs.sains.ac.id/index.php/Jikomti/article/view/97/119>
- [15] M. Fadli and R. A. Saputra, "KLASIFIKASI DAN EVALUASI PERFORMA MODEL RANDOM FOREST UNTUK PREDIKSI STROKE Classification And Evaluation Of Performance Models Random Forest For Stroke Prediction," *J. Tek.*, vol. 12, no. 2, pp. 72–80, 2023, [Online]. Available: <http://jurnal.umt.ac.id/index.php/jt/index>