SENTRI: Jurnal Riset Ilmiah
https://ejournal.nusantaraglobal.ac.id/index.php/sentri
Email: sentri@nusantaraglobal.ac.id
DOI: https://doi.org/10.55681/sentri.v5i2.5858

Vol. 5, No. 2, February 2026
e-ISSN: 2963-1130
pp. 2353-2361

Analisis Komparatif Sentimen Publik terhadap Liputan Media Terkait Aksi Menteri Keuangan Menggunakan Algoritma SVM dan RoBERTa

Budi Santosa[1*], Kardita Magda[1], Ega Budiman[1], Ryan Randy Suryono[1]
[1]Universitas Teknokrat Indonesia, Bandar Lampung, Indonesia

*Corresponding Author's e-mail: budi.santosa@teknokrat.ac.id

Article History:
Received: Januari 21, 2026
Revised: February 19, 2026
Accepted: February 27, 2026

Keywords:
Sentiment Analysis, YouTube Comments, SVM, RoBERTa, Finance Minister

Abstract: *Public opinion on social media is a crucial representation of government policy legitimacy, especially in the fiscal sector. This study intends to provide a comparative investigation of the efficacy of sentiment categorization on YouTube comments pertaining to the activities of the Indonesian Finance Minister by juxtaposing the Support Vector Machine (SVM) algorithm with the RoBERTa Transformer model. A total of 3,780 comments were acquired from national digital media channels. The research method involves intensive text preprocessing, including stemming using the Sastrawi algorithm and lexicon-based labeling. The results showed that the SVM algorithm with TF-IDF features achieved an accuracy of 83.33% and an F1-score of 76.05%. In contrast, the RoBERTa model showed a significantly lower performance with an accuracy of 29.76%. This study concludes that for datasets dominated by neutral sentiments and informal language in specific Indonesian contexts, traditional machine learning like SVM with optimal feature engineering remains more reliable and efficient than complex Transformer models that require more extensive fine-tuning.*

Copyright © 2026, The Author(s).
This is an open access article under the CC–BY-SA license

How to cite: Santosa, B., Magda, K., Budiman, E., & Suryono, R. R. (2026). Analisis Komparatif Sentimen Publik terhadap Liputan Media Terkait Aksi Menteri Keuangan Menggunakan Algoritma SVM dan RoBERTa. *SENTRI: Jurnal Riset Ilmiah*, *5*(2), 2353–2361. https://doi.org/10.55681/sentri.v5i2.5858

PENDAHULUAN

Opini publik yang berkembang pesat di media sosial kini menjadi instrumen krusial bagi legitimasi kebijakan pemerintah, khususnya di sektor fiskal yang dikelola oleh Kementerian Keuangan. Sebagai platform berbasis video dengan interaksi komentar yang masif, YouTube menjadi sumber data utama untuk memahami respons masyarakat terhadap kebijakan strategis yang diambil oleh figur publik seperti Menteri Keuangan (Pramesti, 2025). Perubahan kepemimpinan atau kebijakan fiskal yang signifikan seringkali memicu gelombang diskusi digital yang mencerminkan tingkat kepercayaan dan transparansi tata kelola negara di mata masyarakat luas (Khaidar & Kurnia, 2025).

Namun, menganalisis opini publik dari platform digital seperti YouTube menghadirkan tantangan teknis yang kompleks. Data teks yang dihasilkan seringkali tidak terstruktur, menggunakan bahasa informal, dialek lokal, hingga penggunaan sarkasme yang tinggi (Kamila et al., 2025). Selain itu, distribusi kelas sentimen dalam interaksi media sosial di Indonesia cenderung tidak seimbang (class imbalance), di mana sentimen netral seringkali mendominasi dibandingkan sentimen positif atau negatif yang ekstrem (Alita et al., 2023). Tantangan ini menuntut penggunaan metode komputasi yang tepat untuk dapat menangkap esensi opini publik secara akurat dan efisien.

Dalam perkembangan teknologi pemrosesan bahasa alami (NLP), algoritma machine learning klasik seperti Support Vector Machine (SVM) telah lama dikenal keandalannya dalam menangani klasifikasi teks pendek dengan fitur frekuensi kata (Agustianto et al., 2025). Di sisi lain, munculnya arsitektur berbasis Transformer seperti RoBERTa (Robustly Optimized BERT Approach) menjanjikan kemampuan yang lebih unggul dalam menangkap konteks semantik dan ambiguitas bahasa dibandingkan model tradisional (Syafia et al., 2023). Penerapan model Deep Learning ini diproyeksikan mampu membedah kompleksitas struktural bahasa Indonesia yang digunakan di media sosial secara lebih komprehensif.

Penelitian terdahulu telah banyak mengeksplorasi perbandingan antara model machine learning klasik dan model berbasis Transformer pada berbagai topik, mulai dari ulasan aplikasi hingga isu politik nasional (Putri & Lestari, 2024). Beberapa studi menunjukkan bahwa meskipun model kompleks seperti BERT atau RoBERTa seringkali unggul dalam akurasi, model klasik seperti SVM tetap kompetitif dan lebih efisien pada dataset dengan skala tertentu atau pada kondisi data yang sangat tidak seimbang (Rohman & Agung, 2025). Hal ini memicu kebutuhan untuk menguji kembali performa kedua pendekatan tersebut secara spesifik pada konteks liputan media terkait kebijakan fiskal di Indonesia.

Maka dari itu, penelitian ini dilakukan untuk membandingkan kinerja dan efektivitas algoritma klasifikasi sentimen dalam mengolah data komentar dari platform YouTube terkait aksi dan kebijakan Menteri Keuangan. Dengan membandingkan performa algoritma SVM berbasis TF-IDF dan model RoBERTa, penelitian ini diharapkan dapat memberikan rekomendasi metodologi yang paling efektif dalam memetakan persepsi publik. Melalui tahapan prapemrosesan teks yang intensif dan pengujian metrik performa yang ketat, temuan ini diharapkan menjadi referensi bagi pemerintah dalam melakukan pemantauan media (social media monitoring) guna merespons dinamika opini masyarakat secara tepat waktu dan berbasis data (Ulinuha et al., 2025).

LANDASAN TEORI

Sebagai sub-bidang dari Natural Language Processing (NLP), analisis sentimen difungsikan untuk mengidentifikasi serta mengategorikan opini, manifestasi emosi, dan orientasi sikap publik yang terkandung dalam data tekstual (Rahmadina et al., 2025). Dalam konteks kebijakan publik, analisis ini menjadi instrumen penting bagi pemerintah untuk mengukur tingkat penerimaan masyarakat terhadap regulasi fiskal yang dikelola oleh Kementerian Keuangan (Haliza & Ikhsan, 2025). Secara metodologis, pengolahan data diawali dengan tahap prapemrosesan teks yang intensif untuk membersihkan data dari derau (noise) sebelum masuk ke tahap klasifikasi (Apriansyah et al., 2025).

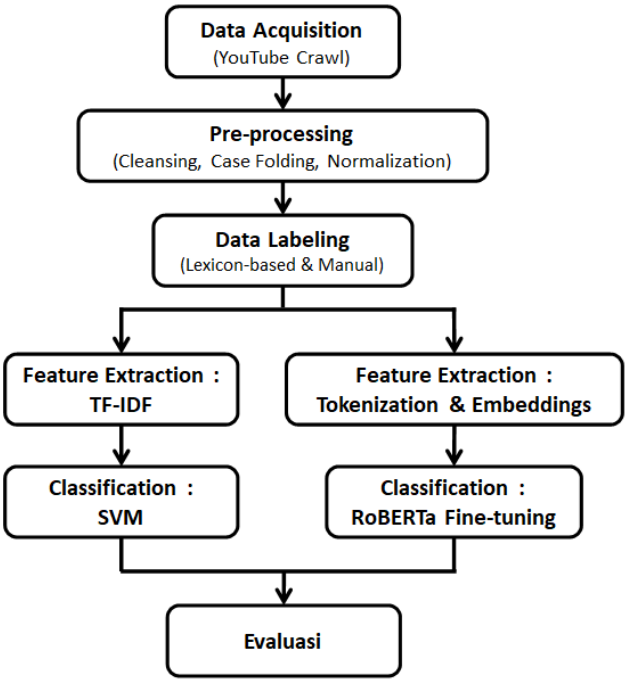
Support Vector Machine (SVM) merupakan algoritma pembelajaran mesin berbasis statistik yang mengandalkan pencarian hyperplane optimal sebagai pemisah antar-kelas dalam ruang dimensi tinggi (Salsabila et al., 2025). Dalam ranah analisis tekstual berbahasa Indonesia, algoritma SVM kerap diintegrasikan dengan teknik Term Frequency-Inverse Document Frequency (TF-IDF) guna mengoptimalkan pembobotan kata-kata yang memiliki signifikansi informasi tinggi (Tanggraeni & Sitokdana, 2022). Penelitian terdahulu menunjukkan bahwa SVM memiliki performa yang sangat stabil pada dataset dengan ukuran menengah dan sering kali mengungguli metode klasifikasi tradisional lainnya (Kusuma & Nugroho, 2021).

RoBERTa (Robustly Optimized BERT Approach) merupakan pengembangan dari arsitektur BERT yang melakukan optimasi pada proses pelatihan dengan menggunakan dataset yang lebih besar dan waktu pelatihan yang lebih lama (Fauzi, 2019). Model ini menggunakan mekanisme transformer yang memungkinkan sistem memahami hubungan kontekstual antar kata secara dua arah (bidirectional), sehingga mampu menangkap nuansa semantik yang lebih kompleks dalam bahasa informal media sosial. Meskipun memerlukan sumber daya komputasi yang tinggi, model berbasis transformer ini sering kali dianggap sebagai standar terkini dalam mencapai akurasi tinggi pada tugas klasifikasi teks (Ulinuha et al., 2025).

Sastrawi Stemmer digunakan dalam penelitian ini sebagai alat bantu untuk proses stemming teks bahasa Indonesia (Salsabila et al., 2025). Tahapan ini memegang peranan krusial dalam mentransformasikan kata berimbuhan ke dalam bentuk dasarnya, sehingga berbagai variasi leksikal dengan akar semantik yang sama dapat dikenali sebagai fitur tunggal yang konsisten oleh algoritma klasifikasi. Penggunaan stemmer yang tepat sangat berpengaruh terhadap pengurangan dimensi fitur dan peningkatan akurasi pada algoritma berbasis frekuensi kata seperti SVM (Agresia et al., 2025).

METODE PENELITIAN

Metodologi penelitian ini dirancang untuk membandingkan efektivitas algoritma machine learning konvensional dengan model deep learning dalam konteks opini publik nasional. Tahapan penelitian digambarkan melalui alur sistematis berikut :



Gambar 1. Diagram Alir Tahapan Penelitian

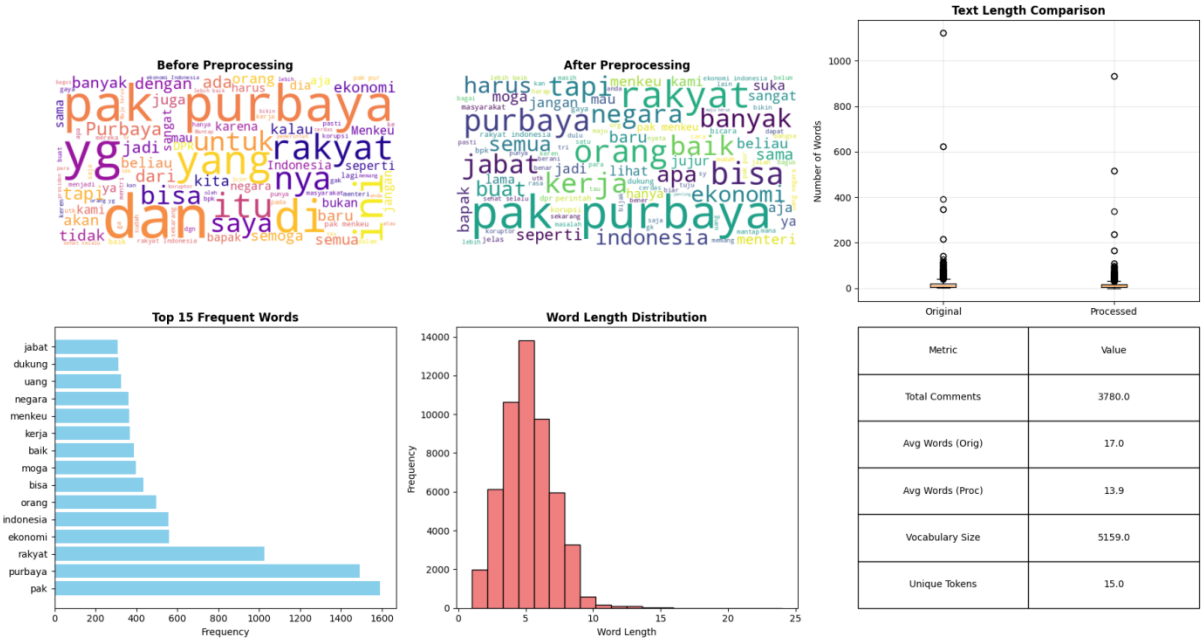
1. Akuisisi Data

Data diperoleh melalui proses ekstraksi komentar pada kanal YouTube media digital nasional terkait kebijakan fiskal Menteri Keuangan dengan total 3.780 komentar unik.

Dataset tersebut selanjutnya diarsip dalam format Comma-Separated Values (CSV) guna memfasilitasi tahapan pemrosesan data berikutnya.

2. Prapemrosesan Teks

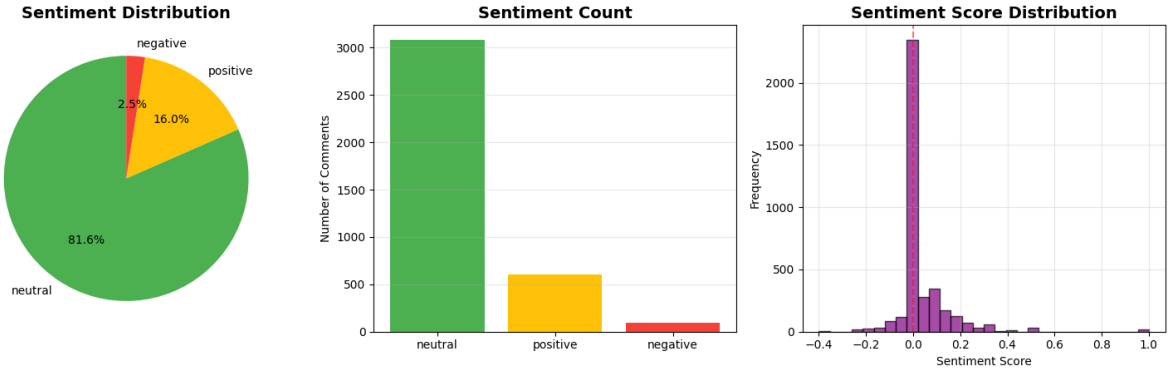
Tahap ini melibatkan pembersihan data (cleaning) dari URL, simbol, dan angka. Selanjutnya, dilakukan penghapusan 65 stopwords bahasa Indonesia untuk mereduksi kata yang tidak memiliki nilai informatif. Proses stemming menggunakan algoritma Sastrawi diterapkan untuk mengubah kata berimbuhan menjadi kata dasar guna meningkatkan konsistensi fitur teks.



Gambar 2. Hasil Pemrosesan Teks

3. Pelabelan dan Ekstraksi Fitur

roses pelabelan data dikerjakan secara otomatis melalui pendekatan leksikon yang telah tervalidasi. Sementara itu, pada algoritma SVM, ekstraksi fitur dilakukan menggunakan metode TF-IDF dengan menerapkan konfigurasi unigram dan bigram untuk merekam konteks kata yang saling berdekatan.



Gambar 3. Hasil Pelabelan Sentimen

4. Implementasi Model

Implementasi model dalam penelitian ini membandingkan dua paradigma berbeda dalam Natural Language Processing (NLP): pendekatan statistik berbasis frekuensi (Support Vector Machine) dan pendekatan berbasis konteks (RoBERTa).

Algoritma SVM diimplementasikan menggunakan kernel linear yang dipilih karena efisiensinya dalam menangani ruang fitur berdimensi tinggi seperti teks. Proses kerja SVM dalam penelitian ini meliputi:

a. Optimal Hyperplane: Model mencari garis pemisah (hyperplane) terbaik yang memaksimalkan jarak (margin) antara data sentimen positif, negatif, dan netral.
b. Regularization Parameter (C): Parameter C diatur pada nilai 1.0 untuk menyeimbangkan antara klasifikasi data latih yang benar dan margin keputusan yang lebih luas guna menghindari overfitting.
c. Class Weighting: Karena adanya ketidakseimbangan kelas (class imbalance), fitur class_weight='balanced' digunakan untuk memberikan bobot lebih besar pada kelas minoritas (positif dan negatif) agar model tidak bias terhadap kelas mayoritas (netral).

Model RoBERTa (Robustly Optimized BERT Approach) merupakan varian dari BERT yang telah dioptimasi melalui pelatihan pada dataset yang lebih besar dengan teknik Dynamic Masking. Implementasi model ini meliputi:

a. Self-Attention Mechanism: RoBERTa menggunakan mekanisme multi-head attention yang memungkinkan model memberikan bobot kepentingan yang berbeda pada setiap kata dalam satu kalimat secara simultan. Hal ini membantu model memahami kata-kata ambigu berdasarkan konteks kata-kata di sekitarnya.
b. Fine-Tuning Architecture: Penelitian ini melakukan fine-tuning pada model roberta-base dengan menambahkan lapisan fully connected di atas representasi hidden state untuk klasifikasi tiga kelas sentimen.
c. Training Hyperparameters: Model dilatih dengan menggunakan batch size sebesar 8, learning rate yang diatur melalui warmup steps sebanyak 100, dan dilatih selama 3 epochs untuk meminimalkan cross-entropy loss.3.4.3. Perbedaan Ekstraksi FiturPerbedaan mendasar dalam implementasi ini terletak pada cara model "melihat" teks. SVM melihat teks sebagai kumpulan kata independen yang diberi bobot melalui TF-IDF (pendekatan bag-of-words), sedangkan RoBERTa melihat teks sebagai urutan token yang memiliki hubungan semantik satu sama lain (word embeddings), sehingga mampu menangkap makna di balik struktur kalimat yang kompleks.

HASIL DAN PEMBAHASAN

Hasil pengujian performa model menunjukkan perbedaan yang signifikan antara metode berbasis frekuensi kata dengan model berbasis konteks semantik pada dataset ini.

1. Analisis Deskriptif Data

Sebelum pengujian, dilakukan analisis terhadap sebaran kelas sentimen untuk memahami karakteristik opini publik yang masuk.

Tabel. 1 Distribusi Sentimen Publik pada Dataset

No	Sentimen	Jumlah Komentar	Persentase
1	Netral	3.083	81.56%
2	Positif	604	15,98%
3	Negatif	93	2,46%

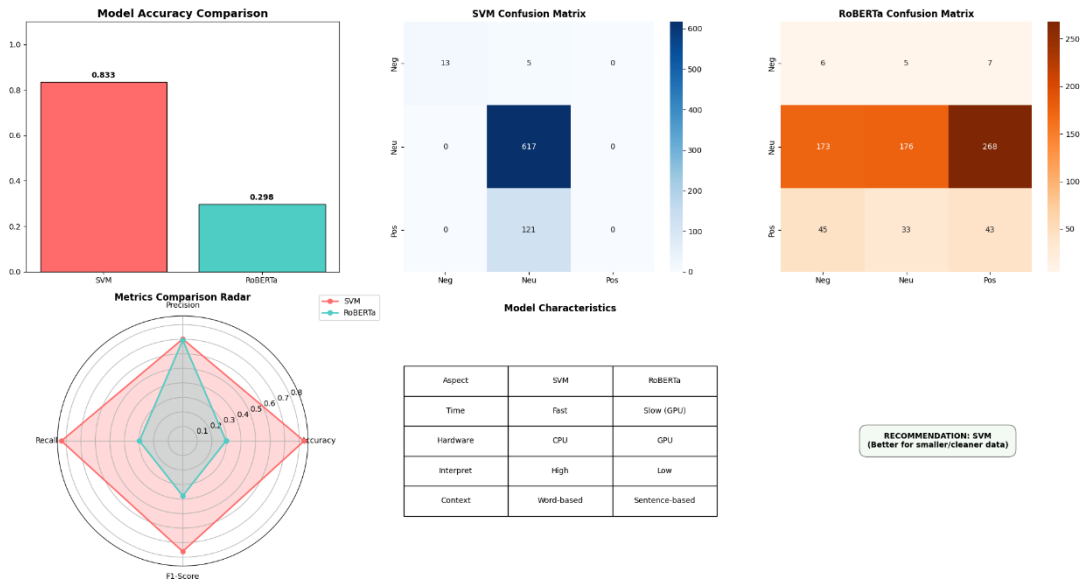
Dominasi kelas Netral (81,56%) mengindikasikan bahwa mayoritas interaksi masyarakat bersifat observatif. Namun, distribusi yang sangat timpang ini (extreme class imbalance) menjadi tantangan besar. Meskipun demikian, penggunaan teknik class balancing pada tahap prapemrosesan terbukti membantu model dalam mengenali ciri khas unik dari kelas minoritas.

2. Perbandingan Performa Model

Berdasarkan hasil pengujian yang dievaluasi menggunakan confusion matrix, diperoleh parameter performa berupa akurasi dan F1-score dengan rincian sebagai berikut:

Tabel. 2 Hasil Pengujian Performa SVM vs RoBERTa

No	Model	Accuracy	Precision	Recall	F1-Score
1	Support Vector Machine (SVM)	0.8333	0.7015	0.8333	0.7605
2	RoBERTa (Fine-tuned)	0.2976	0.6935	0.2976	0.3782



Gambar 4. Evaluasi Model : SVM dan RoBERTa

3. Analisis Keunggulan SVM

Algoritma SVM mencapai akurasi 83,33%, yang menunjukkan efektivitas tinggi dalam memetakan teks pendek media sosial. Keberhasilan ini didorong oleh implementasi feature engineering yang optimal menggunakan TF-IDF dengan kombinasi N-Gram (2.000 unigram dan 3.000 bigram). Temuan ini selaras dengan studi terdahulu yang menegaskan bahwa penggunaan fitur N-gram memegang peranan vital bagi model linier dalam mengidentifikasi frasa bermakna ganda pada data tekstual informal (Kusuma & Nugroho, 2021).

4. Analisis Rendahnya Performa RoBERTa

Sebaliknya, model RoBERTa memberikan akurasi yang rendah (29,76%). Fenomena ini disebabkan oleh beberapa faktor teknis:

a. **Kesenjangan Domain**: Model transformer standar seringkali mengalami kesulitan saat dihadapkan pada dialek komentar YouTube Indonesia yang sarat dengan singkatan dan bahasa gaul (Agustianto et al., 2025).
b. **Kebutuhan Data**: Model kompleks seperti RoBERTa memerlukan dataset latih yang jauh lebih besar dan seimbang untuk dapat melakukan fine-tuning secara efektif pada konteks lokal (Ulinuha et al., 2025).

KESIMPULAN

Berlandaskan pada hasil eksperimen dan studi komparatif mengenai sentimen publik terhadap pemberitaan media terkait kebijakan Menteri Keuangan, penelitian ini merumuskan sejumlah kesimpulan fundamental sebagai berikut:

1. Efektivitas Algoritma Klasifikasi

Penelitian ini membuktikan bahwa algoritma Support Vector Machine (SVM) dengan fitur TF-IDF/N-Gram memiliki keunggulan yang signifikan dibandingkan model RoBERTa dalam mengklasifikasikan opini publik pada dataset ini. SVM berhasil mencapai tingkat akurasi sebesar 83,33% dan F1-score sebesar 76,05%. Hal ini menunjukkan bahwa untuk dataset teks bahasa Indonesia yang telah melalui prapemrosesan intensif (pembersihan 65 stopwords dan stemming Sastrawi), model machine learning tradisional tetap menjadi pilihan yang sangat andal dan efisien.

2. Tantangan Model Transformer

Model RoBERTa menunjukkan performa yang rendah dengan akurasi hanya mencapai 29,76% dan F1-score 37,82%. Rendahnya performa ini disebabkan oleh karakteristik data komentar YouTube yang didominasi oleh kelas netral (81,56%) serta adanya ambiguitas bahasa informal dan sarkasme yang sulit dipetakan oleh model transformer tanpa fine-tuning pada korpus bahasa lokal yang lebih masif. Ketimpangan distribusi data (class imbalance), terutama pada kelas negatif yang hanya berjumlah 93 komentar, menjadi faktor penghambat utama bagi generalisasi model deep learning.

3. Relevansi Pilihan Metodologi

Penggunaan fitur TF-IDF dengan konfigurasi 2.000 unigram dan 3.000 bigram terbukti krusial dalam meningkatkan performa klasifikasi SVM. Strategi ini memungkinkan model untuk menangkap konteks kata yang berdekatan secara lebih efektif dibandingkan hanya menggunakan unigram tunggal. Oleh karena itu, penelitian ini menyimpulkan bahwa

untuk implementasi sistem pemantauan media sosial (social media monitoring) di instansi pemerintah yang memerlukan kecepatan inferensi tinggi dan interpretabilitas model yang baik, algoritma SVM lebih direkomendasikan untuk digunakan dalam skala produksi.

4. Implikasi dan Rekomendasi (Future Works)

Temuan ini memberikan kontribusi pada pengembangan sistem analisis sentimen kebijakan publik di Indonesia. Untuk penelitian mendatang, disarankan beberapa pengembangan berikut:

a. Penerapan teknik augmentasi data atau transfer learning menggunakan model bahasa yang lebih spesifik seperti IndoBERT untuk mengatasi keterbatasan data.
b. Eksplorasi penggunaan word embedding lain seperti Word2Vec untuk memperkaya ekstraksi fitur pada model linear.
c. Perluasan dataset dari berbagai platform media sosial lainnya guna mendapatkan representasi opini publik yang lebih komprehensif terkait isu-isu fiskal nasional.

DAFTAR REFERENSI

Agresia, V., Suryono, R. R., Indonesia, U. T., Ratu, L., & Lampung, K. B. (2025). *COMPARISON OF SVM , NAÏVE BAYES , AND LOGISTIC REGRESSION ALGORITHMS FOR SENTIMENT ANALYSIS OF FRAUD AND BOTS IN PURCHASING CONCERT TICKET KOMPARASI ALGORITMA SVM , NAÏVE BAYES , DAN LOGISTIC REGRESSION UNTUK ANALISIS SENTIME*.

Agustianto, B., Musyafa, A., & Waskita, A. A. (2025). *Analisis Sentimen Publik Terhadap Enterprise Resource Planning (ERP) Di Media Sosial X Menggunakan Model Roberta Dan Twitbert*. *8*(2), 142–152.

Alita, D., Indonesia, U. T., Classifier, N. B., Machine, V., & Mining, T. (2023). *Sentimen Analisis Vaksin Covid-19 Menggunakan Naive Bayes Dan Support Vector Machine*. *1*(1), 1–12.

Apriansyah, F. M., Ramadhan, T. I., Hidayat, C. R., & Karta, A. (2025). *Perbandingan IndoBERT dan IndoRoBERTa Untuk Analisis Sentimen Pada Film Dokumenter Dirty Vote*. *10*(3), 593–605. https://doi.org/10.30591/jpit.v10i3.8607

Fauzi, M. A. (2019). *Word2Vec model for sentiment analysis of product reviews in Indonesian language*. *9*(1), 525–530. https://doi.org/10.11591/ijece.v9i1.pp525-530

Haliza, D., & Ikhsan, M. (2025). *Sentiment Analysis on Public Perception of the Nusantara Capital on Social Media X Using Support Vector Machine (SVM) and K-Nearest Neighbor (K-NN) Methods*. *9*(3), 716–723.

Kamila, S. A., Wahda, A. W., Febriati, B. N., & Yudhianto, R. B. (2025). *AUGMENTASI GPT-4O DAN FINE-TUNING INDOBERT UNTUK ANALISIS SENTIMEN PUBLIK PADA ISU RESHUFFLE MENTERI KEUANGAN*. *10*(2), 1–17.

Khaidar, A., & Kurnia, S. (2025). *Intelligent Sentiment Mapping on Instagram for Finance Minister Purbaya Yudhi Sadewa Based on Logistic Regression*. *5*(1).

Kusuma, A., & Nugroho, A. (2021). *Analisa Sentimen Pada Twitter Terhadap Kenaikan Tarif Dasar Listrik Dengan Metode Naïve Bayes*. *15*(2), 137–146.

Pramesti, R. F. (2025). *ANALISIS EFISIENSI APBN ERA PRABOWO : KAJIAN EKONOMI DAN ANALISIS SENTIMEN PUBLIK*. *8*(2), 1147–1161.

Putri, I., & Lestari, M. T. (2024). *ANALISIS SENTIMEN OPINI PUBLIK TERHADAP LEMBAGA MEDIA TWITTER*. *2*(1), 20–22.

Rahmadina, A. P., Setiawan, N. Y., & Rusydi, A. N. (2025). *ANALISIS SENTIMEN TERHADAP KEBIJAKAN KENAIKAN PPN 12 % MENGGUNAKAN SUPPORT VECTOR MACHINE (SVM)*. *9*(9), 1–10.

Rohman, S., & Agung, I. W. P. (2025). *KOMPARASI ALGORITMA INDOBERT, SVM, DAN RANDOM FOREST DALAM ANALISIS SENTIMEN PROGRAM MAKAN BERGIZI GRATIS*. *9*(6), 9464–9471.

Salsabila, A., Priyatna, B., & Hananto, A. (2025). *Komparasi Kinerja Model Naive Bayes , SVM , dan BERT dalam Klasifikasi Sentimen Ulasan Pada Aplikasi YUMMY*. *4*(2), 42–47.

Syafia, A. N., Hidayattullah, M. F., & Suteddy, W. (2023). *Studi Komparasi Algoritma SVM Dan Random Forest Pada Analisis Sentimen Komentar Youtube BTS*. *8*(3), 207–212.

Tanggraeni, A. I., & Sitokdana, M. N. N. (2022). *Analisis Sentimen Aplikasi E-Government Pada Google Play Menggunakan Algoritma Naïve Bayes*. *9*(2), 785–795.

Ulinuha, A., Majid, E., Nuari, R., Indonesia, U. T., & Lampung, B. (2025). *PERFORMANCE COMPARISON OF BERT METRICS AND CLASSICAL MACHINE LEARNING MODELS (SVM , NAIVE BAYES) FOR SENTIMENT ANALYSIS PERBANDINGAN KINERJA METRIK BERT DAN MODEL MACHINE LEARNING KLASIK (SVM , NAIVE BAYES) UNTUK*. *10*(2), 741–752.