



ANALISIS SENTIMEN OPINI MASYARAKAT TERHADAP KEBIJAKAN KENAIKAN HARGA PERTAMAX MENGGUNAKAN METODE SUPPORT VECTOR MACHINE

Haban Al-ma'rufi ^{1*}

¹ Program Studi Teknik Informatika, Universitas Sains Al-Qur'an

E-mail: habanal0987@gmail.com

ABSTRACT

While the Pertamina price adjustment policy frequently generates public debate on social media, the use of YouTube text comments for data-driven policy evaluation remains limited. This study aims to classify public sentiment regarding the Pertamina price hike and evaluate the reliability of the Support Vector Machine (SVM) algorithm. A total of 2,403 initial comments were extracted via Instant Data Scraper from official YouTube news channels. Following preprocessing (cleansing, case folding, tokenizing, stopword removal, and Sastrawi stemming), 2,359 clean data points were categorized into three sentiment classes: Negative, Neutral, and Positive. The data was transformed using TF-IDF weighting and split with an 80:20 ratio. Classification modeling was conducted using SVM and evaluated via a confusion matrix. The results showed that the SVM model achieved an accuracy of 86.65% (weighted accuracy 87%), precision of 0.87, recall of 0.87, and F1-score of 0.86. Sentiment distribution was dominated by negative responses at 55.53% (1,310 comments), neutral at 35.40% (835 comments), and positive at 9.07% (214 comments). These findings demonstrate SVM's effectiveness in mapping unstructured public opinion for real-time policy monitoring.

Keywords: *Instant Data Scraper, Pertamina, sentiment analysis, Support Vector Machine, TF-IDF, YouTube*

ABSTRAK

Meskipun kebijakan penyesuaian harga Pertamina menimbulkan beragam reaksi publik di media sosial, analisis berbasis data terhadap komentar di YouTube belum banyak dimanfaatkan untuk evaluasi kebijakan. Penelitian ini bertujuan mengklasifikasikan sentimen opini masyarakat terkait kenaikan harga Pertamina dan menguji keandalan algoritma *Support Vector Machine* (SVM). Data sekunder sebanyak 2.403 komentar dikumpulkan menggunakan *Instant Data Scraper* dari saluran berita resmi YouTube. Setelah melalui praproses (pembersihan, *case folding*, *tokenizing*, *stopword removal*, dan *stemming* Sastrawi), diperoleh 2.359 data bersih yang dikelompokkan ke dalam tiga kelas sentimen (Negatif, Netral, Positif). Data ditransformasikan menggunakan TF-IDF dan

Article History

Received: Juli 2026

Reviewed: Juli 2026

Published: Juli 2026

Plagiarism Checker No
237.u873247.fh8374

Copyright : Author
Publish by : Kohesi



This work is licensed
under a [Creative Commons Attribution-NonCommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/)



dibagi dengan rasio 80:20. Pemodelan diklasifikasikan dengan SVM dan dievaluasi menggunakan *confusion matrix*. Hasil pengujian menunjukkan model SVM mencapai akurasi sebesar 86,65% (*weighted accuracy* 87%), *precision* 0,87, *recall* 0,87, dan *F1-score* 0,86. Distribusi sentimen didominasi oleh kelas negatif sebesar 55,53% (1.310 komentar), netral 35,40% (835 komentar), dan positif 9,07% (214 komentar). Temuan ini membuktikan efektivitas SVM dalam memetakan opini tidak terstruktur dan berpotensi diterapkan sebagai *early warning system* bagi pemangku kebijakan.

Kata Kunci: *analisis sentimen, Instant Data Scraper, Pertamina, Support Vector Machine, TF-IDF, YouTube*

1. PENDAHULUAN

Perubahan harga Pertamina selalu menjadi topik hangat yang memicu perdebatan di tengah masyarakat. Sebagai bahan bakar non-subsidi yang diandalkan kelompok kelas menengah dan industri transportasi, komoditas ini sangat memengaruhi hajat hidup orang banyak. Fluktuasi harga minyak mentah di pasar global sendiri kerap memaksa pemerintah bersama PT Pertamina (Persero) menyesuaikan tarif secara berkala demi menyelamatkan anggaran negara. Dampaknya, setiap kali kenaikan harga diumumkan, kolom komentar di platform digital seperti YouTube langsung dibanjiri beragam reaksi publik, mulai dari kritik tajam, ungkapan kecewa, hingga sikap maklum atas situasi ekonomi. Sayangnya, tumpukan opini digital yang bernilai tinggi ini belum dimanfaatkan secara optimal oleh pihak akademisi maupun lembaga terkait untuk mengevaluasi dampak psikososial kebijakan tersebut. Oleh sebab itu, riset analisis sentimen terhadap komentar masyarakat mengenai kenaikan harga Pertamina sangat penting dilakukan demi memetakan dinamika respons publik secara otomatis, objektif, dan terukur.

Penelitian mengenai analisis sentimen terhadap kebijakan publik dan komoditas energi terus berkembang pesat seiring meningkatnya penerapan teknik text mining dan pembelajaran mesin (Liu, 2012). Beberapa studi terdahulu menunjukkan bahwa platform digital seperti X (Twitter) atau Instagram menjadi sumber data sekunder yang kaya untuk menangkap polaritas opini publik (Jain & Katkar, 2015). Dalam konteks metodologi, algoritma klasifikasi berbasis terbimbing (*supervised learning*) kerap digunakan untuk mengelompokkan teks ke dalam kelas sentimen positif dan negatif. Proses pra-proses data teks—seperti *case folding*, *cleansing*, *tokenization*, *stopword removal*, dan *stemming*—serta metode pembobotan kata menggunakan *Term Frequency-Inverse Document Frequency* (TF-IDF) terbukti menjadi tahapan standar yang sangat menentukan kualitas fitur teks sebelum dimasukkan ke dalam model komputasi.

Dalam konteks empiris komparasi algoritma, banyak peneliti mengkaji performa berbagai model *machine learning* untuk data teks berbahasa Indonesia. Penelitian terdahulu memperlihatkan bahwa algoritma seperti Naïve Bayes dan K-Nearest Neighbor sering kali menghasilkan akurasi yang cukup baik untuk dataset berskala kecil [4], namun performanya cenderung menurun atau kurang stabil ketika dihadapkan pada karakteristik data teks opini publik yang memiliki tingkat



ambiguitas tinggi, penggunaan bahasa slang, dan pola kalimat yang tidak terstruktur. Sebaliknya, beberapa studi mengenai optimasi klasifikasi teks menegaskan bahwa pendekatan yang memanfaatkan ruang hiperplana (hyperplane) terpisah memiliki keunggulan inheren [10] dalam menangani data dimensi tinggi seperti teks hasil ekstraksi TF-IDF. Hal ini memperlihatkan bahwa pemilihan algoritma yang tepat menjadi kunci utama, di mana karakteristik model sangat memengaruhi tingkat presisi dan reliabilitas hasil pemetaan sentimen yang diperoleh.

Meskipun penelitian terdahulu telah membuktikan efektivitas *machine learning* dalam memetakan opini publik, masih terdapat beberapa celah penting yang perlu diperhatikan. Sebagian besar kajian sentimen kebijakan ekonomi di Indonesia cenderung menggunakan algoritma klasifikasi probabilistik sederhana yang rentan terhadap *overfitting* ketika jumlah kosakata meluas, atau sebaliknya, menggunakan model ansambel (*ensemble learning*) yang sangat kompleks sehingga membutuhkan komputasi tinggi dan kurang praktis untuk diimplementasikan sebagai alat monitoring harian. Di samping itu, kajian yang secara khusus menempatkan algoritma *Support Vector Machine* (SVM) sebagai model utama untuk menganalisis sentimen masyarakat terkait isu spesifik kenaikan harga Pertamina—dengan pengujian metrik evaluasi yang mendalam dan visualisasi teks yang komprehensif—masih relatif terbatas. Dengan kata lain, masih dibutuhkan penelitian terarah yang berfokus pada pembangunan model prediksi teks yang kokoh, stabil pada dimensi tinggi, dan aplikatif untuk mendukung evaluasi kebijakan energi secara objektif.

Penelitian ini menawarkan kebaruan pada penerapan algoritma *Support Vector Machine* (SVM) dengan pembobotan TF-IDF sebagai model utama untuk mengklasifikasikan analisis sentimen opini masyarakat terhadap kebijakan kenaikan harga Pertamina secara spesifik. Berbeda dari penelitian sebelumnya yang banyak menitikberatkan pada eksplorasi multi-model atau skenario praproses yang berbelit-belit, studi ini diarahkan pada pembangunan model klasifikasi multikelas (negatif, netral, dan positif) yang fokus, berkinerja tinggi pada teks dimensi tinggi, dan realistis untuk mendukung analisis sentimen berbasis data teks. Kontribusi penelitian ini terletak pada upaya menjembatani teknik komputasi *text mining* dengan kebutuhan analisis kebijakan publik, sehingga *output* penelitian berupa *confusion matrix*, metrik akurasi, dan visualisasi *word cloud* tidak hanya bernilai secara teknis, tetapi juga dapat dimanfaatkan oleh pemangku kebijakan sebagai instrumen pemantauan (*monitoring*) respons masyarakat yang valid.

Berdasarkan latar belakang, kajian literatur, dan kesenjangan penelitian tersebut, penelitian ini bertujuan untuk menganalisis sentimen opini masyarakat terhadap kebijakan kenaikan harga Pertamina berdasarkan data komentar YouTube menggunakan metode *Support Vector Machine* (SVM). Secara lebih spesifik, penelitian ini diarahkan untuk melakukan ekstraksi dan praproses data opini teks dari media digital, menerapkan pembobotan TF-IDF dan pemodelan klasifikasi SVM, mengevaluasi kinerja model melalui metrik klasifikasi (*accuracy*, *precision*, *recall*), serta memetakan kecenderungan respons publik melalui visualisasi data ilmiah yang representatif.

2. METODE PENELITIAN

2.1 Jenis dan Desain Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan desain eksperimen komputasional. Pendekatan ini dipilih karena penelitian berfokus pada pengolahan data teks opini masyarakat



tidak terstruktur (*unstructured text*) untuk ditransformasikan ke dalam representasi numerik dan dibangun menjadi model prediksi klasifikasi sentimen menggunakan algoritma *Support Vector Machine* (SVM).

2.2 Lokasi dan Waktu Penelitian

Penelitian dilaksanakan secara daring (komputasional) dengan mengeksplorasi data publik pada kolom komentar platform YouTube. Proses pengumpulan data, praproses, pelabelan, pemodelan, dan evaluasi performa algoritma dilakukan pada bulan Mei-Juli 2026.

2.3 Populasi, Sampel, atau Sumber Data

Populasi dalam penelitian ini mencakup seluruh komentar publik berbahasa Indonesia pada video bertema kebijakan kenaikan harga Pertamina di platform YouTube. Sumber data yang digunakan berupa data sekunder yang diekstraksi dari kolom komentar saluran (*channel*) berita nasional resmi di YouTube.

Teknik pengambilan sampel dilakukan dengan pendekatan *purposive sampling*, di mana penentuan sampel didasarkan pada tiga kriteria utama, yaitu: (1) komentar ditulis pada video berita atau diskusi terkait kenaikan harga Pertamina dari saluran berita nasional resmi; (2) komentar menggunakan bahasa Indonesia atau bahasa daerah yang lazim dipahami secara nasional; serta (3) komentar tidak berupa spam, promosi produk, atau teks kosong.

Berdasarkan kriteria tersebut, jumlah data awal yang berhasil dikumpulkan menggunakan alat bantu *scraper* adalah sebanyak 2.403 komentar. Setelah melalui proses penyaringan, penghapusan data duplikat, dan pembersihan (*data cleaning*), diperoleh data akhir sebanyak 2.359 komentar bersih yang layak digunakan sebagai dataset penelitian. Distribusi perolehan data komentar bersih dari masing-masing saluran berita YouTube disajikan pada Tabel 1.

Tabel 1 Distribusi Dataset Komentar Bersih per Channel YouTube

No	Channel Youtube	Jumlah Komentar Bersih
1	METRO TV	215
2	KOMPAS TV	791
3	INEWS	553
4	TVONENEWS	662
5	CNBC	138
Total		2359

2.4 Variabel atau Fokus yang Diteliti

Variabel dependen/output/target dalam penelitian ini adalah kelas sentimen opini masyarakat yang diklasifikasikan menjadi tiga kelas (multikelas), yaitu sentimen negatif (-1), sentimen netral (0), dan sentimen positif (1). Variabel independen/input/fitur berupa bobot kata dari teks komentar hasil ekstraksi menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF).

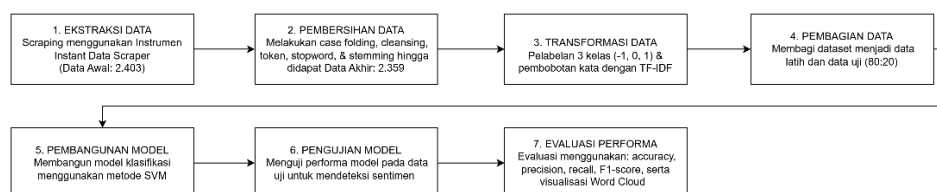
2.5 Instrumen atau Alat Pengumpulan Data

Instrumen Pengumpulan Data: Penarikan data sekunder dilakukan secara semi-otomatis menggunakan alat bantu ekstensi penjelajah web Instant Data Scraper. Alat ini digunakan untuk mengekstrak seluruh teks komentar, nama pengguna (yang kemudian dianonimkan), dan waktu penayangan dari laman YouTube ke dalam format tabular (*Comma-Separated Values / CSV*).

Instrumen Pengolahan Data: Pengolahan data dan pemodelan dilakukan menggunakan perangkat lunak Python dengan dukungan pustaka (*library*) seperti Pandas dan NumPy untuk strukturasi data, Sastrawi untuk penanganan bahasa (*preprocessing* teks bahasa Indonesia), serta Scikit-learn untuk proses TF-IDF, pembagian dataset, dan pemodelan SVM.

2.6 Prosedur Penelitian

Tahapan penelitian secara sistematis disajikan pada Gambar 1. Proses diawali dengan mengekstrak data komentar dari YouTube menggunakan Instant Data Scraper untuk mendapatkan 2.403 data awal. Data mentah tersebut kemudian melalui tahap pembersihan data (*preprocessing*) meliputi *case folding*, *cleansing*, *tokenization*, *stopword removal*, dan *stemming* berbasis Sastrawi hingga diperoleh 2.359 data akhir. Langkah selanjutnya adalah melakukan pelabelan manual ke dalam tiga kelas sentimen: negatif (-1), netral (0), dan positif (1). Data teks kemudian ditransformasikan ke bentuk numerik menggunakan pembobotan TF-IDF. Dataset selanjutnya dibagi dengan proporsi 80:20 menggunakan *stratified split*, menghasilkan 1.886 baris data latih (*training data*) dan 472 baris data uji (*testing data*) sesuai dengan kebutuhan pemodelan algoritma. Model prediksi sentimen kemudian dibangun menggunakan algoritma *Support Vector Machine* (SVM) pada data latih, yang selanjutnya dievaluasi pada data uji untuk mengukur performa pengujiannya.



Gambar 1 Alur Penelitian Analisis Sentimen SVM dengan Instant Data Scraper

2.7 Teknik Analisis Data

Analisis data dilakukan dengan pendekatan klasifikasi *machine learning* menggunakan algoritma *Support Vector Machine* (SVM) multikelas. Kinerja model dievaluasi secara matematis menggunakan pengujian *confusion matrix* untuk menghitung nilai metrik *accuracy*, *precision*, *recall*, dan *F1-score* untuk masing-masing kelas sentimen (Negatif, Netral, Positif). Selain itu, analisis data diperkuat dengan visualisasi *Word Cloud* untuk mengidentifikasi kecenderungan kata kunci yang paling dominan muncul pada masing-masing kelompok sentimen tersebut.

3. HASIL DAN PEMBAHASAN

Hasil penelitian disusun mulai dari gambaran data, performa pemodelan *Support Vector Machine* (SVM), hingga pembahasan komprehensif terkait temuan sentimen opini masyarakat terhadap kebijakan kenaikan harga Pertamina.



3.1 Deskripsi Data

Dataset yang digunakan dalam penelitian ini berjumlah 2.359 *record* komentar bersih yang diekstraksi dari saluran YouTube berita nasional resmi dan telah memenuhi kriteria sampel. Berdasarkan hasil pelabelan target, data dibagi ke dalam tiga kelas sentimen, yaitu Negatif (-1), Netral (0), dan Positif (1). Distribusi data target dari masing-masing saluran berita disajikan pada Tabel 2.

Tabel 2 Distribusi Kelas Sentimen Opini Masyarakat per Channel YouTube

No	Channel Youtube	Negatif (-1)	Netral (0)	Positif (1)	Total Bersih	Persentase
1	KOMPAS TV	437	294	60	791	33,53%
2	TVONENEWS	374	221	67	662	28,06%
3	INEWS	284	220	49	553	23,44%
4	METRO TV	142	51	22	215	9,11%
5	CNBC	73	49	16	138	5,85%
Total		1.310	835	214	2.359	100,00%

Berdasarkan Tabel 2, jumlah sentimen negatif jauh lebih dominan dibandingkan sentimen netral dan positif, yaitu mencapai 1.310 komentar (55,53%). Sentimen netral berada pada urutan kedua dengan 835 komentar (35,40%), sedangkan sentimen positif memiliki proporsi paling sedikit yaitu 214 komentar (9,07%). Dominasi kelas negatif mengindikasikan bahwa kebijakan penyesuaian harga Pertamina secara umum direspons secara kritis dan cenderung memicu penolakan dari publik di media sosial YouTube.

3.2 Pemodelan Support Vector Machine (SVM)

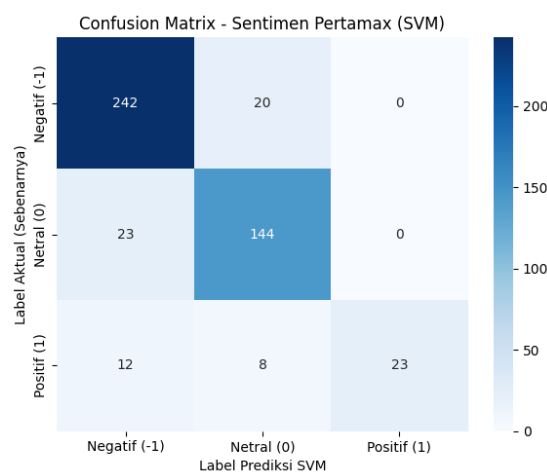
Model prediksi dibangun menggunakan algoritma *Support Vector Machine* (SVM) multikelas dengan teknik pembobotan kata TF-IDF. Pembagian dataset dilakukan dengan rasio 80% data latih (1.886 baris) dan 20% data uji (472 baris). Evaluasi performa model pada data uji diukur menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score*. Hasil evaluasi klasifikasi disajikan pada Tabel 3.

Tabel 3 Hasil Evaluasi Performa Model SVM

Kelas Sentimen	Precision	Recall	F1-Score	Support
Negatif (-1)	0,87	0,92	0,90	262
Netral (0)	0,84	0,86	0,85	167
Positif (1)	1,00	0,53	0,70	43
Akurasi Akhir SVM			86,65%	472
Macro Average	0,90	0,77	0,81	472
Weighted Average	0,87	0,87	0,86	472

Berdasarkan Tabel 3, model SVM menunjukkan performa klasifikasi yang sangat baik dengan tingkat Akurasi Akhir sebesar 86,65% (atau *weighted accuracy* 87%). Nilai *precision* untuk kelas Positif mencapai nilai sempurna (1,00), yang berarti tidak ada opini negatif/netral yang salah diprediksi sebagai positif. Kelas Negatif menunjukkan performa paling seimbang dan tangguh dengan *recall* 0,92 dan *F1-score* 0,90 pada 262 sampel data uji. Rerata berbobot (*weighted average*) untuk *precision*, *recall*, dan *F1-score* secara konsisten berada pada rentang 0,86-0,87, membuktikan bahwa model memiliki keandalan tinggi dalam mengenali opini publik.

Untuk memberikan gambaran yang lebih rinci mengenai distribusi prediksi riil dan tingkat kesalahan model, hasil klasifikasi disajikan melalui *Confusion Matrix* pada Gambar 2.



Gambar 2 *Confusion Matrix* Model SVM Klasifikasi Sentimen Pertamax

Berdasarkan Gambar 2, model berhasil mengklasifikasikan 242 data negatif, 144 data netral, dan 23 data positif secara tepat (*True Positives* pada diagonal utama). Kesalahan klasifikasi mayoritas terjadi pada kelas positif di mana dari 43 data aktual, terdapat 12 data yang terprediksi sebagai negatif dan 8 data sebagai netral. Namun demikian, secara keseluruhan jumlah prediksi benar (409 dari 472 sampel data uji) jauh lebih dominan dibanding kesalahan prediksi, sehingga membuktikan bahwa kombinasi TF-IDF dan SVM sangat efektif dalam memetakan pola opini tidak terstruktur.

3.3 Pembahasan

Hasil penelitian menunjukkan bahwa algoritma *Support Vector Machine* (SVM) berbasis pembobotan kata TF-IDF mampu memprediksi dan mengelompokkan sentimen opini masyarakat terkait kenaikan harga Pertamina dengan tingkat akurasi mencapai 86,65%. Temuan ini mempertegas bahwa ekstraksi opini digital dari kolom komentar YouTube dapat dimanfaatkan secara objektif untuk menangkap respon nyata masyarakat secara cepat dan otomatis.

Dominasi sentimen negatif (55,53%) dalam dataset mengindikasikan bahwa kebijakan kenaikan harga Pertamina memicu dampak psikosial berupa kekecewaan dan penurunan daya beli yang dirasakan publik. Kondisi ini terjadi karena Pertamina merupakan komoditas energi vital yang berpengaruh langsung terhadap pengeluaran harian masyarakat. Temuan ini sejalan dengan



studi-studi terdahulu mengenai analisis sentimen kebijakan publik yang menunjukkan bahwa isu ekonomi dan energi cenderung menghasilkan polaritas negatif yang tinggi di media sosial.

Dari sisi metodologi, keunggulan performa SVM ($F1$ -score 0,90 pada kelas dominan) disebabkan oleh kemampuan algoritma ini dalam menangani ruang fitur berdimensi tinggi (*high-dimensional feature space*) hasil transformasi TF-IDF tanpa mengalami *overfitting*. Nilai *recall* kelas positif yang relatif lebih rendah (0,53) diduga terjadi akibat adanya ketidakseimbangan kelas (*class imbalance*), di mana sampel data positif jauh lebih sedikit (214 data) dibanding data negatif (1.310 data).

KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa algoritma *Support Vector Machine* (SVM) berbasis pembobotan kata TF-IDF berhasil diterapkan untuk mengklasifikasikan sentimen opini masyarakat terhadap kebijakan kenaikan harga Pertamina pada media sosial YouTube dengan performa yang sangat baik. Model klasifikasi multikelas yang dibangun mampu mencapai tingkat akurasi sebesar 86,65% pada data uji.

Hasil analisis distribusi data menunjukkan bahwa mayoritas opini masyarakat tergolong dalam sentimen negatif (55,53%), diikuti oleh sentimen netral (35,40%), dan sentimen positif (9,07%). Temuan ini membuktikan bahwa kebijakan kenaikan harga Pertamina secara umum direspons secara kritis dan memicu keresahan publik di media sosial. Hal ini menegaskan bahwa ekstraksi data opini digital dari kolom komentar YouTube dapat dimanfaatkan sebagai dasar analisis prediktif yang objektif untuk mengukur respons masyarakat terhadap kebijakan publik secara otomatis dan *real-time*.

Secara praktis, model dan metodologi yang dihasilkan berpotensi dimanfaatkan oleh pemangku kebijakan (pemerintah dan BUMN terkait) sebagai instrumen *early warning system* atau evaluasi dampak sosial-ekonomi sebelum dan sesudah kebijakan publik diterapkan, sehingga komunikasi publik dapat dirancang dengan lebih proaktif dan tepat sasaran.

Berdasarkan keterbatasan yang ditemukan dalam penelitian ini, beberapa saran dapat dipertimbangkan untuk pengembangan berikutnya. Dari aspek metodologi dan data, penelitian selanjutnya disarankan untuk menerapkan teknik penanganan ketidakseimbangan kelas (*imbalanced data*) seperti SMOTE guna meningkatkan sensitivitas pada kelas minoritas, serta memperluas sumber data ke platform media sosial lain seperti X atau Instagram. Selain itu, pemrosesan teks dapat diperdalam dengan menambahkan penanganan kata tidak baku (*slang*) dan sarkasme, serta mengeksplorasi penggunaan metode *word embedding* modern untuk dibandingkan dengan algoritma SVM. Sementara dari sisi implementasi praktis, hasil penelitian ini dapat dikembangkan lebih lanjut ke dalam bentuk *dashboard* pemantauan sentimen berbasis web agar dapat dimanfaatkan oleh unit kehumasan instansi terkait dalam memantau dinamika opini publik secara berkelanjutan.



DAFTAR PUSTAKA

- [1] Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273-297.
<https://link.springer.com/article/10.1007/BF00994018>
- [2] Feldman, R., & Sanger, J. (2007). *The text mining handbook: Advanced approaches in analyzing unstructured data*. Cambridge University Press.
<https://www.cambridge.org/core/books/text-mining-handbook/0634B1DF14259CB43FCCF28972AE4382>
- [3] Joachims, T. (1998). *Text categorization with support vector machines: Learning with many relevant features*. Springer.
<https://dl.acm.org/doi/10.1007/BFb0026683>
- [4] Liu, B. (2012). *Sentiment analysis and opinion mining*. Morgan & Claypool Publishers.
<https://dl.acm.org/doi/10.5555/3019323>
- [5] Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to information retrieval*. Cambridge University Press.
<https://nlp.stanford.edu/IR-book/>
- [6] Salton, G., & Buckley, C. (1988). Term-weighting approaches in automatic text retrieval. *Information Processing & Management*, 24(5), 513-523.
<https://www.sciencedirect.com/science/article/abs/pii/0306457388900210>
- [7] Scikit-learn Developers. (2024). *Scikit-learn: Machine learning in Python*.
<https://scikit-learn.org/stable/>
- [8] Sastrawi Project. (2016). *Sastrawi: Indonesian stemming library*.
<https://github.com/sastrawi/sastrawi>
- [9] Instant Data Scraper. (2023). *Instant data scraper documentation*.