

Pemodelan Sentimen Komentar Youtube Terkait Kebijakan Hilirisasi Nikel Menggunakan Algoritma Supervised Learning

Oleh :

Ditta Adelia (221080200087)

Dosen Pembimbing :

Uce Indahyanti

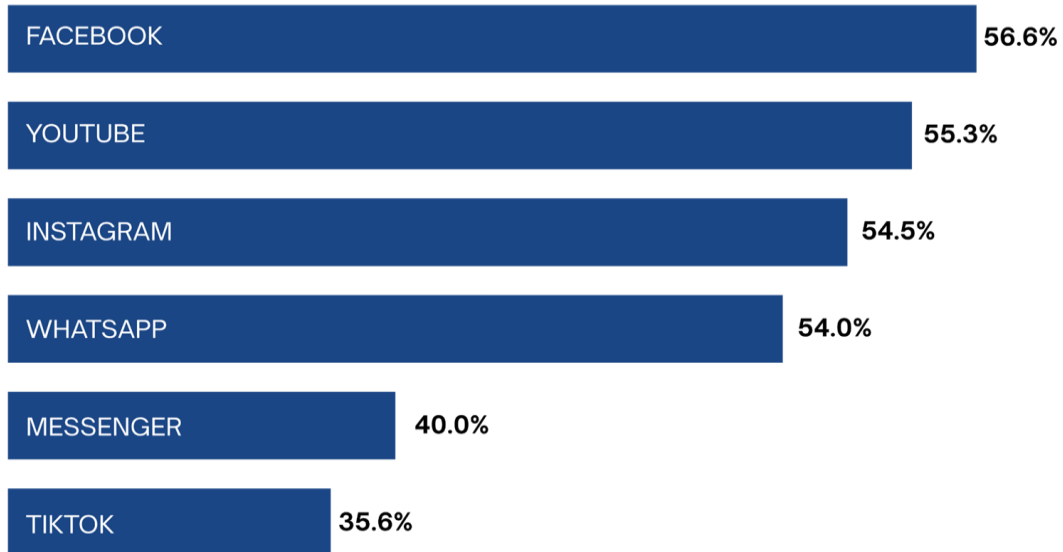
Program Studi Informatika

Universitas Muhammadiyah Sidoarjo

April, 2026

Pendahuluan

Penggunaan Media Sosial di Indonesia



Menurut data *We Are Social*, YouTube adalah platform terpopuler kedua di Indonesia dengan 139 juta pengguna.

Indonesia memiliki 22% cadangan nikel dunia. Sejak 2020, pemerintah melarang ekspor bijih nikel mentah untuk meningkatkan nilai tambah nasional.

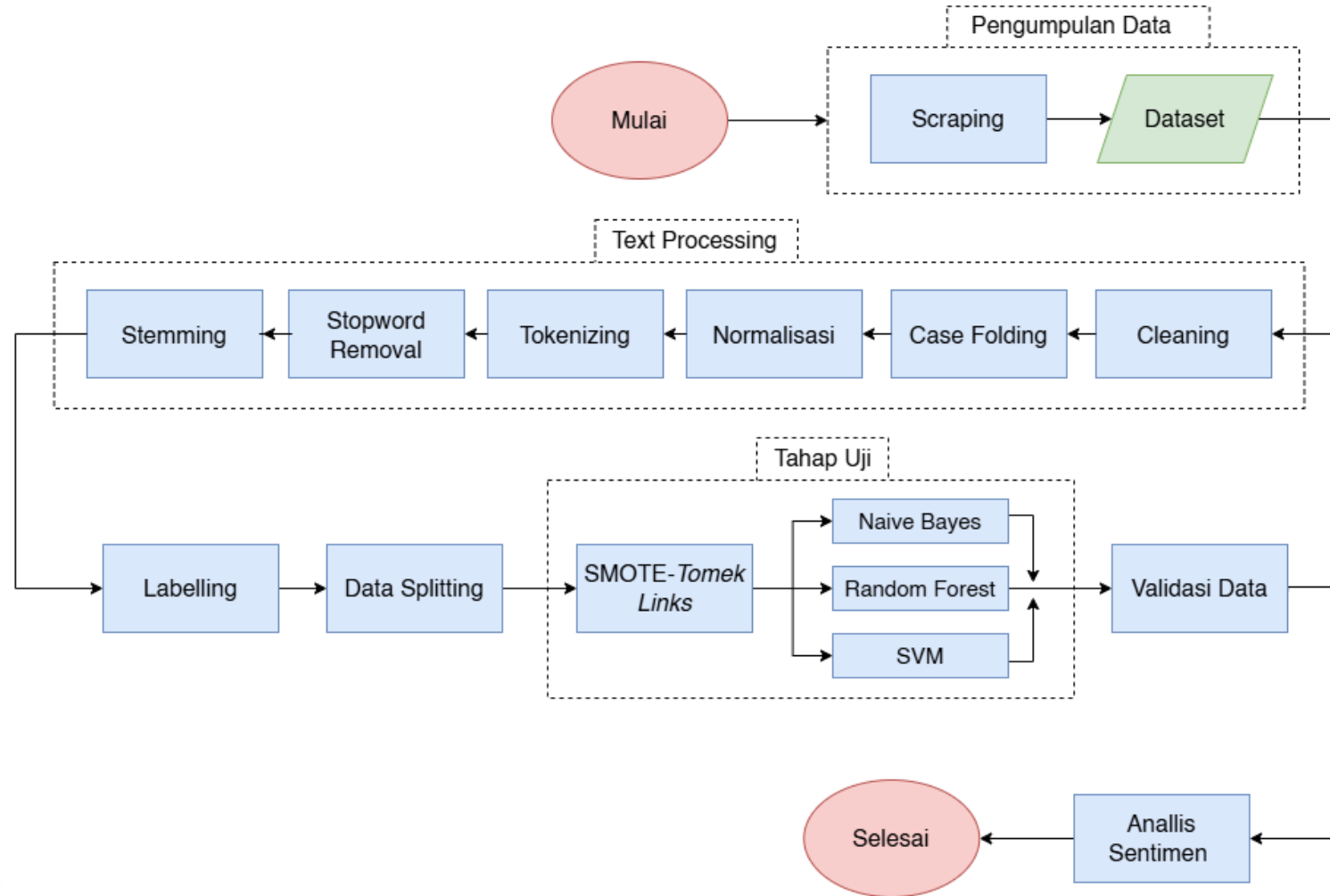
Kebijakan ini memicu pro-kontra: sebagian mendukung demi kedaulatan ekonomi, sebagian mengkritisi kerusakan lingkungan dan dominasi asing.

YouTube (139 juta pengguna aktif di Indonesia) menjadi ruang diskusi publik yang kaya, namun data komentarnya tidak seimbang sehingga diperlukan pendekatan Supervised Learning + SMOTE-Tomek Links.

Petanyaan Penelitian (Rumusan Masalah)

Rumusan masalah yang diangkat dalam penelitian ini adalah:
“Bagaimana membangun model klasifikasi sentimen komentar YouTube terkait kebijakan hilirisasi nikel dengan mengatasi masalah ketidakseimbangan data dan membandingkan performa algoritma supervised learning?”

Metode



Metode

Pengumpulan Data

51 video YouTube terkait kebijakan hilirisasi nikel pada September 2019 sampai Agustus 2025. Total 15.014 komentar mentah.

Text Processing (6 Tahap)

1) Cleaning 2) Case Folding 3) Normalisasi 4) Tokenizing 5) Stopword Removal 6) Stemming

Pelabelan Sentimen

Menggunakan kamus InSet (Indonesia Sentiment Lexicon) secara otomatis. Hasil: 7.850 data bersih - 6.362 negatif (81,04%) dan 1.488 positif (18,96%).

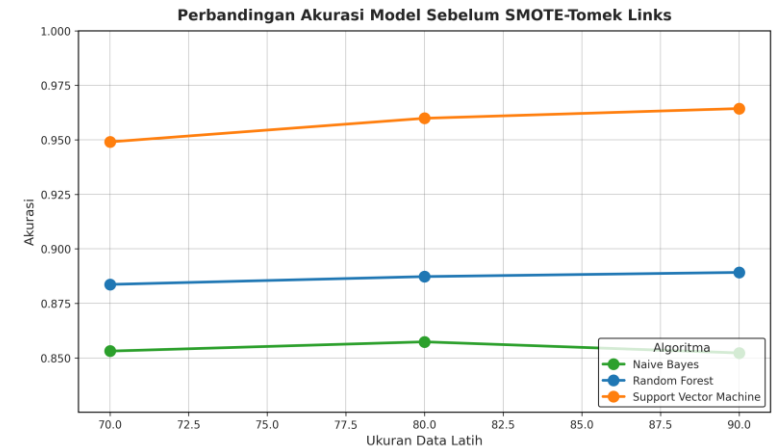
Pembagian Data & Ekstraksi Fitur

Rasio: 70:30 | 80:20 | 90:10. Representasi numerik teks menggunakan TF-IDF. Penyeimbangan data training dengan SMOTE-Tomek Links.

Hasil (1)

✓ Hasil Klasifikasi Skenario 1

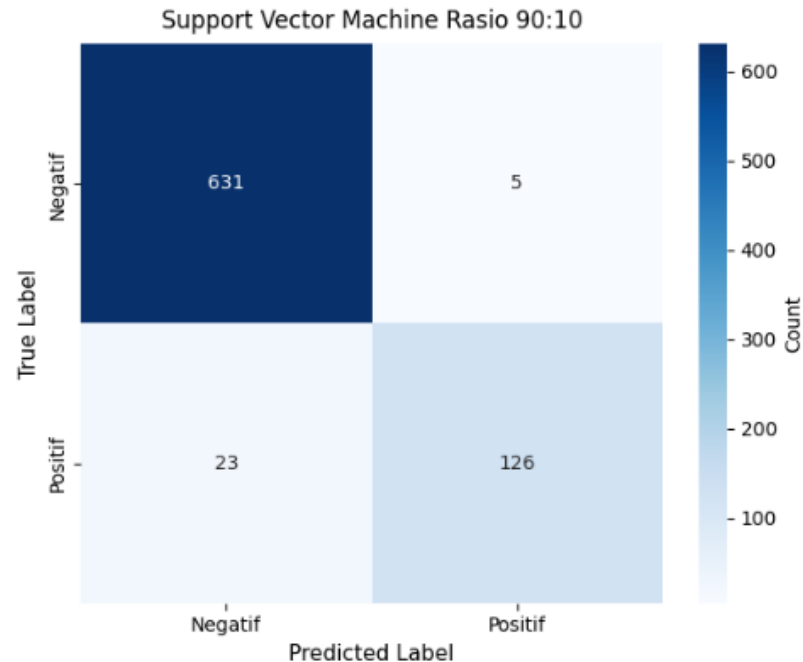
Rasio	Algoritma	Accuracy	Precision	Recall	F1-Score
90:10	Naive Bayes	85,22%	87,50%	85,22%	81,13%
	Random Forest	88,92%	89,11%	88,92%	87,42%
	Support Vector Machine	96,43%	96,43%	96,43%	96,34%
80:20	Naive Bayes	85,73%	87,87%	85,73%	82,01%
	Random Forest	88,73%	88,81%	88,73%	87,22%
	Support Vector Machine	95,99%	95,96%	95,99%	95,88%
70:30	Naive Bayes	85,31%	87,25%	85,31%	81,34%
	Random Forest	88,37%	87,97%	88,37%	87,03%
	Support Vector Machine	94,90%	94,84%	94,90%	94,74%



SVM mencapai akurasi tertinggi 96,43% pada rasio 90:10.

Hasil (2)

✓ Masalah Bias Model Terbaik Skenario 1

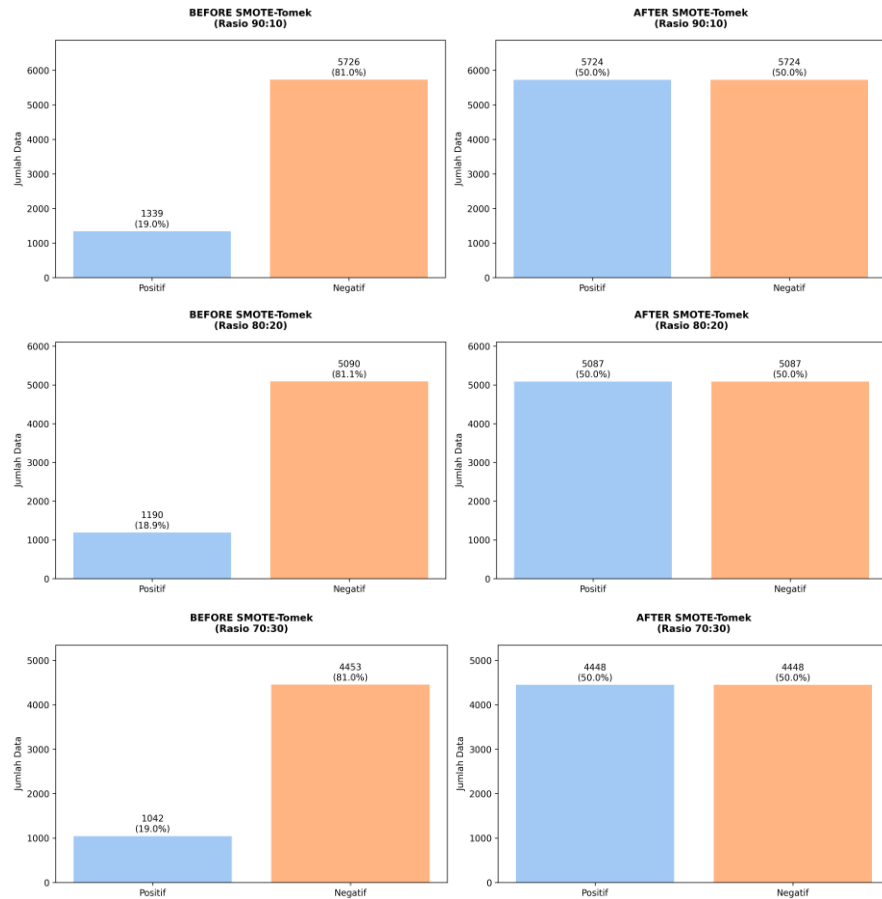


Algoritma	Recall		Selisih
	Negatif	Positif	
Naive Bayes	100%	22%	78%
Random Forest	99%	46%	53%
SVM	99%	85%	14%

Selisih recall terbesar terdapat pada Naive Bayes (78%), menunjukkan model bias terhadap kelas negatif.

Hasil (3)

✓ Penerapan *SMOTE – Tomek Links* pada Data Latih



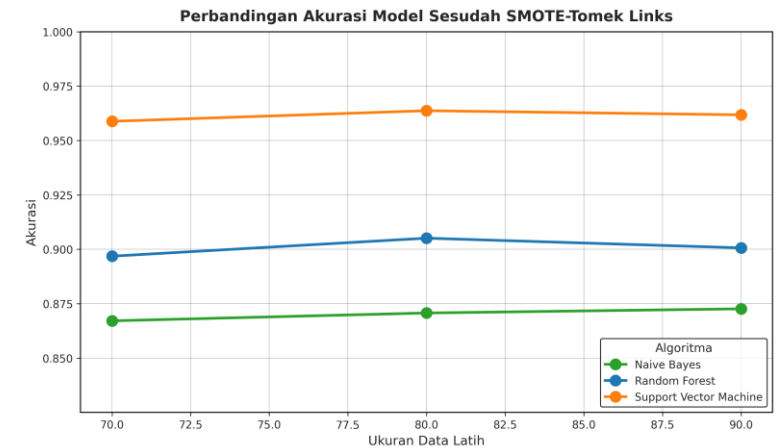
Rasio	Kelas	SMOTE-Tomek Links	
		Sebelum	Sesudah
90 : 10	Negatif	5.726	5.724
	Positif	1.339	5.724
80 : 20	Negatif	5.090	5.087
	Positif	1.190	5.087
70 : 30	Negatif	4.453	4.448
	Positif	1.042	4.448

Penerapan **SMOTE-Tomek Links** pada ketiga rasio berhasil menyeimbangkan distribusi data menjadi **rasio 1:1**

Hasil (4)

✓ Hasil Klasifikasi Skenario 2

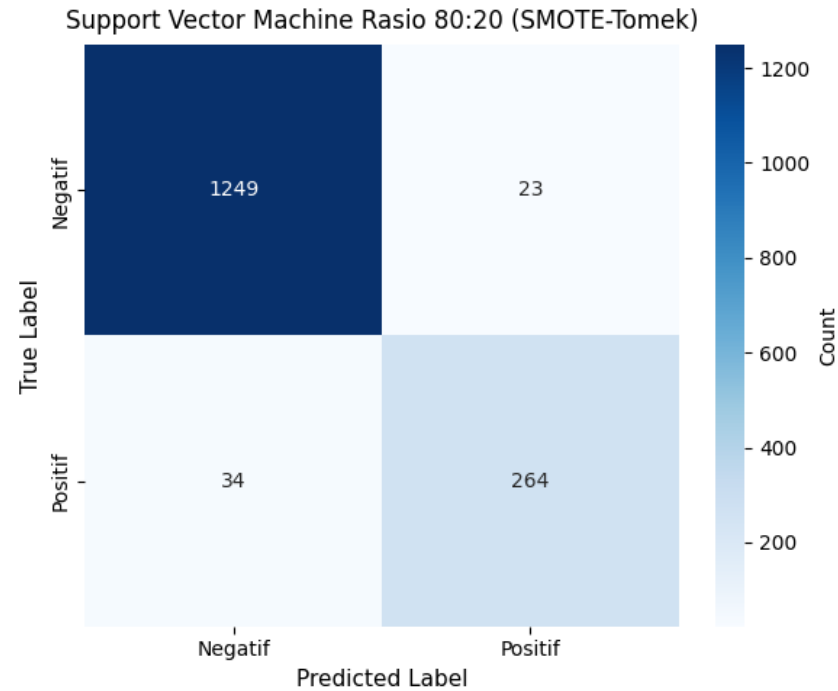
Rasio	Algoritma	Accuracy	Precision	Recall	F1-Score
90:10	Naive Bayes	87,26%	87,13%	87,26%	87,19%
	Random Forest	90,06%	89,61%	90,06%	89,48%
	Support Vector Machine	96,18%	96,14%	96,18%	96,15%
80:20	Naive Bayes	87,07%	86,89%	87,07%	86,98%
	Random Forest	90,51%	90,09%	90,51%	90,06%
	Support Vector Machine	96,37%	96,33%	96,37%	96,34%
70:30	Naive Bayes	86,71%	86,50%	86,71%	86,60%
	Random Forest	89,68%	89,19%	89,68%	89,25%
	Support Vector Machine	95,88%	95,84%	95,88%	95,85%



Stabilitas akurasi, tetap terjaga (96,37%) pada metode SVM dengan rasio 80:20.

Hasil (5)

✓ Evaluasi Perbaikan dan Penurunan Bias



Algoritma	Recall		Selisih
	Negatif	Positif	
Naive Bayes	92%	64%	28%
Random Forest	97%	64%	33%
SVM	98%	89%	9%

Recall positif meningkat pada semua algoritma, Hal ini membuktikan *SMOTE-Tomek Links* berhasil mengurangi bias dan menghasilkan model yang lebih seimbang dalam mendeteksi kedua kelas sentimen.

Hasil Analisis Sentimen

Word Cloud Positif



Dukungan terhadap kemajuan industri, lapangan kerja, dan kedaulatan ekonomi nasional.

Word Cloud Negatif



Kritik terhadap kerusakan lingkungan, ketimpangan ekonomi, dan dominasi modal asing.

Pembahasan

Performa Algoritma

SVM unggul karena mampu memisahkan kelas data dengan hyperplane optimal di ruang berdimensi tinggi. Naive Bayes mengalami bias terparah karena asumsi independensi fitur tidak terpenuhi pada data teks.

Efektivitas SMOTE-Tomek Links

SMOTE mensintesis data minoritas dari 1.190 menjadi 5.087 sampel. Tomek Links membersihkan noise batas keputusan. Model lebih seimbang tanpa mengorbankan akurasi global (turun hanya 0,06%).

Analisis Sentimen Publik

Negatif (81,04%): kata dominan "rusak", "asing", "lingkungan", "pejabat" - kekhawatiran kerusakan lingkungan, keterlibatan asing, dan ketimpangan ekonomi.

Positif (18,96%): kata dominan "maju", "kerja", "usaha", "hilirisasi" - dukungan kemajuan industri dan kedaulatan negara.

Temuan Penting Penelitian

- ✓ **Dominasi Sentimen Negatif:** 81,04% komentar YouTube bersifat negatif, mayoritas menyoroti kerusakan lingkungan, dominasi asing, dan ketimpangan distribusi manfaat ekonomi.
- ✓ **SVM Algoritma Terbaik:** SVM konsisten mengungguli Naive Bayes dan Random Forest di semua skenario. Akurasi tertinggi 96,37% pada rasio 80:20 dengan SMOTE-Tomek Links.
- ✓ **SMOTE-Tomek Links Efektif Kurangi Bias:** Selisih recall antar kelas berhasil direduksi dari 14% menjadi 9%. Recall kelas positif meningkat signifikan: 85% → 89% (SVM), 22% → 64% (NB), 46% → 64% (RF).
- ✓ **Dataset YouTube Lebih Representatif:** 7.850 komentar dari 51 video menghasilkan gambaran opini publik yang lebih komprehensif dibanding studi sebelumnya berbasis Twitter.

Manfaat Penelitian

- ✓ Memberikan wawasan berbasis data mengenai persepsi publik terhadap kebijakan hilirisasi nikel, sehingga pemerintah dapat menyesuaikan komunikasi dan implementasi kebijakan sesuai harapan masyarakat.
- ✓ Membuktikan efektivitas metode hibrida *SMOTE-Tomek Links* dalam mengatasi *imbalanced dataset* pada analisis sentimen teks berbahasa Indonesia, sebagai referensi penelitian selanjutnya.
- ✓ Mencoba metode *Deep Learning* seperti LSTM atau BERT untuk menangkap konteks kalimat yang lebih kompleks, serta memperluas sumber data ke platform media sosial lain (Twitter, TikTok, Instagram).

Referensi

- [1] W. P. Sahputra, B. A. Badia, M. I. Putra, F. C. Putra, dan A. A. Aji, "Rekayasa Proses Ekstraksi dan Pengolahan Bijih Nikel: Teknologi, Tantangan, dan Prospek Masa Depan," *KAPALAMADA J. Multidisipliner*, vol. 4, no. 02, hal. 243–255, 2025, doi: 10.62668/kapalamada.v4i02.1546.
- [2] R. B. Santoso, D. F. Moenardy, R. Muttaqin, dan D. Saputera, "Pilihan Rasional Indonesia dalam Kebijakan Larangan Ekspor Bijih Nikel," *Indones. Perspect.*, vol. 8, no. 1, hal. 154–179, 2023.
- [3] M. Hidayat dan U. Budiyanto, "Sentimen Analisis Tentang Hilirisasi Industri Berdasarkan Opini Masyarakat Di Twitter Menggunakan Metode K-Nearest Neighbor," *Pros. Semin. Nas. ...*, vol. 2, no. September, hal. 826–835, 2023.
- [4] Norlaila, W. W. Winarno, dan E. T. Luthfi, "Analisis Sentimen Masyarakat Tentang Tambang Di Indonesia Pada Twitter Menggunakan Data Mining," *JlPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.)*, vol. 9, no. 3, hal. 1091–1099, 2024, doi: 10.29100/jipi.v9i3.5402.
- [5] A. A. Ningtyas, A. Solichin, dan 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, hal. 9, 2023, doi: 10.36080/bit.v20i1.2317.
- [6] W. A. Social, "Digital 2025: Indonesia," *We Are Social & Meltwater*.
- [7] M. B. Al Hakiki dan Y. Darmi, "Penerapan Algoritma Machine Learning SVM dan NBC pada Sentimen Analisis Komentar Youtube Program Pengaduan Masyarakat Laporan Mas Wapres," *J. Ilmu Multidisiplin*, vol. 4, no. 1, hal. 396–410, 2025, doi: 10.38035/jim.v4i1.884.
- [8] M. I. Prayugah, U. Indahyanti, dan N. Ariyanti, "Analisis sentimen publik pada pemerintah dalam serangan ransomware dengan pendekatan smote," vol. 8, no. 2, hal. 333–343, 2024, doi: 10.35145/joisie.v8i2.4764.
- [9] A. B. Pratama dan D. Febriawan, "Analisis Sentimen Terkait Hilirisasi Industri Pada Opini Masyarakat X dengan Menggunakan Naive Bayes," *J. Inf. Syst. Res.*, vol. 6, no. 2, hal. 1444–1453, 2025, doi: 10.47065/josh.v6i2.6811.

Referensi

- [10] F. Destiyanti, A. I. Hadiana, dan F. R. Umbara, "Penerapan Metode Support Vector Machine dan SMOTE untuk Klasifikasi Sentimen Publik Terhadap Polisi Republik Indonesia," *Jumanji*, vol. 8, no. 1, hal. 1–15, 2024.
- [11] A. Miftahusalam, A. F. Nuraini, A. A. Khoirunisa, dan H. Pratiwi, "Perbandingan Algoritma Random Forest, Naïve Bayes, dan Support Vector Machine Pada Analisis Sentimen Twitter Mengenai Opini Masyarakat Terhadap Penghapusan Tenaga Honorer," *Semin. Nas. Off. Stat.*, vol. 2022, no. 1, hal. 563–572, 2022, doi: 10.34123/semnasoffstat.v2022i1.1410.
- [12] K. Marzuki, L. G. Rady Putra, H. Hairani, L. Z. A. Mardedi, dan J. X. Guterres, "Performance Improvement of The Random Forest Method Based on Smote-Tomek Link on Lombok Tourism Analysis Sentiment," *J. Bumigora Inf. Technol.*, vol. 5, no. 2, hal. 151–158, 2024, doi: 10.30812/bite.v5i2.3166.
- [13] A. Nurhopipah dan C. Magnolia, "Perbandingan Metode Resampling pada Imbalanced Dataset untuk Klasifikasi Komentar Program MBKM," *JUPIKOM (Jurnal Publ. Ilmu Komput. dan Multimedia)*, vol. 1, no. 2, hal. 9–22, 2022.
- [14] A. A. Nurrahman, M. Mauladi, dan A. Rahman, "Analisis Sentimen Masyarakat terhadap Kenaikan Harga Bahan Bakar Minyak Menggunakan Support Vector Machine dan SMOTE," *J. Tek. Inform.*, vol. 4, no. 2, hal. 396–410, 2025, doi: <https://doi.org/10.56211/sudo.v4i2.908>.
- [15] P. W. S. Aji, Suprianto, dan R. Dijaya, "Prediksi Penyakit Stroke Menggunakan Metode Random Forest." *Kesatria: Jurnal Penerapan Sistem Informasi (Komputer dan Manajemen)* 4.4 (2023): 916-924.
- [16] E. Tohidi, R. P. Herdiansyah, E. Wahyudin, dan Kaslani, "Analisa sentimen komentar video YouTube di channel TVOneNews tentang calon presiden Prabowo Subianto menggunakan Support Vector Machine," *JATI (Jurnal Mhs. Tek. Inform.)*, vol. 8, no. 1, hal. 660–667, 2024.
- [17] Y. Findawati, U. Indahyanti, Y. Rahmawati, dan R. Puspitasari, "Sentiment Analysis of Potential Presidential Candidates 2024: A Twitter-Based Study," *Acad. Open*, vol. 8, no. 1, hal. 1–17, 2023, doi: 10.21070/acopen.8.2023.7138.
- [18] H. Bichri, A. Chergui, dan M. Hain, "Investigating the Impact of Train / Test Split Ratio on the Performance of Pre-Trained Models with Custom Datasets," *IJACSA*, vol. 15, no. 2, hal. 331–339, 2024.
- [19] B. W. Rauf, "Sentimen Analisis Pertambangan Di Konawe Utara Dengan Metode Naïve Bayes," *Pros. Semin. Nas. Pemanfaat. Sains dan Teknol. Inf.*, vol. 1, no. 1, hal. 97–102, 2023.

