

Peningkatan Kinerja Analisis Sentimen Bioetanol Menggunakan SMOTE pada Komparasi Model *Machine Learning*

Improving Bioethanol Sentiment Analysis Performance Using SMOTE in Machine Learning Model Comparison

¹Rajhu Ilham Pradana*, ²Jasmir Jasmir, ³Gunardi Gunardi

^{1,2,3}Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Dinamika Bangsa
^{1,2,3}Jl. Jendral Sudirman Thehok - Jambi 36138 Indonesia

e-mail: rajhuilham@gmail.com, ijay_jasmir@yahoo.com, gun4rdi.sj@gmail.com

(received: 9 April 2026, revised: 23 May 2026, accepted: 24 May 2026)

Abstrak

Analisis sentimen terhadap kebijakan publik di media sosial krusial untuk evaluasi pemerintah, namun sering terkendala oleh dataset yang sangat tidak seimbang. Penelitian ini bertujuan untuk mengatasi tantangan tersebut pada studi kasus sentimen kebijakan bioetanol BBM di YouTube, di mana data bersih setelah pra-pemrosesan sebanyak 2.409 komentar didominasi oleh sentimen Negatif (1.430), diikuti sentimen Netral (734), dan sangat sedikit sentimen Positif (245). Kinerja model *Machine Learning* (ML) klasik terdegradasi parah akibat ketidakseimbangan ini, terutama dalam mendeteksi kelas minoritas. Penelitian ini menerapkan metode pembobotan TF-IDF untuk ekstraksi fitur, dilanjutkan dengan teknik *oversampling* SMOTE guna menyeimbangkan data latih (1.927 sampel) sebelum melakukan komparasi kinerja tiga algoritma ML, yaitu *Logistic Regression*, *Support Vector Machine*, dan *LightGBM*. Hasil evaluasi pada data uji (482 sampel) menunjukkan bahwa penerapan SMOTE secara signifikan meningkatkan kemampuan model dalam mengenali kelas “Positif”. Model *LightGBM* dengan SMOTE menunjukkan kinerja tertinggi dengan akurasi 64,11%, di mana penerapan SMOTE terbukti mampu melonjakkan F1-Score kelas minoritas dari *baseline* 18,18% menjadi 35,29%. Hasil ini membuktikan bahwa penanganan data tidak seimbang adalah langkah krusial untuk analisis sentimen yang valid.

Kata kunci: analisis sentimen, bioetanol, data tidak seimbang, *LightGBM*, SMOTE

Abstract

Sentiment analysis of public policy on social media is crucial for government evaluation, but is often hampered by highly imbalanced datasets. This study aims to address this challenge in a case study of sentiment toward bioethanol fuel policy on YouTube, where the pre-processed dataset of 2,409 comments is dominated by Negative sentiment (1,430), followed by Neutral sentiment (734), and very few Positive sentiments (245). The performance of classical Machine Learning (ML) models is severely degraded due to this imbalance, particularly in detecting the minority class. This study applied the TF-IDF weighting method for feature extraction, followed by the SMOTE oversampling technique to balance the training data (1,927 samples) before comparing the performance of three ML algorithms: Logistic Regression, Support Vector Machine, and LightGBM. Evaluation results on the test data (482 samples) show that the application of SMOTE significantly improves the model's ability to recognize the “Positive” class. The LightGBM model with SMOTE demonstrated the highest performance with an accuracy of 64.11%, where the application of SMOTE was proven to boost the minority class F1-Score from a baseline of 18.18% to 35.29%. These results demonstrate that handling imbalanced data is a crucial step for valid sentiment analysis.

Keywords: sentiment analysis, bioethanol, imbalanced data, *LightGBM*, SMOTE

1 Pendahuluan

Di era transformasi digital, platform media sosial seperti YouTube telah berevolusi menjadi instrumen vital bagi masyarakat untuk menyuarakan aspirasi terhadap kebijakan publik secara terbuka dan real-time [1]. Fenomena ini menghasilkan data tekstual masif yang, jika diolah dengan tepat, dapat menjadi indikator krusial bagi pemerintah dalam mengevaluasi penerimaan sebuah kebijakan [2]. Saat ini, kebijakan mandatori pencampuran bioetanol 10% (E10) ke dalam BBM menjadi isu strategis yang memicu diskusi intensif dan terpolarisasi, serupa dengan polemik kenaikan harga BBM yang memicu kekhawatiran dampak ekonomi luas [3]. Oleh karena itu, analisis sentimen menjadi pendekatan urgen untuk memetakan opini publik secara objektif.

Tinjauan terhadap literatur terkini menunjukkan dominasi penerapan algoritma *Machine Learning* klasik dalam analisis sentimen kebijakan publik, seperti *Support Vector Machine* (SVM) dan *Logistic Regression* (LR) [4]. Namun, terdapat kesenjangan metodologis yang signifikan pada mayoritas penelitian terdahulu, terutama yang terindeks pada jurnal lokal tingkat bawah, yang cenderung mengabaikan karakteristik alami data media sosial yang sangat tidak seimbang [5], [6]. Penelitian oleh [7], [8] mengkritisi bahwa penerapan algoritma standar pada dataset timpang tanpa penanganan khusus akan menghasilkan model yang bias, di mana model cenderung memprediksi kelas mayoritas dengan benar namun gagal total dalam mengenali kelas minoritas. Hal ini mengakibatkan metrik akurasi yang dihasilkan menjadi semu (*misleading*) dan tidak merepresentasikan kinerja model yang sebenarnya dalam mendeteksi aspirasi minoritas yang seringkali memuat kritik penting bagi pemerintah.

Penelitian ini hadir untuk menjawab kelemahan tersebut dengan mengusulkan pendekatan komparatif yang lebih *robust* menggunakan teknik *Synthetic Minority Over-sampling Technique* (SMOTE) untuk merekonstruksi keseimbangan data latih. Fokus utama penelitian ini adalah membuktikan secara empiris apakah integrasi SMOTE mampu meningkatkan validitas klasifikasi sentimen pada kasus kebijakan bioetanol dengan membandingkan performa model linear SVM dan LR terhadap algoritma *gradient boosting* modern, yaitu LightGBM. Hasil penelitian ini diharapkan memberikan kontribusi berupa model analisis sentimen yang lebih adil dan akurat dalam menangkap opini positif masyarakat yang sering kali terabaikan dalam dataset yang didominasi oleh sentimen negatif.

2 Tinjauan Literatur

Analisis sentimen atau *opinion mining* telah berkembang pesat sebagai cabang dari *Natural Language Processing* (NLP) untuk mengidentifikasi polaritas opini dalam teks [9]. Dalam konteks kebijakan publik di Indonesia, penelitian terdahulu [10], [11], [12] telah memanfaatkan data Twitter dan YouTube untuk memetakan preferensi masyarakat. Namun, tantangan utama yang sering ditemukan adalah penggunaan bahasa tidak baku dan distribusi kelas yang sangat tidak merata, seperti yang terlihat pada studi kasus kebijakan bioetanol di mana sentimen negatif mendominasi secara ekstrem. Pengabaian terhadap masalah ini pada penelitian lokal sering kali menurunkan validitas temuan penelitian.

Dalam klasifikasi teks, *Support Vector Machine* (SVM) kerap dianggap superior karena kemampuannya dalam menangani ruang fitur berdimensi tinggi [13]. Studi oleh [14], [15], [16] menunjukkan bahwa SVM memberikan akurasi yang tinggi pada data ulasan yang bersih, namun efektivitasnya berkurang pada data yang sangat *noisy*. Di sisi lain, *Logistic Regression* (LR) menawarkan model linear yang lebih sederhana namun sering kali memberikan performa yang kompetitif dan lebih mudah diinterpretasikan dibandingkan model yang lebih kompleks [1], [17]. Perkembangan terkini juga mulai mengarah pada penggunaan model *ensemble* seperti LightGBM yang berbasis *Gradient Boosting Decision Tree* (GBDT) [18]. LightGBM dinilai unggul dalam efisiensi komputasi dan kemampuan menangkap pola non-linear, yang dibuktikan dalam beberapa penelitian terbaru mencapai tingkat akurasi yang sangat tinggi pada klasifikasi emosi yang kompleks [19], [20].

Untuk mengatasi bias terhadap kelas mayoritas, teknik *oversampling* seperti SMOTE menjadi standar dalam penelitian data mining modern [21]. SMOTE bekerja dengan menciptakan sampel sintesis baru, bukan sekadar menduplikasi data, sehingga memberikan variasi informasi pada kelas minoritas [22]. Penelitian yang terbit di [10], [23], [24] membuktikan bahwa integrasi SMOTE pada model pohon keputusan dan SVM mampu meningkatkan nilai F1-Score dan AUC secara drastis pada dataset yang sangat timpang. Penelitian ini akan mengeksplorasi lebih jauh efektivitas SMOTE ketika

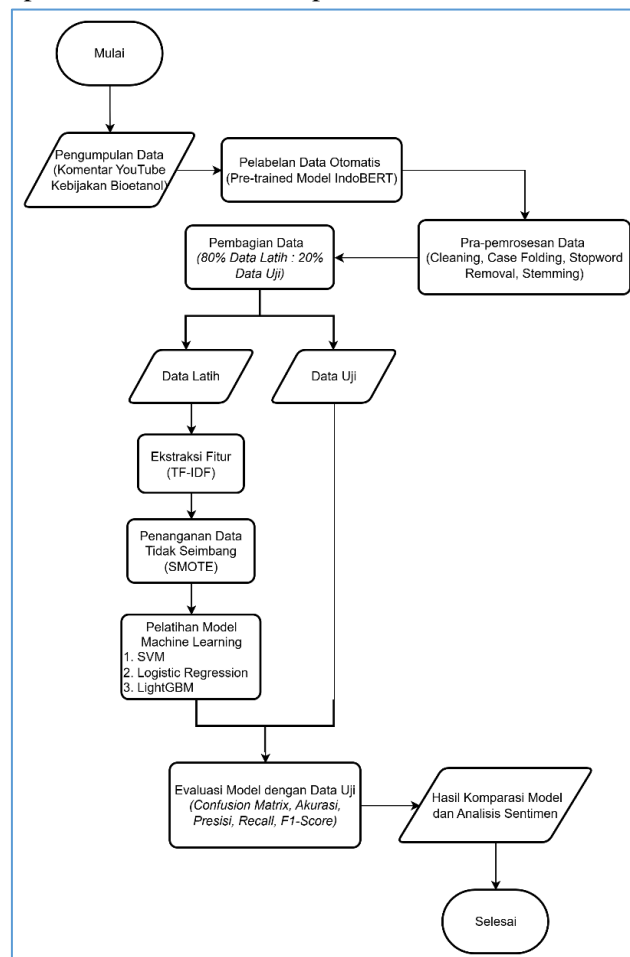
dikombinasikan dengan metode pembobotan fitur TF-IDF, yang secara teoretis mampu menonjolkan kata-kata unik namun penting dalam dokumen [25]. Fokus penelitian ini pada perbandingan model linear dan *boosting* menggunakan SMOTE diharapkan dapat menunjukkan model mana yang paling tangguh dalam menjaga keseimbangan antara akurasi global dan kemampuan deteksi kelas minoritas.

Penggunaan metode pembobotan fitur *Term Frequency-Inverse Document Frequency* (TF-IDF) dalam penelitian ini didasarkan pada efektivitasnya dalam menyeleksi kata-kata yang memiliki bobot informatif tinggi dalam sebuah dokumen teks [26]. Berbeda dengan metode *Count Vectorizer* yang hanya menghitung frekuensi kemunculan kata, TF-IDF memberikan penalti pada kata-kata yang terlalu sering muncul (*common words*) dan memberikan bobot lebih pada istilah unik yang merepresentasikan sentimen spesifik [27]. Implementasi TF-IDF pada data komentar kebijakan publik terbukti mampu meningkatkan akurasi klasifikasi karena kemampuannya menangkap esensi kritik masyarakat di tengah *noise* data media sosial yang tinggi [13].

Selain aspek akurasi, efisiensi komputasi menjadi pertimbangan utama dalam memilih algoritma untuk data berskala besar [28]. LightGBM, sebagai pengembangan dari *Gradient Boosting Decision Tree* (GBDT), menawarkan keunggulan signifikan dalam hal kecepatan pelatihan dan penggunaan memori yang rendah melalui teknik *Gradient-based One-Side Sampling* (GOSS) [29], [18]. Dalam beberapa studi komparatif, LightGBM terbukti mampu memproses jutaan baris data jauh lebih cepat dibandingkan SVM yang memiliki kompleksitas komputasi kuadratik terhadap jumlah sampel [30], [19]. Hal ini menjadikan kombinasi LightGBM dan teknik *sampling* seperti SMOTE sebagai solusi yang sangat relevan untuk sistem analisis sentimen yang memerlukan pemrosesan data secara *real-time* atau berkala.

3 Metode Penelitian

Metodologi penelitian ini dirancang secara sistematis untuk memastikan proses yang transparan dan dapat direplikasi. Alur penelitian diilustrasikan pada Gambar 1.



Gambar 1 Tahapan penelitian

3.1 Pengumpulan dan Pelabelan Data

Data mentah dikumpulkan dengan teknik *crawling* dari komentar publik pada video berita di aplikasi YouTube yang membahas topik "campuran etanol 10% BBM". Proses ini menghasilkan 2,416 data komentar. Karena data label asli tidak valid, pelabelan ulang dilakukan secara otomatis menggunakan model sentimen *pre-trained* Bahasa Indonesia. Penggunaan *pre-trained* model IndoBERT dipilih untuk pelabelan otomatis karena kemampuannya dalam menangkap nuansa semantik dan konteks bahasa Indonesia yang kompleks pada komentar media sosial, yang terbukti lebih akurat dibandingkan pendekatan berbasis kamus (*lexicon*) dalam menangani bahasa informal atau *slang* [31].

3.2 Pra-pemrosesan Data

Data teks mentah dari komentar YouTube sangat "kotor" (*noisy*) dan perlu dibersihkan melalui beberapa tahapan pra-pemrosesan data [32], [33], [34]:

1. *Cleaning*: Menghapus elemen yang tidak relevan (URL, @username, #, angka, simbol).
2. *Case Folding*: Mengubah seluruh teks menjadi format huruf kecil.
3. *Stopword Removal*: Menghapus kata-kata umum (misalnya, 'yang', 'di', 'dari') menggunakan *library* Sastrawi.
4. *Stemming*: Mengubah kata-kata berimbuhan ke bentuk kata dasarnya (misalnya, pemerintah dari pemerintahan) menggunakan *library* Sastrawi. Setelah proses ini, data yang kosong atau gagal diproses dihapus, menghasilkan 2.409 data bersih.

3.3 Ekstraksi Fitur dan Pembagian Data

Dataset bersih yang berjumlah 2.409 sampel terlebih dahulu dibagi menjadi data latih (80% / 1.927 sampel) dan data uji (20% / 482 sampel) menggunakan *stratified sampling* untuk menjaga proporsi kelas. Setelah data terpisah, proses ekstraksi fitur *Term Frequency-Inverse Document Frequency* (TF-IDF) diterapkan. Pembentukan kosa kata (*fitting*) TF-IDF dilakukan secara eksklusif hanya pada data latih. Data uji kemudian ditransformasi berdasarkan kosa kata tersebut. Pendekatan ini secara ketat mencegah terjadinya kebocoran data (*data leakage*) yang dapat merusak validitas evaluasi. Secara matematis, pembobotan TF-IDF dihitung menggunakan rumus (1) berikut:

$$w_{i,j} = tf_{i,j} \times \log\left(\frac{N}{df_i}\right) \quad (1)$$

Di mana $w_{i,j}$ merepresentasikan bobot akhir dari kata i pada dokumen j . Variabel $tf_{i,j}$ menunjukkan frekuensi kemunculan kata, sedangkan $\log\left(\frac{N}{df_i}\right)$ adalah *inverse document frequency* yang mengukur seberapa jarang kata tersebut muncul di seluruh dataset sebanyak 2.409 komentar. Dataset kemudian dibagi menjadi data latih (80% / 1.927 sampel) dan data uji (20% / 482 sampel) menggunakan *stratified sampling* untuk menjaga proporsi kelas.

3.4 Skenario Pemodelan

Untuk mengatasi masalah data tidak seimbang, penelitian ini menguji skenario penanganan secara sekuensial setelah ekstraksi fitur. Teknik *Synthetic Minority Over-sampling Technique* (SMOTE) diaplikasikan secara ketat hanya pada data latih yang telah diformat dalam bentuk matriks TF-IDF (1.927 sampel) untuk mencegah kebocoran informasi ke data uji. Pembentukan sampel sintetis ini bertujuan menyeimbangkan distribusi sebelum proses induksi oleh algoritma klasifikasi. Penelitian ini membandingkan kinerja tiga model *Machine Learning* yang dilatih menggunakan data hasil SMOTE, yaitu :

1. Skenario 1: Melakukan ekstraksi fitur TF-IDF secara eksklusif pada data latih, dilanjutkan dengan penyeimbangan kelas menggunakan SMOTE, dan diakhiri dengan pelatihan model klasifikasi *Logistic Regression* (LR).
2. Skenario 2: Melakukan ekstraksi fitur TF-IDF secara eksklusif pada data latih, dilanjutkan dengan penyeimbangan kelas menggunakan SMOTE, dan diakhiri dengan pelatihan model klasifikasi *Support Vector Machine* (SVM).

- Skenario 3: Melakukan ekstraksi fitur TF-IDF secara eksklusif pada data latih, dilanjutkan dengan penyeimbangan kelas menggunakan SMOTE, dan diakhiri dengan pelatihan model klasifikasi *gradient boosting* LightGBM.

3.5 Evaluasi Model

Metodologi evaluasi pada penelitian ini dirancang untuk mengukur sejauh mana model mampu mengenali opini publik secara akurat, terutama pada kelas minoritas yang sering kali terabaikan dalam dataset yang tidak seimbang. Seluruh *pipeline* dievaluasi kinerjanya menggunakan data uji sebanyak 482 sampel yang bersifat independen dan tidak pernah terpapar oleh data sintesis dari proses SMOTE. Metrik evaluasi utama yang digunakan diturunkan dari *Confusion Matrix* melalui perhitungan matematis persamaan (2) – (5) sebagai berikut:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (2)$$

$$Precision = \frac{TP}{TP + FP} \quad (3)$$

$$Recall = \frac{TP}{TP + FN} \quad (4)$$

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (5)$$

Dalam konteks analisis sentimen kebijakan bioetanol ini, *True Positive* (TP) merepresentasikan jumlah komentar yang benar diklasifikasikan sesuai label aslinya, sementara *True Negative* (TN) adalah data yang benar dideteksi sebagai bukan kelas tersebut. *False Positive* (FP) dan *False Negative* (FN) menunjukkan tingkat kesalahan model dalam memprediksi kelas sentimen. Penggunaan metrik Recall dan F1-Score menjadi sangat krusial dalam penelitian ini karena memberikan gambaran performa yang lebih valid pada dataset yang didominasi secara ekstrem oleh 1.430 komentar negatif diikuti sentimen Netral (734), dan sangat sedikit sentimen Positif (245). Hal ini bertujuan untuk memastikan bahwa suara atau aspirasi positif masyarakat tetap tertangkap secara adil oleh sistem klasifikasi tanpa terjebak dalam bias mayoritas.

4 Hasil dan Pembahasan

Bagian ini menyajikan hasil dari setiap tahapan metodologi, dari analisis data mentah hingga hasil akhir komparasi model.

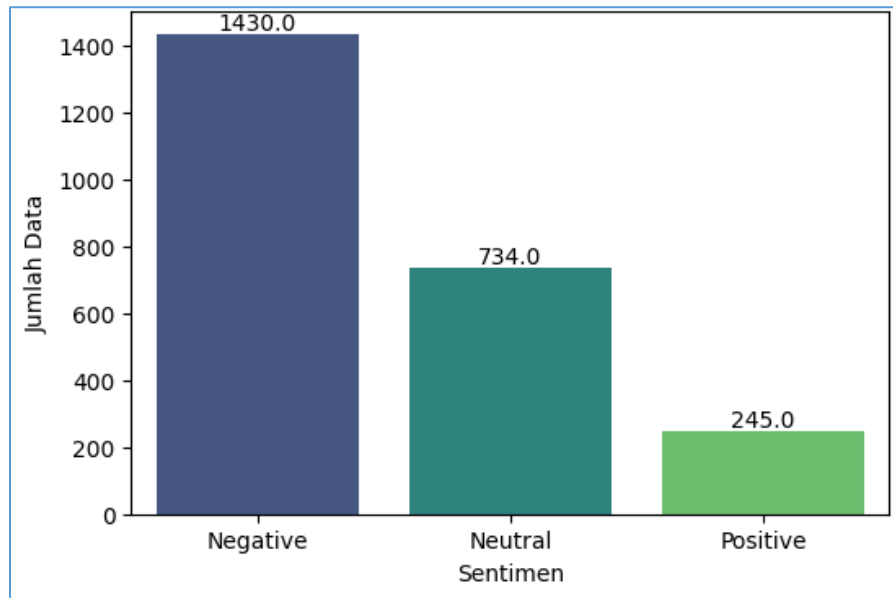
4.1 Hasil Pra-pemrosesan dan Analisis Data

Tahapan *preprocessing* data merupakan bagian krusial pada proses analisis sentimen. Proses ini dilakukan agar nanti data dapat terstruktur yang kemudian dapat digunakan dalam menganalisis sentimen. Setelah pra-pemrosesan, dihasilkan 2.409 data valid. Tabel 1 menyajikan contoh transformasi teks.

Tabel 1 Hasil pra-pemrosesan

No	Teks Asli	Teks Setelah Pra-Pemrosesan
1	BAHLIL di campur etanol 10% pasti bagus tu buat mesin diesel bang 🤔	bahlil campur etanol 10% pasti bagus buat mesin diesel bang
2	Bahlil di campur etanol pasti jadi goblin 😄	bahlil campur etanol pasti jadi goblin
3	Indonesia mmg LUARBIASA 😊	indonesia memang luar biasa
4	Goblin berulah mulu 😄😄	goblin berulah mulu
5	MALUKU MERESAHKAN	maluku meresahkan

Analisis distribusi sentimen menunjukkan ketidakseimbangan data parah, seperti pada Gambar 2.



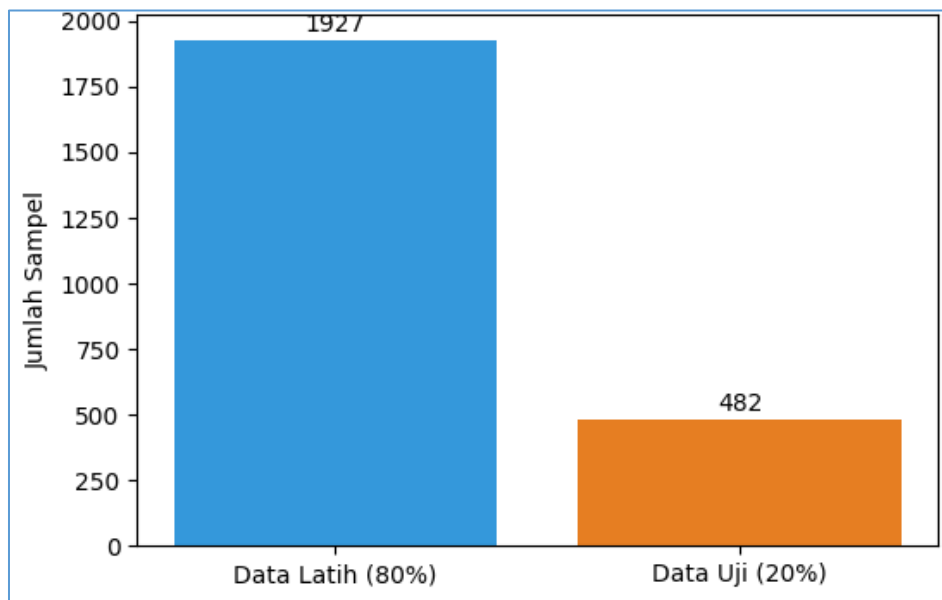
Gambar 2 Distribusi sentimen hasil pelabelan

Dari Gambar 2, terlihat sentimen Negatif (1.430 komentar) sangat mendominasi sebesar 59,4%, sementara Positif (245 komentar) adalah kelas minoritas ekstrem sebesar 10,2% dan kelas Netral sebanyak (734 komentar) sebesar 30,4%. Visualisasi *word cloud* menunjukkan kata-kata dominan dalam diskusi yang terjadi, yang dapat anda lihat pada Gambar 3.



Gambar 3 Word cloud kata kunci dominan

Gambar 3 menunjukkan frekuensi tinggi kata-kata seperti *bbm*, *rakyat*, *etanol*, *pemerintah*, dan *naik*, yang mengindikasikan fokus utama kekhawatiran publik. Dataset kemudian dibagi menjadi data latih dan uji (Gambar 4).



Gambar 4 Proporsi pembagian dataset

4.2 Analisis Dampak Kinerja: *Baseline* (Tanpa SMOTE) vs. Dengan SMOTE

Untuk menjawab tantangan ketidakseimbangan kelas yang ditemukan pada tahap analisis data, penelitian ini melakukan pengujian komparatif dalam dua skenario eksperimen. Skenario pertama *baseline* menggunakan data asli tanpa penanganan khusus, sedangkan Skenario kedua menerapkan teknik *Synthetic Minority Over-sampling Technique* (SMOTE) pada data latih. Evaluasi difokuskan secara kritis pada metrik F1-Score untuk kelas minoritas “Positif”, mengingat akurasi global seringkali memberikan gambaran yang bias pada data timpang. Tabel 2 memperlihatkan perbandingan kinerja ketiga algoritma *Logistic Regression*, SVM, dan LightGBM pada kedua skenario tersebut.

Tabel 2 Perbandingan kinerja model *baseline* vs. model dengan SMOTE

No	Model	Skenario	Akurasi	Presisi (Positif)	Recall (Positif)	F1-Score (Positif)
1	LightGBM	<i>Baseline</i>	66,60	83,33	10,20	18,18
2	LightGBM	Smote	64,11	33,96	36,73	35,29
3	SVM	<i>Baseline</i>	67,22	64,29	18,37	28,57
4	SVM	Smote	63,49	34,62	36,73	35,64
5	LR	<i>Baseline</i>	62,03	40,00	24,49	30,38
6	LR	Smote	62,03	39,02	32,65	35,56

Berdasarkan Tabel 2, terlihat adanya fenomena “*Accuracy Paradox*” pada skenario *baseline*. Meskipun algoritma SVM dan LR menghasilkan akurasi global yang cukup tinggi mendekati 70%, nilai Recall dan F1-Score untuk kelas Positif sangat rendah yaitu di bawah 30%. Hal ini mengindikasikan bahwa model *baseline* mengalami bias mayoritas, model cenderung memprediksi hampir seluruh data sebagai sentimen negatif untuk meminimalkan *loss*, namun gagal mengenali aspirasi positif masyarakat.

Penerapan SMOTE terbukti memberikan dampak signifikan dalam memperbaiki kelemahan tersebut. Meskipun terdapat sedikit trade-off pada presisi di model linear LR dan SVM, terjadi lonjakan drastis pada Recall dan F1-Score kelas Positif di seluruh algoritma. Secara khusus, algoritma LightGBM menunjukkan peningkatan kinerja (*improvement*) yang paling substansial akibat penerapan SMOTE, di

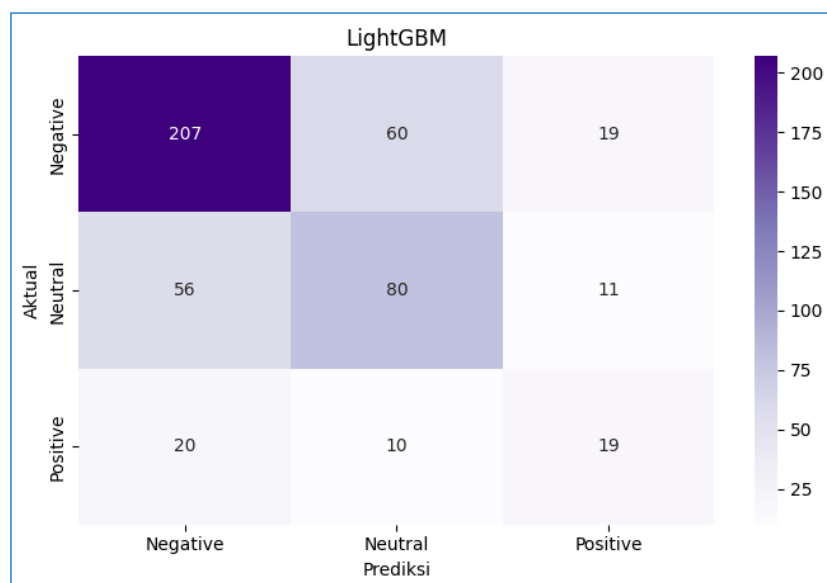
mana nilai F1-Score kelas Positif melonjak drastis dari 18,18% menjadi 35,29%. Meskipun nilai F1-Score Positif absolut tertinggi dicapai oleh model SVM dengan SMOTE (35,64%), LightGBM terbukti sebagai algoritma yang paling diuntungkan dari teknik penyeimbangan data ini. Hal ini membuktikan bahwa rekonstruksi keseimbangan data latih menggunakan SMOTE sangat krusial untuk menghasilkan model klasifikasi yang adil dan valid.

4.3 Pembahasan Hasil

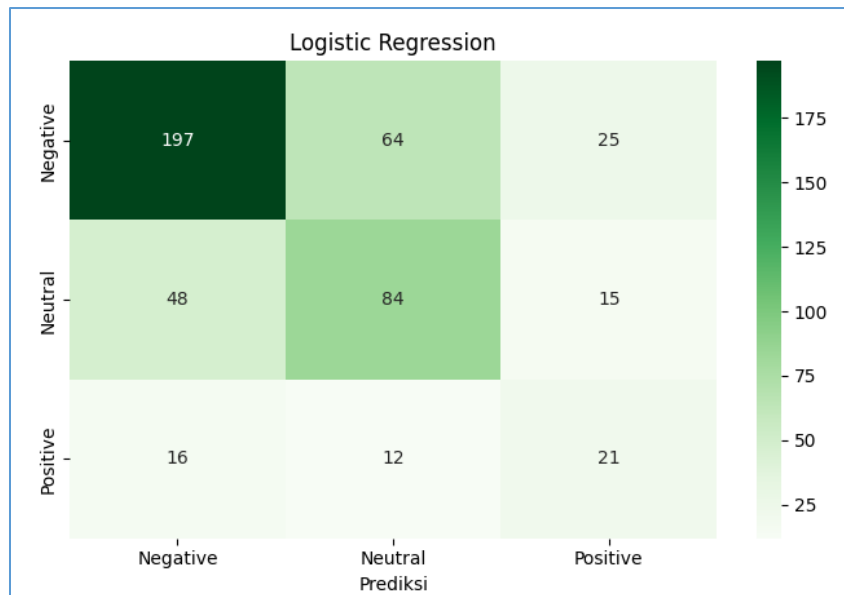
Berdasarkan Tabel 2, evaluasi kinerja model perlu ditinjau dari dua perspektif: performa global dan kemampuan spesifik deteksi kelas minoritas. Secara performa global, LightGBM dengan SMOTE muncul sebagai model dengan kinerja keseluruhan paling seimbang, mencapai akurasi tertinggi sebesar 64,11% dan mengungguli model lain dalam F1-Score rata-rata tertimbang. Ini menunjukkan bahwa arsitektur *boosting* modern mampu menangkap pola kompleks secara komprehensif pada data teks TF-IDF. Namun, secara spesifik dalam hal mengenali aspirasi kelas minoritas (Positif), model linear klasik SVM dengan SMOTE terbukti sedikit lebih tajam dengan capaian F1-Score Positif tertinggi sebesar 35,64%, mengungguli capaian LightGBM (35,29%).

Namun, temuan paling signifikan dari penelitian ini adalah dampak SMOTE terhadap pengenalan kelas minoritas. Pada eksperimen awal tanpa SMOTE, F1-Score untuk kelas Positif mendapatkan nilai 18,18% untuk model LightGBM. Setelah penerapan SMOTE, F1-Score Positif pada model LightGBM meningkat drastis menjadi 35,29%. Peningkatan dramatis ini membuktikan bahwa SMOTE berhasil menyeimbangkan dataset dan melatih model untuk mengenali pola sentimen positif yang sebelumnya diabaikan.

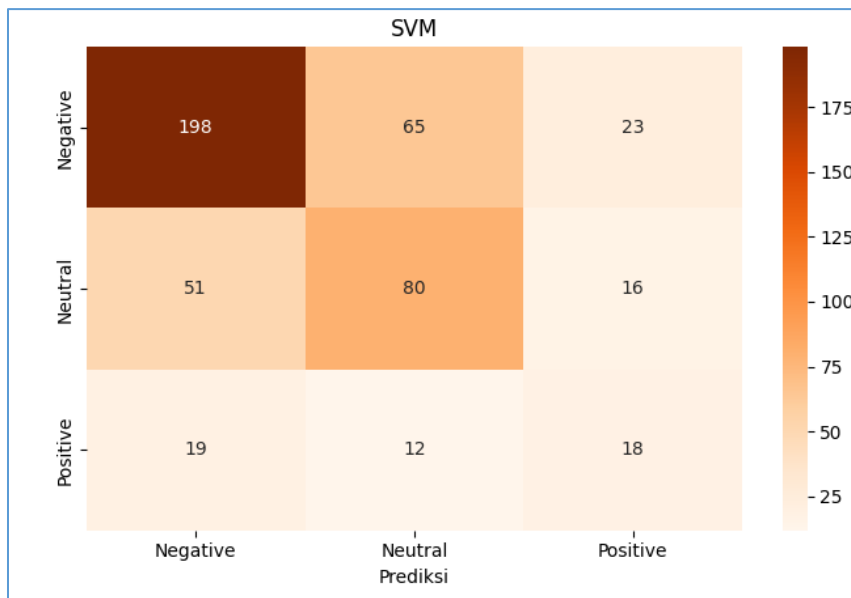
Fenomena peningkatan performa yang drastis setelah penerapan SMOTE ini memperkuat argumen bahwa model *Machine Learning* standar akan selalu gagal memberikan interpretasi yang adil jika dilatih pada dataset yang timpang. Hasil penelitian ini, di mana LightGBM dengan SMOTE mencapai akurasi tertinggi sebesar 64,11% , sejalan dengan temuan pada penelitian terdahulu [35], [36] yang menyatakan bahwa arsitektur *boosting* modern lebih adaptif terhadap data sintesis dibandingkan model linear klasik seperti *Logistic Regression*. Peningkatan F1-Score kelas 'Positif' dari 18,18% menjadi 35,29% membuktikan bahwa model tidak lagi terjebak dalam *Accuracy Paradox*, di mana akurasi tinggi sering kali hanya menutupi kegagalan model dalam mendeteksi kelas minoritas. Temuan ini memberikan validasi empiris bahwa dalam evaluasi kebijakan publik, metrik F1-Score atau Recall jauh lebih krusial daripada akurasi global guna memastikan suara atau aspirasi positif masyarakat tetap tertangkap oleh sistem. Analisis pola kesalahan pada *Confusion Matrix* Gambar 5, 6, dan 7 mengkonfirmasi hal ini.



Gambar 5 *Confusion matrix* (LightGBM + SMOTE)



Gambar 6 Confusion matrix (LR + SMOTE)



Gambar 7 Confusion matrix (SVM + SMOTE)

Dari Gambar 5 LightGBM, terlihat bahwa model sekarang berhasil memprediksi 19 komentar sebagai Positif dengan benar. Meskipun masih ada kesalahan klasifikasi beberapa Netral dan Negatif salah diprediksi sebagai Positif, model tidak lagi buta terhadap kelas minoritas. Ini merupakan peningkatan substansial dibandingkan model tanpa SMOTE yang hampir seluruhnya menebak Negatif atau Netral.

5 Kesimpulan

Penelitian ini menyimpulkan bahwa teknik SMOTE efektif mengatasi degradasi kinerja akibat ketidakseimbangan data ekstrem pada analisis sentimen kebijakan bioetanol. SMOTE berhasil mengeliminasi bias prediksi terhadap kelas "Negatif", yang ditandai dengan lonjakan F1-Score kelas "Positif" secara drastis dari 18,18% menjadi 35,29% pada LightGBM. Meskipun model SVM dengan SMOTE mencatat F1-Score Positif absolut tertinggi (35,64%), LightGBM terbukti paling optimal secara performa keseluruhan dengan akurasi 64,11%. Secara praktis, metodologi ini dapat dimanfaatkan oleh pemerintah sebagai instrumen evaluasi kebijakan publik yang lebih berimbang untuk menangkap aspirasi minoritas. Namun, penelitian ini memiliki keterbatasan pada cakupan data yang hanya

<http://sistemasi.ftik.unisi.ac.id>

bertumpu pada satu *platform* (YouTube) dan kelemahan inheren TF-IDF terhadap kompleksitas konteks bahasa (seperti sarkasme). Oleh karena itu, penelitian selanjutnya disarankan untuk menggunakan sumber data silang *platform* (misalnya X/Twitter dan TikTok) serta mengeksplorasi arsitektur *Deep Learning* (seperti *Transformer* atau LLM berbahasa Indonesia) guna menyempurnakan deteksi klasifikasi sentimen secara lebih mendalam.

Referensi

- [1] H. Gonaygunta, G. S. Nadella, and K. Meduri, "Utilizing logistic regression in machine learning for categorizing social media advertisement," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 37, no. 3, p. 1954, Mar. 2025, doi: 10.11591/ijeecs.v37.i3.pp1954-1963.
- [2] W. Maharani and A. Z. Latifa, "Analyzing Public Sentiment on the Relocation of Indonesia's Capital to Kalimantan as the Ibu Kota Nusantara Using Logistic Regression," *Jurnal Teknik Informatika (Jutif)*, vol. 6, no. 2, pp. 575–592, Apr. 2025, doi: 10.52436/1.jutif.2025.6.2.4230.
- [3] B. Darmawan, A. Dwi Laksito, M. Resa, A. Yudianto, and A. Sidauruk, "Analisis Perbandingan Ekstraksi Fitur Teks pada Sentimen Analisis Kenaikan Harga BBM," *Krea-TIF: Jurnal Teknik Informatika*, vol. 11, no. 1, pp. 53–63, 2023, doi: 10.32832/krea-tif.v11i1.13819.
- [4] A. Handayanto, K. Latifa, N. D. Saputro, and R. R. Waliyansyah, "Analisis dan Penerapan Algoritma Support Vector Machine (SVM) dalam Data Mining untuk Menunjang Strategi Promosi," *JUITA: Jurnal Informatika*, vol. 7, no. 2, pp. 71–79, 2019, doi: 10.30595/juita.v7i2.4378.
- [5] I. M. Karo Karo and H. Hendriyana, "Klasifikasi Penderita Diabetes menggunakan Algoritma Machine Learning dan Z-Score," *Jurnal Teknologi Terpadu*, vol. 8, no. 2, pp. 94–99, 2022, doi: 10.54914/jtt.v8i2.564.
- [6] U. Khaira, R. Johanda, P. E. P. Utomo, and T. Suratno, "Sentiment Analysis Of Cyberbullying On Twitter Using SentiStrength," *Indonesian Journal of Artificial Intelligence and Data Mining*, vol. 3, no. 1, p. 21, 2020, doi: 10.24014/ijaidm.v3i1.9145.
- [7] E. Rohaeti and A. Andriyati, "Comparative Study of Predictive Classification Models on Data with Severely Imbalanced Predictors," *JUITA : Jurnal Informatika*, vol. 12, no. 1, p. 121, May 2024, doi: 10.30595/juita.v12i1.21491.
- [8] B. Solihah, A. Azhari, and A. Musdholifah, "The Empirical Comparison of Machine Learning Algorithm for the Class Imbalanced Problem in Conformational Epitope Prediction," *JUITA: Jurnal Informatika*, vol. 9, no. 1, p. 131, May 2021, doi: 10.30595/juita.v9i1.9969.
- [9] N. L. W. S. R. Ginantra, C. P. Yanti, G. D. Prasetya, I. B. G. Sarasvananda, and I. K. A. G. Wiguna, "Analisis Sentimen Ulasan Villa di Ubud Menggunakan Metode Naive Bayes, Decision Tree, dan K-NN," *Jurnal Nasional Pendidikan Teknik Informatika (JANAPATI)*, vol. 11, no. 3, pp. 205–215, Dec. 2022, doi: 10.23887/janapati.v11i3.49450.
- [10] I. Fadil, R. Surya Manggala, E. Firmansyah, and M. A. Helmiawan, "Performance Optimization of Support Vector Machine with SMOTE for Multiclass Stunting Prediction in Sumedang District, Indonesia," *Jurnal Teknik Informatika (Jutif)*, vol. 6, no. 4, pp. 2917–2928, Sep. 2025, doi: 10.52436/1.jutif.2025.6.4.4843.
- [11] M. Parhusip, S. Sudianto, and T. G. Laksana, "Sentiment Analysis of the Public Towards the Kanjuruhan Tragedy with the Support Vector Machine Method," *JUITA : Jurnal Informatika*, vol. 11, no. 2, p. 241, Nov. 2023, doi: 10.30595/juita.v11i2.17405.
- [12] H. Atsqalani, N. Hayatin, and C. S. K. Aditya, "Sentiment Analysis from Indonesian Twitter Data Using Support Vector Machine And Query Expansion Ranking," *Jurnal Online Informatika*, vol. 7, no. 1, pp. 116–122, Jun. 2022, doi: 10.15575/join.v7i1.669.
- [13] A. Irianti, H. Halimah, S. Sutedi, and M. Agariana, "Integration of BERT and SVM in Sentiment Analysis of Twitter/X Regarding Constitutional Court Decision No. 60/PUU-XXII/2024," *Jurnal Teknik Informatika (Jutif)*, vol. 6, no. 2, pp. 469–482, Apr. 2025, doi: 10.52436/1.jutif.2025.6.2.4068.
- [14] G. V. Gopal and G. R. M. Babu, "An ensemble feature selection approach using hybrid kernel based SVM for network intrusion detection system," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 23, no. 1, p. 558, Jul. 2021, doi: 10.11591/ijeecs.v23.i1.pp558-565.

- [15] C. Adiwinata and A. Afiyati, "Sentiment Analysis of X Users Toward Electric Motorcycles Using SVM and BERT Algorithms," *JUITA: Jurnal Informatika*, vol. 13, no. 2, pp. 119–126, Aug. 2025, doi: 10.30595/juita.v13i2.26152.
- [16] Indra, Agus Umar Hamdani, Suci Setiawati, Zena Dwi Mentari, and Mauridhy Hery Purnomo, "Comparison of K-NN, SVM, and Random Forest Algorithm for Detecting Hoax on Indonesian Election 2024," *Jurnal Nasional Pendidikan Teknik Informatika (JANAPATI)*, vol. 13, no. 1, pp. 166–179, Mar. 2024, doi: 10.23887/janapati.v13i1.76079.
- [17] A. S. Sunge, S. Suzanna, and H. M. Mardi Putra, "Interpretable Machine Learning for Employee Recruitment Prediction Using Boruta, CatBoost, Lasso, Logistic Regression, NLP, and RFE Feature Selection," *Jurnal Teknik Informatika (Jutif)*, vol. 6, no. 4, pp. 2153–2170, Aug. 2025, doi: 10.52436/1.jutif.2025.6.4.4810.
- [18] K. L. Ong, C. P. Lee, H. S. Lim, and K. M. Lim, "Speech emotion recognition with light gradient boosting decision trees machine," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 13, no. 4, p. 4020, Aug. 2023, doi: 10.11591/ijece.v13i4.pp4020-4028.
- [19] W. Wiharto, A. Abdurrahman, and U. Salamah, "Detection of COVID-19 based on cough sound and accompanying symptom using LightGBM algorithm," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 38, no. 2, p. 940, May 2025, doi: 10.11591/ijeecs.v38.i2.pp940-949.
- [20] M. Mar'atuttahirah, K. Tunnisa, D. F. R. Ra, H. Najwa, and J. Fahrival, "MSMEs Recommendation System using Item-Based Collaborative Filtering and LightGBM Machine Learning," *Jurnal Teknik Informatika (Jutif)*, vol. 6, no. 5, pp. 3832–3843, Oct. 2025, doi: 10.52436/1.jutif.2025.6.5.5292.
- [21] S. Uyun and E. Sulistyowati, "Feature selection for multiple water quality status: Integrated bootstrapping and SMOTE approach in imbalance classes," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 10, no. 4, p. 4331, Aug. 2020, doi: 10.11591/ijece.v10i4.pp4331-4339.
- [22] D. Z. Abidin, M. Rosario, A. Sadikin, N. Nurhadi, and J. Jasmir, "Improving Term Deposit Customer Prediction Using Support Vector Machine with SMOTE and Hyperparameter Tuning in Bank Marketing Campaigns," *Jurnal Teknik Informatika (Jutif)*, vol. 6, no. 3, pp. 1267–1278, Jun. 2025, doi: 10.52436/1.jutif.2025.6.3.4585.
- [23] A. F. Rozi, A. Wibowo, and B. Warsito, "Resampling Technique for Imbalanced Class Handling on Educational Dataset," *JUITA: Jurnal Informatika*, vol. 11, no. 1, p. 77, May 2023, doi: 10.30595/juita.v11i1.15498.
- [24] I. K. A. Purnawan, A. D. Wibawa, A. Kurniawati, and M. H. Purnomo, "Optimizing Diabetic Neuropathy Severity Classification Using Electromyography Signals Through Synthetic Oversampling Techniques," *Jurnal Nasional Pendidikan Teknik Informatika (JANAPATI)*, vol. 13, no. 3, pp. 681–690, Dec. 2024, doi: 10.23887/janapati.v13i3.85675.
- [25] A. Ridho Lubis, M. K. M. Nasution, O. Salim Sitompul, and E. Muisa Zamzami, "The effect of the TF-IDF algorithm in times series in forecasting word on social media," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 22, no. 2, p. 976, May 2021, doi: 10.11591/ijeecs.v22.i2.pp976-984.
- [26] A. D. Sripamuji, I. Ramadhanti, R. R. Amalia, J. Saputra, and B. Prihatnowo, "Penerapan Algoritma Support Vector Machine dan Multi-Layer Perceptron pada Klasifikasi Topik Berita," *Jurnal Nasional Pendidikan Teknik Informatika (JANAPATI)*, vol. 11, no. 2, pp. 84–91, 2022, doi: 10.23887/janapati.v11i2.44151.
- [27] S. Defit and G. W. Nurcahyo, "Product Codefication Accuracy with Cosine Similarity and Weighted Term Frequency and Inverse Document Frequency (TF-IDF)," *Journal of Applied Engineering and Technological Science*, vol. 2, no. 2, pp. 14–21, 2021, doi: 10.37385/jaets.v2i2.
- [28] F. Muttakin, N. Andrika, and S. Salsabila, "Sentiment Analysis of Shoe Product Reviews on Indonesian E-Commerce Platform Using Lexicon Based and Support Vector Machine," *Jurnal Teknik Informatika (Jutif)*, vol. 6, no. 2, pp. 839–854, Apr. 2025, doi: 10.52436/1.jutif.2025.6.2.3800.

- [29] M. Sopiyan, F. Fauziah, and Y. F. Wijaya, “Fraud Detection Using Random Forest Classifier, Logistic Regression, and Gradient Boosting Classifier Algorithms on Credit Cards,” *JUITA: Jurnal Informatika*, vol. 10, no. 1, p. 77, May 2022, doi: 10.30595/juita.v10i1.12050.
- [30] N. A. Syam, N. Arifin, W. Firgiawan, and M. F. Rasyid, “Comparison of SVM and Gradient Boosting with PCA for Website Phishing Detection,” *Jurnal Teknik Informatika (Jutif)*, vol. 6, no. 2, pp. 691–708, Apr. 2025, doi: 10.52436/1.jutif.2025.6.2.4344.
- [31] A. R. P. Dewi, S. Riyadi, C. Darmajati, N. A. M. Isa, and A. D. Andriyani, “Sentiment Analysis of Pro-Israel Product Boycott Action Using IndoBERT Method on Unbalanced Data,” *JUITA: Jurnal Informatika*, vol. 13, no. 2, pp. 187–197, Aug. 2025, doi: 10.30595/juita.v13i2.25976.
- [32] N. Hadi and D. Sugiarto, “Analisis Sentimen Pembangunan IKN pada Media Sosial X Menggunakan Algoritma SVM, Logistic Regression dan Naïve Bayes,” *Jurnal Informatika: Jurnal Pengembangan IT*, vol. 10, no. 1, pp. 37–49, 2025, doi: 10.30591/jpit.v10i1.7106.
- [33] S. Lonang, A. Yudhana, and M. K. Biddinika, “Analisis Komparatif Kinerja Algoritma Machine Learning untuk Deteksi Stunting,” *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 7, no. 4, p. 2109, Oct. 2023, doi: 10.30865/mib.v7i4.6553.
- [34] H. Imaduddin, F. Y. A’la, and Y. S. Nugroho, “Sentiment Analysis in Indonesian Healthcare Applications using IndoBERT Approach,” *International Journal of Advanced Computer Science and Applications*, vol. 14, no. 8, pp. 113–117, 2023, doi: 10.14569/IJACSA.2023.0140813.
- [35] C. G. L. Pringandana and K. Kusnawi, “A Comparative Analysis of Hyperparameter-Tuned XGBoost and LightGBM for Multiclass Rainfall Classification in Jakarta,” *Jurnal Teknik Informatika (Jutif)*, vol. 6, no. 4, pp. 2467–2483, Aug. 2025, doi: 10.52436/1.jutif.2025.6.4.4965.
- [36] S. Widodo, F. S. Utomo, and Berlilana, “A Comprehensive Evaluation of CatBoost and LightGBM Algorithms for Honorarium Prediction on Categorical Datasets with Class Imbalance,” *JUITA: Jurnal Informatika*, vol. 13, no. 3, pp. 359–370, Nov. 2025, doi: 10.30595/juita.v13i3.27363.