Building of Informatics, Technology and Science (BITS)
Volume 8, No 1, June 2026 Page: 485−496
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.10095

Klasifikasi Sentimen Program Makan Bergizi Gratis di Platform X dengan TF-IDF dan Class-Weighted LinearSVC

Muhammad Syihabuddin Musyaffa, Novita Kurnia Ningrum*

Fakultas Ilmu Komputer, Program Studi Teknik Informatika, Universitas Dian Nuswantoro, Semarang, Indonesia
Email: [1]111202214156@mhs.dinus.ac.id, [2,*]novita.kn@dsn.dinus.ac.id
Email Penulis Korespondensi: novita.kn@dsn.dinus.ac.id
Submitted: **29/05/2026**; Accepted: **30/06/2026**; Published**: 30/06/2026**

Abstrak−Penelitian ini mengkaji optimasi klasifikasi sentimen publik terkait Program Makan Bergizi Gratis pada Platform X dengan fokus pada karakter data media sosial yang ringkas, tidak formal, mengandung noise, dan memiliki ketidakseimbangan kelas. Data dikumpulkan melalui proses scraping pada periode 5 Januari 2025 sampai 26 Februari 2026 dan menghasilkan 7.659 unggahan awal. Setelah melalui seleksi, preprocessing, cleaning, dan labeling, sebanyak 5.307 teks digunakan sebagai dataset pemodelan, dengan komposisi 2.183 sentimen positif, 2.391 sentimen netral, dan 733 sentimen negatif. Teks diubah menjadi fitur numerik menggunakan Term Frequency-Inverse Document Frequency dengan 5.000 fitur dan konfigurasi n-gram satu sampai dua. Penelitian ini mengevaluasi Multinomial Naive Bayes, Logistic Regression, LinearSVC, dan Random Forest sebagai model pembanding. Optimasi dilakukan menggunakan GridSearchCV, sedangkan class_weight balanced diterapkan untuk meningkatkan kemampuan model dalam mengenali kelas negatif yang jumlahnya lebih kecil. Hasil evaluasi menunjukkan bahwa LinearSVC dengan class_weight balanced memberikan kinerja paling seimbang, dengan accuracy 0,9011, macro F1-score 0,8989, weighted F1-score 0,9011, recall negatif 0,8367, dan F1-score negatif 0,8913. Kontribusi penelitian ini terletak pada penekanan evaluasi kelas minoritas melalui recall dan F1-score negatif, sehingga model dapat menjadi komponen awal pemantauan opini publik digital yang lebih peka terhadap kritik dan keluhan.

Kata Kunci: GridSearchCV; Ketidakseimbangan Kelas; Klasifikasi Sentimen; LinearSVC; TF-IDF

Abstract−This study investigates the optimization of public sentiment classification related to the Free Nutritious Meal Program on Platform X by focusing on social media data that are brief, informal, noisy, and class-imbalanced. The data were collected through scraping from January 5, 2025, to February 26, 2026, producing 7,659 initial posts. After selection, preprocessing, cleaning, and labeling, 5,307 texts were used as the modeling dataset, consisting of 2,183 positive, 2,391 neutral, and 733 negative sentiments. The texts were transformed into numerical features using Term Frequency-Inverse Document Frequency with 5,000 features and unigram-bigram settings. This study evaluated Multinomial Naive Bayes, Logistic Regression, LinearSVC, and Random Forest as comparison models. Optimization was performed using GridSearchCV, while class_weight balanced was applied to improve the model's ability to identify the smaller negative class. The evaluation results show that LinearSVC with class_weight balanced produced the most balanced performance, with an accuracy of 0.9011, macro F1-score of 0.8989, weighted F1-score of 0.9011, negative recall of 0.8367, and negative F1-score of 0.8913. The contribution of this study lies in emphasizing minority-class evaluation through negative recall and F1-score, making the model a preliminary component for digital public opinion monitoring that is more sensitive to criticism and complaints.

Keywords: Class Imbalance; GridSearchCV; LinearSVC; Sentiment Classification; TF-IDF

1. PENDAHULUAN

Program Makan Bergizi Gratis merupakan salah satu kebijakan publik yang berkaitan langsung dengan pemenuhan gizi, kesehatan anak, pendidikan, dan peningkatan kualitas sumber daya manusia. Pelaksanaan program makan di lingkungan sekolah berpotensi mendukung pemenuhan kebutuhan gizi dan proses pembelajaran siswa, tetapi keberhasilannya juga dipengaruhi oleh kesesuaian menu, pemerataan distribusi, kesiapan penyedia makanan, keamanan pangan, dan pengawasan pelaksanaan [1], [2]. Besarnya perhatian masyarakat terhadap program tersebut menimbulkan percakapan yang intensif di Platform X dalam bentuk dukungan, kritik, laporan kegiatan, pemberitaan, pengalaman penerima manfaat, maupun keluhan mengenai pelaksanaan program. Teks pada Platform X cenderung pendek, tidak formal, mengandung singkatan, tagar, *mention*, URL, emoji, serta konteks yang tidak selalu dinyatakan secara lengkap. Banyaknya unggahan dan cepatnya perubahan topik menyebabkan pemantauan opini secara manual menjadi kurang efektif. Selain itu, distribusi sentimen yang tidak seimbang dapat menyebabkan kritik dan keluhan sebagai kelas minoritas kurang terdeteksi apabila pemilihan model hanya didasarkan pada *accuracy*. Kondisi tersebut menunjukkan perlunya klasifikasi sentimen yang mampu mengolah teks media sosial sekaligus memberikan perhatian lebih besar terhadap sentimen negatif. Sentimen negatif penting karena dapat menjadi sinyal awal mengenai persoalan kualitas makanan, distribusi, anggaran, kesiapan dapur penyedia, kasus keracunan, atau tata kelola program. Oleh karena itu, penelitian ini dilakukan untuk menghasilkan klasifikasi sentimen yang tidak hanya memiliki performa keseluruhan yang baik, tetapi juga mampu mengenali kelas negatif secara lebih sensitif.

Sejumlah penelitian telah menerapkan *machine learning* untuk menganalisis sentimen masyarakat terhadap Program Makan Bergizi Gratis dan isu kebijakan publik lainnya. Saleh dan Imanda membandingkan *Naive Bayes* dan *Support Vector Machine* pada data Platform X, sedangkan Aprianti et al. menggunakan *Naive Bayes* dengan representasi TF-IDF untuk mengklasifikasikan sentimen mengenai Program Makan Bergizi Gratis [3], [4]. Kedua penelitian tersebut menunjukkan bahwa metode klasifikasi teks klasik dapat digunakan untuk membaca opini masyarakat di Platform X, tetapi fokus utamanya masih berada pada perbandingan algoritma dan performa umum model. Najib et al. menggunakan *Naive Bayes* dan teknik *resampling* pada komentar YouTube mengenai Program

Building of Informatics, Technology and Science (BITS)
Volume 8, No 1, June 2026 Page: 485−496
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.10095

MBG, sehingga menunjukkan bahwa ketidakseimbangan kelas juga menjadi persoalan dalam klasifikasi sentimen topik tersebut [5]. Pada domain isu sosial lainnya, Kono et al. membandingkan IndoBERT, *Logistic Regression*, dan *Linear SVM*, yang memperlihatkan bahwa representasi kontekstual memiliki keunggulan dalam memahami makna teks, tetapi model linear tetap dapat digunakan sebagai pembanding yang kompetitif [6]. Penelitian lain menunjukkan bahwa kombinasi TF-IDF dengan *Support Vector Machine* maupun model klasifikasi klasik masih relevan untuk data teks berdimensi tinggi karena prosesnya efisien, dapat direplikasi, dan memiliki fitur yang relatif mudah dianalisis [7], [8], [9]. Sementara itu, penelitian mengenai ketidakseimbangan kelas menunjukkan bahwa evaluasi model perlu mempertimbangkan *macro F1-score*, *recall*, dan *F1-score* kelas minoritas, bukan hanya *accuracy* [10], [11], [12]. Berdasarkan perbandingan tersebut, penelitian ini tidak hanya membandingkan algoritma atau menambahkan *GridSearchCV*, tetapi secara khusus menguji pembobotan kelas pada model linear dan memilih model berdasarkan kemampuannya mengenali kelas negatif. Pendekatan ini digunakan untuk melengkapi penelitian sebelumnya yang belum menempatkan *class imbalance* sebagai dasar utama dalam pemilihan model sentimen MBG. Secara lebih spesifik, gap penelitian ini terletak pada belum kuatnya penempatan class imbalance sebagai dasar pemilihan model final pada sentimen MBG, karena sebagian studi MBG masih menekankan perbandingan algoritma dan accuracy [3], [4], studi lain membahas resampling pada sumber data YouTube [5], sedangkan studi imbalance pada domain berbeda belum langsung menjawab kebutuhan pemantauan kritik publik terhadap MBG [10], [11], [12].

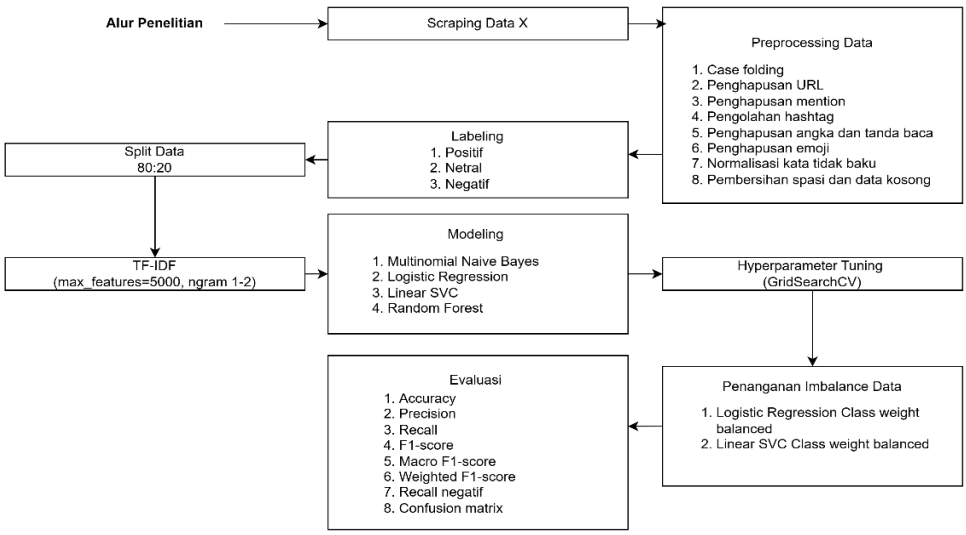
Selain itu, penelitian sentimen kebijakan publik perlu mempertimbangkan karakter percakapan digital yang cepat berubah dan sering dipengaruhi oleh peristiwa tertentu. Pada topik Program Makan Bergizi Gratis, kritik publik tidak selalu muncul dalam jumlah dominan, tetapi dapat memuat informasi penting mengenai kualitas layanan, distribusi, keamanan pangan, dan tata kelola program. Oleh karena itu, pendekatan klasifikasi yang hanya menonjolkan performa agregat belum cukup untuk mendukung pemantauan opini publik. Penelitian ini menempatkan kelas negatif sebagai perhatian utama agar model tidak hanya membaca kecenderungan umum percakapan, tetapi juga mampu membantu menemukan sinyal evaluatif yang memerlukan pemeriksaan lebih lanjut. Penelitian ini bertujuan membangun dan mengevaluasi model klasifikasi sentimen Program Makan Bergizi Gratis pada Platform X dengan representasi TF-IDF, empat model baseline, optimasi GridSearchCV, dan penerapan class_weight balanced pada model linear.

Kontribusi penelitian ini adalah memperjelas posisi model linear berbobot kelas sebagai pendekatan yang tidak hanya mengejar performa umum, tetapi juga memperhatikan recall negatif dan F1-score negatif sebagai indikator penting pada data tidak seimbang. Dengan demikian, penelitian ini diharapkan dapat melengkapi penelitian terdahulu sekaligus menyediakan dasar awal bagi sistem monitoring opini publik digital yang mampu menandai peningkatan kritik dan keluhan untuk pemeriksaan manusia.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Penelitian ini menggunakan pendekatan kuantitatif untuk mengolah opini publik dalam bentuk teks. Proses penelitian disusun secara berurutan agar setiap tahap dapat diperiksa kembali, mulai dari pengambilan data, penyaringan unggahan, pembersihan teks, pemberian label sentimen, pembagian data, pembentukan fitur, pelatihan model, optimasi parameter, penanganan kelas tidak seimbang, sampai evaluasi performa. Alur tersebut digunakan agar pemilihan model final didasarkan pada hasil eksperimen yang terukur, bukan hanya pada satu algoritma tertentu. Rangkaian proses penelitian disajikan pada Gambar 1.



Gambar 1. Alur penelitian klasifikasi sentimen Program Makan Bergizi Gratis

Building of Informatics, Technology and Science (BITS)
Volume 8, No 1, June 2026 Page: 485−496
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.10095

Pada tahap pertama, data awal diperoleh dari Platform X dengan kata kunci yang berkaitan dengan Program Makan Bergizi Gratis. Setelah data terkumpul, unggahan yang kosong, duplikat, tidak relevan, atau tidak layak dianalisis dikeluarkan dari dataset. Teks yang tersisa kemudian melalui tahap preprocessing untuk mengurangi gangguan khas media sosial. Setiap teks diberi label positif, netral, atau negatif, lalu dataset dibagi menjadi data latih dan data uji menggunakan rasio 80:20 secara stratified. Fitur teks dibentuk menggunakan TF-IDF dengan 5.000 fitur dan n-gram satu sampai dua. Empat model digunakan sebagai pembanding awal, yaitu Multinomial Naive Bayes, Logistic Regression, LinearSVC, dan Random Forest. Setelah itu, GridSearchCV digunakan untuk mencari parameter terbaik, sedangkan class_weight balanced diterapkan pada model linear untuk membantu mendeteksi kelas negatif minoritas.

2.2 Pengumpulan Data, Seleksi, dan Labeling

Data penelitian dikumpulkan dari Platform X melalui proses scraping dengan kata kunci yang relevan dengan Program Makan Bergizi Gratis dan MBG. Periode data mencakup unggahan sejak 5 Januari 2025 sampai 26 Februari 2026. Dari proses awal diperoleh 7.659 unggahan. Setelah dilakukan seleksi, preprocessing, cleaning, dan labeling, dataset final yang digunakan untuk pemodelan berjumlah 5.307 teks. Informasi yang dapat mengarah pada identitas pengguna, tautan unggahan, dan data personal tidak disajikan dalam artikel karena fokus penelitian berada pada isi teks dan pola sentimen, bukan pada profil individu pengguna.

Label sentimen diklasifikasikan menjadi tiga kategori. Kelas positif mencerminkan dukungan, apresiasi, harapan, atau penilaian baik terhadap program. Kelas netral berisi unggahan informatif, berita, laporan kegiatan, atau pernyataan yang tidak memperlihatkan sikap evaluatif yang kuat. Kelas negatif mencakup kritik, keluhan, penolakan, atau penilaian kurang baik terhadap pelaksanaan program. Pembagian ini digunakan untuk menjaga konsistensi proses klasifikasi. Distribusi data akhir disajikan pada Tabel 1.

Pelabelan dilakukan otomatis pada clean_text menggunakan aturan leksikal dan skor sentimen, dengan kelas ditentukan oleh skor dominan. Proses ini tidak melibatkan ahli bahasa atau dua anotator independen seperti Cohen's Kappa. Label diperlakukan sebagai weak labels, sedangkan sarkasme, ironi, kutipan, dan makna implisit belum ditangani secara khusus.

Tabel 1. Distribusi Label Dataset Final

Sentimen	Jumlah	Persentase (%)
Positif	2.183	41,13
Netral	2.391	45,05
Negatif	733	13,81
Total	5.307	100,00

2.3 Preprocessing Teks

Preprocessing dilakukan untuk mengubah teks mentah menjadi data yang lebih konsisten dan siap diproses oleh model. Teks pada Platform X memiliki tingkat noise yang tinggi karena dapat memuat URL, mention, tagar, emoji, angka, tanda baca, kata tidak baku, serta variasi penulisan. Pada penelitian ini, preprocessing mencakup case folding, penghapusan URL, penghapusan mention, pengolahan hashtag, penghapusan angka dan tanda baca, penghapusan emoji, normalisasi kata tidak baku, perapian spasi, serta penghapusan data kosong setelah tahap pembersihan. Contoh perubahan teks pada setiap tahap preprocessing ditampilkan pada Tabel 2.

Tabel 2. Contoh Tahapan Preprocessing

Tahap	Sebelum	Sesudah
Case folding	Kabar Bahagia! Bersama Program Makan Bergizi Gratis	kabar bahagia bersama program makan bergizi gratis
Hapus URL	Program Makan Bergizi Gratis https://t.co/GsZ0uQcRxj	program makan bergizi gratis
Hapus mention	@akun1 @akun2 Oppa mari saya jelaskan isu-isu parah pemerintah Indonesia	oppa mari saya jelaskan isu isu parah pemerintah indonesia
Pengolahan hashtag	#FaktaIndonesia #MakanBergiziGratis	fakta indonesia makan bergizi gratis
Hapus angka dan tanda baca	Program Januari 2025 picu ratusan kasus keracunan siswa	program januari picu ratusan kasus keracunan siswa
Hapus emoji	meningkatkan kualitas gizi anak Indonesia [emoji]	meningkatkan kualitas gizi anak indonesia
Normalisasi kata	tdk optimal krn hulunya dsitu	tidak optimal karena hulunya dsitu

2.4 Pembagian Data dan Ekstraksi Fitur TF-IDF

Dataset akhir dibagi menjadi data latih dan data uji dengan proporsi 80:20 menggunakan teknik *stratified split*. Teknik ini digunakan agar distribusi sentimen positif, netral, dan negatif pada data uji tetap mengikuti komposisi dataset

Building of Informatics, Technology and Science (BITS)
Volume 8, No 1, June 2026 Page: 485−496
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.10095

keseluruhan. Jumlah data latih adalah 4.245 teks, sedangkan data uji berjumlah 1.062 teks. Setelah tahap *preprocessing*, setiap teks dikonversi ke bentuk numerik menggunakan *Term Frequency-Inverse Document Frequency* (TF-IDF). TF-IDF memberi bobot lebih tinggi pada term yang penting dalam suatu dokumen, tetapi tidak terlalu sering muncul pada seluruh korpus. Bobot fitur dinyatakan dengan rumus $w(t,d) = tf(t,d) \times \log(N/df(t))$, dengan $tf(t,d)$ sebagai frekuensi term dalam dokumen, N sebagai jumlah dokumen, dan $df(t)$ sebagai jumlah dokumen yang memuat term tersebut. Konfigurasi TF-IDF pada penelitian ini menggunakan max_features=5000 dan ngram_range=(1,2). Pembatasan 5.000 fitur dilakukan untuk mengambil term yang paling informatif dari korpus. Penggunaan unigram dan bigram membantu model mengenali kata tunggal serta frasa pendek yang relevan, seperti "makan bergizi", "program makan", "gratis mbg", atau "partai merah". Proses ekstraksi fitur menghasilkan matriks data latih berukuran 4.245 x 5.000 dan matriks data uji berukuran 1.062 x 5.000. Karena setiap unggahan hanya mengandung sebagian kecil dari keseluruhan kosakata, matriks yang dihasilkan bersifat *sparse*. Secara metodologis, proses prapengolahan dan representasi teks mengikuti prinsip *natural language processing*, yaitu mengubah teks menjadi vektor numerik agar dapat diolah oleh algoritma *machine learning* [13], [14].

Untuk membatasi data leakage, duplikat dihapus sebelum split, TF-IDF di-fit hanya pada data latih, dan data uji hanya ditransformasi serta digunakan untuk evaluasi akhir. GridSearchCV juga dijalankan pada data latih.

Tabel 3. Pembagian Data dan Konfigurasi TF-IDF

Komponen	Nilai
Data latih	4.245 teks
Data uji	1.062 teks
Total data	5.307 teks
Jumlah fitur TF-IDF	5.000
N-gram	(1,2)
Rasio pembagian data	80:20 stratified split

2.5 Model Klasifikasi, Tuning, dan Evaluasi

Empat algoritma digunakan sebagai baseline, yaitu Multinomial Naive Bayes, Logistic Regression, LinearSVC, dan Random Forest. Multinomial Naive Bayes dipilih karena sederhana dan banyak digunakan pada klasifikasi teks berbasis TF-IDF [3], [9]. Logistic Regression digunakan sebagai model linear pembanding karena stabil pada data berdimensi tinggi dan sering diterapkan dalam analisis sentimen [6], [12], [15]. LinearSVC digunakan karena sesuai untuk representasi teks sparse hasil TF-IDF dan mampu membentuk batas keputusan linear pada ruang fitur yang besar [4], [7], [8]. Random Forest ditempatkan sebagai pembanding berbasis ensemble pohon keputusan yang juga digunakan dalam klasifikasi sentimen [10], [16]. Setelah baseline diperoleh, GridSearchCV diterapkan untuk mencari parameter terbaik pada setiap model. Selanjutnya, class_weight balanced diuji pada LinearSVC dan Logistic Regression karena kedua model tersebut mendukung mekanisme pembobotan kelas. Pemilihan model linear tersebut didasarkan pada prinsip klasifikasi terawasi, yaitu pembelajaran pola dari data berlabel melalui representasi numerik dan optimasi parameter [13], [14].

Evaluasi model dilakukan menggunakan accuracy, precision, recall, F1-score, macro F1-score, weighted F1-score, recall negatif, F1-score negatif, dan confusion matrix. Macro F1-score digunakan karena dataset tidak seimbang sehingga setiap kelas perlu memperoleh bobot evaluasi yang setara. Weighted F1-score digunakan untuk menggambarkan performa keseluruhan dengan tetap mempertimbangkan jumlah data pada masing-masing kelas. Recall negatif dan F1-score negatif menjadi indikator penting karena kelas negatif merupakan kelas minoritas yang berisi kritik publik. Skenario pengujian disajikan pada Tabel 4, sedangkan parameter terbaik hasil GridSearchCV ditampilkan pada Tabel 5.

Tabel 4. Skenario Pengujian Model

Skenario	Tujuan	Model
Baseline	Mengetahui performa awal	MNB, LR, LinearSVC, RF
GridSearchCV	Mencari parameter terbaik	MNB, LR, LinearSVC, RF
Class weight balanced	Menangani kelas negatif minoritas	LinearSVC dan LR
Model final	Memilih model paling seimbang	Kandidat terbaik

Tabel 5. Parameter Terbaik Hasil GridSearchCV

Model	Parameter Terbaik	CV Weighted F1
Random Forest	max_depth=None; min_samples_split=5; n_estimators=200	0,8959
LinearSVC	C=1	0,8880
Logistic Regression	C=10; max_iter=1000; solver=lbfgs	0,8856
Multinomial Naive Bayes	alpha=0,1	0,7984

Building of Informatics, Technology and Science (BITS)
Volume 8, No 1, June 2026 Page: 485−496
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.10095

3. HASIL DAN PEMBAHASAN

3.1 Karakteristik Dataset

Dataset final terdiri atas 5.307 teks. Sentimen netral berjumlah 2.391 data atau 45,05%, sentimen positif berjumlah 2.183 data atau 41,13%, dan sentimen negatif berjumlah 733 data atau 13,81%. Distribusi ini menunjukkan bahwa percakapan Program Makan Bergizi Gratis pada Platform X lebih banyak berupa informasi, berita, laporan kegiatan, atau pernyataan yang tidak memuat sikap evaluatif yang kuat. Sentimen positif berada pada posisi kedua dan menunjukkan adanya dukungan masyarakat terhadap tujuan program, terutama terkait pemenuhan gizi, bantuan bagi siswa, dan harapan terhadap kualitas sumber daya manusia.

Kelas negatif memiliki proporsi paling kecil, tetapi tetap penting secara substantif. Kritik pada kelas negatif dapat memuat isu kualitas makanan, kasus keracunan, distribusi, kesiapan dapur, anggaran, maupun tata kelola pelaksanaan. Dari sisi machine learning, proporsi kelas negatif yang lebih kecil menciptakan masalah ketidakseimbangan kelas. Jika model terlalu dipengaruhi oleh kelas mayoritas, maka kritik publik berisiko diprediksi sebagai netral atau positif. Oleh karena itu, evaluasi penelitian ini tidak hanya menggunakan accuracy, tetapi juga mempertimbangkan macro F1-score, recall negatif, dan F1-score negatif.

Tabel 6. Distribusi Train-Test Secara Stratified

Kelas	Data Latih	Data Uji	Total
Positif	1.746	437	2.183
Netral	1.913	478	2.391
Negatif	586	147	733
Total	4.245	1.062	5.307

3.2 Representasi TF-IDF dan Term Dominan

Hasil pembentukan fitur TF-IDF menunjukkan bahwa representasi teks menghasilkan matriks berdimensi tinggi dan bersifat sparse. Pada data latih, matriks yang terbentuk berukuran 4.245 x 5.000, sedangkan pada data uji matriks berukuran 1.062 x 5.000. Jumlah fitur 5.000 diperoleh dari term paling informatif setelah pembatasan max_features dan penggunaan n-gram satu sampai dua. Informasi ukuran matriks, nilai non-zero, tingkat sparsity, dan konfigurasi n-gram disajikan pada Tabel 7.

Tabel 7. Hasil Pembentukan Matriks TF-IDF

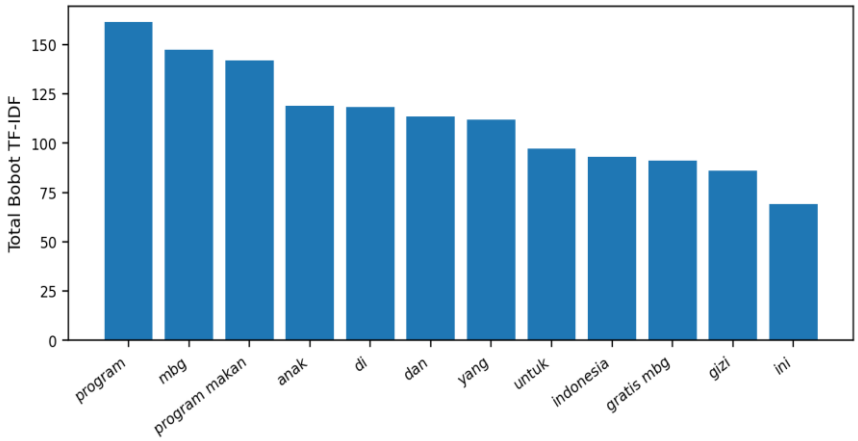
Komponen	Nilai
Jumlah fitur	5.000
Matriks latih	4.245 x 5.000
Matriks uji	1.062 x 5.000
Non-zero latih	108.525
Non-zero uji	25.277
Sparsity latih	0,9949
Sparsity uji	0,9952
N-gram	(1,2)

Nilai sparsity yang sangat tinggi menunjukkan bahwa sebagian besar elemen pada matriks bernilai nol. Kondisi ini wajar pada data teks pendek karena satu unggahan hanya memuat sebagian kecil dari keseluruhan kosakata yang tersedia. Karakteristik tersebut memperkuat alasan penggunaan model linear, terutama LinearSVC, karena model linear umumnya stabil pada representasi vektor berdimensi tinggi dan sparse.

Tabel 8. Dua Belas Term dengan Bobot TF-IDF Tertinggi

Term	Total Bobot	Rata-rata
program	161,3080	0,0380
mbg	147,2910	0,0347
program makan	141,9130	0,0334
anak	118,9830	0,0280
di	118,2900	0,0279
dan	113,4530	0,0267
yang	112,0810	0,0264
untuk	97,1860	0,0229
indonesia	92,9770	0,0219
gratis mbg	91,1680	0,0215
gizi	86,1960	0,0203
ini	69,0390	0,0163

Building of Informatics, Technology and Science (BITS)
Volume 8, No 1, June 2026 Page: 485−496
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.10095



Gambar 2. Term dengan total bobot TF-IDF tertinggi pada data latih

Tabel 8 dan Gambar 2 menunjukkan bahwa term program, mbg, program makan, anak, indonesia, gratis mbg, dan gizi menjadi penanda utama dalam korpus. Term tersebut merepresentasikan pokok percakapan mengenai kebijakan, sasaran program, dan isu gizi. Kemunculan bigram seperti program makan dan gratis mbg menunjukkan bahwa penggunaan n-gram (1,2) membantu menangkap frasa pendek yang lebih informatif dibandingkan hanya mengandalkan kata tunggal.

Tabel 9. Term TF-IDF Dominan Berdasarkan Kelas Sentimen

Positif	Netral	Negatif
program (0,0475)	program (0,0335)	keracunan (0,0652)
anak (0,0451)	di (0,0319)	mbg (0,0475)
program makan (0,0417)	program makan (0,0305)	korupsi (0,0373)
indonesia (0,0393)	mbg (0,0280)	yang (0,0366)
mbg (0,0377)	yang (0,0243)	partai merah (0,0358)
dan (0,0343)	dan (0,0217)	partai (0,0353)
dukung (0,0329)	gratis mbg (0,0202)	merah (0,0352)
untuk (0,0327)	tidak (0,0187)	prabowo (0,0347)

Analisis per kelas memperlihatkan perbedaan pola leksikal yang lebih substantif. Kelas positif banyak berkaitan dengan term program, anak, dukung, dan indonesia yang merepresentasikan dukungan atau harapan terhadap manfaat program. Kelas netral didominasi term informatif seperti program, di, mbg, dan gratis mbg yang cenderung muncul pada unggahan berita, laporan kegiatan, atau informasi kebijakan. Kelas negatif memiliki pola berbeda karena term keracunan, korupsi, partai merah, partai, merah, dan prabowo lebih menonjol. Pola tersebut menunjukkan bahwa fitur TF-IDF dapat membantu model memisahkan kelas berdasarkan kata dan frasa yang mencerminkan konteks opini. Namun, term partai merah dan prabowo menunjukkan potensi bias konteks politik. Term tersebut tidak dapat langsung ditafsirkan sebagai indikator negatif terhadap substansi MBG tanpa pemeriksaan konteks unggahan.

Tabel 10. Contoh Bobot TF-IDF pada Dokumen Uji

Label	Cuplikan teks	Bobot TF-IDF tertinggi
Positif	pengalihan fungsi itu terjadi karena sekolah sekolah penerima manfaat makan bergizi gratis mbg di berbagai wilayah yang terdampak...	terdampak banjir=0,273; wilayah yang=0,266; fungsi=0,261; yang terdampak=0,256; berbagai wilayah=0,234; sekolah=0,232
Netral	mengutamakan evaluasi demi perbaikan adalah solusi bijak semoga program makan bergizi gratis semakin terdepan dalam memenuhi kebutuhan rakyat...	mengutamakan=0,357; bijak=0,357; memenuhi kebutuhan=0,294; semoga=0,288; perbaikan=0,285; semakin=0,281
Negatif	lah kocak di sebuah resto saja kalo ada ulat atau kecoa satu itu kan kejadian sama sekali tidak...	saja=0,363; lah=0,324; ulat=0,208; tidak pernah=0,204; sama sekali=0,204; ada=0,201

Contoh pada Tabel 10 menunjukkan bahwa bobot TF-IDF tertinggi tidak selalu berasal dari kata yang paling umum, tetapi dari kata atau frasa yang paling khas pada dokumen tertentu. Pada teks negatif, term seperti ulat, tidak pernah, dan sama sekali menjadi sinyal leksikal yang berkaitan dengan keluhan. Pada teks netral, term seperti mengutamakan, bijak, dan perbaikan muncul dalam konteks informatif atau evaluatif ringan. Dengan demikian, hasil TF-IDF tidak hanya berfungsi sebagai masukan numerik, tetapi juga memberikan dasar interpretatif untuk memahami fitur yang digunakan model dalam proses klasifikasi.

Building of Informatics, Technology and Science (BITS)
Volume 8, No 1, June 2026 Page: 485−496
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.10095

3.3 Hasil Baseline dan Hyperparameter Tuning

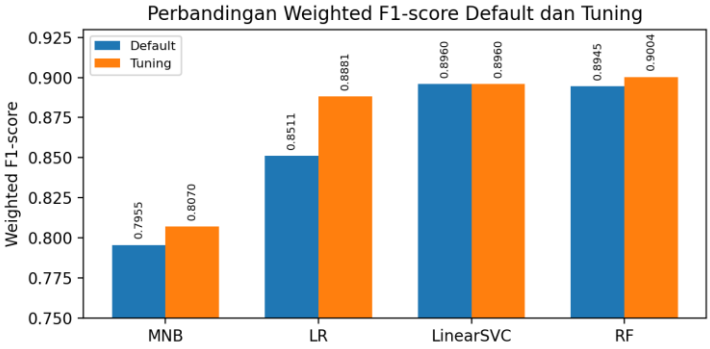
Pengujian awal menunjukkan bahwa LinearSVC memperoleh performa terbaik pada skenario *baseline*, dengan *accuracy* 0,8964 dan *weighted F1-score* 0,8960. Random Forest menghasilkan nilai yang hampir sama, yaitu *accuracy* 0,8955 dan *weighted F1-score* 0,8945. Logistic Regression memperoleh *weighted F1-score* 0,8511, sedangkan Multinomial Naive Bayes menghasilkan *weighted F1-score* 0,7955. Hasil ini menunjukkan bahwa model linear berbasis margin lebih sesuai untuk data TF-IDF yang berdimensi tinggi dan *sparse*. Naive Bayes tetap digunakan sebagai pembanding dasar, tetapi asumsi independensi antarfitur membuat performanya lebih terbatas pada teks media sosial yang pendek, informal, dan memiliki banyak variasi frasa.

Proses *hyperparameter tuning* dengan GridSearchCV meningkatkan performa beberapa model. Random Forest menjadi model terbaik setelah tuning berdasarkan *accuracy*, yaitu 0,9011, dengan *weighted F1-score* 0,9004. Logistic Regression juga meningkat dari *weighted F1-score* 0,8511 menjadi 0,8881. Multinomial Naive Bayes naik dari 0,7955 menjadi 0,8070 setelah pengaturan nilai alpha. Pada LinearSVC, hasil tidak berubah karena parameter terbaik tetap menghasilkan performa yang sama dengan model awal. Hasil ini menunjukkan bahwa tuning dapat memperbaiki performa sebagian model, tetapi dampaknya tetap dipengaruhi oleh karakteristik data dan representasi fitur [17], [18]. Namun, Random Forest tuning belum dipilih sebagai model final karena nilai *recall* negatifnya masih lebih rendah dibandingkan LinearSVC balanced. Oleh karena itu, pemilihan model final tetap mempertimbangkan kemampuan mendeteksi kelas negatif, bukan hanya performa keseluruhan.

Sebagai pemeriksaan awal overfitting, selisih weighted F1 antara cross-validation dan data uji hanya 0,0080 pada LinearSVC dan 0,0045 pada Random Forest. Namun, generalisasi lintas periode belum diuji.

Tabel 11. Hasil Evaluasi Model Baseline dan Tuning

Model	Skenario	Accuracy	Precision	Recall	Weighted F1
LinearSVC	Default	0,8964	0,9008	0,8964	0,8960
Random Forest	Default	0,8955	0,9015	0,8955	0,8945
Logistic Regression	Default	0,8531	0,8674	0,8531	0,8511
Multinomial NB	Default	0,7994	0,8156	0,7994	0,7955
Random Forest	Tuning	0,9011	0,9061	0,9011	0,9004
LinearSVC	Tuning	0,8964	0,9008	0,8964	0,8960
Logistic Regression	Tuning	0,8889	0,8933	0,8889	0,8881
Multinomial NB	Tuning	0,8079	0,8112	0,8079	0,8070



Gambar 3. Perbandingan weighted F1-score model default dan tuning

3.4 Pengaruh Class Weight Balanced

Pengujian class_weight balanced dilakukan untuk melihat dampak pembobotan kelas terhadap data yang tidak seimbang. Pada penelitian sentimen, kelas minoritas sering kurang terdeteksi apabila model terlalu dipengaruhi oleh kelas mayoritas. Oleh karena itu, strategi pembobotan kelas digunakan agar kesalahan pada kelas negatif memperoleh perhatian lebih besar selama proses pelatihan. Temuan penelitian terdahulu juga menunjukkan bahwa penanganan kelas minoritas melalui resampling maupun pembobotan dapat meningkatkan sensitivitas model terhadap kelas yang jumlah datanya lebih kecil [5], [11], [12].

Pada LinearSVC, penggunaan class_weight balanced meningkatkan accuracy dari 0,8964 menjadi 0,9011. Macro F1-score naik dari 0,8898 menjadi 0,8989, sedangkan weighted F1-score meningkat dari 0,8960 menjadi 0,9011. Perubahan paling penting terlihat pada kelas negatif, yaitu recall negatif meningkat dari 0,7823 menjadi 0,8367 dan F1-score negatif meningkat dari 0,8679 menjadi 0,8913. Hasil ini menunjukkan bahwa pembobotan kelas membantu model mengenali lebih banyak unggahan negatif tanpa menurunkan performa umum.

Pengujian tambahan terhadap LinearSVC balanced dengan tuning parameter C menghasilkan nilai evaluasi yang sama dengan LinearSVC balanced, yaitu accuracy 0,9011, macro F1-score 0,8989, weighted F1-score 0,9011, recall negatif 0,8367, dan F1-score negatif 0,8913. Oleh karena itu, hasil LinearSVC balanced tuning tidak ditampilkan sebagai baris terpisah pada Tabel 12 agar tidak terjadi pengulangan nilai yang identik.

Building of Informatics, Technology and Science (BITS)
Volume 8, No 1, June 2026 Page: 485−496
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.10095

Pola peningkatan juga terlihat pada Logistic Regression setelah class_weight balanced dan tuning C diterapkan. Pada skenario default, Logistic Regression memperoleh recall negatif 0,6259 dan F1-score negatif 0,7699. Setelah pembobotan kelas dan tuning, recall negatif meningkat menjadi 0,8231 dan F1-score negatif menjadi 0,8832. Hal ini mengindikasikan bahwa distribusi kelas menjadi faktor penting dalam kinerja model. Class_weight balanced tidak menambah data sintetis, tetapi mengubah bobot kesalahan sehingga kelas minoritas memperoleh perlakuan lebih proporsional dalam pelatihan.

Tabel 12. Hasil Model Linear dengan Class Weight Balanced

Model	Accuracy	Macro F1	Weighted F1	Recall Negatif	F1 Negatif
LinearSVC default	0,8964	0,8898	0,8960	0,7823	0,8679
LinearSVC balanced	0,9011	0,8989	0,9011	0,8367	0,8913
Logistic Regression default	0,8531	0,8329	0,8511	0,6259	0,7699
Logistic Regression balanced tuning	0,9002	0,8964	0,9001	0,8231	0,8832

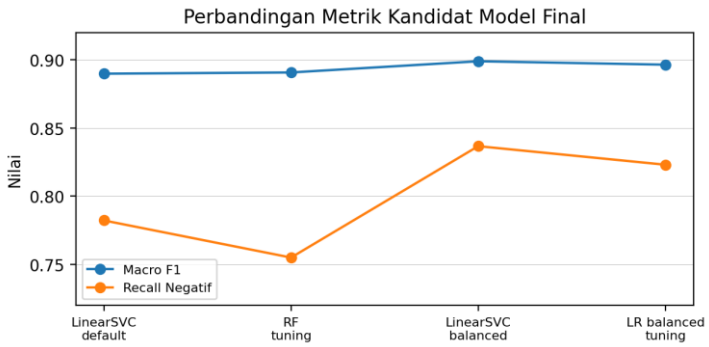
3.5 Pemilihan Model Final

Berdasarkan perbandingan kandidat model, LinearSVC dengan class_weight balanced dipilih sebagai model final. Model ini memiliki accuracy yang sama dengan Random Forest tuning, yaitu 0,9011. Namun, LinearSVC balanced memberikan macro F1-score, weighted F1-score, recall negatif, dan F1-score negatif yang lebih tinggi. Jika dibandingkan dengan Logistic Regression balanced tuning, LinearSVC balanced juga tetap unggul pada metrik utama. Temuan ini memperlihatkan bahwa pemilihan model tidak cukup hanya didasarkan pada accuracy, terutama ketika dataset memiliki kelas minoritas yang penting untuk dideteksi.

Secara teknis, LinearSVC sesuai dengan karakteristik data TF-IDF karena model ini bekerja dengan batas keputusan linear pada ruang fitur berdimensi besar dan sparse. Pada teks pendek, sebagian besar fitur bernilai nol sehingga model linear sering lebih stabil dibandingkan model berbasis pohon. Penerapan class_weight balanced membuat LinearSVC lebih peka terhadap kelas negatif dengan menaikkan penalti kesalahan pada kelas minoritas. Kombinasi tersebut menghasilkan keseimbangan antara performa keseluruhan dan kemampuan mendeteksi kritik publik.

Tabel 13. Perbandingan Kandidat Model Final

Model	Accuracy	Macro F1	Weighted F1	Recall Negatif	F1 Negatif
LinearSVC default	0,8964	0,8898	0,8960	0,7823	0,8679
Random Forest tuning	0,9011	0,8907	0,9004	0,7551	0,8571
LinearSVC balanced	0,9011	0,8989	0,9011	0,8367	0,8913
LR balanced tuning	0,9002	0,8964	0,9001	0,8231	0,8832



Gambar 4. Perbandingan metrik kandidat model final

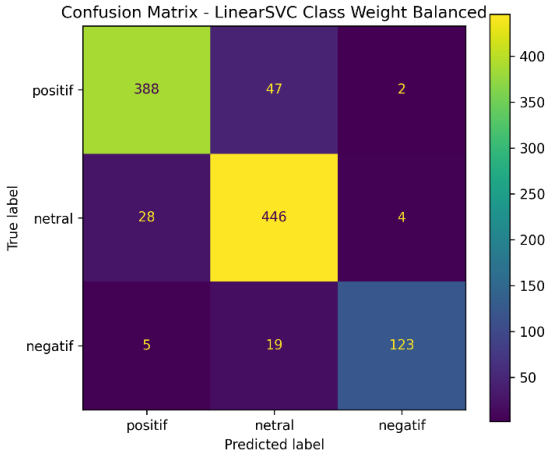
3.6 Confusion Matrix, Classification Report, dan Analisis Kesalahan

Confusion matrix LinearSVC balanced memperlihatkan bahwa 388 dari 437 data positif berhasil diprediksi sebagai positif. Pada kelas netral, 446 dari 478 data terklasifikasi dengan benar. Pada kelas negatif, 123 dari 147 data berhasil dikenali sebagai negatif, sedangkan 5 data diprediksi sebagai positif dan 19 data diprediksi sebagai netral. Kesalahan terbesar pada kelas negatif terjadi ketika teks negatif diprediksi sebagai netral. Kondisi ini dapat terjadi karena sebagian kritik ditulis dalam bentuk laporan atau informasi faktual sehingga secara leksikal mendekati kelas netral.

Tabel 14. Confusion Matrix LinearSVC Balanced

Aktual / Prediksi	Positif	Netral	Negatif
Positif	388	47	2
Netral	28	446	4
Negatif	5	19	123

Building of Informatics, Technology and Science (BITS)
Volume 8, No 1, June 2026 Page: 485−496
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.10095



Gambar 5. Confusion matrix model LinearSVC balanced

Classification report menunjukkan bahwa model final memperoleh precision 0,9216, recall 0,8879, dan F1-score 0,9044 pada kelas positif. Pada kelas netral, model menghasilkan precision 0,8711, recall 0,9331, dan F1-score 0,9010. Pada kelas negatif, precision mencapai 0,9535, recall 0,8367, dan F1-score 0,8913. Precision negatif yang tinggi menunjukkan bahwa prediksi negatif cenderung tepat, sedangkan peningkatan recall negatif menunjukkan bahwa class_weight balanced membantu model menemukan lebih banyak unggahan yang benar-benar berisi kritik.

Tabel 15. Classification Report LinearSVC Balanced

Kelas	Precision	Recall	F1-score	Support
Positif	0,9216	0,8879	0,9044	437
Netral	0,8711	0,9331	0,9010	478
Negatif	0,9535	0,8367	0,8913	147

Analisis kesalahan memperlihatkan bahwa model klasik berbasis TF-IDF masih memiliki keterbatasan dalam memahami konteks semantik. Teks negatif dapat masuk ke kelas netral apabila kritik disampaikan dalam bentuk narasi informatif. Sebaliknya, teks netral dapat diprediksi positif atau negatif ketika memuat kata dengan muatan evaluatif yang kuat. Hal ini menunjukkan bahwa TF-IDF efektif menangkap pola kata dan frasa [8], [9], [19], tetapi belum sepenuhnya menangkap sarkasme, ironi, kutipan, atau makna implisit. Oleh karena itu, model berbasis representasi kontekstual seperti IndoBERT dan transformer dapat diuji pada penelitian berikutnya [20], [21], [22], [23]. Kesalahan kemungkinan dipengaruhi berita faktual, dukungan implisit, negasi, atau sarkasme. Mitigasinya meliputi penyaringan relevansi, pemeriksaan prediksi bermargin rendah, dan pengujian representasi kontekstual.

Tabel 16. Contoh Kesalahan Klasifikasi Model Final

Aktual	Prediksi	Cuplikan Teks
Negatif	Netral	kritik program ditulis seperti laporan berita sehingga dekat dengan kelas netral
Positif	Netral	dukungan program tidak dinyatakan secara eksplisit dalam teks
Netral	Positif	informasi kegiatan memuat kata bernada dukungan
Netral	Negatif	laporan informatif memuat istilah kerugian atau pemotongan anggaran

3.7 Pembahasan

Dibandingkan dengan penelitian terdahulu pada topik Program Makan Bergizi Gratis, penelitian ini memberi penguatan pada rancangan eksperimen, jumlah model yang diuji, proses optimasi, dan perhatian terhadap kelas negatif minoritas. Saleh dan Imanda meneliti sentimen publik terkait Program Makan Gratis pada Platform X dengan membandingkan Naive Bayes dan Support Vector Machine [3], Aprianti et al. menggunakan Naive Bayes dan TF-IDF untuk menganalisis sentimen Program Makan Bergizi Gratis pada media sosial X [4]. Najib et al. mengkaji persepsi publik terhadap Program MBG melalui komentar YouTube dengan Naive Bayes dan teknik resampling [5]. Penelitian ini memperluas pendekatan tersebut dengan membandingkan empat model baseline, menerapkan GridSearchCV pada seluruh model, serta menguji class_weight balanced pada dua model linear. Kajian Barus et al., Budaya dan Suniantara, serta Johnson dan Khoshgoftaar juga menunjukkan bahwa penanganan *imbalance* penting dalam klasifikasi teks dan analisis sentimen [10], [11], [12], Sementara itu, Kono et al. memperlihatkan bahwa model linear masih kompetitif terhadap pendekatan *deep learning* pada isu sosial tertentu [6].

Jika dibandingkan secara naratif, posisi penelitian ini berbeda dari Saleh dan Imanda[3] serta Aprianti et al. [4] yang sama-sama membahas sentimen program makan pada Platform X, tetapi belum menjadikan kinerja kelas negatif sebagai fokus utama pemilihan model. Penelitian ini juga berbeda dari Najib et al. [5] yang menggunakan komentar

Building of Informatics, Technology and Science (BITS)
Volume 8, No 1, June 2026 Page: 485−496
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.10095

YouTube dan teknik resampling, karena penelitian ini mempertahankan distribusi data asli serta menerapkan class_weight balanced pada LinearSVC dan Logistic Regression. Perbandingan tersebut menunjukkan bahwa kebaruan penelitian ini berada pada penanganan ketidakseimbangan kelas melalui pembobotan model linear, bukan hanya pada penggunaan TF-IDF atau GridSearchCV.

Pada domain lain, Barus et al. [10] dan Budaya dan Suniantara [11] menunjukkan pentingnya penanganan imbalance pada klasifikasi sentimen, sedangkan Johnson dan Khoshgoftaar [12] menegaskan bahwa kelas minoritas sering membutuhkan perhatian khusus dalam proses pelatihan dan evaluasi. Kono et al. [6], Putri et al. [20], Jim et al. [21], Arasy et al. [22], dan Andhika et al. [23] menunjukkan bahwa pendekatan kontekstual atau model pembelajaran lebih lanjut dapat membantu memahami makna teks yang kompleks. Sementara itu, Firmanda dan Suryono [24] memperlihatkan relevansi analisis sentimen kebijakan publik di Platform X, tetapi objek kebijakannya berbeda dari MBG. Dengan demikian, temuan penelitian ini menempatkan LinearSVC balanced sebagai pendekatan yang efisien, mudah direplikasi, dan tetap sensitif terhadap kritik publik sebagai kelas minoritas.

Implikasi praktis dari penelitian ini berkaitan dengan pemanfaatan model LinearSVC balanced untuk pemantauan opini publik digital. Model tersebut dapat digunakan sebagai dasar klasifikasi otomatis terhadap unggahan baru setelah teks melalui tahap *preprocessing* dan transformasi TF-IDF dengan *vectorizer* yang sama. Hasil prediksi dapat ditampilkan dalam bentuk dashboard untuk memantau distribusi sentimen dari waktu ke waktu. Kenaikan sentimen negatif pada periode tertentu dapat menjadi sinyal awal untuk memeriksa isu seperti kualitas makanan, distribusi, kesiapan dapur penyedia, anggaran, atau kasus keracunan. Sentimen netral dapat menunjukkan dominasi arus informasi dan pemberitaan, sedangkan sentimen positif dapat membantu melihat aspek program yang diterima masyarakat.

Secara ilmiah, kontribusi penelitian ini terletak pada penggunaan metrik yang tidak hanya menilai performa umum, tetapi juga memperhatikan kelas minoritas. Dalam konteks kebijakan publik, kritik tidak selalu menjadi kelas mayoritas, tetapi tetap dapat memuat informasi evaluatif yang penting. Oleh karena itu, *recall* negatif dan *F1-score* negatif digunakan untuk memastikan model tidak mengabaikan data yang berisi kritik. Hasil penelitian ini juga menunjukkan bahwa model *machine learning* klasik masih relevan ketika ukuran dataset belum terlalu besar, kebutuhan replikasi tinggi, dan interpretasi fitur masih diperlukan [8], [9], [19]. Pendekatan transformer dapat menjadi pengembangan berikutnya, tetapi LinearSVC tetap menawarkan efisiensi dan kemudahan implementasi [6], [20], [21], [22], [23].

3.8 Ancaman Validitas

Penelitian ini memiliki beberapa keterbatasan yang perlu diperhatikan. Pertama, data hanya berasal dari Platform X, sehingga hasil klasifikasi belum dapat mewakili seluruh opini masyarakat Indonesia. Karakter pengguna, pemilihan kata kunci, serta periode pengambilan data dapat memengaruhi isi korpus. Kedua, proses labeling berpengaruh langsung terhadap kualitas model. Label yang tidak konsisten dapat membuat model mempelajari pola yang kurang tepat, sehingga kriteria pelabelan perlu dijaga dan data ambigu perlu diperiksa ulang. Ketiga, representasi TF-IDF masih berfokus pada pola kata dan frasa, sehingga belum mampu memahami konteks semantik secara mendalam. Sarkasme, ironi, kutipan, dan makna tersirat masih berpotensi menimbulkan kesalahan prediksi. Keempat, penelitian ini belum memasukkan analisis perubahan sentimen berdasarkan waktu. Padahal, opini publik terhadap Program Makan Bergizi Gratis dapat berubah mengikuti pemberitaan, peristiwa baru, atau dinamika pelaksanaan program.

Temuan penelitian ini juga menunjukkan bahwa evaluasi pada kelas minoritas perlu diperhatikan dalam analisis sentimen kebijakan publik. Pada kasus Program Makan Bergizi Gratis, sentimen negatif tidak dapat dianggap sebagai gangguan data karena berisi kritik, keluhan, atau sinyal masalah yang dapat menjadi masukan kebijakan. Oleh karena itu, model final dipilih bukan hanya karena nilai akurasinya tinggi, tetapi juga karena mampu menjaga keseimbangan antara performa keseluruhan dan kemampuan mengenali kritik publik. Pendekatan ini dapat dijadikan dasar untuk penelitian sentimen pada isu kebijakan lain yang memiliki pola percakapan digital serupa.

4. KESIMPULAN

Penelitian ini membangun model klasifikasi sentimen publik terhadap Program Makan Bergizi Gratis pada Platform X menggunakan pendekatan text mining dan machine learning. Data awal terdiri atas 7.659 unggahan pada periode 5 Januari 2025 sampai 26 Februari 2026. Setelah melalui seleksi, preprocessing, cleaning, dan labeling, sebanyak 5.307 teks digunakan sebagai dataset final, terdiri atas 2.183 sentimen positif, 2.391 sentimen netral, dan 733 sentimen negatif. Distribusi tersebut menunjukkan adanya ketidakseimbangan kelas, terutama pada kelas negatif, sehingga evaluasi model perlu memperhatikan metrik yang sensitif terhadap kelas minoritas. Representasi teks dibangun menggunakan TF-IDF dengan max_features=5000 dan ngram_range=(1,2), menghasilkan matriks latih berukuran 4.245 x 5.000 dan matriks uji berukuran 1.062 x 5.000. Hasil baseline menunjukkan LinearSVC sebagai model default terbaik dengan accuracy 0,8964 dan weighted F1-score 0,8960. Setelah tuning, Random Forest mencapai accuracy 0,9011 dan weighted F1-score 0,9004, tetapi recall negatifnya masih lebih rendah dibandingkan model linear berbobot. Pengujian class_weight balanced menunjukkan bahwa LinearSVC balanced memberikan performa paling seimbang dengan accuracy 0,9011, macro F1-score 0,8989, weighted F1-score 0,9011, recall negatif 0,8367, dan F1-score negatif 0,8913. Berdasarkan hasil tersebut, LinearSVC dengan class_weight balanced dipilih sebagai model

Building of Informatics, Technology and Science (BITS)
Volume 8, No 1, June 2026 Page: 485−496
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.10095

final karena mampu meningkatkan deteksi sentimen negatif tanpa menurunkan performa keseluruhan. Temuan ini dapat digunakan sebagai dasar pengembangan sistem pemantauan opini publik digital terhadap Program Makan Bergizi Gratis. Secara praktis, model dapat diintegrasikan melalui alur pengumpulan unggahan baru, preprocessing, transformasi TF-IDF, klasifikasi otomatis, visualisasi dashboard, dan penandaan kenaikan sentimen negatif untuk pemeriksaan manusia. Secara ilmiah, penelitian ini menegaskan bahwa evaluasi klasifikasi sentimen kebijakan publik perlu memperhatikan kelas minoritas agar kritik yang bernilai evaluatif tidak tertutup oleh dominasi kelas positif atau netral. Penelitian berikutnya dapat memperluas sumber data, memperkuat validasi labeling, menambahkan analisis temporal, serta menguji model berbasis transformer untuk menangkap konteks semantik secara lebih mendalam. Model ini berpotensi menjadi komponen awal pemantauan untuk menandai peningkatan sentimen negatif bagi pemeriksaan manusia, bukan sebagai dasar tunggal penilaian kebijakan. Pengujian transformer diposisikan sebagai pembanding konteks, bukan pengganti mutlak LinearSVC.

REFERENCES

[1] Ulfatul Karomah; Fani Cahya Wahyuni; Yunita Dewi Trisnasari, "Program Penyelenggaraan Makan Siang Sekolah: Studi Literatur tentang Dampak Kesehatan, Hambatan dan Tantangan," *Salus Cultura: Jurnal Pembangunan Manusia dan Kebudayaan*, vol. 4, no. 1, pp. 91–103, Jun. 2024, doi: 10.55480/saluscultura.v4i1.188.

[2] S. Nur, M. Irmanto, and M. S. Fatiah, "Status Gizi Siswa Sekolah Dasar Sebelum Program Makan Bergizi Gratis di SD Negeri Inpres Skouw Sae Kota Jayapura," *Jurnal SAGO Gizi dan Kesehatan*, vol. 6, no. 3, pp. 616–623, 2025, doi: 10.30867/gikes.v6i3.2818.

[3] M. Farhan Saleh and R. Imanda, "Public Sentiment Analysis of the Free Meal Program: A Comparison of Naive Bayes and Support Vector Machine Methods on the Twitter (X) Social Media Platform," 2025. doi: 10.30871/jaic.v9i1.8895.

[4] N. Nyoman Aprianti *et al.*, "Public Sentiment Analysis of the Free Nutritious Meals Program (MBG) on Social Media X Using the Naive Bayes Method," 2025. doi: 10.30871/jaic.v9i6.11420.

[5] L. Najib, A. A. Mahfudh, and S. Bakhri, "Analisis Sentimen Persepsi Publik Terhadap Program MBG Pada Komentar YouTube Menggunakan Naïve Bayes dan Resampling," *Technology and Science (BITS)*, vol. 7, no. 4, pp. 2467–2478, 2026, doi: 10.47065/bits.v7i4.9400.

[6] M. F. Kono, I. N. Fajri, and Y. Pristyanto, "Public Sentiment Analysis on Corruption Issues in Indonesia Using IndoBERT Fine-Tuning, Logistic Regression, and Linear SVM," 2025. doi: 10.30871/jaic.v9i5.10537.

[7] F. Rifaldy, Y. Sibaroni, and S. S. Prasetiyowati, "Effectiveness of Word2Vec and TF-IDF in Sentiment Classification on Online Investment Platforms Using Support Vector Machine," *JIPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 10, no. 2, pp. 863–874, Mar. 2025, doi: 10.29100/jipi.v10i2.6055.

[8] F. Syarifuddin and D. Kusumaningsih, "Analisis Perbandingan Algoritma Naive Bayes dan Support Vector Machine dengan Pendekatan TF-IDF Sebagai Klasifikasi Perintah Suara," *Technology and Science (BITS)*, vol. 7, no. 1, pp. 44–53, 2025, doi: 10.47065/bits.v7i1.7160.

[9] S. Khoerunnisa, D. F. Shiddieq, and D. Nurhayati, "Penerapan Algoritma Naive Bayes dengan Teknik TF-IDF dan Cross Validation untuk Analisis Sentimen Terhadap Starlink," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 5, no. 2, pp. 566–577, Mar. 2025, doi: 10.57152/malcom.v5i2.1852.

[10] H. Barus, I. N. Fajri, and Y. Pristyanto, "Sentiment Classification Analysis of Tokopedia Reviews Using TF-IDF, SMOTE, and Traditional Machine Learning Models," 2025. doi: 10.30871/jaic.v9i5.10524.

[11] I. G. B. A. Budaya and I. K. P. Suniantara, "Comparison of Sentiment Analysis Algorithms with SMOTE Oversampling and TF-IDF Implementation on Google Reviews for Public Health Centers," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 4, no. 3, pp. 1077–1086, Jul. 2024, doi: 10.57152/malcom.v4i3.1459.

[12] J. M. Johnson and T. M. Khoshgoftaar, "Survey on deep learning with class imbalance," *J. Big Data*, vol. 6, no. 1, Dec. 2019, doi: 10.1186/s40537-019-0192-5.

[13] D. Jurafsky and J. H. Martin, "Speech and Language Processing An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition with Language Models," 2026.

[14] M. Peter, D. A. Aldo, F. Cheng, and S. Ong, "MATHEMATICS FOR MACHINE LEARNING," 2020. doi: 10.1017/9781108679930.

[15] D. Y. Saraswati, M. R. Handayani, K. Umam, and I. Mustofa, "Sentiment Classification of MyPertamina Reviews Using Naïve Bayes and Logistic Regression," 2025. doi: 10.30871/jaic.v9i4.9723.

[16] R. E. Nurfirdaus, M. Agung Barata, and I. W. Prastya, "Comparison of SVM and Random Forest for TikTok E10 Fuel Sentiment Analysis," 2026. doi: 10.30871/jaic.v10i2.12395.

[17] W. Wijiyanto, A. I. Pradana, S. Sopingi, and V. Atina, "Teknik K-Fold Cross Validation untuk Mengevaluasi Kinerja Mahasiswa," *Jurnal Algoritma*, vol. 21, no. 1, May 2024, doi: 10.33364/algoritma/v.21-1.1618.

[18] E. Elgeldawi, A. Sayed, A. R. Galal, and A. M. Zaki, "Hyperparameter tuning for machine learning algorithms used for arabic sentiment analysis," *Informatics*, vol. 8, no. 4, Dec. 2021, doi: 10.3390/informatics8040079.

[19] R. Mulyawan, H. Naparin, and W. M. Fatihia, "Comparison of Text Vectorization Methods for IMDB Movie Review Sentiment Analysis Using SVM," 2025. doi: 10.30871/jaic.v9i5.10372.

[20] D. Ismiyana Putri, A. Nurul Alfian, M. Yudhi Putra, and P. Dwi Mulyo, "IndoBERT Model Analysis: Twitter Sentiments on Indonesia's 2024 Presidential Election," 2024. doi: 10.30871/jaic.v8i1.7440.

[21] J. R. Jim, M. A. R. Talukder, P. Malakar, M. M. Kabir, K. Nur, and M. F. Mridha, "Recent advancements and challenges of NLP-based sentiment analysis: A state-of-the-art review," Mar. 01, 2024, *Elsevier Ltd*. doi: 10.1016/j.nlp.2024.100059.

Building of Informatics, Technology and Science (BITS)
Volume 8, No 1, June 2026 Page: 485−496
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.10095

[22] A. Arasy, S. Agustian, L. Handayani, and I. Iskandar, "Klasifikasi Sentimen Menggunakan Metode Multilayer Perceptron dengan Fitur TF-IDF," *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 5, no. 3, pp. 908–919, Jun. 2025, doi: 10.57152/malcom.v5i3.2052.

[23] F. Rafiandi Andhika, W. Witanti, and P. N. Sabrina, "Analisis Sentimen Menggunakan Metode IndoBERT pada Ulasan Aplikasi Zoom Menggunakan Fitur Ekstrasi GloVe," vol. 9, no. 2, pp. 439–448, 2025, doi: 10.47002/metik.v9i2.1098.

[24] F. Firmanda and R. R. Suryono, "Analisis Sentimen Publik Terhadap Danantara di Media Sosial X Menggunakan Naïve Bayes dan Support Vector Machine," *Technology and Science (BITS)*, vol. 7, no. 2, 2025, doi: 10.47065/bits.v7i2.7250.