Indonesian Journal of Multidisciplinary on Social and Technology
Homepage: https://journal.ilmudata.co.id/index.php/ijmst
Vol. 4 No. 3 (2026) pp: 9958-9966
P-ISSN: 2986-6790, e-ISSN: 2986-6782

Analisis Sentimen terhadap Program Makan Bergizi Gratis di Platform X dengan SVM

Arya Ghufron Al Bayhaqy Izhar[1], Haryanto[2], Harun Al Rasyid[3]
[1]Sistem Informasi, Teknik dan Informatika, Universitas Bina Sarana Informatika
[2,3]Universitas Bina Sarana Informatika
[1]aryaghufron76@gmail.com, [2]haryanto.hyt@bsi.ac.id, [3]harun.har@bsi.ac.id

Abstrak

Program Makan Bergizi Gratis (MBG) merupakan program nasional yang memicu beragam tanggapan masyarakat di media sosial, termasuk di platform X. Penelitian ini bertujuan mengklasifikasikan sentimen publik terhadap program MBG di platform X menggunakan algoritma Support Vector Machine (SVM) serta mengevaluasi model yang dihasilkan. Sebanyak 8.820 tweet yang terbit pada Februari 2024 sampai April 2026 dikumpulkan melalui web scraping dengan Tweet Harvest. Teks diproses melalui tahap cleaning, case folding, tokenizing, filtering stopword, dan stemming, kemudian diberi label positif, negatif, atau netral menggunakan leksikon InSet sehingga diperoleh 5.022 tweet negatif, 3.283 positif, dan 515 netral. Fitur diekstraksi dengan TF-IDF dan data dibagi menjadi data latih dan data uji dengan rasio 80:20. Synthetic Minority Over-sampling Technique (SMOTE) diterapkan hanya pada data latih untuk mengatasi ketidakseimbangan kelas sehingga data latih bertambah dari 7.056 menjadi 12.054 tweet, sedangkan data uji tidak diubah, dan model SVM dibangun dengan kernel linear. Evaluasi pada 1.764 tweet data uji menggunakan confusion matrix menghasilkan akurasi 80,95%. Model bekerja baik pada kelas negatif (F1-score 87,59%) dan kelas positif (F1-score 80,09%), tetapi lemah pada kelas netral (recall 23,30%; F1-score 22,64%). Hasil ini menunjukkan bahwa SVM dengan fitur TF-IDF mampu mengklasifikasikan sentimen terhadap program MBG di platform X dengan baik pada kelas dominan, sedangkan pengenalan kelas netral yang minoritas masih perlu ditingkatkan.

Kata kunci: Analisis Sentimen, Program Makan Bergizi Gratis, Platform X, Support Vector Machine, TF-IDF, InSet Lexicon.

Abstract

Free Nutritious Meals (Makan Bergizi Gratis, MBG) is a national program that has drawn diverse public responses on social media, including on platform X. This study aims to classify public sentiment toward the MBG program on platform X using the Support Vector Machine (SVM) algorithm and to evaluate the resulting model. A total of 8,820 tweets posted between February 2024 and April 2026 were collected through web scraping with Tweet Harvest. The texts were preprocessed through cleaning, case folding, tokenizing, stopword filtering, and stemming, then labeled as positive, negative, or neutral using the InSet lexicon, which yielded 5,022 negative, 3,283 positive, and 515 neutral tweets. Features were extracted with TF-IDF, and the data were divided into training and testing sets at an 80:20 ratio. The Synthetic Minority Over-sampling Technique (SMOTE) was applied only to the training data to address class imbalance, increasing them from 7,056 to 12,054 tweets while the test data were left unchanged, and the SVM model was built with a linear kernel. Evaluation on 1,764 test tweets using a confusion matrix produced an accuracy of 80.95%. The model performed well on the negative class (F1-score 87.59%) and the positive class (F1-score 80.09%), but poorly on the neutral class (recall 23.30%; F1-score 22.64%). These results indicate that SVM with TF-IDF features can classify sentiment toward the MBG program on platform X well for the dominant classes, whereas recognition of the neutral minority class still needs improvement.

Keywords: Sentiment Analysis, Free Nutritious Meals Program, Platform X, Support Vector Machine, TF-IDF, SMOTE

1. Pendahuluan

Program Makan Bergizi Gratis (MBG) diimplementasikan sebagai kebijakan strategis berskala nasional dalam rangka mengoptimalkan standar pemenuhan gizi peserta didik serta mendukung pencapaian visi Indonesia Emas 2045 (Agustini, 2025) . Peluncuran program ini diproyeksikan memberikan dampak positif terhadap perbaikan status gizi, kedisiplinan kehadiran, capaian akademik, dan perluasan akses pendidikan (Theoline et al., 2025)

Pemenuhan gizi harian merupakan faktor mendasar yang menunjang kesiapan dan fokus belajar siswa di sekolah (Hasibuan et al., 2025) . Meskipun demikian, kesiapan pelaksanaan dan efektivitas distribusi bantuan di lapangan

masih memicu dinamika serta kritik dari sebagian masyarakat. Respons dan perbedaan pandangan publik mengenai kebijakan ini tercermin secara terbuka di media sosial, yang menjadi indikator awal dalam mengukur persepsi masyarakat (Adib et al., 2024).

Perbedaan tanggapan masyarakat membentuk pola opini publik yang belum terpetakan secara sistematis, di mana informasi kecenderungan sentimen selama ini masih terbatas pada pengamatan umum. Platform X dimanfaatkan secara luas oleh masyarakat untuk menyampaikan dukungan, kritik, maupun saran terhadap kebijakan pemerintah. Karakteristik data opini publik di platform X yang berulang, bervolume besar, dan terus bertambah menuntut adanya pendekatan komputasi berbasis data teks guna memetakan kecenderungan sentimen secara terukur dan terstruktur.

Dalam analisis sentimen berbasis teks, algoritma *Support Vector Machine* (SVM) dikenal memiliki keandalan yang tinggi untuk tugas klasifikasi. Beberapa studi terdahulu telah membuktikan efektivitas pemodelan SVM dalam memetakan persepsi publik lintas domain, seperti analisis evaluasi layanan Indihome(Diana Puspitasari & Tata Sutabri, 2024) , persepsi terhadap figur presiden dan wakil presiden (Ratnaswari et al., 2025) hingga respons masyarakat terhadap isu bencana megathrust di Indonesia (Christanto & Timur Samuel, 2025). Rangkaian penelitian tersebut secara konsisten menunjukkan bahwa integrasi pemrosesan teks (*text preprocessing*), ekstraksi fitur berbasis Term Frequency-Inverse Document Frequency (TF-IDF), serta evaluasi pemodelan menggunakan *confusion matrix* mampu mengklasifikasikan data opini publik ke dalam polaritas sentimen secara akurat (Puspitasari & Sutabri, 2024; Ratnaswari et al., 2024; Christanto & Samuel, 2025).

Meskipun penerapan SVM untuk pemetaan opini publik telah banyak dilakukan, riset komputasi yang secara spesifik membedah klasifikasi sentimen masyarakat di platform X terkait kebijakan Program Makan Bergizi Gratis masih sangat terbatas dalam literatur ilmiah. Padahal, sebagai program prioritas nasional, dinamika tanggapan publik terhadap Program MBG di media sosial bergerak secara masif dan memerlukan evaluasi berbasis data. Kekosongan kajian ini menjadi landasan penting digelarnya penelitian untuk memetakan polaritas sentimen masyarakat secara objektif dan terukur.

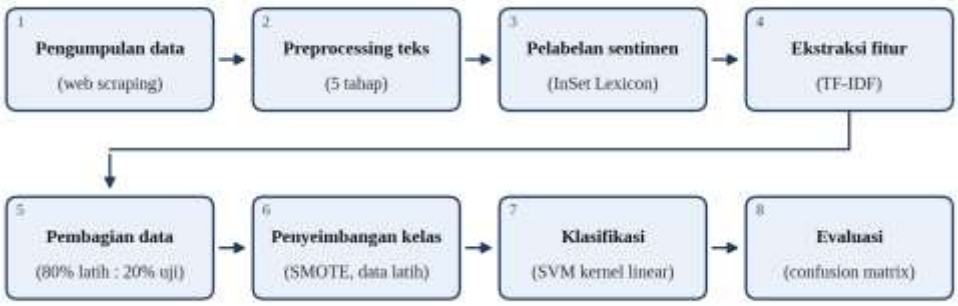
Penelitian ini bertujuan untuk mengumpulkan data opini publik terkait Program MBG di platform X melalui metode *scraping* berbasis kueri spesifik, melakukan pra-pemrosesan teks (*text preprocessing*), mengimplementasikan algoritma *Support Vector Machine* (SVM) untuk mengklasifikasikan sentimen ke dalam kelas positif, negatif, dan netral, serta mengevaluasi performa pemodelan menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score*.

Pendekatan metode yang diusulkan meliputi alur komputasi sistematis: akuisisi data opini melalui *web scraping*, *text preprocessing* (meliputi *cleaning*, *case folding*, *tokenizing*, *filtering*, dan *stemming*), pelabelan sentimen otomatis memanfaatkan InSet *Lexicon*, ekstraksi fitur numerik berbasis pembobotan TF-IDF, penanganan ketidakseimbangan data (*imbalanced data*) menggunakan *Synthetic Minority Over-sampling Technique* (SMOTE), klasifikasi dengan pemodelan SVM, serta evaluasi model melalui *confusion matrix*. Kontribusi dari penelitian ini diharapkan dapat menyajikan gambaran komprehensif mengenai respons publik terhadap Program MBG sekaligus memperkaya literatur pemanfaatan analitik sentimen berbasis pembelajaran mesin dalam mengevaluasi kebijakan publik di ranah digital.

2. Metode Penelitian

Penelitian ini dilaksanakan dalam delapan tahap yang ditunjukkan pada Gambar 1. Data *tweet* dikumpulkan dan diproses menjadi teks bersih, diberi label sentimen dengan leksikon InSet, dan diubah menjadi representasi numerik dengan TF-IDF. Matriks yang dihasilkan dibagi menjadi data latih dan data uji, kelas pada data latih diseimbangkan dengan SMOTE, lalu model SVM dilatih dan dievaluasi pada data uji.

DOI: https://doi.org/10.69693/ijmst.v4i3.14348
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)



Gambar 1. Alur penelitian

2.1. Pengumpulan Data

Data dikumpulkan dari platform X dengan teknik *web scraping* menggunakan bahasa pemrograman Python dan pustaka *Tweet* Harvest, yaitu pustaka yang mengumpulkan *tweet* berdasarkan kata kunci tertentu. Teknik *scraping* dipilih karena memungkinkan pengambilan data teks dalam jumlah besar secara otomatis(Chrisinta & Simarmata, 2024). Kata kunci yang digunakan adalah "Makan Bergizi Gratis", "MBG", dan beberapa kata lain yang berkaitan dengan program tersebut. Periode pengambilan data adalah Februari 2024 sampai April 2026, agar data mencerminkan dinamika opini publik yang berkembang terkait program ini.

Hasil *scraping* tersimpan dalam dataset dengan sejumlah atribut, antara lain identitas pengguna, nama pengguna, *URL* gambar, dan *URL tweet*. Analisis hanya memakai kolom *full_text* karena kolom ini memuat isi opini yang akan diklasifikasikan. Dataset yang ditetapkan sebagai data utama berjumlah 8.820 *tweet*.

2.2. *Text preprocessing*

Preprocessing bertujuan membersihkan *noise* dan menstrukturkan teks agar kualitas analisis meningkat (Rahayu et al., 2023) . Penelitian ini menerapkan lima tahap yang dirangkum pada Tabel 1. Tahap *cleaning* menghapus *URL*, *mention*, tanda pagar, entitas HTML, angka, dan karakter selain huruf serta spasi berlebih. Seluruh teks kemudian diubah menjadi huruf kecil (*case folding*) dan dipisah menjadi token berdasarkan spasi (*tokenizing*).

Tabel 1 Tahapan *text preprocessing*

Tahap	Proses yang dilakukan
Cleaning	Penghapusan *URL*, *mention*, tanda pagar, entitas HTML, angka, karakter selain huruf, dan spasi berlebih
Case folding	Konversi seluruh teks menjadi huruf kecil
Tokenizing	Pemisahan teks menjadi token berdasarkan spasi
Filtering	Penghapusan *stopword* dengan daftar *stopword* Sastrawi; kata negasi tidak, bukan, belum, jangan, dan kurang dipertahankan
Stemming	Pengembalian setiap token ke bentuk kata dasar dengan stemmer Sastrawi

Case folding menyeragamkan huruf agar kata yang sama tidak dihitung sebagai dua entitas hanya karena perbedaan huruf kapital. *Tokenizing* mengubah kalimat menjadi unit kata yang dapat diolah secara mandiri pada tahap berikutnya. Seluruh tahap dijalankan berurutan pada kolom *full_text* sehingga keluaran satu tahap menjadi masukan tahap berikutnya.

Pada tahap *filtering*, kata negasi dikeluarkan dari daftar *stopword* karena kata tersebut membalik makna kata sentimen di sekitarnya. Tahap terakhir, *stemming*, mengembalikan kata berimbuhan ke bentuk dasar sehingga variasi bentuk kata yang bermakna serupa dikelompokkan menjadi satu akar kata.

Tabel 2 menunjukkan keluaran setiap tahap pada satu contoh *tweet*. Keluaran *cleaning* pada contoh ini menghilangkan tanda pagar beserta isinya dan menggabungkan kata bertanda hubung.

DOI: https://doi.org/10.69693/ijmst.v4i3.14348
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

Tabel 2 Contoh keluaran setiap tahap *text preprocessing*

Tahap	Keluaran pada *tweet* contoh
Tweet asli	Program Makan Bergizi Gratis menjadi bukti keseriusan Pemerintah Prabowo-Gibran dalam menuntaskan malnutrisi #MBG #MakanBergiziGratis https://t.co/j0UXLkz59p
Cleaning	Program Makan Bergizi Gratis menjadi bukti keseriusan Pemerintah PrabowoGibran dalam menuntaskan malnutrisi
Case folding	program makan bergizi gratis menjadi bukti keseriusan pemerintah prabowogibran dalam menuntaskan malnutrisi
Tokenizing	[program, makan, bergizi, gratis, menjadi, bukti, keseriusan, pemerintah, prabowogibran, dalam, menuntaskan, malnutrisi]
Filtering	[program, makan, bergizi, gratis, bukti, keseriusan, pemerintah, prabowogibran, menuntaskan, malnutrisi]
Stemming	[program, makan, gizi, gratis, bukti, serius, perintah, prabowogibran, tuntas, malnutrisi]

2.3. Pelabelan Sentimen dengan InSet

Label sentimen ditentukan dengan pendekatan berbasis leksikon menggunakan InSet (Indonesia Sentiment *Lexicon*). Menurut (Nadira et al., 2023), leksikon ini berisi 3.609 kata bersentimen positif dan 6.609 kata bersentimen negatif, tersedia terbuka di GitHub, dan banyak dipakai pada penelitian sentimen berbahasa Indonesia. Pada implementasi penelitian ini, kamus positif dan kamus negatif digabung menjadi satu kamus yang memuat 10.218 kata.

Setiap *tweet* yang telah di*preprocessing* dinilai dengan menjumlahkan bobot kata sentimen yang dikenali dalam kamus. *Tweet* dengan total skor lebih dari nol diberi label positif, kurang dari nol diberi label negatif, dan sama dengan nol diberi label netral. Hasil pelabelan ini menjadi label target pada klasifikasi SVM. Tabel 3 memberi contoh penerapan aturan tersebut pada dua *tweet*.

Tabel 3 Contoh pelabelan *tweet* dengan kamus InSet

Tweet	Kata sentimen (skor)	Total skor	Label
Setuju! Kalau anak kenyang dan makan makanan bergizi gratis bisa dikontrol. Doa gw Indonesia maju, anak² gak mampu tingkat SMP sudah boarding aja free dari Negara.	setuju (+3), maju (+4), kenyang (+2), bergizi (+3)	+12	Positif
Tanpa korupsi, program MBG ini hal kecil gak akan tumbalin anggaran prioritas lain. Di Indonesia bukan programnya yang buruk, tapi budaya korupsi yang parah.	korupsi (−5), buruk (−4), korup (−3), parah (−3)	−15	Negatif

2.4. Ekstraksi Fitur TF-IDF dan Pembagian Data

Ekstraksi fitur mengubah teks menjadi bentuk numerik agar dapat diproses algoritma *machine learning* (Hasanah & Sari, 2024). Teks hasil *stemming* yang telah berlabel diubah menjadi matriks numerik dengan pembobotan *Term Frequency–Inverse Document Frequency* (TF-IDF), yang bobot termnya meningkat jika sering muncul pada suatu dokumen tetapi jarang muncul pada dokumen lain (Wibowo et al., 2024) , dengan rumus:

$$\mathrm{TF}-\mathrm{IDF}(d,t) = f(d,t) \times \log\left(\frac{N}{\mathrm{df}(t)}\right) \qquad (1)$$

dengan f(d,t) adalah frekuensi kemunculan term t pada dokumen d, N adalah jumlah dokumen, dan df(t) adalah jumlah dokumen yang memuat term t . Pembobotan dilakukan dengan kelas *TfidfVectorizer* dari Scikit-learn dengan parameter bawaan terhadap seluruh 8.820 dokumen sebelum data dibagi, sehingga terbentuk matriks berukuran 8.820 dokumen × 12.154 fitur.

Matriks tersebut dibagi menjadi data latih dan data uji dengan rasio 80:20 memakai fungsi *train_test_split* dari Scikit-learn, dengan *random_state*, 42 agar pembagian dapat diulang dan stratifikasi berdasarkan label agar proporsi kelas terjaga. Pembagian ini menghasilkan 7.056 data latih dan 1.764 data uji.

2.5. Penyeimbangan Kelas dengan SMOTE

Pelabelan menghasilkan distribusi kelas yang timpang, dan ketimpangan seperti ini cenderung menurunkan kinerja model karena bias terhadap kelas mayoritas (Pulungan et al., 2024). Untuk mengatasinya diterapkan *Synthetic Minority Over-sampling Technique* (SMOTE), yang membangkitkan sampel sintetis kelas minoritas melalui

DOI: https://doi.org/10.69693/ijmst.v4i3.14348
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

pendekatan tetangga terdekat, bukan sekadar menduplikasi data. Sampel sintetis dibentuk dengan rumus (Amanda Prawita Ningrum et al., 2024) :

$$x_{\text{baru}} = x_i + \lambda(x_{nn} - x_i) \quad (2)$$

dengan x_baru adalah sampel sintetis, x_i adalah sampel kelas minoritas, x_nn adalah tetangga terdekat x_i, dan λ adalah bilangan acak antara 0 dan 1. SMOTE dijalankan dengan pustaka imbalanced-learn (*random_state*, 42) hanya pada data latih. Data uji tidak diseimbangkan agar evaluasi tetap mencerminkan distribusi data yang sebenarnya.

2.6. Klasifikasi SVM dan Evaluasi Model

SVM adalah metode *supervised learning* yang memisahkan kelas dengan membangun *hyperplane* secara optimal (Siti Aisah et al., 2024). Fungsi keputusannya dinyatakan sebagai (I Wayan Budi Suryawan et al., 2023) :

$$f(x) = w \cdot x + b \quad (3)$$

dengan x adalah data masukan, w adalah bobot, dan b adalah bias. Posisi data terhadap *hyperplane*, yaitu tanda f(x), menentukan kelasnya. Data teks berdimensi tinggi, seperti matriks TF-IDF pada penelitian ini, membuat SVM banyak direkomendasikan untuk analisis sentimen, dan fungsi *kernel* memungkinkan pemisahan kelas non-linear dengan memetakan data ke ruang berdimensi lebih tinggi. Model dibangun dengan kelas *SVC* dari Scikit-learn menggunakan *kernel* linear dan *random_state*, 42, sedangkan parameter lain memakai nilai bawaan. Model dilatih pada data latih hasil SMOTE untuk mengklasifikasikan *tweet* ke dalam kelas negatif, netral, dan positif. Konfigurasi eksperimen dirangkum pada Tabel 4.

Tabel 4 Konfigurasi eksperimen

Tahap	Pustaka	Konfigurasi
Pengumpulan data	*Tweet* Harvest	Kata kunci "Makan Bergizi Gratis", "MBG", dan lainnya; Februari 2024 – April 2026
Filtering dan *stemming*	Sastrawi	*Stopword* bawaan tanpa lima kata negasi; stemmer bawaan
Pelabelan	Kamus InSet	Skor total > 0 positif, < 0 negatif, = 0 netral
Ekstraksi fitur	Scikit-learn	*TfidfVectorizer*, parameter bawaan
Pembagian data	Scikit-learn	*train_test_split*: *test_size* 0,2; *random_state*, 42; stratifikasi berdasarkan label
Penyeimbangan kelas	imbalanced-learn	SMOTE, *random_state*, 42, hanya pada data latih
Klasifikasi	Scikit-learn	*SVC*: *kernel* linear; *random_state*, 42

Kinerja model dievaluasi dengan *confusion matrix*, yaitu tabel yang membandingkan hasil prediksi dengan data aktual (Valerian et al., 2025). Dari matriks tersebut dihitung *accuracy*, *precision*, *recall*, dan *F1-score* dengan rumus berikut(Haikal et al., 2024) :

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (4)$$

$$Precision = \frac{TP}{TP+FP} \quad (5)$$

$$Recall = \frac{TP}{TP+FN} \quad (6)$$

$$\text{F1} = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (7)$$

dengan TP adalah data yang diprediksi benar pada kelas yang dievaluasi, TN adalah data yang diprediksi benar pada kelas lainnya, FP adalah data kelas lain yang keliru diprediksi sebagai kelas yang dievaluasi, dan FN adalah data kelas yang dievaluasi yang keliru diprediksi sebagai kelas lain. *Precision*, *recall*, dan *F1-score* dihitung per kelas dari nilai TP, FP, dan FN pada *confusion matrix*, dengan Microsoft Excel, lalu dibandingkan dengan *classification report* Scikit-learn sebagai pemeriksaan silang. *Accuracy* dihitung dari jumlah prediksi benar dibagi seluruh data uji.

DOI: https://doi.org/10.69693/ijmst.v4i3.14348
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

3. Hasil dan Pembahasan

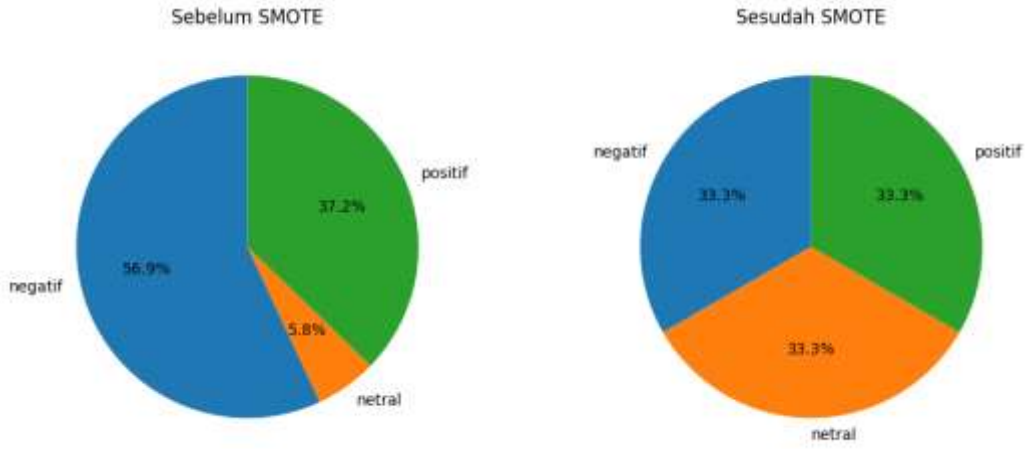
3.1. Distribusi Data dan Hasil Pelabelan

Dari 8.820 *tweet*, pelabelan dengan InSet menghasilkan 5.022 *tweet* negatif (56,94%), 3.283 positif (37,22%), dan 515 netral (5,84%). Distribusi per kelas pada keseluruhan data, data latih, dan data uji disajikan pada Tabel 5. Sentimen negatif merupakan kelas terbesar, sedangkan kelas netral hanya sekitar 6% dari seluruh data. Ketimpangan ini berisiko menurunkan kinerja model pada kelas minoritas.

Tabel 5 Distribusi kelas sentimen pada seluruh data, data latih, dan data uji

Kelas	Seluruh data	Data latih (sebelum SMOTE)	Data latih (sesudah SMOTE)	Data uji
Negatif	5.022	4.018	4.018	1.004
Netral	515	412	4.018	103
Positif	3.283	2.626	4.018	657
Total	8.820	7.056	12.054	1.764

Data latih 7.056 *tweet* (80%) dan data uji 1.764 *tweet* (20%). Persentase kelas pada data latih sebelum SMOTE adalah 56,9% negatif, 37,2% positif, dan 5,8% netral. Setelah SMOTE, jumlah data latih menjadi 12.054 *tweet* dengan proporsi masing-masing kelas 33,3%, seperti terlihat pada Gambar 2, sedangkan data uji tidak berubah.



Gambar 2. Persentase label sebelum dan sesudah SMOTE

3.2. Kinerja Model SVM

Model SVM yang dilatih pada data latih hasil SMOTE menghasilkan *accuracy* sebesar 80,95%, yaitu 1.428 dari 1.764 data uji terklasifikasi dengan benar. Nilai evaluasi per kelas dirangkum pada Tabel 6, sedangkan rincian prediksi ditunjukkan oleh *confusion matrix* pada Gambar 3.

Tabel 6 Hasil evaluasi model SVM per kelas

Kelas	TP	FP	FN	*Precision* (%)	*Recall* (%)	*F1-score* (%)
Negatif	889	137	115	86,65	88,55	87,59
Netral	24	85	79	22,02	23,30	22,64
Positif	515	114	142	81,88	78,39	80,09

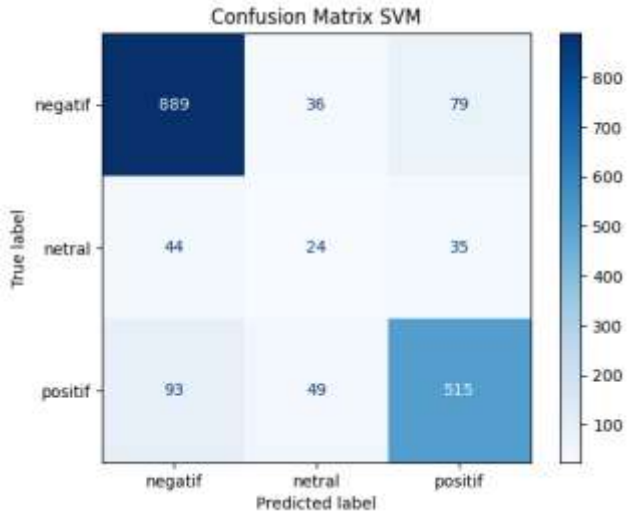
Rata-rata makro *precision*, *recall*, dan *F1-score* pada *classification report* masing-masing 0,64; 0,63; dan 0,63, sedangkan rata-rata tertimbang masing-masing 0,81. Hasil perhitungan Excel dibandingkan dengan *classification report* pada Tabel 7.

DOI: https://doi.org/10.69693/ijmst.v4i3.14348
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

Tabel 7 Perbandingan *classification report* dan perhitungan Excel

Metrik	*Classification report*	Perhitungan Excel
Accuracy	81%	80,95%
Precision negatif	87%	86,65%
Recall negatif	89%	88,55%
F1-score negatif	88%	87,59%
Precision netral	22%	22,02%
Recall netral	23%	23,30%
F1-score netral	23%	22,64%
Precision positif	82%	81,88%
Recall positif	78%	78,39%
F1-score positif	80%	80,09%

Kedua sumber konsisten. Selisih yang tampak hanya berasal dari pembulatan pada *classification report*, sedangkan perhitungan Excel menampilkan persentase lebih rinci; misalnya *accuracy* 80,95% tampil sebagai 0,81.



Gambar 3. *Confusion matrix* model SVM

3.3. Analisis Kesalahan Klasifikasi

Kinerja model sangat berbeda antarkelas. Kelas negatif, yang terbesar, memperoleh *F1-score* 87,59%, dan kelas positif 80,09%. Pada kelas netral, model hanya mengenali 24 dari 103 *tweet* aktual (*recall* 23,30%), sedangkan 44 *tweet* diprediksi negatif dan 35 positif. *Precision* kelas netral juga rendah (22,02%) karena 85 *tweet* dari kelas lain, yaitu 36 negatif dan 49 positif, keliru diprediksi netral. Kesalahan terbanyak berikutnya terjadi antara kelas negatif dan positif: 79 *tweet* negatif diprediksi positif dan 93 *tweet* positif diprediksi negatif.

Sebaran prediksi pada setiap kelas aktual dirangkum pada Tabel 8. Sebanyak 76,70% *tweet* netral aktual salah diklasifikasikan, terutama sebagai negatif (42,72%), sedangkan 88,55% *tweet* negatif dan 78,39% *tweet* positif diklasifikasikan dengan benar. *Tweet* negatif yang keliru lebih sering diprediksi positif (7,87%) daripada netral (3,59%), sementara *tweet* positif yang keliru lebih sering diprediksi negatif (14,16%) daripada netral (7,46%).

Tabel 8 Sebaran prediksi model pada setiap kelas aktual, jumlah *tweet* (persentase dari kelas aktual)

Kelas aktual	Prediksi negatif	Prediksi netral	Prediksi positif
Negatif (1.004)	889 (88,55%)	36 (3,59%)	79 (7,87%)
Netral (103)	44 (42,72%)	24 (23,30%)	35 (33,98%)
Positif (657)	93 (14,16%)	49 (7,46%)	515 (78,39%)

Rendahnya kinerja pada kelas netral konsisten dengan ketimpangan kelas yang dilaporkan pada Subbab 3.1. Kelas ini hanya 5,84% dari seluruh data, dan SMOTE yang menaikkannya menjadi 4.018 sampel latih bekerja dengan sampel sintetis dari 412 *tweet* asli, sehingga belum cukup untuk mengangkat pengenalan kelas netral. Dengan

DOI: https://doi.org/10.69693/ijmst.v4i3.14348
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

demikian, *accuracy* 80,95% terutama ditopang oleh kelas negatif dan positif, dan hasil ini tidak boleh dibaca sebagai kemampuan yang merata pada ketiga kelas.

3.4. Implikasi dan Keterbatasan

Hasil ini perlu ditafsirkan dengan mempertimbangkan cara pelabelan. Label berasal dari leksikon InSet, bukan dari anotasi manual, sehingga evaluasi mengukur kemampuan model mereproduksi pelabelan berbasis leksikon. Dominasi sentimen negatif pada Subbab 3.1 pun merupakan hasil pelabelan tersebut. Penelitian ini juga belum melakukan optimasi parameter SVM dan belum membandingkan SVM dengan algoritma lain.

Meski begitu, hasilnya menunjukkan bahwa kombinasi TF-IDF, SMOTE pada data latih, dan SVM *kernel* linear dapat memetakan kecenderungan sentimen publik terhadap kebijakan MBG di platform X pada kelas dominan. Pemetaan semacam ini berpotensi menjadi masukan awal bagi pemantauan opini publik dan evaluasi kebijakan, dengan catatan bahwa pengenalan sentimen netral perlu diperbaiki sebelum model dipakai untuk menggambarkan seluruh spektrum opini.

3.5. Temuan terhadap Pertanyaan Penelitian

Terhadap tujuan pertama dan kedua, penelitian ini menunjukkan bahwa *tweet* berbahasa Indonesia tentang MBG dapat dikumpulkan secara otomatis dalam jumlah besar, yaitu 8.820 *tweet* untuk periode Februari 2024 sampai April 2026, dan disiapkan melalui lima tahap *preprocessing* sehingga siap diberi label dan diubah menjadi fitur TF-IDF berukuran 12.154 fitur.

Terhadap pertanyaan ketiga dan keempat, SVM dengan *kernel* linear yang dilatih pada data latih hasil SMOTE menghasilkan *accuracy* 80,95%, dengan *F1-score* 87,59% pada kelas negatif, 80,09% pada kelas positif, dan 22,64% pada kelas netral. Dengan demikian model sudah layak untuk membedakan sentimen negatif dan positif, tetapi belum dapat diandalkan untuk mengenali sentimen netral pada data dengan ketimpangan kelas seperti ini.

4. Kesimpulan

Penelitian ini mengklasifikasikan sentimen publik terhadap program Makan Bergizi Gratis di platform X menggunakan SVM. Sebanyak 8.820 *tweet* periode Februari 2024 sampai April 2026 berhasil dikumpulkan dengan *web scraping*, diproses melalui *cleaning*, *case folding*, *tokenizing*, *filtering*, dan *stemming*, lalu diberi label dengan InSet menjadi 5.022 *tweet* negatif, 3.283 positif, dan 515 netral. Dominasi sentimen negatif pada hasil pelabelan ini menjadi temuan utama tentang opini publik dalam data yang dianalisis. Model SVM dengan fitur TF-IDF, pembagian data 80:20, dan SMOTE pada data latih mencapai *accuracy* 80,95% pada data uji. Kinerjanya baik pada kelas negatif (*F1-score* 87,59%) dan positif (80,09%), tetapi lemah pada kelas netral (*recall* 23,30%; *F1-score* 22,64%). Penelitian selanjutnya disarankan membandingkan SVM dengan algoritma lain seperti Naïve Bayes, Random Forest, LSTM, dan BERT, menerapkan optimasi parameter atau teknik *oversampling* lain untuk meningkatkan pengenalan kelas netral, serta menambahkan topic modeling dan analisis tren sentimen untuk mengungkap topik utama di balik opini publik. Penelitian ini juga menegaskan bahwa data media sosial dapat menjadi dasar pemantauan sentimen publik sebagai masukan bagi evaluasi kebijakan.

Reference

1. Adib, K., Handayani, M. R., Yuniarti, W. D., & Umam, K. (2024). Opini Publik Pasca-Pemilihan Presiden: Eksplorasi Analisis Sentimen Media Sosial X Menggunakan SVM. *SINTECH (Science and Information Technology) Journal*, *7*(2), 80–91. https://doi.org/10.31598/sintechjournal.v7i2.1581
2. Agustini, U. (2025). Efektivitas dan Tantangan Kebijakan Program Makan Bergizi Gratis sebagai Intervensi Pendidikan di Indonesia. *Jurnal Kiprah Pendidikan*, *4*(3), 362–368. https://doi.org/10.33578/kpd.v4i3.p362-368
3. Amanda Prawita Ningrum, Sri Winarno, & Vincentius Praskatama. (2024). Klasifikasi Kualitas Biji Kedelai Menggunakan Transfer Learning Convolutional Neural Network Dan SMOTE. *Journal of Applied Computer Science and Technology*, *5*(2), 155–164. https://doi.org/10.52158/jacost.v5i2.1002
4. Chrisinta, D., & Simarmata, J. E. (2024). Eksplorasi Teknik Web Scraping pada Data Mining: Pendekatan Pencarian Data Berbasis Python. *Faktor Exacta*, *17*(1). https://doi.org/10.30998/faktorexacta.v17i1.22393
5. Christanto, A., & Timur Samuel, Y. (2025). Analisis Sentimen Isu Megathrust Indonesia Di Twitter Menggunakan Support Vector Machine Dan Naive Bayes. *Jurnal Ilmiah Sistem Informasi (JISI)*, *4*(1), 1–10. https://doi.org/10.24127/jisi.v4i1.8453
6. Diana Puspitasari, & Tata Sutabri. (2024). Analisis Sentimen Berdasarkan pada Twitter (X) terhadap Layanan Indihome Menggunakan Algoritma Support Vector Machine (SVM). *JUMINTAL: Jurnal Manajemen Informatika Dan Bisnis Digital*, *3*(2), 58–71. https://doi.org/10.55123/jumintal.v3i2.4449

DOI: https://doi.org/10.69693/ijmst.v4i3.14348
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)

7. Haikal, M., Martanto, M., & Hayati, U. (2024). ANALISIS SENTIMEN TERHADAP PENGGUNAAN APLIKASI GAME ONLINE PUBG MOBILE MENGGUNAKAN ALGORITMA NAIVE BAYES. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *7*(6), 3275–3281. https://doi.org/10.36040/jati.v7i6.8174
8. Hasanah, A. N., & Sari, B. N. (2024). ANALISIS SENTIMEN ULASAN PENGGUNA APLIKASI JASA OJEK ONLINE MAXIM PADA GOOGLE PLAY DENGAN METODE NAÏVE BAYES CLASSIFIER. *Jurnal Informatika Dan Teknik Elektro Terapan*, *12*(1). https://doi.org/10.23960/jitet.v12i1.3628
9. Hasibuan, N. H., Armanda, T. B., Aulya, F., Fauzi, M. I., & Ivanna, J. (2025). Analisis Kebijakan Program Makan Bergizi Gratis (MBG) terhadap Peningkatan Motivasi dan Konsentrasi Belajar Siswa di SMK Negeri 6 Medan. *Jurnal Pendidikan Tambusai*, *9*(1), 10611–10615. https://doi.org/10.31004/jptam.v9i1.26343
10. I Wayan Budi Suryawan, Nengah Widya Utami, & Ketut Queena Fredlina. (2023). ANALISIS SENTIMEN REVIEW WISATAWAN PADA OBJEK WISATA UBUD MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE. *Jurnal Informatika Teknologi Dan Sains*, *5*(1), 133–140. https://doi.org/10.51401/jinteks.v5i1.2242
11. Nadira, A., Setiawan, N. Y., & Purnomo, W. (2023). ANALISIS SENTIMEN PADA ULASAN APLIKASI MOBILE BANKING MENGGUNAKAN METODE NAÏVE BAYES DENGAN KAMUS INSET. *INDEXIA*, *5*(01), 35. https://doi.org/10.30587/indexia.v5i01.5138
12. Pulungan, M. P., Purnomo, A., & Kurniasih, A. (2024). Penerapan SMOTE untuk Mengatasi Imbalance Class dalam Klasifikasi Kepribadian MBTI Menggunakan Naive Bayes Classifier. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, *11*(5), 1033–1042. https://doi.org/10.25126/jtiik.2024117989
13. Rahayu, K., Fitria, V., Septhya, D., Rahmaddeni, R., & Efrizoni, L. (2023). Klasifikasi Teks untuk Mendeteksi Depresi dan Kecemasan pada Pengguna Twitter Berbasis Machine Learning. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, *3*(2), 108–114. https://doi.org/10.57152/malcom.v3i2.780
14. Ratnaswari, S., Wibowo, N. C., & Kartika, D. S. Y. (2025). ANALISIS SENTIMEN MENGGUNAKAN METODE LEXICON-BASED DAN SUPPORT VECTOR MACHINE PADA PRESIDEN DAN WAKIL PRESIDEN INDONESIA PERIODE 2024–2029. *Jurnal Informatika Dan Teknik Elektro Terapan*, *13*(1). https://doi.org/10.23960/jitet.v13i1.5604
15. Siti Aisah, I., Irawan, B., & Suprapti, T. (2024). ALGORITMA SUPPORT VECTOR MACHINE (SVM) UNTUK ANALISIS SENTIMEN ULASAN APLIKASI AL QUR'AN DIGITAL. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *7*(6), 3759–3765. https://doi.org/10.36040/jati.v7i6.8263
16. Theoline, E., Santosa, H., & Susanto, T. T. D. (2025). Manajemen Kolaboratif Program Makan Bergizi Gratis Sebagai Upaya Pencegahan Stunting di Sekolah Dasar: Tinjauan Literatur Review. *Prosiding Fakultas Ilmu Pendidikan Universitas Negeri Jakarta*, *3*(1), 506–514. https://journal.unj.ac.id/unj/index.php/semnas-ps/article/view/63699
17. Valerian, F., Syarief, M., & Abdul Fatah, D. (2025). KLASIFIKASI TINGKAT OBESITAS MENGGUNAKAN METODE GBM DAN CONFUSION MATRIX. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *9*(2), 2242–2249. https://doi.org/10.36040/jati.v9i2.13062
18. Wibowo, I. S., Witanti, A., & Susilawati, I. (2024). Keyword Extraction Judul Berita Online Di Indonesia Menggunakan Metode TF-IDF. *JATISI*, *11*(1). https://doi.org/https://doi.org/10.35957/jatisi.v11i1.6718

DOI: https://doi.org/10.69693/ijmst.v4i3.14348
Lisensi: Creative Commons Attribution 4.0 International (CC BY 4.0)