JURNAL JTIK
(JURNAL TEKNOLOGI INFORMASI DAN KOMUNIKASI)
KITA Komunitas Informasi Teknologi Aceh
OPEN ACCESS
E-ISSN: 2580-1643
Volume 11, Nomor 1, Januari 2027 | Pages: 54-63 | DOI: https://doi.org/10.35870/jtik.v11i1.7407

Analisis Sentimen Masyarakat terhadap Kebijakan Program Makan Bergizi Gratis (MBG) di X Menggunakan Metode *Support Vector Machine* (SVM)

Mohammad Farhan Fatah [1*], A. Hamdani [2], Farihin Lazim [3]

[1*,2] Program Studi Teknologi Informasi, Fakultas Sains dan Teknologi, Universitas Ibrahimy, Kota Situbondo, Provinsi Jawa Timur, Indonesia.
[3] Program Studi Ilmu Komputer, Fakultas Sains dan Teknologi, Universitas Ibrahimy, Kota Kota Situbondo, Provinsi Jawa Timur, Indonesia.

Corresponding Email: farhanfatah097@gmail.com [1*]

Article History: Received: 4 June 2026, Revision: 8 July 2026, Accepted: 14 July 2026, Available Online: 1 January 2027.

Abstract

The Free Nutritious Meal (MBG) policy has elicited diverse responses on social media platform X, necessitating sentiment analysis to objectively understand public opinion. This study employs Support Vector Machine (SVM) with TF-IDF weighting to classify 3,796 opinion data into positive, negative, and neutral sentiments. Results show the SVM model achieves an accuracy of 76.05%. Public perception is predominantly positive (1,431 data), followed by neutral (1,258), and negative (1,107). Word Cloud analysis indicates optimism based on hopes for improved student nutrition, while negative opinions highlight concerns about food safety and budget transparency. In conclusion, MBG enjoys strong public support, but the government should prioritize food safety and transparent financial management to reduce public unrest. Identifying negative sentiments provides a basis for critical evaluation and future policy improvements.

Keyword: Sentiment Analysis; Free Nutritious Meal; Support Vector Machine; Twitter; TF-IDF.

Abstrak

Kebijakan Program Makan Bergizi Gratis (MBG) memunculkan beragam respons di media sosial X, sehingga analisis sentimen diperlukan untuk memahami opini publik secara objektif. Penelitian ini menggunakan algoritma Support Vector Machine (SVM) dengan pembobotan TF-IDF untuk mengklasifikasikan 3.796 data opini menjadi sentimen positif, negatif, dan netral. Hasil pengujian menunjukkan akurasi sebesar 76,05%. Persepsi masyarakat didominasi oleh sentimen positif (1.431 data), diikuti netral (1.258), dan negatif (1.107). Analisis Word Cloud menunjukkan bahwa optimisme didasarkan pada harapan peningkatan gizi pelajar, sementara opini negatif menyoroti isu keamanan pangan dan transparansi anggaran. Kesimpulannya, MBG mendapatkan dukungan kuat, namun pemerintah perlu fokus pada keamanan dan kejelasan pengelolaan dana untuk mengurangi keresahan masyarakat.

Kata Kunci: Analisis Sentimen; Makan Bergizi Gratis; Support Vector Machine; Twitter; TF-IDF.

E-ISSN: 2580-1643

1. Pendahuluan

Perkembangan teknologi informasi telah mentransformasi platform media sosial sebagai wadah diskursus digital yang dinamis. Hal ini secara signifikan memegang peranan krusial dalam mengarahkan maupun memengaruhi persepsi khalayak terkait berbagai isu publik. (Husada & Paramita, 2021). X (*Twitter*) merupakan salah satu platform yang banyak digunakan untuk menyampaikan gagasan, pengalaman, dan tanggapan terhadap berbagai peristiwa melalui unggahan singkat yang bersifat *real-time*. Untuk menganalisis bagaimana masyarakat memandang suatu kebijakan pemerintah, platform X menawarkan sumber data yang sangat kaya akibat besarnya aktivitas harian penggunanya. Representasi nyata dari kebijakan yang kini menyita atensi publik secara masif di platform tersebut adalah implementasi Program Makan Bergizi Gratis (MBG) merupakan sebuah inisiatif penyediaan pangan sehat bagi kelompok spesifik yang diusung untuk mendongkrak mutu gizi siswa. Secara lebih luas, program ini menjadi instrumen penting dalam memperkuat fondasi pembangunan sumber daya manusia nasional (Prastiawan & Yuniarto, 2025). Sejak implementasinya, program tersebut memunculkan beragam respons di media sosial. Sebagian masyarakat memberikan dukungan karena dinilai mampu membantu pemenuhan kebutuhan gizi anak dan mengurangi risiko stunting, sedangkan sebagian lainnya mempertanyakan efektivitas pelaksanaan, kesiapan infrastruktur, mekanisme distribusi, serta keberlanjutan program dalam jangka panjang (Nugroho *et al.*, 2025).

Perbedaan pandangan tersebut menunjukkan pentingnya memahami kecenderungan opini publik terhadap kebijakan MBG secara lebih objektif dan terukur. Banyaknya opini yang diunggah pengguna media sosial menghasilkan data teks dalam jumlah besar dan tidak terstruktur sehingga sulit dianalisis secara konvensional. Untuk mengatasi permasalahan tersebut, analisis sentimen digunakan sebagai salah satu teknik dalam *text mining* yang bertujuan mengidentifikasi dan mengklasifikasikan opini ke dalam kategori *positif*, *negatif*, dan netral (Ansori & Holle, 2022). Meskipun penelitian terdahulu telah memetakan opini publik terhadap program gizi nasional, Sebagian besar literatur mengabaikan tingginya ambiguitas semantik pada platform X, sehingga diperlukan pendekatan klasifikasi berdasarkan margin optimal yang mampu memisahkan polaritas sentimen secara presisi di tengah riuhnya bias opini politik. Algoritma populer seperti *Naïve Bayes* terbukti memiliki kelemahan saat menangani data ulasan publik, di mana akurasinya hanya mampu mencapai 68% (Indarwati & Februariyanti, 2023). Selain itu, metode *Random Forest* (akurasi 93%) juga memiliki performa di bawah *Support Vector Machine* (SVM) yang unggul dengan akurasi mencapai 97% (Sidauruk *et al.*, 2023). Oleh karena itu, Penelitian ini memilih algoritma *Support Vector Machine* (SVM) dengan fungsi *Kernel Radial Basis Function* (RBF) yang terbukti sangat tangguh dalam mencari *hyperplane* optimal untuk memisahkan polaritas sentimen, serta tahan terhadap *noise* data teks berdimensi tinggi hasil pembobotan TF-IDF (Styawati *et al.*, 2021). Keunggulan metode SVM telah dibuktikan dalam berbagai penelitian terdahulu. Penelitian yang dilakukan oleh Vina Fitriyana dkk. membuktikan kapabilitas metode SVM dalam mengklasifikasikan sentimen pengguna aplikasi Jamsostek Mobile, di mana model tersebut sukses memberikan *accuracy* sebesar 96%, *precision* 92%, *recall* 96%, dan F1-*score* 94% (Fitriyana *et al.*, 2023). Hasil serupa juga dilaporkan oleh Rani Yunita dan Mia Kamayani pada analisis sentimen kebijakan penghapusan kewajiban skripsi di *Twitter*, yang memperoleh *accuracy* sebesar 80%, *recall* 83%, *precision* 76%, dan F1-*score* 79%, serta menunjukkan performa yang lebih baik dibandingkan metode *Naïve Bayes* (Yunita & Kamayani, 2023). Meskipun berbagai penelitian telah berhasil menerapkan SVM pada analisis sentimen di media sosial, mengingat terbatasnya jumlah studi terdahulu yang memfokuskan analisis polaritas sentimen terhadap kebijakan MBG menggunakan data X terkini, urgensi riset di ranah ini semakin meningkat. Kajian komprehensif diperlukan untuk menangkap dan menganalisis secara akurat esensi persepsi masyarakat terhadap berjalannya program pemerintah tersebut.

Berdasarkan latar belakang tersebut, rumusan masalah dalam penelitian ini adalah bagaimana penerapan metode *Support Vector Machine* (SVM) dalam menganalisis sentimen opini publik terhadap kebijakan Program Makan Bergizi Gratis (MBG) di X (*Twitter*) sehingga dapat memberikan gambaran persepsi masyarakat? Untuk menjawab permasalahan tersebut, penelitian ini bertujuan untuk menganalisis dan mengklasifikasikan opini publik ke dalam kategori positif, negatif, dan netral menggunakan metode SVM. Penelitian ini juga dirancang untuk mengetahui kecenderungan sentimen dominan Masyarakat dan mengevaluasi kinerja model SVM menggunakan metrik akurasi, presisi, recall dan *Confusion Matrix*. Hasil penelitian diharapkan dapat memberikan informasi objektif mengenai persepsi masyarakat yang dapat dijadikan bahan evaluasi bagi pemerintah dalam meningkatkan efektivitas pelaksanaan program.

E-ISSN: 2580-1643

2. Metode Penelitian

Jenis Penelitian

Studi ini dirancang sebagai penelitian kuantitatif deskriptif melalui pendekatan *text mining* dan *machine learning*. Fokus utamanya adalah mengestraksi serta mengklasifikasikan respons opini publik di platform X (*Twitter*) mengenai kebijakan Program Makan Bergizi Gratis (MBG). Dalam membangun model klasifikasi sentimen ini, algoritma SVM diterapkan sebagai metode utama karena kapabilitasnya yang dinilai sangat efektif dalam menangani karakteristik data teks tidak terstruktur (Dewi *et al*., 2025).

Teknik Pengumpulan Data

Teknik pengumpulan data pada penelitian ini ditujukan untuk menghimpun data empiris yang dibutuhkan dalam memetakan polarisasi sentimen publik terkait program Makan Bergizi Gratis (MBG) (Maldini & Andryana, 2025). Proses ini dilaksanakan melalui dua pendekatan utama:

1) Pengumpulan Data Sekunder
   Data sekunder dihimpun dari media sosial X (*Twitter*) berupa aktivitas *tweet* pengguna yang berkaitan dengan topik Program Makan Bergizi Gratis. Pengambilan data dilakukan melalui metode *crawling* dengan memanfaatkan kata kunci serta tagar yang representatif. Dataset mentah yang diperoleh kemudian melewati fase *processing* untuk membuang *tweet* yang tidak relevan, membersihkan data ganda, dan menghapus spam demi menjaga validitas hasil analisis.
2) Studi Literatur
   Studi literatur dijalankan dengan mengeksplorasi dan menganalisis berbagai sumber pustaka seperti buku, jurnal ilmiah, serta penelitian sejenis sebelumnya. Pembahasan ditekankan pada konsep analisis sentimen, penerapan model *Support Vector Machine* (SVM), dan dokumentasi kebijakan Program Makan Bergizi Gratis (MBG). Kajian pustaka ini diintegrasikan sebagai landasan teori utama dalam mengeksekusi dan menyusun struktur argumentasi dalam penelitian.

Crawling Data

Pada tahapan ini, data sekunder berupa *tweet* seputar kebijakan Program Makan Bergizi Gratis (MBG) diekstraksi dari *platform* X. Proses penambangan data dijalankan menggunakan *tools crawling* dengan filter kata kunci serta tagar yang relevan. *Output* dari fase ini adalah sebuah dataset awal yang siap digunakan untuk masuk ke tahapan data *preparation* (Putri *et al*., 2022).

Labeling Data

Dataset yang telah dihimpun melalui proses kurasi awal disaring guna mengeliminasi tweet yang tidak sesuai topik, data ganda (duplikat), serta unsur spam. Setelah melewati fase pembersihan, data tersebut dikategorikan ke dalam tiga kelas sentimen, yaitu positif, negatif, dan netral. Pemberian label dilakukan secara manual menggunakan panduan anotasi ketat guna mencegah bias subjektivitas, di mana sentimen positif dukungan atau harapan terhadap keberhasilan program, negatif jika memuat kritik maupun kekhawatiran terkait transparansi anggaran dan keamanan logistik, serta netral untuk teks informatif atau kutipan portal berita tanpa tendensi emosional. Selanjutnya, dataset siap diproses menggunakan *script Python* untuk tahapan ekstraksi fitur (Harahap & Kurniawan, 2024).

Processing Data

Fase *preprocessing* data dilakukan dengan tujuan untuk membersihkan serta mempersiapkan data teks, sehingga data tersebut layak dan siap diolah lebih lanjut pada tahap klasifikasi. (Muhammadin & Sobari, 2021). Tahapan yang dilakukan meliputi:

1) *Case Folding*: Menyeragamkan seluruh huruf dalam teks menjadi huruf kecil (*lowercase*) guna mengeliminasi redundansi kata yang dipicu oleh variasi kapitalisasi huruf kapital.
2) *Cleaning*: Menyingkirkan komponen teks yang tidak bernilai informatif seperti tautan URL, penanda akun (*@username*), tanda baca, simbol khusus, angka, dan emoji menggunakan Regular *Expression* (Regex).
3) Normalisasi: Mentransformasikan kata-kata tidak baku, akronim, bahasa gaul (*slang*), atau kesalahan ketik (*typo*) ke dalam bentuk kamus formal berdasarkan korpus leksikon bahasa Indonesia (contoh: “udh” →” sudah”, “gue” → ”saya”).
4) *Tokenizing*: Memecah rangkaian kalimat utuh atau dokumen teks menjadi satuan kata tunggal (token) yang terpisah agar dapat diidentifikasi bobotnya.
5) *Filtering*: Mengeliminasi kata-kata umum yang frekuensinya tinggi namun tidak memiliki bobot semantik atau makna kontekstual yang signifikan (seperti kata hubung atau kata depan "yang", "di", "dari").
6) *Stemming*: Mengonversi kata yang memiliki afiks (imbuhan awal, sisipan, atau akhiran) kembali ke bentuk kata dasarnya dengan bantuan algoritma pemrosesan bahasa morfemis (contoh: “makanan” → ”makan”, “bergizi” →gizi).

Pembagian Dataset

Setelah melalui tahapan pra-pemrosesan, dataset dipisahkan menjadi data latih (training data) dan data uji (testing data) sebelum masuk ke proses

E-ISSN: 2580-1643

ekstraksi fitur TF-IDF. Pembagian ini dilakukan menggunakan fungsi train_test_split dari library Scikit-Learn dengan proporsi rasio 80:20, di mana 3.036 data (80%) dialokasikan untuk pelatihan dan 760 data (20%) untuk pengujian. Guna menjamin konsistensi urutan pembagian data saat sistem dijalankan ulang, parameter random_state=50 diterapkan. Untuk menangani masalah data yang tidak seimbang (*imbalanced data*), distribusi pada saat data latih telah dipastikan memiliki komposisi yang cukup berimbang sejak proses pelabelan manual (yaitu 1.431 sentimen positif, 1.258 netral, dan 1.107 negatif). Keseimbangan kelas ini secara otomatis memastikan bahwa model latih tidak mengalami bias terhadap satu kelas mayoritas tertentu saat proses pelatihan. (Supian *et al*, 2024).

Ekstraksi Fitur (TF-IDF)

Pascapemrosesan, proses konversi teks menjadi vektor numerik dilakukan menggunakan metode TF-IDF. Parameter ini mengukur bobot pentingnya sebuah kata dengan mengkombinasikan frekuensi lokal pada dokumen dan metrik kelangkaan global di dalam korpus. (Arifin *et al*., 2021). Bobot TF-IDF dihitung menggunakan rumus berikut:

1) *Term Frequency* Adalah frekuesi atau banyaknya kemunculan kata pada dokumen.

$$TF\ (t,d) = f_{t,d} \quad (1)$$

2) *Inverse Document Frequency* yaitu jumlah kemunculan *term* pada keseluruhan dokumen.

$$IDF(t) = \ln\left(\frac{N}{df(t)}\right) + 1 \quad (2)$$

3) Menghitung bobot dari setiap *term* pada dokumen.

$$W\ (t,d)\ =\ TF\ (t,d)\ x\ IDF\ (t,d) \quad (3)$$

***Testing* & Klasifikasi Sentimen**

Proses klasifikasi menggunakan algoritma Support Vector Machine (SVM) yang dioptimasi secara spesifik untuk menangani data teks berdimensi tinggi. Optimasi parameter pada model ini dikonfigurasi menggunakan fungsi *Kernel Radial Basis Function* (RBF), dengan nilai parameter kompleksitas atau *Cost* (C) bernilai 1 untuk mengatur keseimbangan antar margin dan toleransi kesalahan, serta nilai *Gamma* bernilai 1 untuk mengidentifikasikan jangkauan radius pengaruh dari sampel latih. Untuk mengakomodasi tiga kelas sentimen, diterapkan strategi *One-vs-One* (OvO) yang membangun beberapa model biner terpisah, di mana keputusan akhir ditentukan melalui pemungutan suara terbanyak (*majority voting*) (Amrozi *et al*., 2022). Berdasarkan konsep dasar *hyperplane*, fungsi keputusan untuk memisahkan data dirumuskan sebagai berikut:

1) Konsep Dasar Hyperlane

$$f(x) = w^T \emptyset(x) + b \quad (4)$$

2) Pendektan Fungsi Kernel

$$f(x) = \sum_{i=1}^{N} \alpha_i\ yiK(x_i, x) + b \quad (5)$$

Evaluasi Model

Validasi kinerja algoritma pada himpunan data uji dievaluasi melalui instrumen confusion matrix. Penggunaan matriks ini difungsikan untuk menganalisis komparasi secara mendetail antara klasifikasi riil (*ground truth*) dengan luaran prediktif yang dihasilkan oleh model (*predicted labels*). Diagonal utama matriks merepresentasikan jumlah klasifikasi yang tepat, sedangkan elemen di luar diagonal menunjukkan akumulasi kesalahan prediksi (Nurhidayat & Dewi, 2023). Untuk mengukur kinerja pada data multikelas tersebut, nilai *True Positive* (TP), *False Positive* (FP), *False Negative* (FN), dan *True Negative* (TN) dari masing-masing kelas. Tingkat keberhasilan model kemudian dihitung secara kuantitatif menggunakan empat *matrix* evaluasi standar berikut:

1) Akurasi (*Accuracy*)

$$Akurasi = \frac{TP + TN}{(TP + TN + FP + FN)} X100\% \quad (6)$$

2) Presisi (*Precission*)

$$Precission = \frac{TP}{TP + FP} \quad (7)$$

3) *Recall*

$$Recall = \frac{TP}{TP + FN} \quad (8)$$

4) F1-*Score*

$$F1\ = 2x \frac{Precission\ x\ recal}{Precission\ x\ recal} \quad (9)$$

3. Hasil dan Pembahasan

3.1 Hasil

Hasil Pengumpulan dan Palabelan Data

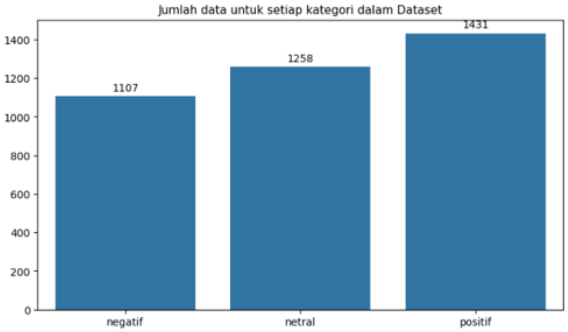
Sumber data primer dalam penelitian ini diperoleh melalui ekstraksi opini publik dari *platform* media sosial X (*Twitter*), yang secara spesifik membahas mengenai implementasi Program kebijakan MBG. Proses penambangan data (*crawling*) dilakukan menggunakan *tools* ekstraksi *tweet-harvest* yang dijalankan pada *platform Google Colab*. Dengan menggunakan kata kunci "makan bergizi gratis", sistem berhasil mengumpulkan data mentah sebanyak 3.796 baris *tweet*.

	created_at	text	user_id_str	label
0	Thu May 01 23:26:30 +0000 2025	@akhdiyatmodjo96 @riseris_risriis @Reborn_Sile...	6.312648e+08	0
1	Thu May 01 23:12:38 +0000 2025	@txtdrimedia Memang paling cocok tuh MSG (Maka...	1.610000e+18	-1
2	Thu May 01 23:11:40 +0000 2025	mbg : makan bergizi gratis makan beracun gratis	1.820000e+18	-1
3	Thu May 01 21:37:41 +0000 2025	Legislator Nilai Program Makan Bergizi Gratis ...	1.280000e+18	1
4	Thu May 01 21:36:32 +0000 2025	Legislator Nilai Program Makan Bergizi Gratis ...	1.280000e+18	1
...	...	...	...	...
3791	Sat Apr 25 21:14:21 +0000 2026	Kedua program itu penting tapi prioritas bisa ...	2.048149e+18	0
3792	Sat Apr 25 21:17:43 +0000 2026	Sumpah anjengg knp sih kalian berdua nakal nga...	2.048149e+18	-1
3793	Sat Apr 25 21:18:50 +0000 2026	Video itu tampaknya edit meme dari pidato Rach...	2.048150e+18	0
3794	Sat Apr 25 21:19:06 +0000 2026	Mau sampai gimana lg ini mbg di pertahanin	2.048150e+18	-1
3795	Sat Apr 25 21:20:21 +0000 2026	jingg stress bgt dah negara apaan jir udah mh ...	2.048150e+18	-1

Gambar 1. Dataset Hasil Pelabelan

E-ISSN: 2580-1643

Selanjutnya, dataset tersebut melalui proses pelabelan secara konvensional untuk mengklasifikasikan opini ke dalam tiga kategori sentimen. Pendekatan konvensional dipilih untuk menangkap konteks dan makna sebenarnya dari bahasa tidak baku atau sarkasme sebelum diproses oleh mesin. Hasil pelabelan menghasilkan distribusi yang seimbang, yaitu 1.431 data untuk sentimen *positif*, 1.258 data untuk sentimen netral, dan 1.107 data untuk sentimen *negatif*. Keseimbangan kelas ini krusial agar model *machine learning* tidak memihak pada kelas mayoritas saat proses pelatihan.



Gambar 2. Jumlah Data Setiap Kategori

Hasil *Processing*

Tahap *processing* dilakukan untuk mengeliminasi elemen tidak relevan (*noise*) serta mentransformasikan data teks mentah menjadi format yang lebih terstruktur. Rangkaian prosedur ini mencakup enam tahapan utama, yaitu: *case folding*, *cleaning*, normalisasi, *tokenizing*, *filtering* (*stopword removal*), dan *stemming*.

1) *Case Folding*

Tahapan ini berfungsi untuk menstandardisasi seluruh teks dengan mengonversi karakter huruf kapital menjadi huruf kecil (*lowercase*). Melalui penyeragaman ini, sistem dapat mengidentifikasi kata-kata yang identik secara konsisten dan menghindari redundansi fitur yang dipicu oleh variasi penulisan kapitalisasi.

Tabel 1. *Case Folding*

Teks Awal	Sesudah *Case Folding*
Gue udh angkat tangan bgt deh ampun. jujurrrr ada beban moril tersendiri buat gue sebagai lulusan gizi (karna namanya makan berGIZI gratis)	Gue udh angkat tangan bgt deh ampun. jujurrrr ada beban moril tersendiri buat gue sebagai lulusan gizi (karna namanya makan bergizi gratis)
Makan Bergizi Gratis diyakini sebagai langkah progresif tuntaskan masalah Malnutrisi #dukungMBG https://t.co/Tn66IZmmHo	Makan bergizi gratis diyakini sebagai langkah progresif tuntaskan masalah malnutrisi #dukungmbg https://t.co/tn66izmmho

2) *Cleaning*

Prosedur ini bertujuan untuk mengeliminasi seluruh elemen *non*-alfabet yang tidak memiliki nilai informatif bagi pemodelan. Mengingat karakteristik data yang bersumber dari platform X, proses pembersihan dikonsentrasikan pada penghapusan angka, tanda baca, tautan URL, emoji, serta simbol khusus seperti penanda akun (*mention*) dan tagar (*hashtag*).

Tabel 2. *Cleaning*

Case Folding	Sesudah *Cleaning*
gue udh angkat tangan bgt deh ampun. jujurrrr ada beban moril tersendiri buat gue sebagai lulusan gizi (karna namanya makan bergizi gratis)	Gue udh angkat tangan bgt ampun jujurrrr ada beban moril tersendiri buat gue sebagai lulusan gizi karna namanya makan bergizi gratis
makan bergizi gratis diyakini sebagai langkah progresif tuntaskan masalah malnutrisi #dukungmbg https://t.co/tn66izmmho	Makan bergizi gratis diyakini sebagai langkah progresif tuntaskan masalah malnutrisi

3) Normalisasi

Tahap penyeragaman untuk menangani karakteristik bahasa informal, singkatan, dan kesalahan pengetikan (*typo*) yang lazim dijumpai pada media sosial. Dengan memanfaatkan kamus leksikon (*slang dictionary*), Kosakata tidak baku yang terdapat di dalam dataset disesuaikan secara sistematis menjadi bentuk standar agar selaras dengan kaidah ejaan resmi bahasa Indonesia.

E-ISSN: 2580-1643

Tabel 3. Normalisasi

Cleaning	Sesudah Normalisasi
Gue udh angkat tangan bgt ampun jujurrrr ada beban moril tersendiri buat gue sebagai lulusan gizi karna namanya makan bergizi gratis	Saya sudah angkat tangan sangat deh ampun jujur ada beban moril tersendiri buat saya sebagai lulusan gizi karena namanya makan bergizi gratis
Makan bergizi gratis diyakini sebagai langkah progresif tuntaskan masalah malnutrisi	Makan bergizi gratis diyakini sebagai langkah progresif tuntaskan masalah malnutrisi

4) *Tokenizing*
Tahapan segmentasi yang memecah konstruksi kalimat utuh menjadi unit kata tunggal atau token. Proses pemisahan ini memanfaatkan spasi antar-kata sebagai pembatas (*delimiter*), sehingga algoritma dapat mengeksekusi analisis data secara tekstual kata demi kata.

Tabel 4. *Tokenizing*

Normalisasi	Sesudah *Tokenizing*
saya sudah angkat tangan sangat deh ampun jujur ada beban moril tersendiri buat saya sebagai lulusan gizi karena namanya makan bergizi gratis	[‘saya’, ‘sudah’, ‘angkat’, ‘tangan’, ‘sangat’, ‘deh’, ‘ampun’, ‘jujur’, ‘ada’, ‘beban’, ‘moril’, ‘tersendiri’, ‘buat’, ‘saya’, ‘sebagai’, ‘lulusan’, ‘gizi’, ‘karena’, ‘namanya’, ‘makan’, ‘bergizi’, ‘gratis’]
makan bergizi gratis diyakini sebagai langkah progresif tuntaskan masalah malnutrisi	[‘makan’, ‘bergizi’, ‘gratis’, ‘diyakini’, ‘sebagai’, ‘langkah’, ‘progresif’, ‘tuntaskan’, ‘masalah’, ‘malnutrisi’]

5) *Filtering*
Tahapan ini bertujuan untuk menyeleksi dan mengeliminasi kata-kata umum (*stopwords*) yang memiliki frekuensi kemunculan tinggi namun tidak membawa bobot informasi yang spesifik terhadap arah sentimen. Reduksi terhadap kata hubung maupun kata ganti ini dilakukan secara sistematis guna meningkatkan efisiensi komputasi pada proses pembobotan fitur TF-IDF.

Tabel 5. *Filtering*

Tokenizing	Sesudah *Filtering*
[‘saya’, ‘sudah’, ‘angkat’, ‘tangan’, ‘sangat’, ‘deh’, ‘ampun’, ‘jujur’, ‘ada’, ‘beban’, ‘moril’, ‘tersendiri’, ‘buat’, ‘saya’, ‘sebagai’, ‘lulusan’, ‘gizi’, ‘karena’, ‘namanya’, ‘makan’, ‘bergizi’, ‘gratis’]	[‘angkat’, ‘tangan’, ‘deh’, ‘ampun’, ‘jujur’, ‘beban’, ‘moril’, ‘tersendiri’, ‘lulusan’, ‘gizi’, ‘namanya’, ‘makan’, ‘bergizi’, ‘gratis’]
[‘makan’, ‘bergizi’, ‘gratis’, ‘diyakini’, ‘sebagai’, ‘langkah’, ‘progresif’, ‘tuntaskan’, ‘masalah’, ‘malnutrisi’]	[‘makan’, ‘bergizi’, ‘gratis’, ‘diyakini’, ‘langkah’, ‘progresif’, ‘tuntaskan’, ‘masalah’, ‘malnutrisi’]

6) *Stemming*
Tahapan terakhir dari *processing* adalah *stemming*, yakni proses morfologi untuk menghilangkan seluruh imbuhan (awalan, sisipan, maupun akhiran) pada kata. Proses ini akan mengembalikan setiap kata menjadi bentuk kata dasarnya (*root word*) agar variasi kata dapat dikenali sebagai satu fitur yang sama oleh SVM.

Tabel 6. *Stemming*

Filtering	Sesudah *Stemming*
[‘angkat’, ‘tangan’, ‘deh’, ‘ampun’, ‘jujur’, ‘beban’, ‘moril’, ‘tersendiri’, ‘lulusan’, ‘gizi’, ‘namanya’, ‘makan’, ‘bergizi’, ‘gratis’]	[‘angkat’, ‘tangan’, ‘deh’, ‘ampun’, ‘jujur’, ‘beban’, ‘moril’, ‘sendiri’, ‘lulus’, ‘gizi’, ‘nama’, ‘makan’, ‘gizi’, ‘gratis’]
[‘makan’, ‘bergizi’, ‘gratis’, ‘diyakini’, ‘langkah’, ‘progresif’, ‘tuntaskan’, ‘masalah’, ‘malnutrisi’]	[‘makan’, ‘gizi’, ‘gratis’, ‘yakin’, ‘langkah’, ‘progresif’, ‘tuntas’, ‘masalah’, ‘malnutrisi’]

Ekstraksi Fitur (TF-IDF)
Tahap selanjutnya setelah data teks dibersihkan adalah proses ekstraksi fitur, di mana teks diubah menjadi vektor numerik dengan menerapkan metode TF-IDF. Algoritma pembobotan ini mengalkulasi bobot relevansi tiap kata dalam suatu *tweet* berdasarkan kemunculannya di tingkat dokumen dan keseluruhan korpus. Tabel 7 menampilkan sebagian sampel data sebagai contoh visualisasi dari representasi fitur tersebut.

E-ISSN: 2580-1643

Tabel 7. Dokumen Pembobotan Kata

Dokumen	Teks Hasil *Processing*
D1	[mari, dukung, makan, gizi, gratis, masa, depan, tanpa, lapar]
D2	[makan, gizi, gratis, yakin, langkah, progresif, tuntas, masalah, malnutrisi]

Berdasarkan sampel dokumen tersebut, dilakukan perhitungan bobot akhir dengan mengalikan nilai TF dengan IDF seperti yang direpresentasikan dalam Tabel 8.

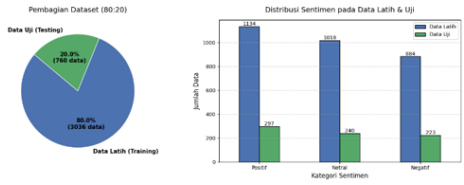
Tabel 8. Hasil Pembobotan Kata

Dokumen	Hasil Pembobotan
D1	[2,3863, 2,3863, 1,000, 1,000, 1,000, 2,3863, 2,3863, 2,3863, 2,3863]
D2	[1,000, 1,000, 1,000, 2,3863, 2,3863, 2,3863, 2,3863, 2,3863, 2,3863]

Berdasarkan hasil perhitungan, kata-kata yang sangat umum dan muncul di hampir seluruh dokumen (seperti "makan", "gizi", dan "gratis") mendapatkan penalti berupa nilai IDF yang rendah (sebesar 1,0000) karena perannya sebagai pembeda antar dokumen berkurang. Sebaliknya, kata-kata yang spesifik dan unik pada sentimen tertentu mendapatkan nilai bobot yang jauh lebih tinggi (mencapai 2,3863). Nilai perkalian TF-IDF inilah yang kemudian membentuk *matrix* fitur definitif sebagai masukan (*input*) utama algoritma SVM.

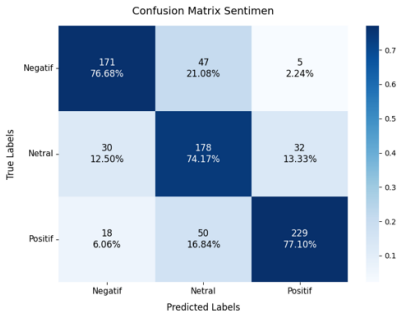
Ekstraksi Fitur (TF-IDF)

Sebagai implementasi dari tahapan metodologi, visualisasi proporsi pembagian 3.796 dataset menjadi data latih dan data uji disajikan pada Gambar 3 berikut.



Gambar 3. Hasil Pembagian Dataset

Model *Support Vector Machine* (SVM) yang telah dilatih menggunakan proporsi dan parameter yang telah ditetapkan tersebut, selanjutnya dievaluasi menggunakan 760 data uji. Pengujian performa model dalam mengenali sentimen aktual dibandingkan dengan hasil prediksi divisualisasikan menggunakan *Confusion Matrix* pada Gambar 4.



Gambar 4. *Confusion Matrix*

Hasilnya, algoritma SVM mampu menebak 578 data secara tepat, yang menghasilkan tingkat Akurasi (*Accuracy*) keseluruhan sebesar 76,05% (atau 76%). Secara lebih rinci, performa model pada masing-masing kelas adalah:

Tabel 9. *Clasification Report*

Kelas Sentimen	Presisi (*Precission*)	*Recall*	F1-*Score*
Positif	86%	77%	81%
Negatif	78%	77%	77%
Netral	65%	74%	69%
Akurasi (*Accuracy*)		76%	

Berdasarkan Tabel 9, akurasi 76,05% yang diperoleh didukung oleh metrik performa yang solid. Kelas positif mencatat *Precision* 86%, *Recall* 77%, dan *F1-Score* 81%, mengindikasikan keandalan system mendeteksi dukungan publik. Performa konsisten juga terlihat pada kelas negatif (*Precision* 78%, *Recall* 77%, *F1-Score* 77%). Adapun kelas netral memiliki *Precision* terendah 65% dan *Recall* 74%, yang secara statistik wajar akibat kemiripan struktur kosakata teks berita dengan kelas bersentimen. Secara keseluruhan, perolehan *F1-Score* di atas 69% membuktikan efektivitas algoritma SVM dalam mengklasifikasikan ketiga polaritas opini secara signifikan.

Visualisasi Hasil

Berdasarkan nilai evaluasi yang didapatkan, model SVM terbukti sangat andal (dengan akurasi 76%) dalam mengenali pola opini masyarakat terkait kebijakan MBG. Model memiliki tingkat presisi tertinggi pada kelas *positif* (86%), yang mengindikasikan bahwa sistem sangat akurat dalam mendeteksi dukungan publik. Sebaliknya, kesalahan prediksi paling banyak terjadi pada kelas

E-ISSN: 2580-1643

netral (presisi 65%) akibat kemiripan struktur teks informatif portal berita dengan kelas bersentimen. Untuk memahami lebih dalam mengenai konteks perbincangan publik dari prediksi model tersebut, pemetaan frekuensi kata divisualisasikan melalui *Word Cloud*. Hasil visualisasi membedah fokus narasi masyarakat pada tiap kategori.

1) Sentimen *Positif*
   Selain nama program, *Word Cloud* didominasi oleh kata "dukung", "sehat", "anak", dan "sekolah". Hal ini menunjukkan bahwa optimisme masyarakat terbangun dari keyakinan terhadap manfaat nyata program dalam peningkatan mutu kesehatan dan gizi pelajar.

Gambar 5. *Word Cloud* Sentimen *Positif*

2) Sentimen *Negatif*
   Visualisasi menunjukkan kemunculan mencolok pada kata "racun", "korban", "anggar", "korupsi", dan "rakyat". Temuan krusial ini merepresentasikan ketakutan masif masyarakat terhadap isu keamanan pangan dan potensi keracunan, diiringi dengan sikap kritis terhadap transparansi pengelolaan anggaran pemerintah. Munculnya kata kunci bernada kecurigaan seperti "racun" dan "korupsi" dalam kluster sentimen negatif menunjukkan adanya krisis kepercayaan publik terhadap kesiapan logistik lokal, yang menuntut pemerintah untuk mengadopsi sistem pengawasan rantai pasok makanan yang tersertifikasi secara terbuka.

Gambar 6. *Word Cloud* Sentimen *Negatif*

3) Sentimen Netral
   Narasi berpusat pada kata "sekolah", "siswa", "layanan", "distribusi", dan "pantau". Kosakata ini mencerminkan penyebaran informasi prosedural mengenai target penerima manfaat, logistik, dan pengawasan tanpa tendensi emosional.

Gambar 7. *Word Cloud* Sentimen Netra

Secara keseluruhan, sentimen masyarakat cenderung *positif* dengan harapan yang tinggi, namun pemerintah perlu menaruh perhatian serius pada isu transparansi anggaran dan standar keamanan pangan untuk meminimalisir keresahan yang mendominasi opini *negatif* publik. Sebagai bentuk evaluasi kritis terhadap hasil pengujian, performa model SVM dengan akurasi 76,05% pada penelitian ini sejalan dengan penelitian sentimen menggunakan metode serupa oleh Rani Yunita (2023) yang berada di angka 80%, meskipun sedikit di bawah performa pada pengujian ulasan aplikasi oleh Vina Fitriyana dkk. (2023) yang mencapai 96%. Variasi akurasi ini sangat dipengaruhi oleh tingginya ambiguitas semantik dan dinamika bahasa gaul (*slang*) yang spesifik pada topik opini kebijakan MBG. Perlu diakui pula bahwa penelitian ini memiliki batasan sistem yang tidak dapat dihindari sepenuhnya. Pertama, terdapat potensi bias subjektivitas yang membayangi dataset karena proses pelabelan data (anotasi sentimen) dikerjakan secara manual oleh peneliti. Kedua, representasi data pada penelitian ini hanya terbatas pada ruang lingkup pengguna platform X (*Twitter*), sehingga belum mewakili keragaman opini populasi dari platform media sosial lainnya.

3.2 Pembahasan

Hasil pengujian menunjukkan bahwa model Support Vector Machine (SVM) mampu mencapai tingkat akurasi sebesar 76,05%, yang menunjukkan kinerja yang cukup baik dalam mengklasifikasi opini masyarakat terhadap kebijakan Program Makan Bergizi Gratis (MBG) di platform X. Performa ini sejalan dengan penelitian yang dilakukan oleh Rani Yunita (2023), yang mendapatkan akurasi sekitar 80% dalam analisis sentimen kebijakan di Twitter, meskipun sedikit lebih tinggi dari hasil penelitian ini. Kinerja model yang cukup baik ini juga didukung oleh metrik presisi dan recall yang menunjukkan kemampuan SVM dalam mendeteksi polaritas opini secara akurat, terutama pada kategori positif dengan presisi 86%. Hasil ini konsisten dengan studi oleh Vina Fitriyana *et al.* (2023), yang melaporkan

E-ISSN: 2580-1643

bahwa SVM mampu memberikan akurasi tinggi dalam analisis sentimen aplikasi mobile, yaitu mencapai 96%. Analisis Word Cloud memperlihatkan bahwa sentimen positif didominasi kata-kata seperti "dukung", "sehat", dan "anak", yang menunjukkan bahwa masyarakat memiliki harapan besar terhadap manfaat program dalam meningkatkan mutu gizi dan kesehatan siswa (Yusuf *et al.*, 2022). Sebaliknya, sentimen negatif menampilkan kata-kata seperti "racun", "korupsi", dan "anggar", yang mencerminkan kekhawatiran masyarakat terhadap isu keamanan pangan dan transparansi pengelolaan dana (Khofifah & Indarwati, 2023). Temuan ini sejalan dengan penelitian oleh Dewi *et al.* (2025), yang menegaskan bahwa opini negatif seringkali berkaitan dengan ketidakpercayaan terhadap aspek keamanan dan pengelolaan program pemerintah. Meskipun demikian, tingkat akurasi yang diperoleh masih menunjukkan adanya ruang untuk peningkatan, terutama dalam menangani ambiguitas semantik dan perbedaan konteks opini yang kompleks di media sosial. Oleh karena itu, pengembangan model dengan teknik penyeimbangan data seperti SMOTE dan penggunaan arsitektur deep learning seperti BERT diharapkan dapat meningkatkan performa klasifikasi di masa mendatang (Ridwan *et al.*, 2024; Hidayatulloh *et al.*, 2025).

4. Kesimpulan

Implementasi model *Support Vector Machine* (SVM) untuk menentukan kelas sentimen warganet terhadap Program kebijakan MBG di platform X terbukti andal, yang ditunjukkan melalui perolehan akurasi sebesar 76,05%. Dari total 3.796 data yang diolah, opini publik didominasi oleh sentimen *positif* (1.431 data), yang menandakan tingginya optimisme masyarakat terhadap program ini, disusul oleh sentimen netral (1.258 data) dan *negatif* (1.107 data). Analisis visualisasi *Word Cloud* menunjukkan bahwa dukungan masyarakat didasari oleh harapan akan peningkatan gizi dan kesehatan siswa. Sebaliknya, opini *negatif* terpusat pada kekhawatiran atas keamanan pangan (risiko keracunan) serta transparansi tata kelola anggaran. Sebagai bahan evaluasi kebijakan, pemerintah perlu menaruh perhatian khusus pada pengetatan standar mutu makanan dan transparansi pengelolaan dana guna meredam keresahan publik serta mengoptimalkan keberhasilan pelaksanaan program di masa mendatang. Sebagai rekomendasi untuk pengembangan sistem komputasi selanjutnya, penelitian mendatang disarankan untuk melakukan pengujian teknik penyeimbangan kelas (*handling imbalanced data*) menggunakan metode SMOTE guna mengatasi tingginya kesalahan prediksi pada kelas netral (Ridwan *et al.*, 2024). Selain itu, penggunaan arsitektur berbasis deep learning seperti BERT sangat direkomendasikan untuk meningkatkan akurasi klasifikasi teks bahasa Indonesia yang tidak baku (*slang*) (Hidayatulloh *et al.*, 2025), serta perluasan cakupan pengambilan data ke platform media sosial lain seperti Instagram atau TikTok untuk memperkaya keragaman representasi opini publik.

5. Daftar Pustaka

Amrozi, Y., Yuliati, D., Susilo, A., & Novianto, N. (2022). Klasifikasi jenis buah pisang berdasarkan citra warna dengan metode SVM. *Jurnal SISFOKOM (Sistem Informasi dan Komputer)*, 11(3), 394–399. https://doi.org/10.32736/sisfokom.v11i3.1502.

Ansori, Y., & F. H. H. (2022). Perbandingan metode machine learning dalam analisis sentimen Twitter. *Jurnal Sistem dan Teknologi Informasi*, 10(4), 1–6. https://doi.org/10.26418/justin.v10i4.51784.

Arifin, N., & N. S. (2021). Penerapan algoritma Support Vector Machine (SVM) dengan TF-IDF n-gram untuk text classification. *Satuan Tulisan Riset dan Inovasi Teknologi*, 6(2), 129–136. http://dx.doi.org/10.30998/string.v6i2.10133.

Dewi, A. K., Rahmadani, N. F., Syahputri, R., & Rahma, L. (2025). Deteksi berita hoax pada platform X menggunakan pendekatan text mining dan algoritma machine learning. *Jurnal Teknologi dan Sistem Informasi*, 5(1), 33–46. https://doi.org/10.47709/dsi.v5i1.6011.

Fitriyana, V., Hakim, L., & Rini, D. C. N. (2023). Analisis sentimen ulasan aplikasi Jamsostek Mobile menggunakan metode Support Vector Machine. *Jurnal Buana Informatika*, 14(1), 40–49. https://doi.org/10.24002/jbi.v14i01.6909.

Harahap, E. D., & R. K. (2024). Analisis sentimen komentar terhadap kebijakan pemerintah mengenai Tabungan Perumahan Rakyat (TAPERA) pada aplikasi X menggunakan metode Naïve Bayes. *Jurnal Teknik Informatika Unika ST. Thomas (JTIUST)*, 09(01), 166–175. https://doi.org/10.33365/jtsi.v4i1.2445.

E-ISSN: 2580-1643

Hidayatulloh, S., Muflikhah, L., & R. S. P. (2025). Implementasi embedding IndoBERT dan Support Vector Machine (SVM) untuk analisis sentimen publik terhadap layanan Biznet. *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, 9(9), 1–10.

Husada, H. C., & A. S. P. (2021). Analisis sentimen pada maskapai penerbangan di platform Twitter menggunakan algoritma Support Vector Machine (SVM). *TEKNIKA*, 10(1), 18–26. https://doi.org/10.34148/teknika.v10i1.311.

Indarwati, K., & H. F. (2023). Analisis sentimen terhadap kualitas pelayanan aplikasi Gojek menggunakan metode Naive Bayes classifier. *JATISI (Jurnal Teknik Informatika & Sistem Informasi)*, 10(1). https://doi.org/10.35957/jatisi.v10i1.2643

Maldini, M. A., & S. A. (2025). Analisis sentimen ulasan pengguna aplikasi perbankan menggunakan algoritma Support Vector Machine dan Naive Bayes. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 9(3), 4098–4105. https://doi.org/10.36040/jati.v9i3.13522.

Muhammadin, I. A. S. (2021). Analisis sentimen pada ulasan aplikasi Kredivo dengan algoritma SVM dan NBC. *Jurnal Rekayasa Perangkat Lunak*, 2(2), 85–91. https://doi.org/10.31294/reputasi.v2i2.785.

Nugroho, D. C., Agustina, A. D., Maulana, B., Darussalam, F., & Asyraf, M. A. (2025). Penerapan metode Support Vector Machine untuk analisis sentimen publik pada program Makan Bergizi Gratis. *Infotech Journal*, 11(2), 474–479. https://doi.org/10.31949/infotech.v11i2.16781.

Nurhidayat, R. N., & K. E. D. (2023). Penerapan algoritma K-Nearest Neighbor dan fitur ekstraksi N-gram dalam analisis sentimen berbasis aspek. *Jurnal Ilmiah Komputer dan Informatika*, 12(1), 91–100. https://doi.org/10.34010/komputa.v12i1.9458.

Prastiawan, K. H., & D. Y. (2025). Analisis sentimen publik terhadap program Makan Bergizi Gratis dengan algoritma Naive Bayes. *Journal of Artificial Intelligence and Digital Business (RIGGS)*, 4(4), 5412–5419. https://doi.org/10.31004/riggs.v4i4.3652.

Putri, D. D., Nama, G. F., & W. E. S. (2022). Analisis sentimen kinerja Dewan Perwakilan Rakyat (DPR) pada Twitter menggunakan metode Naive Bayes classifier. *Jurnal Informatika dan Teknik Elektro Terapan (JITET)*, 10(1), 34–40. https://doi.org/10.23960/jitet.v10i1.2262.

Ridwan, E., & M. E. (2024). Penerapan metode SMOTE untuk mengatasi imbalanced data pada klasifikasi ujaran kebencian. *Jurnal Ilmu Komputer (CO-SCIENCE)*, 4(1), 80–88. https://doi.org/10.31294/coscience.v4i1.2990.

Sidauruk, N. B., Riza, N., & R. N. S. F. (2023). Penggunaan metode SVM dan Random Forest untuk analisis sentimen ulasan pengguna terhadap Kai Access di Google Playstore. *Jurnal Mahasiswa Teknik Informatika*, 7(3), 1901–1906. https://doi.org/10.36040/jati.v7i3.6899.

Styawati, N. H., & A. R. I. (2021). Analisis sentimen masyarakat terhadap program Kartu Prakerja di Twitter dengan metode Support Vector Machine. *Jurnal Informatika: Jurnal Pengembangan IT*, 6(3), 150–155. https://doi.org/10.30591/jpit.v6i3.2870.

Supian, B. T. R., Nanda, M., Lusiana, E., & R. (2024). Perbandingan kinerja Naïve Bayes dan SVM pada analisis sentimen Twitter Ibukota Nusantara. *Jurnal Ilmiah Informatika (JIF)*, 12(1), 15–21. https://doi.org/10.33884/jif.v12i01.8721.

Yunita, M. K., & R. (2023). Perbandingan algoritma SVM dan Naïve Bayes pada analisis sentimen kebijakan penghapusan kewajiban skripsi. *Indonesian Journal of Computer Science*, 12(5), 2879–2890. https://doi.org/10.33022/ijcs.v12i5.3415.