

Analisis Sentimen Publik Program Makan Bergizi Gratis (MBG) di Youtube: Perbandingan Kinerja Algoritma *Support Vector Machine (SVM)* dan *Random Forest*

Ratna Yulika Go^{1✉}, Hadasa Sarah Divanda Gaspersz², R. Hendra Kusumawardhana³,
Nur Aeni Hidayah⁴

^{1,2}) Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Esa Unggul

³) Program Studi Teknologi Informasi, Fakultas Ilmu Komputer, Universitas Indonesia,
Jakarta, Indonesia

⁴) Program Studi Sistem Informasi, Fakultas Sains dan Teknologi, Universitas Islam Negeri
Syarif Hidayatullah, Jakarta, Indonesia

¹) ratna.yulika@esaunggul.ac.id[✉], ²) hadasagsz03@student.esaunggul.ac.id,

³) r.hendra.k@gmail.com, ⁴) nur.aeni@uinjkt.ac.id

ABSTRACT

Public attention toward the Free Nutritious Meal Program (Program Makan Bergizi Gratis/MBG) increased following mass food poisoning incidents in various regions. Data from the National Nutrition Agency (Badan Gizi Nasional/ BGN) up to September 2025 recorded 6,517 affected students, while the Indonesian Education Monitoring Network (Jaringan Pemantau Pendidikan Indonesia/JPPPI) reported 11,566 victims as of October 12, 2025. This situation triggered widespread discussion on social media, with YouTube serving as the main platform for public opinion expression. The large number of unstructured comments made manual analysis ineffective, necessitating automated analytical methods. This study aims to analyze public sentiment toward the MBG program and compare the performance of the Support Vector Machine (SVM) and Random Forest algorithms. The research adopts a quantitative approach using 7,522 YouTube comments as the initial dataset, which are processed through data collection, text preprocessing, data labeling, and TF-IDF feature extraction stages. Subsequently, the SVM and Random Forest models are trained and tested to classify sentiment into positive, negative, and neutral categories. Model performance is evaluated using Accuracy, Precision, Recall, and F1-Score metrics. After the preprocessing stage, the number of data used in the sentiment analysis decreased compared to the initial dataset. This reduction resulted from data filtering, where irrelevant, duplicate, or sentiment-unclear data were removed from the dataset. The final cleaned dataset consists of 7,030 YouTube comments regarding the Free Nutritious Meal Program (MBG). The results show that public opinion is dominated by neutral sentiment (77.17% or 5,425 data), followed by negative sentiment (15.21% or 1,069 data), which proportionally exceeds positive sentiment (7.62% or 536 data). In the modeling stage, the Support Vector Machine (SVM) algorithm proved to be more robust and effective in handling imbalanced data compared to Random Forest. However, Random Forest achieved slightly higher accuracy (83.42% compared to 82.71%).

Keywords: Sentiment Analysis, Free Nutritious Meals, Support Vector Machine, Random Forest.

ABSTRAK

Perhatian masyarakat terhadap Program Makan Bergizi Gratis (MBG) meningkat setelah terjadinya insiden keracunan massal di berbagai daerah. Data Badan Gizi Nasional (BGN) hingga September 2025 mencatat 6.517 siswa terdampak keracunan, sementara Jaringan Pemantau Pendidikan Indonesia (JPPPI) melaporkan 11.566 korban per 12 Oktober 2025. Kondisi ini memicu diskusi luas di media sosial YouTube sebagai media utama penyampaian opini publik. Banyaknya komentar yang tidak terstruktur membuat analisis manual menjadi tidak efektif, sehingga diperlukan metode analisis otomatis. Penelitian ini bertujuan untuk menganalisis sentimen publik terhadap Program MBG serta membandingkan kinerja algoritma *Support Vector Machine (SVM)* dan *Random Forest*. Penelitian dilakukan dengan pendekatan kuantitatif menggunakan 7.522 komentar YouTube sebagai data awal yang dianalisis melalui tahapan pengumpulan data, pra-pemrosesan teks, pelabelan data, dan ekstraksi fitur TF-IDF. Selanjutnya, model *Support Vector Machine (SVM)* dan *Random Forest* dilatih dan diuji untuk mengklasifikasikan sentimen ke dalam kategori positif, negatif, dan netral. Evaluasi kinerja model dilakukan menggunakan

metrik Akurasi, Presisi, Recall, dan F1-Score. Setelah tahap *pre-processing*, jumlah data yang digunakan dalam analisis sentimen mengalami pengurangan dibandingkan dengan data awal. Hal ini disebabkan oleh proses penyaringan data (*data filtering*), di mana data yang tidak relevan, terduplikasi, atau tidak mengandung informasi sentimen yang jelas dieliminasi dari dataset. Hasil analisis terdapat 7.030 komentar YouTube yang sudah dibersihkan mengenai Program Makan Bergizi Gratis (MBG) menunjukkan bahwa opini publik didominasi oleh sentimen netral (77,17% atau 5.425 data), disusul sentimen negatif (15,21% atau 1.069 data) yang secara proporsi melampaui sentimen positif (7,62% atau 536 data). Pada tahap pengujian pemodelan, algoritma *Support Vector Machine (SVM)* terbukti lebih tangguh dan efektif menangani *dataset* yang tidak seimbang (*imbalanced*) dibandingkan dengan *Random Forest*. Kendati *Random Forest* mencatat akurasi yang sedikit lebih tinggi (83,42% berbanding 82,71%).

Kata kunci: Analisis Sentimen, Makan Bergizi Gratis, Support Vector Machine, Random Forest

I. PENDAHULUAN

Masalah gizi di Indonesia masih menjadi tantangan serius yang berdampak langsung pada kualitas pendidikan dan kesehatan masyarakat. Data dari Kementerian Kesehatan RI (2023) menunjukkan bahwa angka *stunting* nasional masih berada pada level 21,6%, sebuah angka yang berisiko menghambat perkembangan kognitif dan kemampuan anak dalam menyerap pelajaran [1]. Selain itu, laporan dari *World Health Organization (WHO)* mengindikasikan bahwa kerawanan pangan (*food insecurity*) memengaruhi hampir 25% populasi Indonesia [2], yang berkontribusi pada meluasnya malnutrisi. Kondisi ini tidak hanya menimbulkan masalah kesehatan fisik, tetapi juga berdampak pada kesejahteraan psikologis, di mana ketidakpastian akses terhadap makanan dapat memicu stres dan kecemasan dalam keluarga [3].

Program Makan Bergizi Gratis (MBG) merupakan kebijakan strategis pemerintah yang bertujuan untuk meningkatkan kualitas kesehatan anak sekolah sekaligus mendukung keberlanjutan pendidikan di Indonesia [1]. Program ini dirancang sebagai intervensi gizi untuk mengatasi masalah fundamental bangsa, yaitu ketimpangan gizi dan tingginya angka *stunting*, yang secara langsung memengaruhi perkembangan kognitif dan fisik generasi penerus [2]. Dengan menyediakan asupan gizi yang memadai secara rutin di lingkungan sekolah, program MBG diharapkan dapat meningkatkan konsentrasi belajar, daya tahan tubuh, dan kehadiran siswa, yang pada akhirnya berkontribusi pada peningkatan kualitas sumber daya manusia untuk mewujudkan visi Indonesia Emas 2045 [3].

Program MBG secara resmi diluncurkan pada 6 Januari 2025 dan menunjukkan perkembangan yang pesat [3]. Sasaran program ini adalah ibu hamil, ibu menyusui, balita non-PAUD, serta pelajar di jenjang SD, SMP, SMA, dan SMK. Berdasarkan laporan yang dilansir dari Sekretariat Kabinet Republik Indonesia [4] pada 5 Mei 2025, program ini telah menjangkau lebih dari 3 juta siswa di 1.286 lokasi, diperkirakan akan mencapai 6 juta pada akhir Juni dan 22 juta pada akhir Agustus, dengan target untuk mencakup 82,9 juta siswa pada akhir November 2025. Pelaksanaan program ini melibatkan kerjasama lintas sektor, di mana pemenuhan gizi disediakan melalui Satuan Pelayanan Pemenuhan Gizi (SPPG) yang dikelola langsung oleh pemerintah maupun melalui kemitraan dengan usaha

masyarakat, seperti UMKM, restoran, dan jasa boga. Keterlibatan produsen lokal ini diharapkan tidak hanya memenuhi kebutuhan gizi siswa tetapi juga menggerakkan roda perekonomian daerah [3].

Namun, di balik tujuan mulia dan perkembangannya yang cepat, implementasi program MBG dihadapkan pada tantangan serius terkait keamanan pangan. Sejak diluncurkan, program ini diwarnai oleh serangkaian insiden keracunan massal yang memicu kekhawatiran dan diskusi publik yang intens di berbagai platform digital. Dampak dari insiden ini sangat signifikan, di mana Kompas.com [5] pada 30 September 2025, melaporkan bahwa Badan Gizi Nasional (BGN) menonaktifkan sementara 56 Satuan Pelayanan Pemenuhan Gizi (SPPG) yang diduga terlibat dalam kasus keracunan pada program Makan Bergizi Gratis (MBG), dengan Jawa Barat sebagai provinsi yang memiliki korban keracunan MBG tertinggi yaitu, 4.125 korban. Dilansir dari merdeka.com [6], pada 1 Oktober 2025, data Badan Gizi Nasional (BGN) mencatat sebanyak 6.517 siswa menjadi korban keracunan dalam 75 kasus sepanjang Januari hingga September 2025. Angka yang berbeda dilansir dari Tempo.co [7] pada 13 Oktober 2025, di mana Jaringan Pemantau Pendidikan Indonesia (JPPI) melaporkan bahwa jumlah korban keracunan telah mencapai 11.566 orang hingga 12 Oktober 2025. Serangkaian peristiwa ini sontak menjadi perbincangan hangat di media sosial seperti YouTube, yang telah menjadi platform utama bagi masyarakat untuk mengekspresikan sentimen, opini, dan umpan balik mereka [8].

Sejumlah penelitian sebelumnya telah mengkaji program MBG dari berbagai sudut pandang. Qomarrullah [1], dalam penelitiannya menggunakan pendekatan kualitatif studi kasus untuk menganalisis dampak jangka panjang program terhadap kesehatan dan keberlanjutan pendidikan melalui wawancara mendalam dan observasi. Hasil penelitian ini menemukan bahwa program MBG berpotensi menurunkan angka *stunting* dan malnutrisi serta meningkatkan kehadiran siswa dan capaian akademik mereka. Sementara itu, Suprpto [3], menerapkan analisis kebijakan deskriptif multidimensional yang didukung oleh tinjauan literatur sistematis untuk mengevaluasi implementasi awal program dan membandingkannya dengan praktik internasional. Penelitian ini mengidentifikasi tantangan implementasi

yang signifikan, termasuk kesenjangan dalam pemantauan, masalah keamanan pangan, dan tata kelola program yang lemah. Di sisi lain, penelitian Cahyati [2], menggunakan pendekatan metode campuran (*mixed-methods*) yang mengombinasikan data kesehatan kuantitatif dengan wawancara kualitatif untuk mengukur dampak program terhadap kesehatan fisik dan kesejahteraan psikologis penerima manfaat. Penelitian ini menemukan bahwa program MBG berdampak positif pada kesehatan fisik (pengurangan malnutrisi dan stunting) dan kesejahteraan psikologis (pengurangan stres terkait kerawanan pangan). Selain itu, Virani [9] juga telah menunjukkan relevansi analisis komentar YouTube melalui analisis tematik untuk memahami wacana publik seputar isu kesehatan, yang membuktikan bahwa platform ini adalah sumber data yang kaya untuk penelitian sosial.

Mengingat volume komentar yang masif dan kompleks di media sosial, pendekatan analisis manual menjadi tidak efisien dan rentan terhadap subjektivitas [8]. Oleh karena itu, penelitian ini mengusulkan penggunaan analisis sentimen dengan pendekatan *machine learning*. Pendekatan ini relevan karena mampu mengolah dan mengklasifikasikan data teks dalam jumlah besar secara otomatis. Algoritma *Support Vector Machine (SVM)* dan *Random Forest* dipilih untuk dibandingkan karena telah terbukti efektif dalam klasifikasi sentimen pada konteks ulasan produk makanan di YouTube, di mana *Random Forest* mencapai akurasi 90-92%, sementara *Support Vector Machine (SVM)* mencapai akurasi 93-96%, dengan metrik evaluasi yang jelas seperti akurasi, presisi, *recall*, dan *F1-score* [10]. Penggunaan metode *machine learning* pada komentar YouTube mempunyai pemahaman yang lebih dalam mengenai pola pikir dan opini masyarakat yang beragam [11].

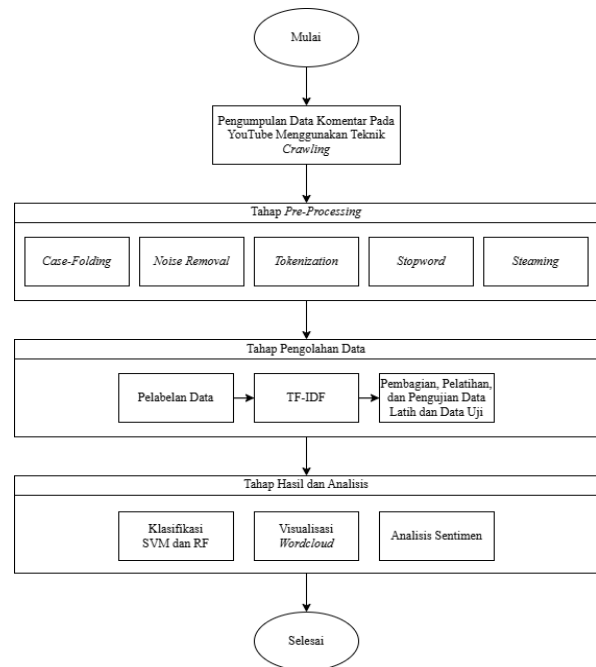
Penelitian ini berbeda dari penelitian-penelitian sebelumnya yang umumnya mengevaluasi dampak Program Makan Bergizi Gratis melalui survei, wawancara, atau analisis kebijakan. Penelitian ini menganalisis persepsi publik secara *real-time* dan dalam skala besar berdasarkan data digital tidak terstruktur dari media sosial. Penelitian ini memilih enam portal berita sebagai sumber data, pemilihan dilakukan dengan mempertimbangkan aspek kredibilitas media, luasnya jangkauan audiens, serta intensitas pemberitaan terkait Program Makan Bergizi Gratis (MBG). *Channel* yang dipilih seperti, KOMPASTV, CNN Indonesia, Harian Kompas, Liputan6, Tribun News, dan Tribun Jakarta Official, termasuk media arus utama yang memiliki jumlah *views* yang besar dan tingkat partisipasi *viewers* yang tinggi, sehingga komentar yang dikumpulkan dapat menggambarkan opini masyarakat secara lebih representatif.

Urgensi penelitian ini terletak pada kebutuhan pemerintah untuk mendapatkan umpan balik yang cepat dan otentik dari masyarakat sebagai bahan evaluasi kebijakan. Jika sentimen negatif publik terutama terkait pada isu keamanan pangan yang tidak diidentifikasi dan ditangani, dampaknya bisa sangat merusak, seperti

menurunnya kepercayaan publik, terhambatnya pencapaian tujuan program, dan bahkan potensi penghentian program akibat tekanan publik. Solusi yang ditawarkan melalui penelitian ini adalah sebuah model klasifikasi sentimen yang akurat untuk memetakan opini masyarakat, sehingga pemerintah dapat merumuskan respons kebijakan yang lebih tepat sasaran dan berbasis data.

II. METODE

Penelitian ini menggunakan pendekatan kuantitatif dengan alur kerja sistematis yang diuraikan dalam kerangka berpikir penelitian. Tahapan penelitian ini mencakup beberapa langkah utama yang saling berurutan seperti ditunjukkan pada Gambar 1:



Gambar 1. Tahapan Penelitian

A. Pengumpulan Data

Teknik pengumpulan data yang digunakan dalam penelitian ini adalah *crawling* data. Proses pengambilan data komentar publik dari platform YouTube dilakukan secara otomatis menggunakan bahasa pemrograman Python [12]. Secara spesifik, proses ini memanfaatkan YouTube Data API untuk mengakses dan mengunduh data komentar dari video-video yang telah ditentukan sebagai sampel penelitian. Populasi dan sample penelitian dapat dilihat pada Tabel 1.

Tabel 1. Populasi dan Sampel Penelitian

Komponen	Penjelasan
Populasi Penelitian	Seluruh komentar YouTube yang membahas Program Makan Bergizi Gratis (MBG) di Indonesia.
Teknik Sampling	<i>Purposive Sampling</i> , yaitu pemilihan sampel secara sengaja berdasarkan relevansi komentar dan kredibilitas kanal.

Kriteria Sampel	1. Komentar berbahasa Indonesia 2. Relevan membahas Program MBG 3. Berasal dari kanal berita resmi 4. Dipublikasikan pada periode Februari – Oktober 2025
Sumber Video/Channel	1. KOMPASTV 2. Harian Kompas 3. Liputan6 4. Tribun Jakarta Official 5. CNN Indonesia 6. Tribun News
Periode Pengambilan Data	September 2025 – Oktober 2025
Jumlah Sampel (Komentar)	7.030 komentar

Pemilihan video dilakukan berdasarkan relevansi konten dengan isu Program Makan Bergizi Gratis (MBG) serta periode publikasi yang sesuai dengan ruang lingkup penelitian. Perbedaan jumlah video pada setiap akun (Tabel 2) disesuaikan dengan intensitas pemberitaan yang tersedia, sehingga data yang dikumpulkan tetap berfokus pada konten yang secara langsung membahas topik penelitian.

Tabel 2. Jumlah Video Sumber Data Penelitian

No.	Channel YouTube	Jumlah Video
1.	KOMPASTV	3
2.	CNN Indonesia	3
3.	Liputan6	1
4.	Tribun Jakarta Official	1
5.	Harian Kompas	1
6.	Tribun News	1
Total Video		10

B. Pre-processing

Data mentah yang kotor dan tidak terstruktur dibersihkan (pra-pemrosesan). Proses ini mencakup normalisasi teks seperti *case folding*, *noise removal*, *tokenization*, *stopword removal*, *stemming* [13]. Berikut (Tabel 3) menjelaskan tahapan pra-pemrosesan data:

Tabel 3. Tahapan Pre-processing

Tahapan	Tujuan/Fungsi
<i>Case Folding</i>	Menyeragamkan seluruh huruf menjadi huruf kecil.
<i>Noise Removal</i>	Menghilangkan elemen pengganggu seperti tanda baca, angka, karakter non-alfabet, dan <i>url</i> .
<i>Tokenization</i>	Menguraikan kalimat menjadi unit kata (token) individual.
<i>Stopword Removal</i>	Menghapus kata-kata umum yang tidak memiliki bobot sentimen (seperti "di", "yang", "dan") menggunakan <i>library NLTK</i> dan Sastrawi.
<i>Stemming</i>	Mereduksi kata-kata berimbuhan ke bentuk kata dasarnya (misal: "makanan" menjadi "makan").

1) Pengolahan Data

Berikut adalah penjelasan dari tahapan pengolahan data [14]:

- a. Pelabelan Data: Proses pelabelan data dilakukan setelah tahap *pre-processing*. Setiap data teks diklasifikasikan ke dalam tiga kategori sentimen, yaitu positif, negatif, dan netral. Pelabelan dilakukan secara manual oleh dua annotator yang memahami konteks bahasa dalam dataset. Sebelum proses pelabelan, annotator diberikan pedoman pelabelan (*annotation guidelines*) yang berisi definisi masing-masing kelas sentimen untuk menjaga konsistensi.
- b. Ekstraksi Fitur TF-IDF: Teks yang sudah bersih diubah menjadi representasi numerik agar dapat diproses oleh algoritma *machine learning*. Penelitian ini menggunakan metode TF-IDF (*Term Frequency-Inverse Document Frequency*) untuk menghitung bobot setiap kata. Perhitungan TF-IDF pada penelitian ini menggunakan *tools* seperti, Python (Bahasa pemrograman), dengan *Library scikit-learn*, dan menggunakan *Google Colab*. Output dari proses TF-IDF berupa matriks numerik yang merepresentasikan bobot kepentingan setiap kata dalam dokumen. Matriks ini digunakan sebagai fitur input pada model *machine learning* untuk melakukan klasifikasi sentimen.
- c. Pembagian Data Latih dan Uji: Dataset yang telah menjadi vektor numerik dibagi menjadi dua bagian: Data Latih (*Training Set*) untuk melatih model, dan Data Uji (*Testing Set*) untuk menguji performa model terhadap data baru seperti terlihat pada Tabel 4.

Tabel 4. Pembagian Data Latih dan Uji

Keterangan	Total Data	Persentase	Komentar (80:20)
Total Dataset	7.030	100%	7.030
Data Latih (<i>Training Set</i>)	7.030	80%	5.624
Data Uji (<i>Testing Set</i>)	7.030	20%	1.406

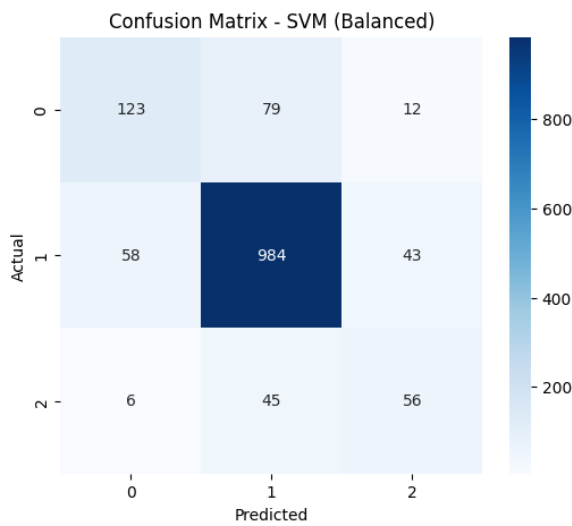
Berikut Dataset dibagi dengan perbandingan 80% data latih dan 20% data uji, untuk diterapkan dalam *machine learning* dengan memastikan model memiliki cukup data untuk belajar sekaligus tetap menyediakan data terpisah untuk menguji kinerjanya. Dengan porsi data latih yang lebih besar, model seperti *Support Vector Machine (SVM)* dan *Random Forest* dapat memahami pola dan bobot fitur TF-IDF dengan lebih baik, terutama karena data teks memiliki banyak fitur. Sementara itu, data uji digunakan untuk melihat seberapa baik model dapat memprediksi data baru yang belum pernah dipelajari sebelumnya.

- d. Pelatihan Model: Data latih digunakan untuk melatih dua algoritma klasifikasi secara terpisah, yaitu *Support Vector Machine (SVM)* dan *Random Forest*.
- e. Pengujian Model: Kedua model yang telah dilatih (*SVM* dan *Random Forest*) diuji kemampuannya

Model *Support Machine Vector (SVM)* menghasilkan akurasi sebesar 82,7% dengan performa yang relatif lebih seimbang antar kelas dibandingkan *Random Forest* (Gambar 4). Tabel 5 menunjukkan, pada kelas netral, model menunjukkan performa sangat baik dengan *recall* 0.91 dan *F1-score* 0.90. Pada kelas negatif, diperoleh *F1-score* sebesar 0.61 dengan *recall* 0.57, menunjukkan kemampuan deteksi yang cukup baik. Sementara itu, pada kelas positif, model mencapai *recall* 0.52 dan *F1-score* 0.51, yang menunjukkan peningkatan kemampuan dalam mendeteksi sentimen positif dibandingkan model lainnya. Nilai *macro F1-score* sebesar 0.67 menunjukkan bahwa model mampu mengklasifikasikan ketiga kelas secara lebih seimbang.

Tabel 5. Hasil Classification Report SVM

	precision	recall	F1-Score	support
negative	0.66	0.57	0.61	214
netral	0.89	0.91	0.90	1085
positif	0.50	0.52	0.51	107
accuracy				0.83
macro avg	0.68	0.67	0.67	1406
weight avg	0.82	0.83	0.82	1406



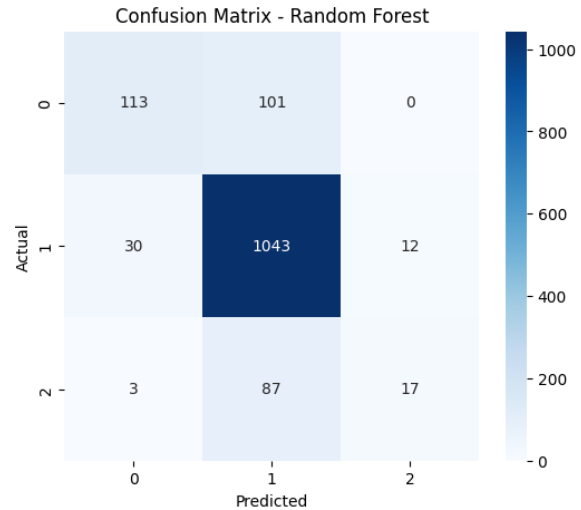
Gambar 4. Confusion Matrix SVM

2) Random Forest

Model *Random Forest* menghasilkan akurasi sebesar 83,4%, dengan performa terbaik pada kelas netral, *recall* 0.96 dan *F1-score* 0.90. Tabel 6 menunjukkan, pada kelas negatif, model menunjukkan performa cukup baik dengan *F1-score* 0.63. Namun, performa pada kelas positif sangat rendah dengan *recall* hanya 0.16 dan *F1-score* 0.25, yang menunjukkan bahwa sebagian besar data positif tidak terdeteksi dengan baik. Nilai *macro F1-score* sebesar 0.59 mengindikasikan bahwa model kurang seimbang dalam mengklasifikasikan seluruh kelas, terutama pada kelas minoritas (Gambar 5).

Tabel 6. Hasil Classification Report Random Forest

	precision	recall	F1-Score	support
negative	0.77	0.53	0.63	214
netral	0.85	0.96	0.90	1085
positif	0.59	0.16	0.25	107
accuracy				0.83
macro avg	0.74	0.55	0.59	1406
weight avg	0.82	0.83	0.81	1406



Gambar 5. Confusion Matrix Random Forest

C. Pembahasan

Berdasarkan hasil evaluasi (Tabel 7), *Random Forest* menghasilkan akurasi sedikit lebih tinggi yaitu, 83,4% dibandingkan *Support Machine Vector (SVM)* yaitu, 82,7%. Namun, jika dianalisis lebih lanjut berdasarkan *macro F1-score*, *Support Machine Vector (SVM)* menunjukkan performa yang lebih seimbang dengan nilai 0.67 dibandingkan *Random Forest* sebesar 0.59. *Random Forest* sangat baik dalam mengklasifikasikan kelas netral, tetapi memiliki kelemahan signifikan pada kelas positif dengan *recall* hanya 0.16. Sebaliknya, *Support Machine Vector (SVM)* mampu mendeteksi kelas positif dengan lebih baik yang dimana *recall* 0.52, serta mempertahankan performa yang stabil pada kelas negatif dan netral. Oleh karena itu, meskipun akurasi *Random Forest* sedikit lebih tinggi, *Support Machine Vector (SVM)* lebih unggul dalam menangani distribusi data yang tidak merata dan memberikan hasil klasifikasi yang lebih seimbang antar kelas sentimen.

Tabel 7. Perbandingan Model Machine Learning

Model	Akurasi	Presisi (Macro Avg)	Recall (Macro Avg)	F1-Score (Macro Avg)
SVM	82,71%	0.68	0.67	0.67
RF	83,42%	0.74	0.55	0.59

Algoritma *Random Forest (RF)* menunjukkan nilai akurasi yang sedikit lebih tinggi, yaitu sebesar 83,4% dibandingkan dengan *Support Vector Machine (SVM)* sebesar 82,7% [16]. Perbedaan ini mengindikasikan bahwa secara umum RF memiliki kemampuan yang sedikit lebih baik dalam mengklasifikasikan data secara keseluruhan. Namun, akurasi sebagai metrik tunggal belum cukup untuk menggambarkan performa model

secara komprehensif, terutama pada dataset dengan distribusi kelas yang tidak seimbang [17].

Jika ditinjau menggunakan metrik macro F1-score, Support Vector Machine (SVM) menunjukkan performa yang lebih unggul dengan nilai 0,67 dibandingkan Random Forest sebesar 0,59. Hal ini menunjukkan bahwa SVM memiliki kemampuan yang lebih baik dalam menjaga keseimbangan performa antar kelas, karena macro F1-score memberikan bobot yang sama pada setiap kelas, tanpa terpengaruh oleh jumlah data pada masing-masing kelas [18].

Analisis lebih lanjut pada metrik per kelas menunjukkan bahwa RF memiliki performa yang sangat baik dalam mengklasifikasikan kelas netral, namun mengalami kelemahan signifikan pada kelas positif dengan nilai recall yang rendah, yaitu hanya sebesar 0,16 [19]. Rendahnya recall ini mengindikasikan bahwa sebagian besar data dengan sentimen positif tidak berhasil terdeteksi oleh model RF, yang disebabkan oleh ketidakseimbangan distribusi data atau karakteristik fitur yang kurang mampu merepresentasikan pola sentimen positif secara optimal [20].

Sebaliknya, SVM mampu mendeteksi kelas positif dengan lebih baik, dengan nilai recall sebesar 0,52. Selain itu, SVM juga menunjukkan performa yang relatif stabil pada kelas negatif dan netral. Hal ini mengindikasikan bahwa SVM lebih robust dalam menangani variasi data dan tidak terlalu bias terhadap kelas mayoritas [21].

Dengan demikian, meskipun RF memiliki keunggulan dalam hal akurasi, SVM lebih direkomendasikan dalam konteks penelitian ini karena kemampuannya dalam menghasilkan klasifikasi yang lebih seimbang antar kelas [22]. Hal ini menjadi penting terutama dalam analisis sentimen, di mana setiap kelas (positif, negatif, dan netral) memiliki peran yang sama penting dalam interpretasi hasil [23].

Hasil penelitian ini juga sejalan dengan beberapa studi sebelumnya yang menunjukkan bahwa SVM cenderung memiliki performa yang stabil dalam tugas klasifikasi teks, termasuk pada data komentar YouTube. Beberapa penelitian melaporkan bahwa SVM mampu mencapai akurasi tinggi, bahkan hingga 93–96% [24], dan sering kali lebih unggul dibandingkan RF dalam konteks data teks yang memiliki dimensi tinggi dan sparsitas yang besar. Sementara itu, RF tetap menjadi algoritma yang kompetitif, terutama dalam menangkap pola umum pada data, meskipun cenderung kurang optimal dalam menangani kelas minoritas [25].

Selain itu, penelitian lain juga menunjukkan bahwa algoritma tradisional seperti SVM dan Random Forest masih relevan dan kompetitif dibandingkan pendekatan yang lebih kompleks, meskipun model berbasis deep learning seperti BERT dan RoBERTa mampu mencapai performa yang lebih tinggi. Namun demikian, penggunaan model tradisional tetap memiliki keunggulan dari sisi efisiensi komputasi dan kemudahan implementasi [26].

IV. KESIMPULAN

Hasil analisis terhadap 7.030 komentar YouTube mengenai Program Makan Bergizi Gratis (MBG) mengungkapkan bahwa opini publik didominasi oleh sentimen netral (77,17% atau 5.425 data), disusul sentimen negatif (15,21% atau 1.069 data) yang secara proporsi melampaui sentimen positif (7,62% atau 536 data). Tingginya respons negatif ini sejalan dengan visualisasi *Wordcloud* yang menyoroti tingginya atensi masyarakat pada isu "anak", "sekolah", "uang", serta keamanan pangan. Pada tahap pengujian pemodelan, algoritma *Support Vector Machine (SVM)* terbukti lebih tangguh dan efektif menangani *dataset* yang tidak seimbang (*imbalanced*) dibandingkan *Random Forest*. Kendati *Random Forest* mencatat akurasi yang sedikit lebih tinggi (83,42% berbanding 82,71%), *SVM* memberikan performa klasifikasi yang jauh lebih seimbang dalam mengenali kelas minoritas, yang dibuktikan dengan perolehan nilai *Macro F1-Score* sebesar 0,67 melampaui *Random Forest* yang hanya mencapai 0,59.

Penelitian ini memiliki beberapa keterbatasan yang perlu diperhatikan dalam menginterpretasikan hasil. Data dikumpulkan pada periode tertentu yang bertepatan dengan puncak isu program MBG, sehingga hasil analisis cenderung merepresentasikan opini publik pada momen tersebut dan tidak dapat digeneralisasikan untuk seluruh periode program. Pada tahap pre-processing, penelitian ini belum melakukan penambahan kata-kata umum terkait topik ke dalam daftar stopwords. Hal ini menyebabkan wordcloud masih menampilkan kata-kata topikal sehingga kurang optimal dalam menonjolkan kata yang merepresentasikan sentimen. Terdapat ketidakseimbangan data (*class imbalance*) dengan dominasi kelas netral (>77%), yang berdampak pada rendahnya performa deteksi kelas minoritas, terutama sentimen positif, dengan recall 0,52 pada Support Vector Machine dan 0,16 pada Random Forest.

Untuk pengembangan penelitian selanjutnya, disarankan melakukan penyempurnaan stopwords pada tahap pre-processing, menerapkan teknik penyeimbangan data seperti SMOTE, serta mengeksplorasi model berbasis deep learning untuk meningkatkan performa klasifikasi sentimen.

REFERENSI

- [1] R. Qomarrullah, S. Suratni, L. W. S, and M. Sawir, "Dampak Jangka Panjang Program Makan Bergizi Gratis terhadap Kesehatan dan Keberlanjutan Pendidikan," *Indonesian Journal of Intellectual Publication*, vol. 5, no. 2, Mar. 2025, doi: 10.51577/ijpublication.v5i2.660.
- [2] N. N. P. Cahyati, "A comprehensive study on MBG (Makan Bergizi Gratis) in the Prabowo-Gibran cabinet: Evaluating the psychological and health impacts of the policy," *Sikontan Journal*, vol. 3, no. 4, pp. 177–186, 2025.
- [3] F. A. Suprpto, E. Praditya, R. M. Dewi, and W. Adiyoso, "A policy implementation review of the free nutritious meal (MBG) program," *J. Indones. Sustain. Dev. Plan.*, vol. 6, no. 2, pp. 297–312, 2025, doi: 10.46456/jisdep.v6i2.798.
- [4] Office of Assistant to Deputy Cabinet Secretary for State Documents & Translation, "President Prabowo: Free Nutritious Meal Program is Investment for Nation's Future," *Cabinet Secretariat of The Republic of Indonesia*, 2025.

- <https://setkab.go.id/en/president-prabowo-free-nutritious-meal-program-is-investment-for-nations-future/> (accessed Oct. 14, 2025).
- [5] J. Carina, "Daftar 56 Dapur MBG yang Dinonaktifkan BGN imbas Kasus Keracunan," Kompas.com, 2025. <https://nasional.kompas.com/read/2025/09/30/00034261/daftar-56-dapur-mbg-yang-dinonaktifkan-bgn-imb-5-kasus-keracunan> (accessed Oct. 22, 2025).
- [6] Idris Rusadi Putra, "Data BGN: 6.517 Siswa Keracunan MBG Selama Januari-September 2025," merdeka.com, 2025. <https://www.merdeka.com/peristiwa/data-bgn-6517-siswa-keracunan-mbg-selama-januari-september-2025-475384-mvk.html/> (accessed Oct. 14, 2025).
- [7] Ervana Trikarinaputri, "JPPI: Korban Keracunan MBG hingga 12 Oktober Capai 11 Ribu Orang," Tempo, 2025. <https://www.tempo.co/politik/jppi-korban-keracunan-mbg-hingga-12-oktober-capai-11-ribu-orang-2078963> (accessed Oct. 14, 2025).
- [8] O. El Azzouzy, T. Chanyour, and S. J. Andaloussi, "Transformer-based models for sentiment analysis of YouTube video comments," *Sci. African*, vol. 29, p. e02836, 2025, doi: 10.1016/j.sciaf.2025.e02836.
- [9] A. Virani, B. Nagra, J. O'Mahony, J. Bacsu, J. K. Ghatore, and S. Panda, "Using YouTube Comments Data to Explore Postpartum Depression in Social Media: An Infodemiology Study," *Int. J. Environ. Res. Public Health*, vol. 21, no. 11, pp. 1–20, 2024, doi: 10.3390/ijerph21111526.
- [10] M. Tsiourlini, K. Tzafilkou, D. Karapiperis, and C. Tjortjis, "Text Analytics on YouTube Comments for Food Products," *Inf.*, vol. 15, no. 10, 2024, doi: 10.3390/info15100599.
- [11] I. Chalkias, K. Tzafilkou, D. Karapiperis, and C. Tjortjis, "Learning Analytics on YouTube Educational Videos: Exploring Sentiment Analysis Methods and Topic Clustering," *Electron.*, vol. 12, no. 18, 2023, doi: 10.3390/electronics12183949.
- [12] L. Ashbaugh and Y. Zhang, "A Comparative Study of Sentiment Analysis on Customer Reviews Using Machine Learning and Deep Learning," *Computers*, vol. 13, no. 12, 2024, doi: 10.3390/computers13120340.
- [13] A. B. Alawi and F. Bozkurt, "A hybrid machine learning model for sentiment analysis and satisfaction assessment with Turkish universities using Twitter data," *Decis. Anal. J.*, vol. 11, no. December 2023, p. 100473, 2024, doi: 10.1016/j.dajour.2024.100473.
- [14] F. Fauzi et al., "Analisis sentimen pengguna YouTube terhadap Program Vaksin COVID-19," pp. 141–148, 2021.
- [15] Munawar, "Extract Transform Loading (ETL) Based Data Quality for Data Warehouse Development," *Proc. 2021 1st Int. Conf. Comput. Sci. Artif. Intell. ICCSAI 2021*, no. October, pp. 373–378, 2021, doi: 10.1109/ICCSAI53272.2021.9609770.
- [16] A. Daza, N. D. González Rueda, M. S. Aguilar Sánchez, W. F. Robles Espíritu, and M. E. Chauca Quiñones, "Sentiment Analysis on E-Commerce Product Reviews Using Machine Learning and Deep Learning Algorithms: A Bibliometric Analysis and Systematic Literature Review, Challenges and Future Works," *Int. J. Inf. Manag. Data Insights*, vol. 4, no. 2, 2024, doi: 10.1016/j.jjime.2024.100267.
- [17] R. D. Cahyani and P. T. Prasetyaningrum, "Sentiment Analysis of User Reviews for AI Applications: Evaluating SVM, Logistic Regression, and Random Forest," *Journal of Information Systems and Informatics*, vol. 8, no. 1, pp. 1–27, 2026, doi: 10.63158/journalisi.v8i1.1366.
- [18] I. Pratistha, A. P. S. Iskandar, E. G. R. Lanang, and N. W. J. K. Dewi, "Comparison of Random Forest and SVM methods in sentiment analysis about electric cars in Indonesia," *Jurnal Mandiri IT*, vol. 14, no. 1, pp. 29–36, 2025, doi: 10.35335/mandiri.v14i1.424.
- [19] B. J. Azzahra, H. A. Id, and K. Agus, "Analisis sentiment terhadap pengguna mobil listrik berdasarkan komentar YouTube menggunakan metode SVM dan teknik ensemble," *Jurnal Pendidikan Informatika dan Sains*, vol. 14, no. 1, pp. 10–23, 2025, doi: 10.31571/saintek.v14i1.8871.
- [20] J. R. Dani and P. Parjito, "Analisis Sentimen Komentar YouTube terhadap Kenaikan Tunjangan DPR RI menggunakan Naïve Bayes, SVM, dan Random Forest," *Building of Informatics, Technology and Science (BITS)*, vol. 7, no. 3, pp. 1512–1524, 2025, doi: 10.47065/bits.v7i3.8513.
- [21] L. Septiana, V. Yasin, and A. Z. Sianipar, "Analyzing public sentiment on YouTube comments regarding the free lunch policy using the Support Vector Machine (SVM) algorithm," *Jurnal Mandiri IT*, vol. 14, no. 1, pp. 1–10, 2024, doi: 10.35335/mandiri.v14i1.434.
- [22] T. A. Khan, R. Sadiq, Z. Shahid, M. M. Alam, and M. B. M. Su'ud, "Sentiment analysis using Support Vector Machine and Random Forest," *Journal of Informatics and Web Engineering*, vol. 3, no. 1, pp. 45–58, 2024, doi: 10.33093/jiwe.2024.3.1.5.
- [23] S. Khomsah, "Sentiment Analysis on YouTube Comments Using Word2Vec and Random Forest," *Telematika: Jurnal Informatika dan Teknologi Informasi*, vol. 18, no. 1, pp. 61–72, 2021, doi: 10.31315/telematika.v18i1.4493.
- [24] B. A. Santoso, B. I. Nugroho, and D. U. Asyfiya, "Perbandingan algoritma Naïve Bayes, SVM, dan Random Forest untuk analisis sentimen komentar politik YouTube," *TIN: Terapan Informatika Nusantara*, vol. 6, no. 4, pp. 320–331, 2025, doi: 10.47065/tin.v6i4.8326.
- [25] E. D. Dewantoro and S. Rakasiwi, "Analisis komparatif kinerja SVM, Random Forest, dan Naïve Bayes untuk klasifikasi sentimen pada komentar YouTube," *BITS Journal*, vol. 7, no. 4, pp. 1603–1610, 2025, doi: 10.47065/bits.v7i4.8959.
- [26] B. K. Syafia, M. F. Hidayattullah, and W. Suteddy, "Studi komparasi algoritma SVM dan Random Forest pada analisis sentimen komentar YouTube BTS," *Jurnal Informatika*, vol. 14, no. 2, pp. 85–94, 2021, doi: 10.30646/sinus.v18i2.457.