

Analisis Sentimen Publik terhadap Program Makan Bergizi Gratis Menggunakan Support Vector Machine: Komparasi Teknik Ekstraksi Fitur pada Data Komentar YouTube

Alex Rikki*¹, Maria Krisnamurti Gea², Indah Sumiati Sinaga³, Yindah Sihole⁴, Lambok Hasibuan⁵, Irwandi Siboro⁶

¹Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Katolik Santo Thomas, Medan, Indonesia,
^{2,3,4,5,6}Mahasiswa Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Katolik Santo Thomas, Medan, Indonesia.

*Corresponding email: alexrikisinaga@gmail.com,
mgea1161@gmail.com, indahsinaga011@gmail.com,
yindahsihole26@gmail.com, hsblambok@gmail.com,
irwandisiboro369@gmail.com

Abstract

Penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap Program Makanan Bergizi Gratis berdasarkan opini yang berkembang di media sosial. Data penelitian diperoleh melalui proses pengumpulan data menggunakan teknik web scraping dengan kata kunci yang relevan, sehingga terkumpul sebanyak 400 komentar. Setelah melalui tahapan preprocessing yang meliputi case folding, normalisasi, pembersihan teks, penghapusan stopword, stemming, dan tokenisasi, diperoleh 242 komentar yang layak digunakan sebagai dataset penelitian. Proses pelabelan sentimen dilakukan secara otomatis menggunakan kamus InSet (Indonesia Sentiment Lexicon) dan diklasifikasikan ke dalam tiga kategori sentimen, yaitu positif, negatif, dan netral. Metode klasifikasi yang digunakan adalah Support Vector Machine (SVM) dengan tiga teknik ekstraksi fitur, yaitu Term Presence, Bag of Words (BoW), dan Term Frequency–Inverse Document Frequency (TF-IDF). Evaluasi performa model dilakukan menggunakan confusion matrix serta 5-Fold Cross Validation yang dikombinasikan dengan teknik Synthetic Minority Oversampling Technique (SMOTE) untuk mengatasi ketidakseimbangan kelas. Hasil analisis sentimen menunjukkan bahwa sentimen netral mendominasi opini masyarakat dengan persentase sebesar 52,89%, diikuti oleh sentimen positif sebesar 31,82% dan sentimen negatif sebesar 15,29%. Temuan ini mengindikasikan bahwa sebagian besar masyarakat bersikap moderat dengan memberikan evaluasi, saran, serta harapan terhadap implementasi Program Makan Bergizi Gratis.

Keywords: Sentiment Analysis, support vector machine, feature extraction; YouTube comments, public opinion

1. PENDAHULUAN

Permasalahan gizi masih menjadi salah satu permasalahan penting dalam pembangunan sumber daya manusia di Indonesia [1]. Kekurangan asupan gizi dapat berdampak pada kesehatan, pertumbuhan fisik, tingkat konsentrasi belajar, serta produktivitas masyarakat, khususnya pada anak usia sekolah [2]. Gizi adalah unsur nutrisi yang dibutuhkan tubuh untuk pertumbuhan, pemeliharaan kesehatan, dan fungsi fisiologis optimal, yang meliputi karbohidrat, protein, lemak, vitamin, mineral, dan air [3]. Pemenuhan gizi yang memadai menjadi faktor krusial bagi tercapainya kualitas hidup dan kesejahteraan masyarakat [4]. Oleh karena itu, pemerintah meluncurkan Program Makanan Bergizi Gratis sebagai upaya untuk meningkatkan kualitas gizi dan kesehatan masyarakat secara merata [5].

Implementasi program MBG memunculkan beragam tanggapan dari masyarakat yang disampaikan melalui media sosial, yang berfungsi sebagai ruang publik digital untuk mengekspresikan opini, kritik, dan dukungan terhadap kebijakan publik [6]. Media sosial seperti YouTube menghasilkan data dalam jumlah besar yang mencerminkan persepsi masyarakat secara spontan dan real time terhadap suatu kebijakan pemerintah [7]. Pemanfaatan data media sosial tersebut dapat dilakukan melalui pendekatan analisis

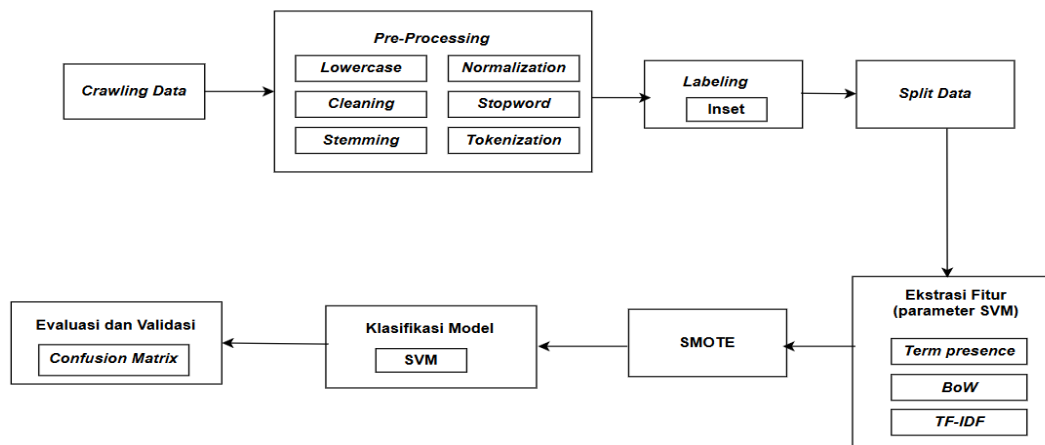
sentimen, yang bertujuan untuk mengklasifikasikan opini masyarakat ke dalam sentimen positif, negatif, dan netral menggunakan teknik *text mining* dan *machine learning* [8].

Beberapa penelitian sebelumnya telah menerapkan analisis sentimen untuk mengkaji respons masyarakat terhadap program makanan bergizi gratis dengan menggunakan algoritma seperti Naïve Bayes, Support Vector Machine, dan IndoBERT [9]. Hasil penelitian terdahulu umumnya berfokus pada pengukuran performa algoritma berdasarkan tingkat akurasi klasifikasi sentimen, tanpa membahas secara mendalam konteks sosial dan implikasi kebijakan dari sentimen yang dihasilkan. Selain itu, pemanfaatan data hasil *web scraping* dari platform YouTube masih relatif terbatas dibandingkan platform media sosial lain, padahal komentar YouTube cenderung lebih panjang dan argumentatif sehingga berpotensi memberikan informasi opini publik yang lebih kaya [10]. Kondisi tersebut menunjukkan adanya kesenjangan penelitian terkait analisis sentimen masyarakat terhadap program makanan bergizi gratis yang dikaji secara khusus menggunakan data hasil web scraping dari YouTube [11].

Urgensi penelitian ini semakin meningkat karena persepsi dan sentimen masyarakat dapat memengaruhi keberhasilan implementasi kebijakan publik, termasuk keberlanjutan program MBG tersebut [12]. Analisis sentimen masyarakat terhadap program MBG di media sosial dapat menjadi sarana evaluasi awal bagi pemerintah dalam mengidentifikasi potensi permasalahan, resistensi, maupun dukungan publik terhadap kebijakan yang sedang berjalan [13]. Berdasarkan latar belakang tersebut, penelitian ini penting dilakukan untuk menganalisis sentimen masyarakat terhadap program makanan bergizi gratis berdasarkan data komentar YouTube yang diperoleh melalui teknik *web scraping* [14]. Berdasarkan uraian di atas rumusan masalah dalam penelitian ini yaitu “bagaimana sentiment masyarakat terhadap program Makan Bergizi Gratis (MBG) yang tercermin dalam komentar pada platform YouTube?”.

2. METODE PENELITIAN

Penelitian ini menerapkan pendekatan supervised learning dengan algoritma Support Vector Machine (SVM), di mana pelabelan awal data dilakukan secara otomatis menggunakan pendekatan berbasis leksikon bahasa Indonesia. Pendekatan pembelajaran supervisi menggunakan algoritma SVM yang dilatih menggunakan dataset berlabel, sedangkan pendekatan berbasis leksikon memanfaatkan kamus sentiment [15]. Seluruh proses pengolahan data pada penelitian ini dilakukan menggunakan Google Colab sebagai lingkungan pengembangan utama. Riset ini meliputi beberapa tahapan yang diperlukan untuk mendapatkan hasil klasifikasi yang ditampilkan secara keseluruhan pada gambar 1 di bawah ini:



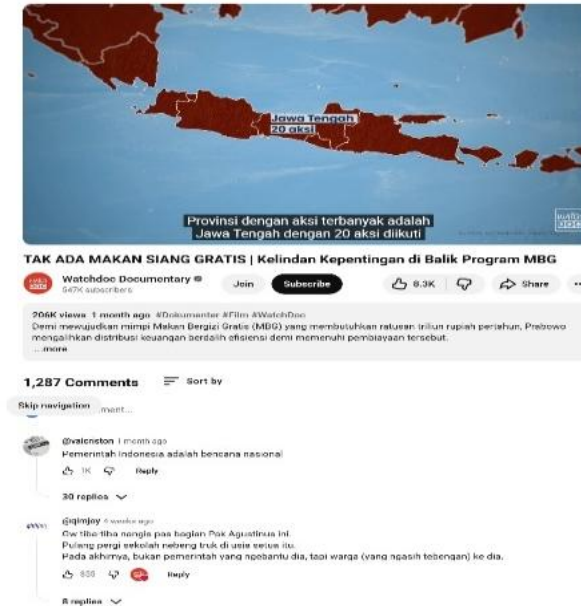
Gambar 1. Tahapan Penelitian

A. Crawling Data

Pengumpulan data pada penelitian ini dilakukan menggunakan teknik *crawling* pada platform YouTube, yaitu proses pengambilan data secara otomatis dengan mengekstraksi komentar pengguna melalui bantuan YouTube Data API menggunakan *script* youtube-comment-harvester yang dikembangkan oleh Helmi Satria [16]. Data dikumpulkan melalui dua metode *crawling* sebagaimana ditampilkan pada Tabel 1. Metode pertama dilakukan dengan pencarian video menggunakan kata kunci “program MBG”,

“makan bergizi gratis”, dan “MBG” dalam rentang waktu 5 Oktober hingga Desember 2025, yang menghasilkan 400 komentar.

Dari total 400 komentar yang berhasil dikumpulkan, dilakukan proses seleksi awal berupa penghapusan komentar duplikat, komentar kosong, dan komentar yang tidak relevan dengan topik penelitian. Setelah proses seleksi dan pembersihan data awal, diperoleh sebanyak 242 komentar yang digunakan sebagai dataset penelitian.



Gambar 2. Crawling data komentar youtube

Unggahan video tersebut dipilih karena memiliki tingkat interaksi yang tinggi, khususnya dalam bentuk komentar yang mencerminkan beragam tanggapan masyarakat terhadap Program Makan Bergizi Gratis (MBG). Banyaknya komentar pada unggahan tersebut menjadikannya sebagai sumber data yang representatif untuk menggambarkan opini publik, baik yang bersifat mendukung, netral, maupun kritis. Oleh karena itu, komentar yang berasal dari video tersebut digunakan sebagai data tambahan guna memperkuat representasi sentimen masyarakat secara aktual dan kontekstual terhadap kebijakan yang sedang dibahas. Rincian data hasil *crawling* dari sumber yang telah ditentukan disajikan pada Tabel 1.

Tabel 1. Hasil Crawling Data

...	created_at	author_account	full_text	...	username
...	2025-11-10T19:13.	UCzm6A9...	lanjutkan seterusnya MBG PROGRAM BAGUS byk yg terbantu gizi	...	@jjofficial...
...	2025-11-06T18:45:	UCUw2oY3....	Ganti program saja gpp kok rkyat pasti setuju...	...	@setiow....
...	2025-09-16T05:55:	UC0j1Rzdk...	Saran saya MBG ke daerah pelosok pak	...	@yatisupr...
...	2025-09-21T23:17.	UC1RgQy....	Proyek MBG Proyek Mudah Korupsi	...	@Nunun...
...	2025-09-27T07:0...	UCgK2Gus...	MBG bubarkan segera	...	@niara...
...	2025-09-18T04:1...	UCC6lr0b-...	Mbg kadang menunya astagfirullah	...	@Mamaa.
...	2025-09-03T16:	UCU-rv7...	Mbg membukan lapangan kerja di semua sektor	...	@HeniIra...

...	2025-08-24T07:	UCAff9y....	Maju terus MBG jgn mundur program ..	@jendrykro...
...	2025-09-01T18:4...	UC_YU4C...	Mbg rugikan kantin dan pedagang keliling	@aditrizqi...
...	2025-08-28T21:	UCjkjizN....	Lanjutkan pak sesekali control menunya agar lebih berkualitas lagi	@SUWATISU...
...	2025-08-28T05:	UC-R5Fn...	bubar aja negara program ga efektif mkin bnyak utang negara	@dzikyrama...
...	2025-08-28T02:	UCA7D...	Saya pribadi lebih baik MBG di stop dan di alihkan ke yg lain	@PenaHitam...
...				
...	2025-08-13T02:	UCi6rjf...	Maju terus MBG untuk generasi emas Indonesia	@pfsofficial
...				

B. Preprocessing

Preprocessing merupakan tahapan awal pengolahan data yang bertujuan untuk membersihkan, mengubah, dan mempersiapkan data mentah agar siap untuk dianalisis atau diproses lebih lanjut [16]. Dataset dalam riset ini awalnya terdiri dari 400 baris, namun setelah melalui tahap preprocessing, jumlah data menjadi 242 baris.

Preprocessing difokuskan pada kolom *full_text*, yang berisi teks asli komentar pengguna pada platform YouTube. Dalam riset ini, preprocessing dilakukan melalui beberapa langkah berurutan, yaitu lowercase, normalization, cleaning, stopwords removal, stemming, dan tokenization. Tahapan preprocessing yang dilakukan adalah sebagai berikut:

- 1) Lowercase: Tahap lowercase bertujuan mengubah seluruh karakter huruf dalam teks menjadi huruf kecil untuk menyeragamkan format penulisan [17]. Proses ini menggunakan fungsi `lower ()` yang merupakan fungsi bawaan Python.

Tabel 2. Hasil Preprocessing Lowercase

No	Full text	lowercase
1.	lanjutkan seterusnya MBG PROGRAM BAGUS byk yg terbantu gizi	lanjutkan seterusnya mbg program bagus banyak yang terbantu gizi...
2.	Ganti program saja gpp kok rkyat pasti setuju...	ganti program saja gapapa kok rakyat pasti setuju...
3.	Saran saya MBG ke daerah pelosok pak	saran saya mbg ke daerah pelosok pak
4.	Proyek MBG Proyek Mudah Korupsi	proyekmbg proyek mudah korupsi
5.	MBG bubarkan segera	mbg bubarkan segera
6.	Mbg kadang menunya astagfirullah	mbg kadang menunya astagfirullah
7.	Mbg membukan lapangan kerja di semua sektor	mbg membukan lapangan kerja di semua sektor
8.	Maju terus MBG jgn mundur program sangat bagus	maju terus mbg jgn mundur program sangat bagus
9.	Mbg rugikan kantin dan pedagang keliling	mbg rugikan kantin dan pedagang keliling

10.	Lanjutkan pak sesekali control menunya agar lebih berkualitas lagi	lanjutkan pak sesekali control menunya agar lebih berkualitas lagi
11.	bubar aja negara program ga efektif mkin bnyak utang negara	bubar aja negara program ga efektif mkin bnyak utang negara
12.	Gurunya tdk diperhatikan anggaran pendidikan habis utk mbg	gurunya tdk diperhatikan anggaran pendidikan habis utk mbg
13.	Saya pribadi lebih baik MBG di stop dan di alihkan ke yg lain	saya pribadi lebih baik mbg di stop dan di alihkan ke yg lain
14.	Maju terus MBG untuk generasi emas Indonesia	maju terus mbg untuk generasi emas Indonesia
15.	kebijakan populis tidak efektif	kebijakan populis tidak efektif
...	...	...
200.	Bahaya laten di Indonesia itu adalah korupsi Semoga dana mbg tdk dikorupsi	bahaya laten di indonesia itu adalah korupsi Semoga dana mbg tdk dikorupsi

- 2) Normalization: Pada tahap normalisasi, seluruh singkatan “MBG” dan variasinya diseragamkan menjadi “makan bergizi gratis” untuk menjaga konsistensi istilah dalam dataset. Proses untuk memperbaiki kata tidak baku dengan membuat kamus kata tidak baku secara manual. Kamus ini disusun berdasarkan observasi data dan referensi bahasa informal yang umum digunakan pada media sosial.

Tabel 3. Hasil Preprocessing Normalization

No	lowercase	normalization
1.	lanjutkan seterusnya mbg program bagus byk yg terbantu gizi...	lanjutkan seterusnya makan bergizi gratis program bagus banyak yang terbantu gizi...
2.	ganti program saja gpp kok rkyat pasti setuju...	ganti program saja tidak apa-apa kok rakyat pasti setuju...
3.	Saran saya MBG ke daerah pelosok pak	saran saya makan bergizi gratis ke daerah pelosok pak
4.	Proyek MBG Proyek Mudah Korupsi	proyek makan bergizi gratis proyek mudah korupsi
5.	MBG bubarkan segera	makan bergizi gratis bubarkan segera
6.	Mbg kadang menunya astagfirullah	makan bergizi gratis kadang menunya astagfirullah
7.	Mbg membukan lapangan kerja di semua sektor	makan bergizi gratis membuka lapangan kerja di semua sektor
8.	Maju terus MBG jgn mundur program sangat bagus	maju terus makan bergizi gratis jangan mundur program sangat bagus
9.	Mbg rugikan kantin dan pedagang keliling	makan bergizi gratis rugikan kantin dan pedagang keliling
10.	Lanjutkan pak sesekali control menunya agar lebih berkualitas lagi	lanjutkan pak sesekali kontrol menunya agar lebih berkualitas lagi
11.	bubar aja negara program ga efektif mkin bnyak utang negara	bubar aja negara program ga efektif makin banyak utang negara

12.	Gurunya tdk diperhatikan anggaran pendidikan habis utk mbg	gurunya tidak diperhatikan anggaran pendidikan habis untuk makan bergizi gratis
13.	Saya pribadi lebih baik MBG di stop dan di alihkan ke yg lain	saya pribadi lebih baik makan bergizi gratis di stop dan di alihkan ke yg lain
14.	Maju terus MBG untuk generasi emas Indonesia	maju terus makan bergizi gratis untuk generasi emas indonesia
15.	kebijakan populis tidak efektif	kebijakan populis tidak efektif
...	...	...
200.	Bahaya laten di Indonesia itu adalah korupsi Semoga dana mbg tdk dikorupsi	bahaya laten di indonesia itu adalah korupsi Semoga dana makan bergizi gratis tidak dikorupsi

- 3) Cleaning: Membersihkan teks dari berbagai elemen yang tidak diperlukan, seperti mention (@username), hashtag (#hashtag), kata "RT" penanda retweet, URL/link, karakter khusus, mengubah multiple spasi menjadi single spasi dan menghapus line break serta menghapus baris data yang kosong maupun duplikasi data [18]. Proses ini menggunakan modul operasi Regular Expression (RegEx) dari Python. Tahap ini tidak lagi mengubah bentuk kata, melainkan hanya menghapus elemen non-tekstual dan karakter yang tidak relevan.

Tabel 4. Hasil Preprocessing Cleaning

No	<i>normalization</i>	<i>cleaning</i>
1.	lanjutkan seterusnya mbg program bagus byk yg terbantu gizi...	lanjutkan seterusnya makan bergizi gratis program bagus banyak yang terbantu gizi...
2.	ganti program saja gpp kok rkyat pasti setuju...	ganti program saja tidak apa-apa kok rakyat pasti setuju...
3.	Saran saya MBG ke daerah pelosok pak	saran saya makan bergizi gratis ke daerah pelosok pak
4.	Proyek MBG Proyek Mudah Korupsi	proyek makan bergizi gratis proyek mudah korupsi
5.	MBG bubarkan segera	makan bergizi gratis bubarkan segera
6.	Mbg kadang menunya astagfirullah	makan bergizi gratis kadang menunya astagfirullah
7.	Mbg membukan lapangan kerja di semua sektor	makan bergizi gratis membukan lapangan kerja di semua sektor
8.	Maju terus MBG jgn mundur program sangat bagus	maju terus makan bergizi gratis jangan mundur program sangat bagus
9.	Mbg rugikan kantin dan pedagang keliling	makan bergizi gratis rugikan kantin dan pedagang keliling
10.	Lanjutkan pak sesekali control menunya agar lebih berkualitas lagi	lanjutkan pak sesekali control menunya agar lebih berkualitas lagi
11.	bubar aja negara program ga efektif mkin bnyak utang negara	bubar aja negara program ga efektif mkin bnyak utang negara
12.	Gurunya tdk diperhatikan anggaran pendidikan habis utk mbg	gurunya tidak diperhatikan anggaran pendidikan habis untuk makan bergizi gratis

13.	Saya pribadi lebih baik MBG di stop dan di alihkan ke yg lain	saya pribadi lebih baik makan bergizi gratis di stop dan di alihkan ke yg lain
14.	Maju terus MBG untuk generasi emas Indonesia	maju terus makan bergizi gratis untuk generasi emas indonesia
15.	kebijakan populis tidak efektif	kebijakan populis tidak efektif
...	...	...
200.	Bahaya laten di Indonesia itu adalah korupsi Semoga dana mbg tdk dikorupsi	bahaya laten di indonesia itu adalah korupsi Semoga dana makan bergizi gratis tidak dikorupsi

- 4) Stopwords Removal: Menghapus kata yang tidak informatif dan tidak memiliki makna pada dokumen, seperti “yang”, “dan”, “untuk”, dan sebagainya menggunakan kamus ID-Stopwords [19]. Penghapusan stopwords dilakukan dengan tetap mempertahankan kata-kata yang berpotensi mengandung makna sentimen agar tidak menghilangkan konteks opini masyarakat.

Tabel 5. Hasil Preprocessing Stopwords Removal

No	cleaning	no_stopwords
1.	lanjutkan seterusnya mbg program bagus byk yg terbantu gizi...	lanjutkan seterusnya makan bergizi gratis program bagus banyak terbantu gizi
2.	ganti program saja gpp kok rkyat pasti setuju...	ganti program saja tidak apa-apa kok rakyat pasti setuju
3.	Saran saya MBG ke daerah pelosok pak	saran saya makan bergizi gratis ke daerah pelosok pak
4.	Proyek MBG Proyek Mudah Kurupsi	proyek makan bergizi gratis proyek mudah korupsi
5.	MBG bubarkan segera	makan bergizi gratis bubarkan segera
6.	Mbg kadang menunya astagfirullah	makan bergizi gratis kadang menunya astagfirullah
7.	Mbg membukan lapangan kerja di semua sektor	makan bergizi gratis membukan lapangan kerja di semua sektor
8.	Maju terus MBG jgn mundur program sangat bagus	maju terus makan bergizi gratis jangan mundur program sangat bagus
9.	Mbg rugikan kantin dan pedagang keliling	makan bergizi gratis rugikan kantin dan pedagang keliling
10.	Lanjutkan pak sesekali control menunya agar lebih berkualitas lagi	lanjutkan pak sesekali control menunya agar lebih berkualitas lagi
11.	bubar aja negara program ga efektif mkin bnyak utang negara	bubar aja negara program ga efektif mkin bnyak utang negara
12.	Gurunya tdk diperhatikan anggaran pendidikan habis utk mbg	gurunya tidak diperhatikan anggaran pendidikan habis untuk makan bergizi gratis
13.	Saya pribadi lebih baik MBG di stop dan di alihkan ke yg lain	saya pribadi lebih baik makan bergizi gratis di stop dan di alihkan ke lain
14.	Maju terus MBG untuk generasi emas Indonesia	maju terus makan bergizi gratis generasi emas indonesia
15.	kebijakan populis tidak efektif	kebijakan populis tidak efektif
...	...	...

200.	Bahaya laten di Indonesia itu adalah korupsi Semoga dana mbg tdk dikorupsi	bahaya laten di indonesia itu adalah korupsi Semoga dana makan bergizi gratis tidak dikorupsi
------	--	---

- 5) Stemming: Proses ini mengkonversi kata-kata dalam tweet menjadi bentuk dasar menggunakan library Sastrawi yang cocok untuk teks berbahasa Indonesia [20].

Tabel 6. Hasil Preprocessing Stemming

No	cleaning	no_stopwords
1.	lanjutkan seterusnya mbg program bagus byk yg terbantu gizi...	lanjutkan seterusnya makan bergizi gratis program bagus banyak terbantu gizi
2.	ganti program saja gpp kok rkyat pasti setuju...	ganti program saja tidak apa-apa kok rakyat pasti setuju
3.	Saran saya MBG ke daerah pelosok pak	saran saya makan bergizi gratis ke daerah pelosok pak
4.	Proyek MBG Proyek Mudah Kurupsi	proyek makan bergizi gratis proyek mudah korupsi
5.	MBG bubarkan segera	makan bergizi gratis bubarkan segera
6.	Mbg kadang menunya astagfirullah	makan bergizi gratis kadang menunya astagfirullah
7.	Mbg membukan lapangan kerja di semua sektor	makan bergizi gratis membukan lapangan kerja di semua sektor
8.	Maju terus MBG jgn mundur program sangat bagus	maju terus makan bergizi gratis jangan mundur program sangat bagus
9.	Mbg rugikan kantin dan pedagang keliling	makan bergizi gratis rugikan kantin dan pedagang keliling
10.	Lanjutkan pak sesekali control menunya agar lebih berkualitas lagi	lanjutkan pak sesekali control menunya agar lebih berkualitas lagi
11.	bubar aja negara program ga efektif mkin bnyak utang negara	bubar aja negara program ga efektif mkin bnyak utang negara
12.	Gurunya tdk diperhatikan anggaran pendidikan habis utk mbg	gurunya tidak diperhatikan anggaran pendidikan habis untuk makan bergizi gratis
13.	Saya pribadi lebih baik MBG di stop dan di alihkan ke yg lain	saya pribadi lebih baik makan bergizi gratis di stop dan di alihkan ke lain
14.	Maju terus MBG untuk generasi emas Indonesia	maju terus makan bergizi gratis generasi emas indonesia
15.	kebijakan populis tidak efektif	kebijakan populis tidak efektif
...	...	...
200.	Bahaya laten di Indonesia itu adalah korupsi Semoga dana mbg tdk dikorupsi	bahaya laten di indonesia itu adalah korupsi Semoga dana makan bergizi gratis tidak dikorupsi

- 6) Tokenization: Proses tokenization yaitu mengonversi teks tweet menjadi kumpulan token (kata) individual untuk memudahkan pemrosesan selanjutnya dengan menggunakan modul nltk tokenize.

Tabel 7. Hasil Preprocessing Tokenization

No	<i>stemming</i>	<i>tokenization</i>
1.	lanjutkan seterusnya mbg program bagus byk yg terbantu gizi...	['lanjutkan', 'seterusnya', 'makan', 'bergizi', 'gratis', 'program', 'bagus', 'banyak', 'yang', 'terbantu', 'gizi', 'sangat', 'penting', 'sumber', 'daya', 'manusia', 'generasi', 'penerus', 'membantu', 'ekonomi', 'bawah', 'membantu', 'loker', 'juga', 'secara', 'agama', 'juga', 'sangat', 'berkah']
2.	ganti program saja gpp kok rkyat pasti setuju...	['ganti', 'program', 'saja', 'tidak', 'apa-apa', 'kok', 'rakyat', 'pasti', 'setuju']
3.	Saran saya MBG ke daerah pelosok pak	["saran", "saya", "makan", "bergizi", "gratis", "ke", "daerah", "pelosok", "pak"]
4.	Proyek MBG Proyek Mudah Korupsi	["proyek", "makan", "bergizi", "gratis", "proyek", "mudah", "kurupsi"]
5.	MBG bubarkan segera	["makan", "bergizi", "gratis", "bubarkan", "segera"]
6.	Mbg kadang menunya astagfirullah	["makan", "bergizi", "gratis", "kadang", "menunya", "astagfirullah"]
7.	Mbg membukan lapangan kerja di semua sektor	["makan", "bergizi", "gratis", "membuka", "lapangan", "kerja", "di", "semua", "sektor"]
8.	Maju terus MBG jgn mundur program sangat bagus	["maju", "terus", "makan", "bergizi", "gratis", "jangan", "mundur", "program", "sangat", "bagus"]
9.	Mbg rugikan kantin dan pedagang keliling	["makan", "bergizi", "gratis", "rugikan", "kantin", "dan", "pedagang", "keliling"]
10.	Lanjutkan pak sesekali control menunya agar lebih berkualitas lagi	["lanjutkan", "pak", "sesekali", "control", "menunya", "agar", "lebih", "berkualitas", "lagi"]
11.	bubar aja negara program ga efektif mkin bnyak utang negara	["bubar", "aja", "negara", "program", "tidak", "efektif", "makin", "banyak", "utang", "negara"]
12.	Gurunya tdk diperhatikan anggaran pendidikan habis utk mbg	["gurunya", "tidak", "diperhatikan", "anggaran", "pendidikan", "habis", "untuk", "makan", "bergizi", "gratis"]
13.	Saya pribadi lebih baik MBG di stop dan di alihkan ke yg lain	["saya", "pribadi", "lebih", "baik", "makan", "bergizi", "gratis", "di", "stop", "dan", "di", "alihkan", "ke", "lain"]
14.	Maju terus MBG untuk generasi emas Indonesia	["maju", "terus", "makan", "bergizi", "gratis", "generasi", "emas", "indonesia"]

15.	kebijakan populis tidak efektif	["kebijakan", "populis", "tidak", "efektif"]
...	...	...
200.	Bahaya laten di Indonesia itu adalah korupsi Semoga dana mbg tdk dikorupsi	["bahaya", "laten", "di", "indonesia", "itu", "adalah", "korupsi", "semoga", "dana", "makan", "bergizi", "gratis", "tidak", "dikorupsi"]

Contoh hasil tokenisasi yang ditampilkan pada tabel merupakan representasi sampel untuk menggambarkan hasil proses tokenisasi secara umum, bukan keseluruhan hasil token dari setiap tahapan prapemrosesan.

A. Labelling

Labelling dilakukan otomatis dengan menggunakan kamus InSet (Indonesia Sentiment Lexicon) merupakan kamus yang memuat bahasa Indonesia berbasis pendekatan leksikon [21]. Setiap kata akan diberikan nilai berdasarkan polaritasnya untuk mengukur respons masyarakat terhadap suatu topik [17]. Kamus InSet merupakan leksikon sentimen bahasa Indonesia yang banyak digunakan dalam penelitian analisis sentimen. Analisis sentimen berbasis leksikon pada riset ini menghasilkan label sentimen yang dikategorikan ke dalam tiga kelas, yaitu positif, negatif, dan netral. Setelah dilakukan preprocessing dan seleksi data, diperoleh sebanyak 242 data yang selanjutnya digunakan dalam proses pelabelan sentimen. Dalam proses ini berdasarkan kamus InSet Lexicon, Proses pelabelan dilakukan berdasarkan skor polaritas yang dihasilkan dari kamus InSet (Indonesia Sentiment Lexicon) bahasa Indonesia.

Tabel 8. Sampel Labeling Dataset

No	Text	Label
1.	['lanjutkan', 'seterusnya', 'makan', 'bergizi', 'gratis', 'program', 'bagus', 'banyak', 'yang', 'terbantu', 'gizi', 'sangat', 'penting', 'sumber', 'daya', 'manusia', 'generasi', 'penerus', 'membantu', 'ekonomi', 'bawah', 'membantu', 'loker', 'juga', 'secara', 'agama', 'juga', 'sangat', 'berkah']	positif
2.	['ganti', 'program', 'saja', 'tidak', 'apa-apa', 'kok', 'rakyat', 'pasti', 'setuju']	positif
3.	["saran", "saya", "makan", "bergizi", "gratis", "ke", "daerah", "pelosok", "pak"]	netral
4.	["proyek", "makan", "bergizi", "gratis", "proyek", "mudah", "kurupsi"]	negatif
5.	["makan", "bergizi", "gratis", "bubarkan", "segera"]	negatif
6.	["makan", "bergizi", "gratis", "kadang", "menunya", "astagfirullah"]	negatif
7.	["makan", "bergizi", "gratis", "membuka", "lapangan", "kerja", "di", "semua", "sektor"]	positif
8.	["maju", "terus", "makan", "bergizi", "gratis", "jangan", "mundur", "program", "sangat", "bagus"]	positif

9.	["makan", "bergizi", "gratis", "negatif", "rugikan", "kantin", "dan", "pedagang", "keliling"]
10.	["lanjutkan", "pak", "sese kali", "netral", "control", "menunya", "agar", "lebih", "berkualitas", "lagi"]
11.	["bubar", "aja", "negara", "negatif", "program", "tidak", "efektif", "makin", "banyak", "utang", "negara"]
12.	["gurunya", "tidak", "negatif", "diperhatikan", "anggaran", "pendidikan", "habis", "untuk", "makan", "bergizi", "gratis"]
13.	["saya", "pribadi", "lebih", "baik", "negatif", "makan", "bergizi", "gratis", "di", "stop", "dan", "di", "alihkan", "ke", "lain"]
14.	["maju", "terus", "makan", "positif", "bergizi", "gratis", "generasi", "emas", "indonesia"]
15.	["kebijakan", "populis", "tidak", "efektif"]
...	...
200.	["bahaya", "laten", "di", "netral", "indonesia", "itu", "adalah", "korupsi", "semoga", "dana", "makan", "bergizi", "gratis", "tidak", "dikorupsi"]

B. Split Data

Pembagian data dilakukan menggunakan teknik stratified split agar proporsi setiap label sentiment (positif, negatif, netral) tetap seimbang pada data latih dan data uji. Metode ini dipilih karena distribusi label pada dataset tidak merata sehingga pembagian acak dapat menimbulkan bias. Proses pembagian menggunakan fungsi `train_test_split` dari `scikit-learn` dengan rasio 80% data latih dan 20% data uji, serta `random_state` untuk menjaga konsistensi hasil. Dari total 242 data, diperoleh 193 data latih dan 49 data uji yang digunakan dalam proses klasifikasi.

Tabel 9. Sampel Split Data

No	Hasil Preprocessing	Label	Jenis Data
3	Indonesia krisis ekonomi rakyat susah makan...	Netral	Training
45	Program makan bergizi membantu anak sekolah	Positif	Training
90	Janji pemerintah sulit ditepati program ini membua.	Negative	Training
122	Semoga program berjalan baik demi generasi....	Positif	Training
150	Masyarakat masih bingung teknis pelaksanaan....	netral	Training
201	Makan bergizi gratis solusi tapi butuh anggaran...	Netral	Testing
77	Rakyat jangan dibohongi janji manis saja.	Negative	Testing
159	Bagus program ini bantu masyarakat miskin...	Positif	Testing
12	Pemerintah harus evaluasi kebijakan secara berka...	Netral	Testing
...		...	...
228	Implementasi program butuh pengawasan agar.....	positif	testing

C. Ekstraksi Fitur

Tahapan ekstraksi fitur dilakukan setelah proses pelabelan data selesai. Pada tahap ini, setiap data teks direpresentasikan dalam bentuk vektor numerik agar dapat digunakan dalam proses klasifikasi. Konversi

ini dilakukan dengan memberikan bobot pada setiap fitur melalui proses pembobotan fitur, sehingga karakteristik teks dapat direpresentasikan secara numerik untuk analisis lebih lanjut [20]. Pada riset ini akan menggunakan tiga teknik ekstraksi fitur yaitu *Term Presence*, *TF-IDF*, dan *Bag of Words*.

Term Presence (TP) adalah teknik yang mengelompokkan dokumen teks berdasarkan keberadaan kata atau fitur dalam suatu korpus, tanpa memperhitungkan jumlah kemunculannya. Sebaliknya, jika fitur tersebut tidak terhitung [22]. Sedangkan, teknik *TF-IDF* menghitung frekuensi kemunculan suatu kata dalam sebuah dokumen lalu membandingkannya dengan kemunculannya di seluruh dokumen yang tersedia. Semakin sering kata muncul dalam dokumen tertentu, semakin tinggi bobotnya dalam analisis teks [16]. Dan teknik *Bag of Words* (BoW) mengabaikan tata bahasa, sintaksis, dan urutan kata, tetapi mempertimbangkan frekuensi kemunculan kata dalam dokumen [20]. Pemilihan ketiganya didasarkan pada kekuatan representasi masing-masing dalam mengonversi data teks menjadi bentuk numerik yang dapat diolah oleh model klasifikasi seperti model SVM. Komparatif ini digunakan untuk mengetahui metode mana yang paling optimal dalam menangkap karakteristik sentimen.

D. Klasifikasi Model

Setelah nilai bobot fitur diperoleh dari hasil pelabelan melalui proses ekstraksi fitur sebagai parameter dalam klasifikasi, tahap berikutnya adalah melakukan proses klasifikasi dengan menggunakan metode Support Vector Machine (SVM). Proses klasifikasi SVM memerlukan nilai bobot untuk setiap kata dalam dataset pelatihan, dan klasifikasi dilaksanakan menggunakan skenario pemodelan dengan *Linear Kernel* SVM [23]. Dalam riset ini, klasifikasi sentimen dilakukan dengan memanfaatkan fungsi *Support Vector Classifier* (SVC) yang tersedia dalam *library scikit-learn (sklearn)* pada Python. SVM dapat digunakan untuk mengklasifikasikan ulasan atau teks ke dalam kategori sentimen positif atau negatif. Dengan belajar dari data pelatihan berlabel, SVM dapat memahami pola pada model [24]. Parameter yang relevan, seperti jenis *Linear Kernel* yang digunakan, telah disesuaikan dengan kebutuhan dan karakteristik data riset. Penggunaan *Linear Kernel* berada dalam dimensi tinggi dan sparse yang memisahkan dua kelas secara linier dalam ruang fitur berdimensi tinggi [25]. Karakteristik ini sesuai dengan representasi data teks hasil ekstraksi fitur *TF-IDF* dan *BoW*. Persamaan (1) menunjukkan rumusan dari Kernel Linear.

$$K(x_i, x_j) = X_i^T \cdot X_j \dots\dots\dots (1)$$

Tabel 9. Sampel Split Data

Prediksi			
		Prediksi Benar	Prediksi Salah
Aktual	Aktual Benar	TP (<i>True Positif</i>)	FN (<i>False Negative</i>)
	Aktual Salah	FP (<i>False Positif</i>)	TN (<i>True Negative</i>)

E. Evaluasi Model

Terakhir menguji performa dari algoritma SVM sebagai klasifikasi model pada riset ini. Tahap pertama menggunakan metode *Confusion Matrix* dengan pendekatan *single train-test split* untuk menghasilkan nilai akurasi, precision, recall, dan F1-Score berdasarkan perbandingan antara hasil prediksi dan data aktual dari model yang dilatih [23]. Implementasi pengujian akan dilakukan dengan membandingkan tiga metode ekstraksi fitur yang berbeda, yaitu *Term Presence*, *Bag of Words*, dan *TF-IDF*. Proses evaluasi dimulai dengan encoding label kategorikal menjadi format numerik menggunakan *LabelEncoder* untuk memastikan konsistensi dalam proses klasifikasi. *Confusion Matrix* yang menampilkan distribusi hasil klasifikasi dalam bentuk true positive, true negative, false positive, dan false negative [26].

Dengan persamaan Accuracy (2), Precision (3), Recall (4), dan F1-Score (5) berikut.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \dots\dots\dots (2)$$

$$Precision = \frac{TP}{TP + FP} \times 100\% \dots\dots\dots (3)$$

$$Recall = \frac{TP}{TP + FN} \times 100\% \dots\dots\dots (4)$$

$$F1 - Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \times 100\% \dots\dots\dots (5)$$

Tahap kedua, evaluasi lanjutan model yang lebih robust dalam mengatasi permasalahan class imbalance, riset ini menerapkan teknik Synthetic Minority Oversampling Technique (SMOTE) yang dikombinasikan dengan validasi KFold Cross Validation sebanyak 5 iterasi.

Teknik Synthetic Minority Oversampling Technique (SMOTE) merupakan metode yang digunakan untuk mengatasi masalah ketidakseimbangan (imbalance) kelas dalam dataset dengan menciptakan sampel sintetis dari kelas minoritas, yang tidak hanya menambah jumlah data, tetapi juga memperkaya variasi representasi untuk meningkatkan kualitas pembelajaran model [27]. K-Fold Cross Validation adalah metode validasi yang membagi dataset secara acak menjadi K partisi. Setiap partisi bergiliran digunakan sebagai data uji, sementara sisanya sebagai data latih, dan proses ini diulang sebanyak K kali [28]. Pada riset ini, menggunakan skema 5-Fold Cross Validation, jadi dataset dibagi menjadi 5 bagian, dan proses pelatihan serta pengujian dilakukan sebanyak 5 iterasi. SMOTE diterapkan hanya pada data latih di setiap iterasi, untuk memastikan tidak terjadi data leakage ke data uji. Hasil evaluasi performa model akan dibahas lebih lanjut pada bagian hasil dan pembahasan.

3. HASIL DAN PEMBAHASAN

3.1. Hasil

Pada penelitian ini, proses pengambilan data (*crawling*) komentar masyarakat terkait Program Makan Bergizi Gratis dilakukan pada platform YouTube dan menghasilkan sebanyak 400 data komentar. Selanjutnya, data tersebut melalui tahapan *preprocessing* yang meliputi *case folding*, *cleaning*, *tokenizing*, *stopword removal*, dan *stemming*, sehingga diperoleh 242 komentar yang layak digunakan dalam proses analisis. Tahap selanjutnya adalah pelabelan sentimen, yang menghasilkan 77 komentar berlabel positif, 37 komentar berlabel negatif, dan 128 komentar berlabel netral. Data hasil pelabelan ini kemudian digunakan sebagai dasar dalam analisis sentimen untuk mengetahui kecenderungan opini masyarakat terhadap Program Makan Bergizi Gratis.

A. Ekstraksi Fitur

Setelah dilakukan proses *split data* dengan proporsi 80:20 antara data latih dan data uji, tahap selanjutnya adalah ekstraksi fitur. Proses ekstraksi fitur pada penelitian ini dilakukan untuk mengubah data teks hasil preprocessing menjadi bentuk numerik. Teknik ekstraksi fitur yang digunakan meliputi Bag of Words (BoW) dan Term Frequency–Inverse Document Frequency (TF-IDF), yang diimplementasikan menggunakan modul CountVectorizer dan TfidfVectorizer dari pustaka *scikit-learn*.

Penggunaan beberapa teknik ekstraksi fitur bertujuan untuk merepresentasikan karakteristik teks komentar masyarakat dari sudut pandang yang berbeda, baik berdasarkan frekuensi kemunculan kata maupun tingkat kepentingan kata dalam dokumen. Hasil dari masing-masing teknik ekstraksi fitur selanjutnya digunakan sebagai representasi fitur dalam proses analisis sentimen masyarakat terhadap Program Makan Bergizi Gratis. Nilai hasil perhitungan dari setiap metode ekstraksi fitur disajikan dalam bentuk sampel pada tabel 10 berikut.

Term Presence

Tabel 10. Sampel Nilai Fitur dari Setiap Teknik Ekstraksi

No	mbg	makan	anak	program	bantu	...
3	0	1	0	0	0	...
45	1	1	1	1	1	...
90	0	0	0	1	0	...
122	0	0	0	1	0	...
150	0	0	0	1	0	...

Bag of Words

201	0	1	0	0	1	...
77			0	0	0	.
						.
						.
159			0	0	1	...
12			0	0	0	...
...		...	...	...	...	...
228			0	0	1	...

No	mbg	makan	anak	program	bantu	...
3	0	1	0	0	0	...
45	1	1	1	1	1	...
90	0	0	0	1	0	...
122	0	0	0	1	0	...
150	0	0	0	1	0	...
201	0	1	0	0	1	...
77	0	0	0	0	0	...
159	0	0	0	1	1	...
12	0	0	0	0	0	...
...	...	...	...	...	...	...
228	0	0	0	1	0	...

TF-IDF

No	mbg	makan	anak	program	bantu	...
3	0.00	0.36	0.00	0.00	0.00	...
45	0.22	0.31	0.33	0.27	0.29	...
90	0.00	0.00	0.00	0.41	0.00	...
122	0.00	0.00	0.00	0.38	0.00	...
150	0.00	0.00	0.00	0.35	0.00	...
201	0.00	0.34	0.00	0.00	0.28	...
77	0.00	0.00	0.00	0.00	0.00	...
159	0.00	0.00	0.00	0.32	0.30	...
12	0.00	0.00	0.00	0.00	0.00	...
...	...	...	...	...	...	...
228	0.00	0.00	0.00	0.37	0.00	...

Secara keseluruhan, proses ekstraksi fitur menghasilkan sebanyak 1.918 fitur kata unik (*unique features*) yang diperoleh dari seluruh komentar youtube pada data latih dan data uji.

B. Klasifikasi Model

Tahap selanjutnya adalah membangun model klasifikasi untuk mengelompokkan sentimen ke dalam tiga kategori, yaitu positif, negatif, dan netral. Pada tahap ini digunakan algoritma Support Vector Machine (SVM) yang memiliki kemampuan baik dalam menangani data hasil ekstraksi fitur teks berdimensi tinggi. Implementasi model klasifikasi dilakukan menggunakan *library* scikit-learn dengan memanfaatkan fungsi *accuracy_score* dari modul *sklearn.metrics* untuk mengukur performa klasifikasi. Proses pelatihan model dilakukan menggunakan data latih yang telah melalui tahap ekstraksi fitur, baik menggunakan teknik Term Presence, Bag of Words (BoW), maupun TF-IDF.

Klasifikasi sentimen dilakukan secara terpisah untuk masing-masing teknik ekstraksi fitur guna mengetahui pengaruh perbedaan representasi fitur terhadap kinerja model SVM dalam mengklasifikasikan sentimen ke dalam tiga kelas. Hasil prediksi dari proses klasifikasi sentimen menggunakan model SVM disajikan pada Tabel 3.2. Parameter regularisasi pada *classifier* tidak diatur secara khusus dan dibiarkan menggunakan nilai *default*, yaitu $C = 1.0$, guna menjaga konsistensi pengujian pada setiap teknik ekstraksi fitur.

Tabel 11. Sampel Hasil Prediksi Klasifikasi

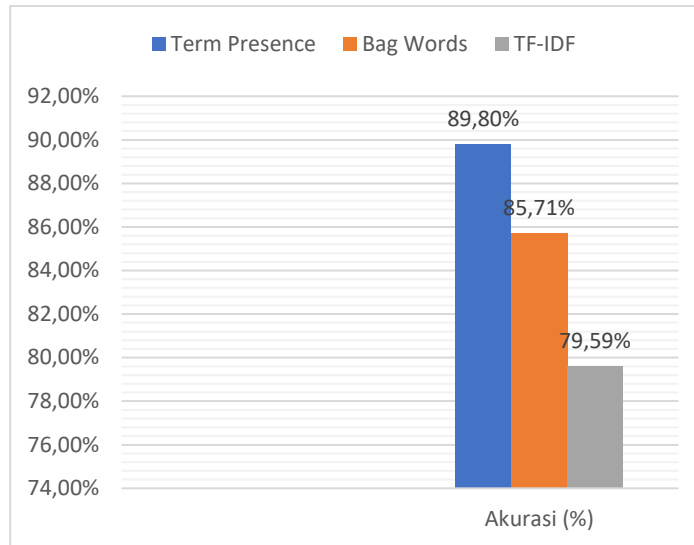
Ekstraksi Fitur	Indeks Data Uji	17	76	47	24	68	21	88	42	...
Term Presence		2	2	1	2	0	1	0	0	...
BoW		2	2	1	2	0	1	0	0	...
TF-IDF		2	2	1	2	0	1	0	0	...

Sementara itu, Tabel 3.3 berikut menyajikan perbandingan skor akurasi klasifikasi sentimen berdasarkan masing-masing teknik ekstraksi fitur.

Tabel 12. Perbandingan Skor Klasifikasi Sentimen

Ekstraksi Fitur	Akurasi (%)
Term Presence	89.80%
Bag of Words (BoW)	85.71%
TF-IDF	79.59%

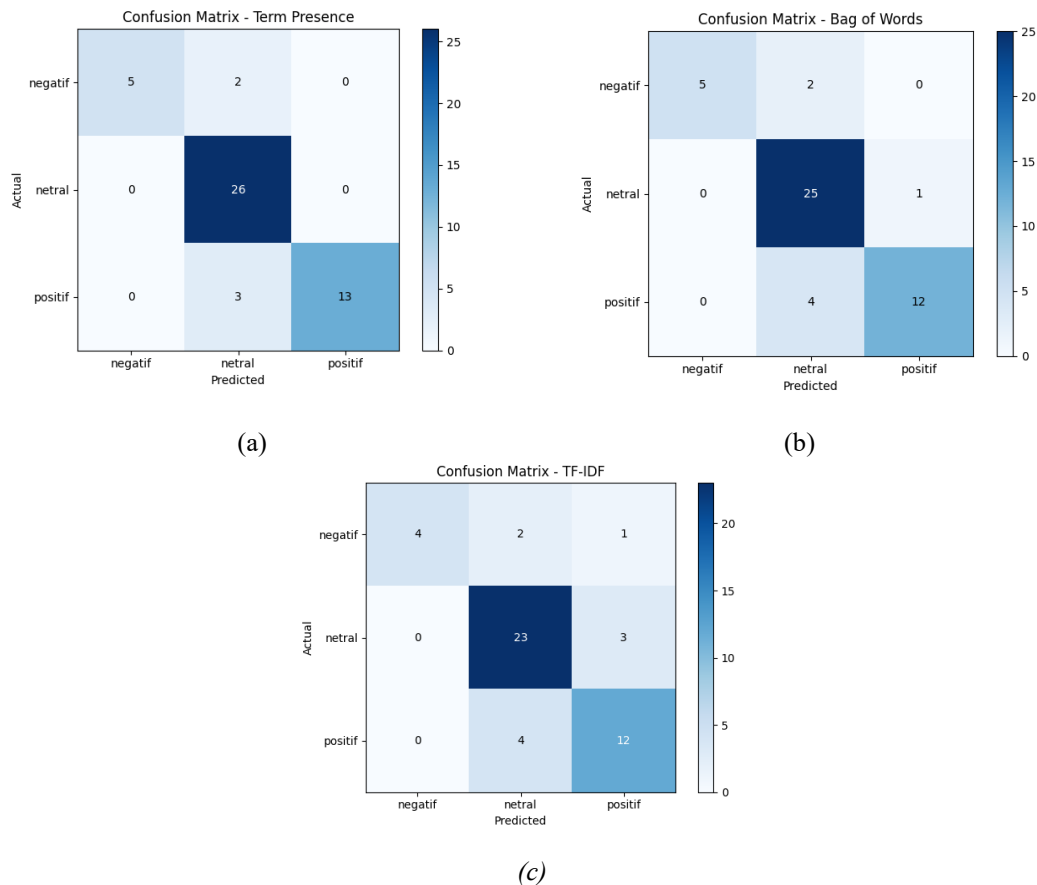
Dari Tabel 3.3, hasil akurasi klasifikasi sentimen menggunakan SVM dari masing-masing ekstraksi fitur menunjukkan bahwa menggabungkan dengan teknik TF-IDF memiliki skor akurasi tertinggi sebesar 89.80%. Teknik *Term Presence* menghasilkan akurasi sebesar 85,71%, dan teknik BoW menghasilkan akurasi sedikit lebih rendah sebesar 79.59%. Berikut gambaran grafik perbandingan skor akurasi klasifikasi sentimen yang disajikan pada Gambar 3 di bawah ini.



Gambar 3. Grafik perbandingan Akurasi

C. Evaluasi Model

Setelah melakukan klasifikasi sentimen menggunakan algoritma SVM, langkah selanjutnya adalah melakukan evaluasi kinerja model dengan menggunakan metode *confusion matrix* dengan pendekatan *single train-test split* berdasarkan masing-masing teknik ekstraksi fitur. Gambar 4 di bawah ini menyajikan visualisasi *confusion matrix* dari hasil klasifikasi yang menunjukkan jumlah prediksi benar (*true positive* dan *true negative*) serta prediksi salah (*false positive* dan *false negative*).



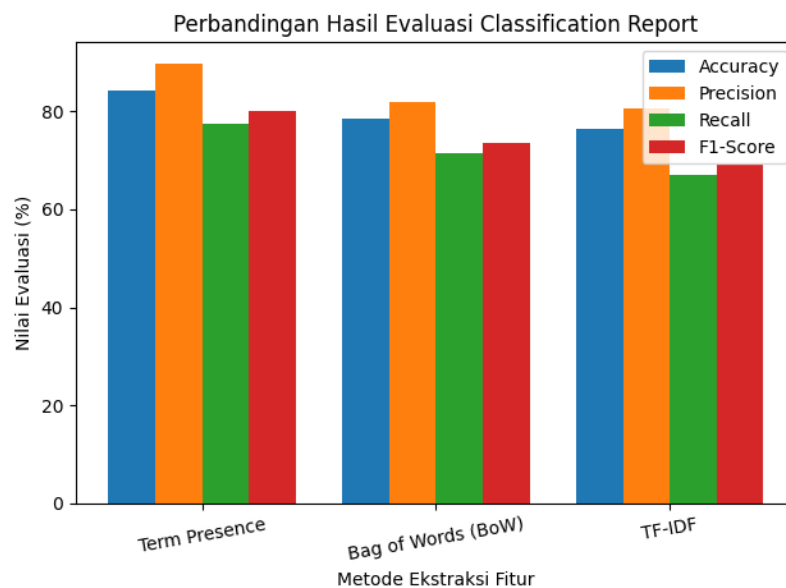
Gambar 4. Visualisasi Confusion Matrix (a) Term Presence, (b) BOW, (C) TF-IDF

- a. Term Presence
Menunjukkan hasil klasifikasi berdasarkan keberadaan kata. Sebagian besar data berhasil diklasifikasikan dengan benar, meskipun masih terdapat beberapa kesalahan prediksi.
- b. Bag of Words (BoW)
Menampilkan hasil klasifikasi berdasarkan frekuensi kata. Model cukup baik mengenali sentimen positif, namun masih terjadi kesalahan pada sentimen negatif.
- c. TF-IDF
Menunjukkan hasil klasifikasi dengan pembobotan kata penting. Prediksi sentimen positif cukup baik, tetapi kesalahan klasifikasi pada sentimen negatif masih terlihat.

Selanjutnya, dilakukan perhitungan metrik evaluasi berupa *Accuracy*, *Precision*, *Recall*, dan *F1-Score* dengan menggunakan modul *classification report*. Nilai-nilai tersebut dirangkum dalam Tabel 13 berikut.

Tabel 13. Hasil Perhitungan classificati

Ekstraksi Fitur	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Term Presence	84.29%	89.64%	77.61%	80.19%
Bag of Words (BoW)	78.50%	81.91%	71.45%	73.52%
TF-IDF	76.40%	80.54%	67.13%	69.16%



Gambar 5. Perbandingan Performa Evaluasi Confusion Matrix

Berdasarkan hasil evaluasi menggunakan *classification report*, dapat disimpulkan bahwa metode Term Presence menghasilkan performa terbaik dalam klasifikasi sentimen dibandingkan dengan dua metode ekstraksi fitur lainnya, yaitu Bag of Words (BoW) dan TF-IDF. Hal ini ditunjukkan oleh nilai accuracy, precision, recall, dan F1-score yang lebih tinggi pada metode Term Presence. Sementara itu, metode BoW dan TF-IDF menunjukkan penurunan performa, khususnya pada metrik recall dan F1-score, yang mengindikasikan keterbatasan model dalam menangkap seluruh data sentimen secara optimal.

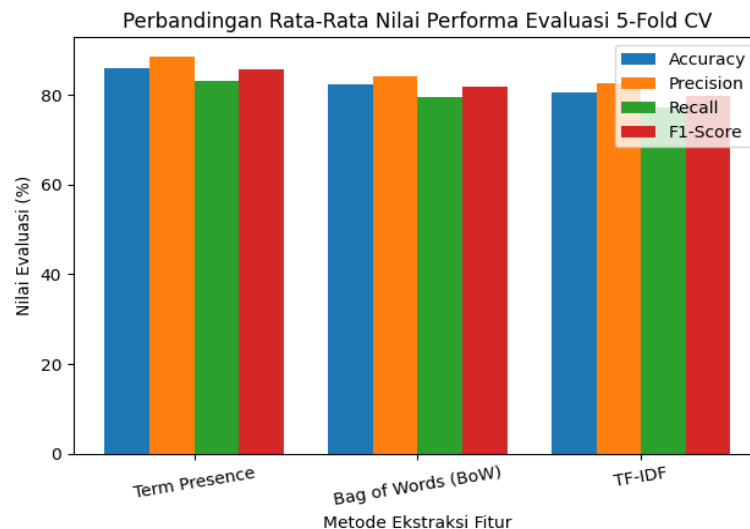
Pada tahap evaluasi model selanjutnya, dilakukan analisis lanjutan untuk memperoleh hasil evaluasi yang lebih akurat dan stabil. Evaluasi ini menggunakan metode 5-Fold Cross Validation yang diimplementasikan dengan fungsi *StratifiedKFold* dari library *scikit-learn*. Dataset dibagi menjadi lima

bagian, di mana setiap bagian secara bergantian digunakan sebagai data uji, sementara empat bagian lainnya digunakan sebagai data latih.

Evaluasi lanjutan ini diterapkan pada data latih hasil pengujian awal yang berjumlah 193 data, dengan skema pembagian 80% data pelatihan dan 20% data pengujian pada setiap iterasi. Namun, dataset menunjukkan adanya ketidakseimbangan kelas (*class imbalance*) yang menyebabkan model cenderung bias terhadap kelas mayoritas. Untuk mengatasi permasalahan tersebut, diterapkan teknik oversampling menggunakan pendekatan SVM-SMOTE melalui library *imbalanced-learn*. Penerapan SVM-SMOTE menghasilkan data sintesis pada area perbatasan antar kelas, sehingga meningkatkan sensitivitas model terhadap kelas minoritas dan memperbaiki performa klasifikasi secara keseluruhan.

Tabel 14. Rata-Rata Nilai Performa Evaluasi 5-Fold CV

Ekstraksi Fitur	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Term Presence	86.12	88.45	83.27	85.78
Bag of Words (BoW)	82.34	84.19	79.65	81.86
TF-IDF	80.47	82.63	77.18	79.81



Gambar 6. Perbandingan Performa Evaluasi SMOTE + 5-Fold CV

Gambar 6 di atas, menyajikan perbandingan hasil evaluasi performa model klasifikasi SVM menggunakan pendekatan 5-Fold Cross Validation yang dikombinasikan dengan SMOTE. Berdasarkan grafik tersebut, metode ekstraksi fitur Term Presence menunjukkan performa terbaik dibandingkan metode Bag of Words (BoW) dan TF-IDF pada seluruh metrik evaluasi. Nilai accuracy, precision, recall, dan F1-score pada metode Term Presence berada pada kisaran 80%–90%, yang menunjukkan bahwa representasi fitur berbasis keberadaan kata mampu memberikan informasi yang efektif dalam proses klasifikasi sentimen.

D. Perbandingan Akurasi Evaluasi

Setelah dilakukan evaluasi model menggunakan dua pendekatan berbeda, yaitu metode evaluasi awal melalui *confusion matrix* dan evaluasi lanjutan menggunakan *5-Fold CV* yang dikombinasikan dengan SMOTE, diperoleh hasil performa yang dapat dibandingkan untuk menilai efektivitas masing-masing pendekatan dalam klasifikasi sentimen. Tabel 15 menampilkan perbandingan akurasi dari kedua evaluasi tersebut.

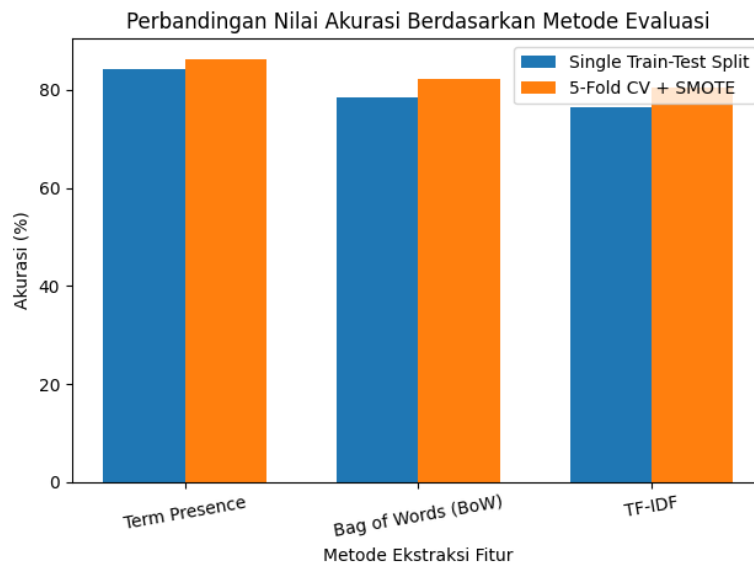
Tabel 15. Perbandingan Nilai Akurasi

Ekstraksi Fitur	Akurasi (%) Confusion Matrix + Single Train-Test	Akurasi (%) Rata-Rata 5-Fold CV + SMOTE
Term Presence	84.29%	86.12%
Bag of Words (BoW)	78.50%	82.34%
TF-IDF	76.40%	80.47%

Hasil evaluasi menunjukkan bahwa pada pendekatan single train-test split, metode ekstraksi fitur Term Presence memperoleh nilai akurasi tertinggi sebesar 84,29%, diikuti oleh Bag of Words (BoW) dengan akurasi 78,50%, dan TF-IDF sebesar 76,40%.

Selanjutnya, ketika evaluasi dilakukan menggunakan rata-rata 5-Fold Cross Validation yang dikombinasikan dengan SMOTE, terjadi peningkatan nilai akurasi pada seluruh metode ekstraksi fitur. Metode Term Presence tetap menunjukkan performa terbaik dengan nilai akurasi 86,12%, disusul oleh BoW sebesar 82,34%, dan TF-IDF sebesar 80,47%.

Peningkatan akurasi pada skema evaluasi 5-Fold Cross Validation + SMOTE menunjukkan bahwa penerapan oversampling mampu membantu model dalam menangani ketidakseimbangan kelas serta meningkatkan kemampuan generalisasi model. Perbandingan nilai akurasi dari masing-masing metode ekstraksi fitur pada kedua pendekatan evaluasi tersebut ditampilkan pada Gambar 7 berikut:



Gambar 7. Grafik Perbandingan Nilai Akurasi Evaluasi

3.2. Pembahasan

Hasil penelitian ini menunjukkan bahwa pemilihan metode ekstraksi fitur memiliki pengaruh signifikan terhadap performa klasifikasi sentimen komentar masyarakat terkait Program Makan Bergizi Gratis (MBG) pada platform YouTube. Temuan ini sejalan dengan penelitian sebelumnya yang menyatakan bahwa representasi fitur merupakan komponen krusial dalam analisis sentimen karena menentukan bagaimana informasi tekstual diterjemahkan ke dalam bentuk numerik yang dapat diproses oleh model machine learning [29].

Distribusi sentimen pada penelitian ini menunjukkan dominasi sentimen netral dibandingkan sentimen positif dan negatif. Temuan ini memiliki perbedaan dengan beberapa penelitian terdahulu yang melaporkan dominasi sentimen positif terhadap kebijakan makan bergizi gratis pada platform media sosial lain seperti X dan Instagram [6], [8], [14]. Perbedaan tersebut diduga disebabkan oleh karakteristik platform YouTube yang memungkinkan pengguna menyampaikan komentar secara lebih panjang, argumentatif, dan

reflektif. Komentar pada YouTube cenderung tidak hanya berisi ekspresi dukungan atau penolakan, tetapi juga memuat saran, evaluasi, serta harapan terhadap implementasi kebijakan, sehingga secara leksikal lebih banyak dikategorikan sebagai sentimen netral.

Dari sisi metode klasifikasi, hasil evaluasi menunjukkan bahwa kombinasi Support Vector Machine (SVM) dengan ekstraksi fitur Term Presence menghasilkan performa yang lebih stabil dibandingkan Bag of Words (BoW) dan TF-IDF, baik pada evaluasi confusion matrix maupun pada skema 5-Fold Cross Validation yang dikombinasikan dengan SMOTE. Temuan ini berbeda dengan beberapa penelitian sebelumnya yang menyatakan bahwa TF-IDF umumnya memberikan performa yang lebih unggul dalam tugas analisis sentimen teks [29]. Namun demikian, hasil penelitian ini sejalan dengan studi yang menunjukkan bahwa pada data media sosial dengan kosakata sentimen yang eksplisit, keberadaan kata kunci lebih berpengaruh dibandingkan bobot frekuensi kata [29].

Perbedaan performa tersebut dapat dijelaskan melalui karakteristik data komentar YouTube yang digunakan dalam penelitian ini. Banyak komentar mengandung kata-kata kunci sentimen yang secara langsung merepresentasikan opini masyarakat, seperti kata yang menunjukkan dukungan, kritik, maupun kekhawatiran terhadap aspek implementasi program MBG. Dalam konteks seperti ini, pendekatan Term Presence mampu menangkap sinyal sentimen secara efektif tanpa dipengaruhi oleh distribusi frekuensi kata yang cenderung tidak merata, sebagaimana terjadi pada metode TF-IDF.

Selain itu, penerapan teknik Synthetic Minority Oversampling Technique (SMOTE) yang dikombinasikan dengan 5-Fold Cross Validation terbukti meningkatkan kestabilan dan generalisasi model. Temuan ini mendukung penelitian sebelumnya yang menyatakan bahwa permasalahan ketidakseimbangan kelas (class imbalance) merupakan tantangan umum dalam analisis sentimen, khususnya pada data media sosial, dan dapat menyebabkan bias model terhadap kelas mayoritas apabila tidak ditangani dengan baik. Dengan penerapan SMOTE, model memperoleh representasi sintesis dari kelas minoritas sehingga mampu mempelajari pola sentimen negatif dan positif secara lebih seimbang.

Hasil evaluasi lanjutan menggunakan cross validation juga menunjukkan bahwa performa model tidak hanya bergantung pada satu skema pembagian data latih dan data uji. Pendekatan ini sejalan dengan penelitian sebelumnya yang menekankan pentingnya validasi berulang untuk memperoleh estimasi performa model yang lebih stabil dan dapat digeneralisasikan. Dengan demikian, hasil penelitian ini memberikan bukti empiris bahwa kombinasi strategi evaluasi yang tepat dan penanganan ketidakseimbangan kelas dapat meningkatkan kualitas analisis sentimen berbasis machine learning.

Secara keseluruhan, penelitian ini tidak hanya mengonfirmasi temuan-temuan sebelumnya terkait pentingnya pemilihan metode ekstraksi fitur dan evaluasi model, tetapi juga memberikan kontribusi baru dengan menunjukkan bahwa efektivitas suatu metode sangat bergantung pada karakteristik data dan platform yang digunakan. Oleh karena itu, hasil penelitian ini dapat menjadi referensi bagi penelitian selanjutnya dalam menentukan pendekatan analisis sentimen yang sesuai, khususnya dalam konteks evaluasi kebijakan publik berbasis opini masyarakat di media sosial.

4. KESIMPULAN

Berdasarkan hasil analisis sentimen masyarakat terhadap Program Makan Bergizi Gratis (MBG) pada platform YouTube, dapat disimpulkan bahwa dari total 242 komentar yang dianalisis, sentimen netral merupakan kategori yang paling dominan dengan persentase sebesar 52,89%, diikuti oleh sentimen positif sebesar 31,82%, dan sentimen negatif sebesar 15,29%. Dominasi sentimen netral menunjukkan bahwa sebagian besar masyarakat belum sepenuhnya bersikap mendukung atau menolak program MBG, melainkan cenderung memberikan evaluasi, saran, serta harapan terkait pelaksanaan dan keberlanjutan program di lapangan.

Dari sisi metodologi, penelitian ini membuktikan bahwa algoritma Support Vector Machine (SVM) mampu mengklasifikasikan sentimen komentar masyarakat dengan baik ketika dikombinasikan dengan teknik ekstraksi fitur yang sesuai. Hasil evaluasi menunjukkan bahwa metode Term Presence menghasilkan performa klasifikasi terbaik dan paling stabil dibandingkan Bag of Words (BoW) dan Term Frequency–Inverse Document Frequency (TF-IDF). Pada evaluasi menggunakan skema 5-Fold Cross Validation yang dikombinasikan dengan Synthetic Minority Oversampling Technique (SMOTE), metode Term Presence

mencapai nilai akurasi rata-rata sebesar 86,12%, lebih tinggi dibandingkan BoW sebesar 82,34% dan TF-IDF sebesar 80,47%.

Penerapan teknik SMOTE terbukti efektif dalam mengatasi permasalahan ketidakseimbangan kelas pada dataset, sehingga meningkatkan kemampuan model dalam mengenali sentimen minoritas, khususnya sentimen negatif. Hal ini menunjukkan bahwa penanganan class imbalance serta pemilihan strategi evaluasi yang tepat memiliki peran penting dalam meningkatkan kualitas dan keandalan model analisis sentimen berbasis machine learning.

Secara keseluruhan, penelitian ini memberikan gambaran kuantitatif mengenai persepsi masyarakat terhadap Program Makan Bergizi Gratis serta menunjukkan bahwa karakteristik platform media sosial dan metode analisis yang digunakan sangat memengaruhi hasil klasifikasi sentimen. Temuan penelitian ini diharapkan dapat menjadi bahan pertimbangan bagi pemerintah dalam mengevaluasi dan menyempurnakan implementasi program MBG, serta menjadi referensi bagi penelitian selanjutnya dengan cakupan data, platform, dan metode yang lebih luas.

REFERENSI

- [1] M. Nango, B. Purwoko, and M. Hazim, "Makan Bergizi Gratis : Strategi Kebijakan Publik Menuju Generasi Emas Indonesia 2045," *Oktober*, vol. 25, no. 2, pp. 2085–0743, 2025, [Online]. Available: <https://doi.org/10.34150/jpak.v22i1.991>
- [2] A. A. Zerga, S. E. Tadesse, F. Y. Ayele, and S. Z. Ayele, "Impact of malnutrition on the academic performance of school children in Ethiopia: A systematic review and meta-analysis," *SAGE Open Med.*, vol. 10, 2022, doi: 10.1177/20503121221122398.
- [3] A. Sarjito, "Free Nutritious Meal Program as a Human Resource Development Strategy to Support National Defence," *Int. J. Adm. Bus. Organ.*, vol. 5, no. 5, pp. 129–141, 2024, doi: 10.61242/ijabo.24.454.
- [4] N. Wulandari, A. Margawati, and Z. Rahfiludin, "The implementation of nutrition improvement programs for underweight children, wasting and stunting in the Department of Health, Central Buton district, Southeast Sulawesi," *J. Gizi Indones. (The Indones. J. Nutr.)*, vol. 9, no. 2, pp. 86–96, 2021, doi: 10.14710/jgi.9.2.86-96.
- [5] S. Zaleha and H. Idris, "Implementation of Stunting Program in Indonesia: a Narrative Review," *Indones. J. Heal. Adm.*, vol. 10, no. 1, pp. 143–151, 2022, doi: 10.20473/jaki.v10i1.2022.143-151.
- [6] J. Ilmiah, S. Informasi, J. Vallen, and A. H. Azizah, "Analisis Sentimen Media Sosial X Program Makanan Sehat Gratis dengan Support Vector Machine dan Naive Bayes tentang berbagai topik . Pendapat dapat mengambil berbagai bentuk , mulai dari kritik , pujian , adalah Twitter (X) , merupakan sosial media yang ," vol. 5, no. September, 2025.
- [7] E. Setiadi and D. A. Sulisty, "Analisis Sentimen Kebijakan Makan Bergizi Gratis Menggunakan IndoBERT dan Machine Learning," vol. 15, no. 3, pp. 507–513, 2025.
- [8] Fikry Destryanto, Putri Rizqiyah, and Petrus Sokibi, "Sentiment Analysis of Public Response to the Free Nutritious Meal Program on Instagram Using IndoBERT," *J. Artif. Intell. Eng. Appl.*, vol. 5, no. 1, pp. 1924–1928, 2025, doi: 10.59934/jaiea.v5i1.1755.
- [9] R. Saputra and F. N. Hasan, "Analisis Sentimen Terhadap Program Makan Siang & Susu Gratis Menggunakan Algoritma Naive Bayes," *J. Teknol. Dan Sist. Inf. Bisnis*, vol. 6, no. 3, pp. 411–419, 2024, doi: 10.47233/jteksis.v6i3.1378.
- [10] H. G. Sinaga and M. H. Syarief, "Impact of Free Nutritious Lunch Program on Student Well-Being and Learning Achievement in Indonesia: Implementation by 2025," *Glob. J. Econ. Financ. Res.*, vol. 02, no. 01, pp. 76–83, 2025, doi: 10.55677/gjefr/09-2025-vol02e1.
- [11] D. W. Brata, W. Purnomo, H. Farisi, and H. Kevin Marcelino Ratu, "Menggali Opini Publik: Sentimen Terhadap Kebijakan Makan Siang Gratis Dengan Supervised Learning," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 12, no. 3, pp. 533–538, 2025, doi: 10.25126/jtiik.20259565.
- [12] M. A. Gumilang, F. Abdillah, M. Y. Amin, and M. Hasan, "Sentiment Analysis of Indonesian Ministries Social Media: Citizen Responses Utilizing TextBlob Analyser," *J. Sositologi*, vol. 23, no. 2, pp. 203–216, 2024, doi: 10.5614/sostek.itbj.2024.23.2.5.

- [13] A. F. Ningrum and I. F. Amri, "Sentiment Analysis Of Public Opinion On Handling Stunting In Indonesia Using Random Forest," *J Stat. J. Ilm. Teor. dan Apl. Stat.*, vol. 17, no. 1, pp. 645–654, 2024, doi: 10.36456/jstat.vol17.no1.a9088.
- [14] N. N. Aprianti *et al.*, "Public Sentiment Analysis of the Free Nutritious Meals Program (MBG) on Social Media X Using the Naive Bayes Method," vol. 9, no. 6, pp. 3929–3936, 2025.
- [15] Sekar Cinta Amaria and Nurtriana Hidayati, "ANALISIS PROGRAM MAKAN BERGIZI GRATIS DENGAN *SUPPORT VECTOR MACHINE* (SVM) PADA APLIKASI X," *JSiI (Jurnal Sist. Informasi)*, vol. 12, no. 2, pp. 16–24, 2025, doi: 10.30656/jsii.v12i2.10708.
- [16] M. Daffa, A. Fahreza, A. Luthfiarta, M. Rafid, M. Indrawan, and A. Nugraha, "JOURNAL OF APPLIED COMPUTER SCIENCE AND TECHNOLOGY (JACOST) Analisis Sentimen : Pengaruh Jam Kerja Terhadap Kesehatan Mental Generasi Z," vol. 5, no. 1, pp. 16–25, 2024.
- [17] I Kadek Nicko Ananda, Ni Putu Novita Puspa Dewi, Ni Wayan Marti, and Luh Joni Erawati Dewi, "Klasifikasi Multilabel Pada Gaya Belajar Siswa Sekolah Dasar Menggunakan Algoritma Machine Learning," *J. Appl. Comput. Sci. Technol.*, vol. 5, no. 2, pp. 144–154, 2024, doi: 10.52158/jacost.v5i2.940.
- [18] P. H. Prastyo, A. S. Sumi, A. W. Dian, and A. E. Permanasari, "Tweets Responding to the Indonesian Government's Handling of COVID-19: Sentiment Analysis Using SVM with Normalized Poly Kernel," *J. Inf. Syst. Eng. Bus. Intell.*, vol. 6, no. 2, p. 112, 2020, doi: 10.20473/jisebi.6.2.112-122.
- [19] N. Anisa, "Penerapan Lexicon Based Untuk Analisis Sentimen pada Game PUBG dengan Ekstraksi Fitur TF-IDF," *J. Pract. Comput. Sci.*, vol. 5, no. 1, pp. 35–41, 2025, doi: 10.37366/jpcs.v5i1.5826.
- [20] M. F. Firdaus, D. E. Ratnawati, and N. Y. Setiawan, "Analisis Sentimen Berbasis Aspek Pada Ulasan Pelanggan Restoran Menggunakan Algoritma Support Vector Machine (Studi Kasus: Depot Bamara)," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 11, no. 6, pp. 1265–1272, 2024, doi: 10.25126/jtiik.2024117564.
- [21] W. F. Abdillah, A. Premana, and R. M. H. Bhakti, "Analisis Sentimen Penanganan Covid-19 dengan Support Vector Machine: Evaluasi Leksikon dan Metode Ekstraksi Fitur," *J. Ilm. Intech Inf. Technol. J. UMUS*, vol. 3, no. 02, pp. 160–170, 2021, doi: 10.46772/intech.v3i02.556.
- [22] N. L. P. R. Dewi, I. N. S. W. Wijaya, I. K. Purnamawan, and N. W. Marti, "Model *Classifer* Judul Berita Pariwisata Indonesia Berdasarkan Sentimen," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 11, no. 1, pp. 117–124, 2024, doi: 10.25126/jtiik.2024117617.
- [23] R. A. Husen, A. Agustin, S. Erlinda, J. Junadhi, and T. Perumal, "Sentiment Analysis of Societal Attitudes Toward the Childfree Lifestyle Using Latent Dirichlet Allocation and Support Vector Machines," *Innov. Res. Informatics*, vol. 7, no. 1, pp. 1–8, 2025, doi: 10.37058/innovatics.v7i1.12005.
- [24] Filan Firmansyah, Saputra Dwi Nurchaya, and Zuhana Realita Alfy, "Implementasi Algoritma Blowfish Pada Sistem Manajemen Surat Dengan Pendekatan Rational Unified Process Yang Ramah Lingkungan," *JSiI (Jurnal Sist. Informasi)*, vol. 11, no. 2, pp. 1–6, 2024, doi: 10.30656/jsii.v11i2.9065.
- [25] S. Suryanto and W. Andriyani, "Sentiment Analysis of X Platform on Viral 'Fufufafa' Account Issue in Indonesia Using SVM," *IJCCS (Indonesian J. Comput. Cybern. Syst.)*, vol. 19, no. 1, p. 95, 2025, doi: 10.22146/ijccs.104158.
- [26] D. R. Nurqotimah, A. N. Khudori, and R. S. Pradini, "JOURNAL OF APPLIED COMPUTER SCIENCE AND TECHNOLOGY (JACOST) Implementasi Algoritma Support Vector Machine (SVM) Untuk Klasifikasi Penyakit Stroke," vol. 5, no. 2, pp. 179–185, 2024.
- [27] M. Samjas, A. Darmawan, A. Info, L. D. Allocation, S. Analysis, and S. V. Machine, "Analysis of Spotify User Sentiment to Improve Customer Satisfaction Using Opinion Mining and Latent Dirichlet Allocation Based on E-Satisfaction Dimensions," vol. 8, no. 3, pp. 776–787, 2025.
- [28] R. R. R. Arisandi, B. Warsito, and A. R. Hakim, "Aplikasi Naïve Bayes Classifier (Nbc) Pada Klasifikasi Status Gizi Balita Stunting Dengan Pengujian K-Fold Cross Validation," *J. Gaussian*, vol. 11, no. 1, pp. 130–139, 2022, doi: 10.14710/j.gauss.v11i1.33991.
- [29] K. Sentimen, A. Davina, M. Putri, N. Sulistianingsih, and R. Rismayati, "JTIM : Jurnal Teknologi

Informasi dan Multimedia Pengaruh Teknik Representasi Teks Bag-of-Words dan TF-IDF,” vol. 7, no. 4, pp. 675–688, 2025.