



# Analisis Sentimen Persepsi Masyarakat pada Media Sosial X Menggunakan Metode *K-Means*, *SVM*, dan Teknik *Smote*

(Studi Kasus: Program Makan Bergizi Gratis)

Rivana Delfalizanty<sup>1\*</sup>, Yulhendri<sup>2</sup>

<sup>1,2</sup> Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Esa Unggul, Indonesia

E-mail: [rdelfalizanty@gmail.com](mailto:rdelfalizanty@gmail.com)<sup>1\*</sup>, [yulhendri@esaunggul.ac.id](mailto:yulhendri@esaunggul.ac.id)<sup>2</sup>

\*Penulis Korespondensi: [rdelfalizanty@gmail.com](mailto:rdelfalizanty@gmail.com)

**Abstract.** The Free Nutritious Meal (MBG) Program has emerged as a major public policy issue frequently discussed on social media platform X (Twitter). Public discourse surrounding this program reflects diverse perspectives, including support, criticism, and broader debates on national nutrition policies. This study seeks to identify the dominant topics within these conversations through *K-Means* clustering and to classify sentiments using the Support Vector Machine (*SVM*) algorithm, enhanced with SMOTE to address class imbalance. A total of 3,053 tweets were collected through crawling. The clustering results revealed three main themes: (1) nutrition fulfillment and food access for priority groups, highlighted by keywords such as nutrition, milk, fish, toddlers, schools, and pesantren; (2) political dynamics and program legitimacy, represented by terms like stunting, support, president, Prabowo, society, and national; and (3) state financial concerns, reflected in words including tax, funds, state budget, cost, people, and expenditure. In sentiment analysis, *SVM* without SMOTE produced accuracies of 65.69%, 67.76%, and 68.12% under 90:10, 80:20, and 70:30 split ratios. After applying SMOTE, accuracy increased to 77.94%, 76.38%, and 75.06%, with an F1-Score of 78%. These results confirm that *K-Means* is effective in identifying discussion topics, while SMOTE enhances *SVM* performance in sentiment classification.

**Keywords:** *K-Means Clustering*; MBG; *Sentiment Analysis*; SMOTE; *SVM*.

**Abstrak.** Program Makan Bergizi Gratis (MBG) menjadi isu kebijakan publik yang banyak diperbincangkan di media sosial X (Twitter). Tweet yang membahas program ini mencerminkan opini publik yang beragam, baik berupa dukungan, kritik, maupun diskusi terkait kebijakan gizi nasional. Penelitian ini bertujuan untuk mengidentifikasi topik utama dalam percakapan publik menggunakan *K-Means clustering* dan menganalisis sentimen tweet dengan algoritma *Support Vector Machine* (*SVM*) yang dikombinasikan dengan SMOTE untuk mengatasi ketidakseimbangan data. Dataset berjumlah 3.053 tweet yang diperoleh melalui proses *crawling*. Hasil klasterisasi *K-Means* menghasilkan tiga topik utama, yaitu: (1) pemenuhan gizi dan distribusi pangan untuk kelompok prioritas dengan kata dominan gizi, susu, ikan, balita, sekolah, dan pesantren; (2) dukungan politik serta legitimasi program melalui kata seperti stunting, dukung, presiden, Prabowo, masyarakat, dan nasional; serta (3) isu beban keuangan negara dan alokasi anggaran dengan kata pajak, dana, biaya, rakyat, dan anggaran. Pada klasifikasi sentimen, akurasi *SVM* tanpa SMOTE mencapai 65,69%, 67,76%, dan 68,12% pada skema 90:10, 80:20, dan 70:30. Setelah penerapan SMOTE, akurasi meningkat menjadi 77,94%, 76,38%, dan 75,06%, dengan F1-Score sebesar 78%. Temuan ini menunjukkan bahwa *K-Means* efektif untuk mengungkap topik pembahasan, sementara SMOTE meningkatkan kinerja *SVM* dalam klasifikasi sentimen.

**Kata kunci:** Analisis Sentimen; *K-Means Clustering*; MBG; SMOTE; *SVM*.

## 1. PENDAHULUAN

Pesatnya perkembangan teknologi informasi dan digitalisasi telah menjadikan internet serta media sosial sebagai kebutuhan yang tidak dapat dipisahkan dalam kehidupan masyarakat modern (Indah Paramitha et al., 2023). Seiring dengan kemajuan teknologi informasi, platform media sosial seperti X, Instagram, dan Facebook mengalami pertumbuhan yang signifikan, menjadikannya pusat aktivitas sosial baru. Melalui platform media sosial tersebut, pengguna tidak hanya terhubung dengan teman dan keluarga, tetapi juga aktif berpartisipasi dalam berbagai diskusi publik, mengekspresikan pendapat, serta memberikan umpan balik terhadap

produk, layanan, dan organisasi (Xu et al., 2022).

*We Are Social Digital* menerbitkan data dalam Suryana & SIK (2024), yang memperlihatkan data pada Januari 2024 bahwa populasi Indonesia terlampir mencapai 278,7 juta jiwa, di mana pada mulai bulan Januari 2024 di antaranya terdapat 185,3 juta sudah terhubung ke internet. Berdasarkan total yang terlihat, ada sejumlah 139 juta jiwa tercatat sebagai pemakai aktif di media sosial, yang sebanding dengan total populasi di Indonesia sebesar 49,9%. Di Indoensia banyak pemakai aktif media sosial yang sedang diantaranya meliputi Facebook dengan 117,6 juta pengguna, diikuti oleh YouTube yang mencapai 139 juta pengguna. Instagram berada di urutan selanjutnya dengan 100,9 juta pengguna, sementara X mencatatkan jumlah pengguna sebesar 24,69 juta.

Pada akhir-akhir tahun terjadi peningkatan penggunaan media sosial, karena penomena ini membawa dampak yang besar kepada kualitas kualitas penyebaran informasi dan opini publik. Kemudahan dalam menuliskan pesan dan opini dalam bentuk status di media sosial membuat pengunaan media soisal menjadi populer di berbagai kalangan masyarakat. Di Indonesia *platform X* menjadi salah satu aplikasi yang menjadi kegemaran dan disukai para kalangan, karena memungkinkan penggunanya berinteraksi secara personal maupun terbuka. Beragam topik mulai dari politik, ekonomi, sosial, budaya, hingga hukum sering kali menjadi bahan perbincangan di platform X (Arsi & Waluyo, 2021, mengutip Teran & Mancera, 2019). Keanekaragaman topik dan usia pengguna menghasilkan berbagai jenis pesan, seperti saran, kritik, atau komentar, yang dapat dianalisis untuk memahami pandangan publik terhadap suatu isu.

Analisis sentimen merupakan metode yang termasuk efektif dalam mengkaji prepsesi publik dengan memanfaatkan data yang diperoleh dari media sosial, memungkinkan melihat sudut pandang yang beragam terhadap beragam jenis konteks, seperti ulasan mengenai barang atau penjualan hingga isu-isu sosial dan politik (Permata, 2024). Dalam konteks ini, salah satu isu yang cukup ramai dibicarakan di media sosial adalah Program Makan Bergizi Gratis (MBG) yang disampaikan oleh Presiden Prabowo Subianto. Melalui analisis sentimen, opini publik terkait program tersebut dapat dikaji untuk memahami persepsi masyarakat terhadap Program Makan Bergizi Gratis (MBG).

Indonesia menghadapi masalah stunting yang serius, dengan lebih dari sepertiga anak yang berumur dibawah 5 tahun sudah merasakan kekurangan tinggi badan akibat gizi tidak baik (Dekasari & Gunawan, 2024). Stunting terjadi akibat kekurangan gizi dalam seribu hari pertama kehidupan (HPK), membuat anak-anak menjadi sering sakit-sakitan dan berdampak pada kualitas di masa dewasa. Secara mendalam, stunting tidak hanya berdampak pada

kesehatan individu, namun juga awal terciptanya hambatan dalam pertumbuhan ekonomi, menaikkan angka kemiskinan, serta membuka lebar ketimpangan sosial di masyarakat (Tarmizi, 2024). Berdasarkan Survei Kesehatan Indonesia 2023, Di Indonesia masih terjadi prevalensi stunting dengan persentase tinggi, yaitu sebesar 21,5% dengan penurunan yang sangat kecil dari angka 21,6% pada 2022. Kondisi ini mengindikasikan bahwa progres penurunan stunting berjalan lambat dan upaya yang dilakukan masih belum dapat meraih tujuan nasional yaitu mencapai target sebesar 14%.

Menghadapi kondisi tersebut, Pemerintah Indonesia melalui Program MBG yang disampaikan oleh Presiden Prabowo Subianto berupaya untuk mengatasi permasalahan gizi dan stunting, terutama pada kelompok rentan seperti ibu hamil , ibu menyusui, anak-anak sekolah, dan balita. Program ini bertujuan untuk memastikan bahwa kebutuhan gizi harian masyarakat, khususnya anak-anak dan ibu, dapat tercukupi sesuai dengan standar Angka Kecukupan Gizi (AKG). Selain itu, Program MBG juga difokuskan pada peningkatan prestasi belajar siswa

melalui penyediaan makanan bergizi di sekolah. Dengan adanya makanan sehat yang diberikan di sekolah, berharap para siswa dapat meningkatkan konsentrasi dan partisipasi dalam kegiatan belajar. Pelaksanaan program ini diharapkan dapat menurunkan angka stunting, memperbaiki daya tahan tubuh anak-anak, serta mendukung peningkatan kualitas sumber daya manusia yang menjadi fondasi utama untuk mewujudkan visi Indonesia Emas 2045.

Tantangan dalam implementasi Program Makan Bergizi Gratis (MBG) mencakup aspek logistik, distribusi, dan penyediaan bahan pangan yang melibatkan berbagai pihak, termasuk petani lokal, UMKM, dan institusi pemerintah. Analisis data serta umpan balik dari masyarakat menjadi langkah penting dalam mengidentifikasi ketidaksesuaian yang terjadi selama pelaksanaan program, sehingga perbaikan dapat dilakukan secara tepat. Dalam pengujian ini, analisis sentimen terhadap persepsi rakyat Indonesia mengenai Program MBG menjadi sangat penting dalam memahami persepsi dan sikap masyarakat perihal program tersebut.

## **2. TINJAUAN PUSTAKA**

### **Analisis Sentimen**

Analisis sentimen atau yang sering disebut sebagai *opinion mining*, merupakan bagian dari cabang *text mining* yang bertujuan untuk mengidentifikasi persepsi atau subjektivitas publik terhadap suatu kejadian, isu atau topik tertentu. Dengan memanfaatkan pendekatan

berbasis teks, analisis sentimen menerapkan metode *machine learning* untuk mengklasifikasikan sentimen yang terkandung dalam berbagai bentuk teks, seperti kalimat, opini, atau dokumen. Sentimen yang teridentifikasi umumnya dikelompokkan ke dalam tiga kategori: positif, negatif, atau netral (Wiliani et al., 2023).

## **Media Sosial**

Media sosial merupakan terminologi populer yang digunakan oleh berbagai generasi. Melalui media sosial, kegiatan yang sebelumnya dilakukan secara langsung atau tatap muka kini dapat dilakukan secara virtual, baik secara sinkron maupun asinkron (Abdillah, 2022). Media sosial berguna sebagai prasarana yang mendukung interaksi antar pengguna dengan kemampuan komunikasi interaktif. Media sosial kerap digunakan sebagai sarana dalam mengembangkan identitas digital seseorang, sekaligus dimanfaatkan oleh pelaku bisnis sebagai sumber berita maupun referensi informasi (Situmorang & Hayati, 2023).

## **Aplikasi X**

Twitter adalah podium media sosial dan *mikroblogging* yang memudahkan untuk berbagi informasi antar penggunanya dan berinteraksi melalui “*tweet*” atau yang dikenal sebagai pesan pendek. Layanan ini berperan sebagai sumber informasi secara langsung dan real-time yang menghubungkan pengguna dengan berbagai berita, opini, dan peristiwa yang sedang ramai diperbincangkan. Pengguna yang telah terdaftar memiliki akses untuk memposting, menyukai, dan me-retweet *tweet*, sementara bagi pemakai yang tidak masuk ke dalam akun dan tidak terdaftar hanya dapat melihat konten tanpa kemampuan untuk berinteraksi. Perusahaan Twitter, Inc. bermakas di San Francisco, California, serta terdapat 25 lebih kantor yang terdapat di seluruh penjuru. Pada awal, Twitter menerapkan batasan 140 karakter untuk setiap tweet, namun sejak November 2017, batas ini ditingkatkan menjadi 280 karakter bagi bahasa selain CJK. Sementara itu, durasi untuk konten audio dan video tetap dibatasi hingga 140 detik bagi sebagian besar akun (Efendi & Lutfianingsih, 2023).

## **Program Makan Bergizi Gratis**

Indonesia diproyeksikan menjadi negara yang tangguh, mandiri, dan inklusif pada tahun 2045, dengan visi mewujudkan kedaulatan, kemajuan, dan keberlanjutan. Salah satu agenda utama dalam pencapaian Indonesia Emas 2045 adalah transformasi sosial, yang mencakup pembangunan di berbagai sektor, termasuk pemerataan layanan kesehatan bagi seluruh masyarakat. Kesehatan dan ketahanan pangan menjadi aspek krusial dalam upaya mewujudkan Indonesia Emas 2045, karena berperan dalam memastikan ketersediaan dan kualitas pangan yang merata di seluruh wilayah. Dengan demikian, kebutuhan gizi masyarakat dapat terpenuhi, sehingga dapat menjalani kehidupan yang sehat, dinamis, dan produktif (Rima Pratiwi

Batubara et al., 2024)

### K-Means Clustering

*K-Means Clustering* merupakan salah satu metode yang dimanfaatkan dalam mengumpulkan informasi ke berbagai klaster berlandaskan kemiripan ciri khas antar informasi. Algoritma *K-Means Clustering* bersifat *hard partition*, artinya setiap data hanya akan menjadi anggota dari satu klaster saja tanpa tumpang tindih dengan klaster lain. Proses pengelompokan diawali dengan penentuan pusat klaster (*centroid*) secara acak, kemudian diikuti dengan penghitungan jarak setiap data terhadap masing-masing pusat klaster. Data akan dikelompokkan ke klaster yang memiliki pusat terdekat. Tahapan ini dijalankan secara iteratif sampai tidak terjadi lagi perubahan yang berarti pada posisi pusat klaster maupun status keanggotaan setiap data. Algoritma *K-Means* sering digunakan karena memiliki proses yang sederhana namun efektif dalam mengelompokkan data secara sistematis (Rahmaddeni et al., 2024).

### SVM (Support Vector Machine)

SVM adalah metode pembelajaran mesin (*machine learning*) yang digunakan secara luas untuk menyelesaikan permasalahan regresi maupun klasifikasi. Dalam konteks pengelompokan data yang melibatkan lebih dari dua kelas, metode SVM menerapkan strategi seperti *One-vs-Rest (OvR)* dan *One-vs- One (OvO)* guna membedakan tiap kategori secara terstruktur dan presisi (Mada Sanjaya, 2024). SVM dirancang untuk mencari *hyperplane* yang paling optimal dalam memisahkan dua kelompok data, dengan memperhitungkan *margin* terbesar antara *hyperplane* dan titik-titik data paling dekat untuk tiap-tiap kelas yang dikenal sebagai *support vector*. Pada dasarnya, SVM merupakan *linear classifier*, namun dikembangkan lebih lanjut dengan konsep *kernel trick* agar dapat menangani data nonlinier dengan cara memproyeksikan data ke ruang berdimensi lebih tinggi sehingga memberikan kesempatan terciptanya pemisahan linier di ruang tersebut (Jollyta et al., 2023).

### SMOTE (Syntethic Minority Oversampling Technique)

SMOTE merupakan pendekatan metode oversampling yang dirancang secara khusus upaya menanggulangi ketidakseimbangan kelas pada data, terutama dalam konteks klasifikasi. Tidak seperti metode oversampling tradisional yang hanya menggandakan sampel dari kelas minoritas, SMOTE menghasilkan data sintetis baru dengan pendekatan interpolatif. Proses ini dilakukan dengan cara membuat titik data baru yang berada di antara dua sampel minoritas yang sudah ada, sehingga informasi yang dihasilkan tidak bersifat duplikatif, namun menghasilkan variasi baru yang masih sejalan dengan distribusi data minoritas (Rianto & Santosa, 2025).

### 3. METODE PENELITIAN

Penelitian ini dilakukan selama 7 bulan. Objek dalam penelitian ini berupa unggahan di media sosial X yang memuat opini publik mengenai Program Makan Bergizi Gratis. Postingan tersebut dianalisis untuk melihat respons dan pandangan masyarakat terhadap program tersebut pada masa awal pemerintahan baru. Peneliti melakukan pencarian dataset yang tepat untuk membantu proses klasifikasi dan klasterisasi model yang direncanakan. Sumber literatur seperti jurnal ilmiah, e-book, artikel, skripsi, dan situs web tepercaya turut dimanfaatkan untuk membangun landasan teori yang kokoh serta memperdalam pemahaman terkait topik yang dikaji.

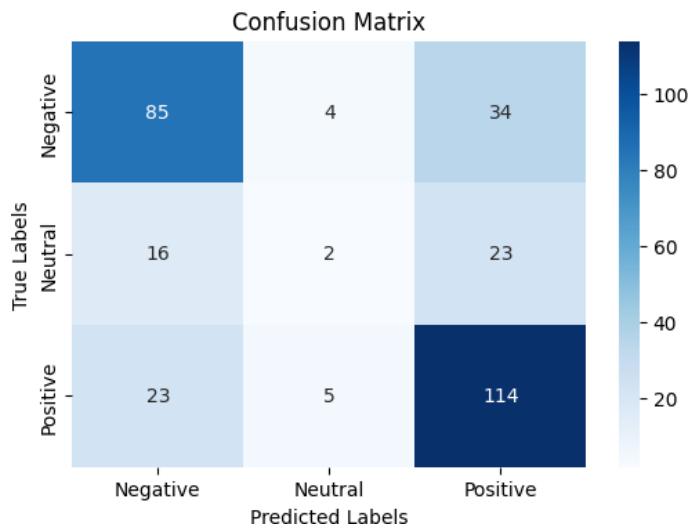
### 4. HASIL DAN PEMBAHASAN

#### Klasifikasi Sentimen

##### *Hasil Klasifikasi SVM*

###### 1. Hasil Klasifikasi SVM dengan Skema 1

Tabel berikut menampilkan confusion matrix yang dihasilkan dari pengujian model dengan pembagian dataset menggunakan rasio 90% untuk data latih dan 10% untuk data uji.



**Gambar 1.** Hasil Confusion Matrix Skema 1.

Hasil analisis terhadap *confusion matrix* menunjukkan bahwa model mampu memprediksi kelas negatif secara akurat sebanyak 85 data. Sementara itu, kelas positif teridentifikasi dengan benar sebanyak 114 data. Untuk kelas netral, hanya terdapat 2 data yang berhasil diprediksi dengan tepat. Berdasarkan *confusion matrix* tersebut, langkah selanjutnya adalah menghitung nilai akurasi model menggunakan rumus berikut.

$$\begin{aligned}
 Akurasi &= \frac{TN + TNt + TP}{TN + FN + TNt + FNt + TP + FP} \\
 &= \frac{85 + 2 + 114}{85 + 4 + 34 + 16 + 2 + 23 + 23 + 5 + 114} = 0.657
 \end{aligned}$$

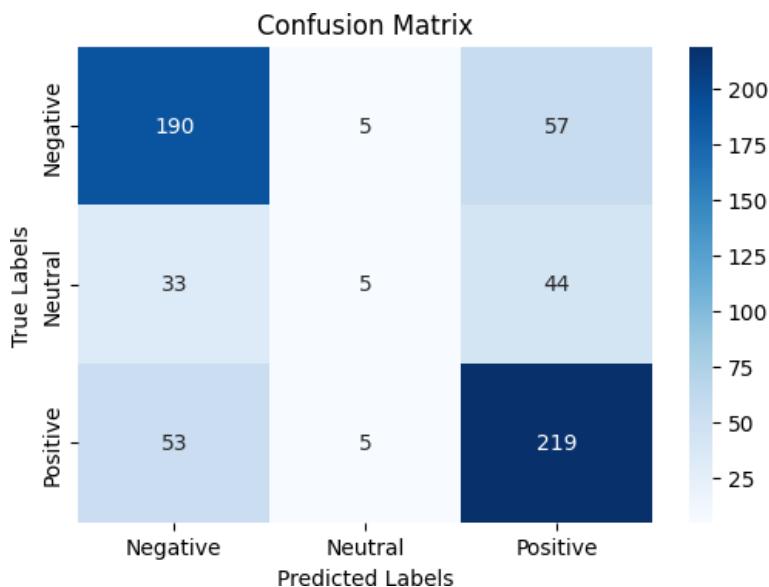
**Tabel 1.** Hasil Evaluasi *Confusion Matrix* Skema 1.

	<i>Precision</i>	<i>Recall</i>	F1 – Score
<i>Negative</i>	0.69	0.69	0.69
Neutral	0.18	0.05	0.08
<i>Positive</i>	0.67	0.80	0.73

Mengacu pada Tabel 1, kinerja model klasifikasi dapat dikatakan cukup baik. Hal ini ditunjukkan oleh nilai *recall* yang tinggi pada kategori sentimen positif, yang menandakan bahwa model memiliki kemampuan yang cukup baik dalam mengenali data dengan sentimen positif maupun negatif. Namun, untuk kategori netral nilai *recall* yang diperoleh sangat rendah, sehingga mengindikasikan bahwa model belum mampu mengklasifikasikan data dengan sentimen netral secara optimal.

## 2. Hasil Klasifikasi SVM Skema 2

Confusion matrix yang dihasilkan dari pengujian model dengan pembagian dataset pada rasio 80% data latih dan 20% data uji ditampilkan pada gambar berikut.

**Gambar 2.** Hasil *Confusion Matrix* Skema 2.

Hasil analisis terhadap confusion matrix menunjukkan bahwa model mampu mengklasifikasikan kelas negatif dengan benar sebanyak 190 data. Untuk kelas positif, prediksi yang tepat berjumlah 219 data, sedangkan pada kelas netral hanya terdapat 5 data yang diprediksi dengan benar. Berdasarkan confusion matrix tersebut, langkah berikutnya adalah menghitung tingkat akurasi model dengan menggunakan rumus sebagai berikut.

$$\begin{aligned}
 Akurasi &= \frac{TN + TNt + TP}{TN + FN + TNt + FNs + TP + FP} \\
 &= \frac{190 + 5 + 219}{190 + 5 + 57 + 33 + 5 + 44 + 53 + 5 + 219} = 0.677
 \end{aligned}$$

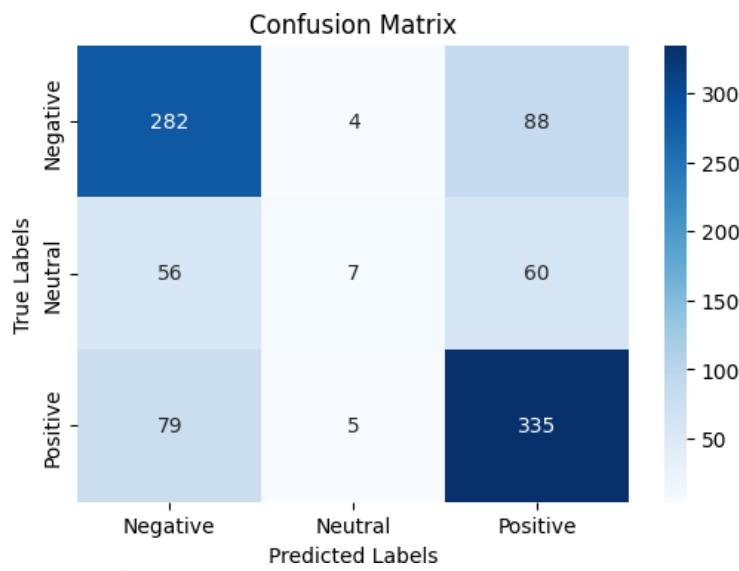
**Tabel 2.** Evaluasi *Confusion Matrix* Skema 2.

	Precision	Recall	F1 – Score
Negative	0.69	0.75	0.72
Neutral	0.33	0.06	0.10
Positive	0.68	0.79	0.73

Mengacu pada Tabel 14, kinerja model klasifikasi dapat dinilai cukup baik. Hal ini ditunjukkan oleh nilai recall yang tinggi pada kategori sentimen positif, yang menandakan bahwa sistem mampu mengenali data dengan sentimen positif maupun negatif secara akurat. Sebaliknya, pada kategori netral, nilai recall sangat rendah, sehingga mengindikasikan bahwa sistem masih kesulitan dalam mengidentifikasi data dengan sentimen netral secara tepat.

### 3. Hasil Klasifikasi SVM Skema 3

*Confusion matrix* hasil pengujian model dengan pembagian dataset menggunakan rasio 70% untuk data latih dan 30% untuk data uji ditampilkan pada bagian berikut.



**Gambar 3.** Hasil *Confusion Matrix* Skema 3.

Hasil evaluasi model klasifikasi sentimen mengungkapkan bahwa terdapat 282 data yang berhasil diidentifikasi dengan benar sebagai ulasan negatif. Pada kategori sentimen positif, sistem mampu mengenali 335 data dengan tepat. Sementara itu, jumlah prediksi yang benar pada kelas netral relatif sedikit, yaitu hanya 7 data. Berdasarkan *confusion matrix* tersebut, langkah selanjutnya adalah menghitung nilai akurasi model menggunakan rumus berikut.

$$\begin{aligned}
 Akurasi &= \frac{TN + TNt + TP}{TN + FN + TNt + FNt + TP + FP} \\
 &= \frac{282 + 7 + 335}{282 + 4 + 88 + 56 + 7 + 60 + 79 + 5 + 335} = 0.691
 \end{aligned}$$

**Tabel 3.** Evaluasi *Confusion Matrix* Skema 3

	<i>Precision</i>	<i>Recall</i>	F1 – Score
<i>Negative</i>	0.68	0.75	0.71
Neutral	0.44	0.06	0.10
<i>Positive</i>	0.69	0.80	0.74

Mengacu pada Tabel 3, kinerja model klasifikasi dapat dianggap cukup baik. Hal ini dibuktikan dengan tingginya nilai recall pada kategori sentimen positif, yang menunjukkan kemampuan sistem dalam mengenali data dengan sentimen positif maupun negatif secara tepat. Namun, pada kategori sentimen netral, nilai recall yang dihasilkan sangat rendah, sehingga menandakan bahwa sistem belum optimal dalam mengidentifikasi data dengan sentimen tersebut.

### Klasifikasi SVM dengan SMOTE

Pada tahap ini, penelitian memanfaatkan metode SMOTE untuk mengatasi ketidakseimbangan kelas dalam dataset sentimen. Teknik ini diterapkan sebelum model dilatih, dengan tujuan menghasilkan jumlah data pada kelas minoritas yang setara dengan kelas mayoritas.

```

from imblearn.over_sampling import SMOTE
# Menerapkan SMOTE
smote = SMOTE()
X_resampled, y_resampled = smote.fit_resample(X_tf_idf, y)
# Menampilkan informasi sebelum dan sesudah oversampling
print("Jumlah sampel sebelum oversampling:", X.shape[0])
print("Jumlah sampel setelah oversampling:", X_resampled.shape[0])

```

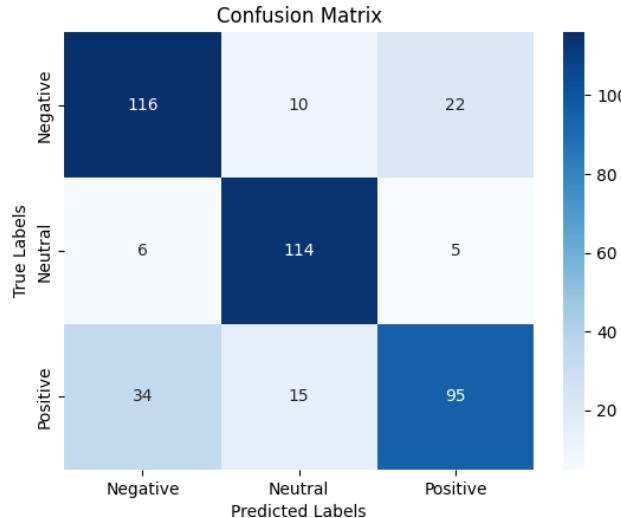
Kode pemrograman diatas menerapkan metode SMOTE dari modul *imblearn.over\_sampling* dalam mengatasi permasalahan ketidakseimbangan kelas pada data. Dalam tahap ini, objek SMOTE dibentuk melalui perintah *smote = SMOTE()*, kemudian fungsi *fit\_resample()* dijalankan pada data fitur (*X\_tf\_idf*) serta label (*y*) untuk menghasilkan data baru (*X\_resampled, y\_resampled*) dengan distribusi kelas yang lebih seimbang.

Pada tahap sebelumnya, dataset telah dipisahkan menjadi data latih dan data uji dengan tiga rasio berbeda, yakni 90:10, 80:20, dan 70:30. Untuk mengatasi masalah ketidakseimbangan kelas, metode SMOTE diterapkan pada data latih sebelum proses

klasifikasi dilakukan dengan algoritma SVM.

#### 4. Hasil Klasifikasi SVM dengan SMOTE Skema 1

Confusion matrix hasil pengujian model dengan pembagian data pada rasio 90% untuk data latih dan 10% untuk data uji ditampilkan pada bagian berikut.



**Gambar 4.** Hasil *Confusion Matrix* SVM dengan SMOTE Skema 1.

Berdasarkan analisis confusion matrix, model klasifikasi sentimen menunjukkan performa yang cukup baik dengan total 325 data dari 417 data uji berhasil diprediksi dengan benar. Untuk kelas Negative, terdapat 116 data yang sesuai dengan label sebenarnya. Pada kelas Neutral, kinerja model juga tergolong baik dengan 114 data terkласifikasi secara tepat. Sementara itu, pada kelas Positive, sebanyak 95 data berhasil diprediksi sesuai label. Setelah memperoleh confusion matrix ini, langkah berikutnya adalah menghitung nilai akurasi model menggunakan persamaan berikut.

$$\begin{aligned}
 \text{Akurasi} &= \frac{TN + TN_t + TP}{TN + FN + TN_t + FN_t + TP + FP} \\
 &= \frac{116 + 114 + 95}{116 + 10 + 22 + 6 + 114 + 5 + 34 + 15 + 95} = 0.779
 \end{aligned}$$

**Tabel 4.** Evaluasi *Confusion Matrix* dengan SMOTE Skema 1.

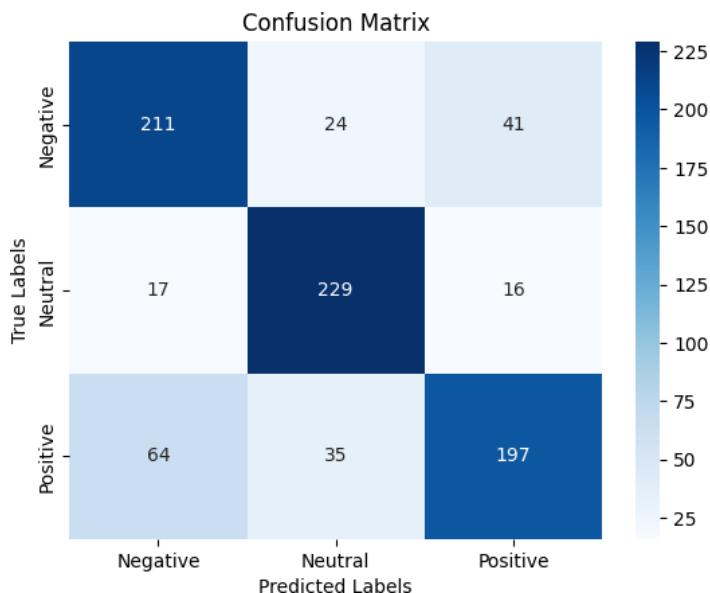
	Precision	Recall	F1 – Score
<i>Negative</i>	0.74	0.78	0.76
<i>Neutral</i>	0.82	0.91	0.86
<i>Positive</i>	0.78	0.66	0.71

Mengacu pada Tabel 16, model klasifikasi sentimen pada skema 1 mencapai nilai F1 sebesar 78%, lebih tinggi dibandingkan model tanpa penerapan SMOTE yang hanya memperoleh nilai F1 sebesar 50%. Temuan ini membuktikan bahwa penggunaan teknik SMOTE mampu meningkatkan kinerja dan akurasi model. Secara keseluruhan, performa model tergolong baik, terlihat dari nilai recall yang tinggi pada kategori sentimen negatif,

positif, maupun netral, yang menandakan kemampuan model dalam mendeteksi data dari ketiga kelas sentimen dengan lebih tepat.

### 5. Hasil Klasifikasi SVM dengan SMOTE Skema 2

Confusion matrix hasil pengujian model dengan pembagian dataset pada rasio 80% data latih dan 20% data uji ditampilkan pada bagian berikut.



**Gambar 5.** Hasil *Confusion Matrix* SVM dengan SMOTE Skema 2.

Berdasarkan hasil confusion matrix, model klasifikasi mampu memberikan prediksi yang akurat dengan 637 data teridentifikasi dengan benar dari seluruh data uji. Untuk kelas Negative, terdapat 211 data yang sesuai dengan label sebenarnya. Pada kelas Neutral, model mencatat jumlah prediksi benar terbanyak, yakni 229 data yang terkласifikasi dengan tepat. Sementara itu, pada kelas Positive, terdapat 197 data yang diprediksi sesuai dengan label. Temuan ini menunjukkan bahwa model memiliki kinerja yang cukup baik dalam mengenali ketiga kelas, dengan performa terbaik pada kelas Neutral. Setelah memperoleh confusion matrix, langkah berikutnya adalah menghitung akurasi model menggunakan persamaan berikut.

**Tabel 5.** Evaluasi *Confusion Matrix* dengan SMOTE Skema 2.

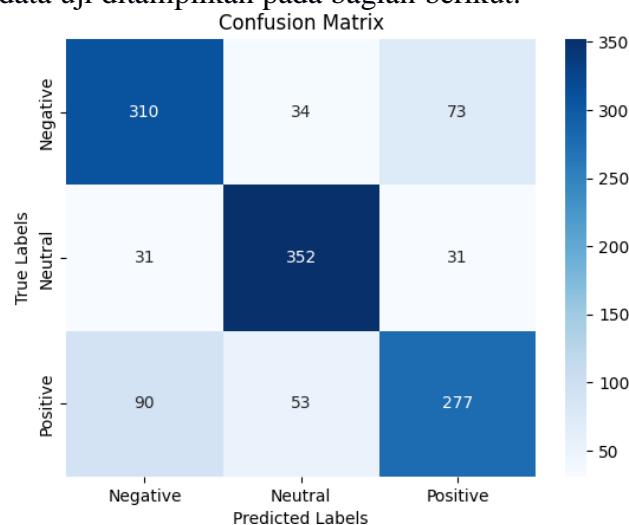
	Precision	Recall	F1 – Score
<i>Negative</i>	0.72	0.76	0.74
Neutral	0.80	0.87	0.83
<i>Positive</i>	0.78	0.67	0.72

Mengacu pada Tabel 5, model klasifikasi sentimen pada skema 2 memperoleh nilai F1 sebesar 76%, yang lebih tinggi dibandingkan model tanpa SMOTE yang hanya mencapai 52%. Hasil ini membuktikan bahwa penggunaan teknik SMOTE efektif dalam meningkatkan

performa sekaligus akurasi model. Secara keseluruhan, kinerja model tergolong baik, yang ditunjukkan oleh tingginya nilai recall pada sentimen Negative, Neutral, dan Positive, sehingga ketiga kelas sentimen dapat teridentifikasi dengan lebih tepat.

## 6. Hasil Klasifikasi SVM dengan SMOTE Skema 3

Confusion matrix dari hasil pengujian model dengan pembagian dataset pada rasio 70% data latih dan 30% data uji ditampilkan pada bagian berikut.



**Gambar 6.** Hasil Confusin Matrix SVM dengan SMOTE Skema 3.

Hasil analisis confusion matrix menunjukkan bahwa model klasifikasi berhasil menghasilkan prediksi yang baik dengan total 939 data terkласifikasi secara benar. Pada kelas negatif, terdapat 310 data yang sesuai dengan label sebenarnya. Untuk kelas netral, jumlah prediksi benar paling tinggi, yakni sebanyak 352 data. Sementara itu, pada kelas positif, terdapat 277 data yang juga berhasil diprediksi dengan tepat.

$$\begin{aligned}
 \text{Akurasi} &= \frac{TN + T_{Nt} + TP}{TN + FN + T_{Nt} + F_{Nt} + TP + FP} \\
 &= \frac{310 + 352 + 277}{310 + 34 + 73 + 31 + 352 + 31 + 90 + 53 + 277} \\
 &= 0.75
 \end{aligned}$$

**Tabel 6.** Evaluasi Confusion Matrix SVM dengan SMOTE Skema 3.

	Precision	Recall	F1 – Score
<i>Negative</i>	0.74	0.78	0.76
<i>Neutral</i>	0.82	0.91	0.86
<i>Positive</i>	0.78	0.66	0.71

Mengacu pada Tabel 18, model klasifikasi sentimen pada skema 3 berhasil meraih nilai F1 sebesar 75%, lebih tinggi dibandingkan model tanpa SMOTE yang hanya memperoleh 52%. Temuan ini mengindikasikan bahwa penerapan teknik SMOTE efektif dalam meningkatkan kinerja sekaligus akurasi model. Secara umum, performa model tergolong baik, yang terlihat dari tingginya nilai recall pada sentimen Negative, Neutral, dan Positive, sehingga

ketiga kelas sentimen tersebut dapat dikenali dengan lebih akurat.

### Hasil Klasifikasi Sentimen

Evaluasi performa model menggunakan pendekatan confusion matrix untuk setiap skema telah dilakukan, dan ringkasan hasilnya disajikan pada tabel berikut.

**Tabel 7.** Rekap Hasil Evaluasi *Confusion Matrix SVM*.

Evaluasi	Skema 1 Rasio data 90:10	Skema 2 Rasio Data 80:20	Skema 3 Rasio Data 70:30
Accuracy	65,69%	67,76%	68,12%
Precision	51%	57%	60%
Recall	51%	54%	54%
F1 – Score	50%	52%	52%

**Tabel 8.** Rekap Hasil *Confusion Matrix SVM* dengan SMOTE.

Evaluasi	Skema 1 Rasio data 90:10	Skema 2 Rasio Data 80:20	Skema 3 Rasio Data 70:30
Accuracy	77,94%	76,38%	75,06%
Precision	78%	76%	75%
Recall	79%	77%	75%
F1 – Score	78%	76%	75%

Berdasarkan tabel hasil pengujian, model SVM mencatat akurasi tertinggi pada skema 3 dengan rasio 70% data latih dan 30% data uji, yaitu sebesar 68,12%. Pada skema 2, akurasi sedikit menurun menjadi 67,76%, sedangkan skema 1 hanya mencapai 65,69%. Nilai precision juga mengalami peningkatan bertahap, dari 51% pada skema 1 menjadi 57% pada skema 2, dan mencapai 60% pada skema 3. Untuk recall dan F1-Score, kenaikan yang terlihat relatif kecil, masing-masing berada pada kisaran 51–54% dan 50–52%. Temuan ini menunjukkan bahwa model SVM cenderung bekerja lebih baik saat porsi data uji lebih besar, meskipun perbedaannya tidak signifikan.

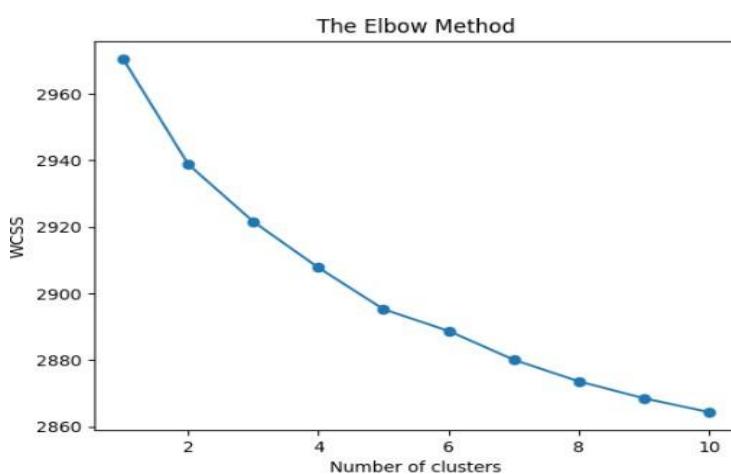
Setelah penerapan metode SMOTE, performa model meningkat cukup signifikan di semua skema. Akurasi tertinggi dicapai pada skema 1 dengan nilai 77,94%, disusul oleh skema 2 dengan 76,38%, dan skema 3 dengan 75,06%. Selain itu, nilai precision, recall, dan F1-Score juga mengalami peningkatan yang cukup besar. Sebagai contoh, pada skema 2, precision meningkat dari 57% menjadi 76%, recall dari 54% menjadi 77%, dan F1-Score dari 52% menjadi 76%. Hasil ini menegaskan bahwa SMOTE mampu menyeimbangkan distribusi data, sehingga model SVM dapat mengenali setiap kelas dengan lebih baik dan menghasilkan prediksi yang lebih akurat.

## Klasterisasi Data

Algoritma K-Means digunakan untuk melakukan pengelompokan data berdasarkan kesamaan nilai fitur numerik pada setiap tweet. Langkah pertama dalam proses ini adalah menentukan jumlah klaster yang paling optimal dengan metode Elbow, yaitu dengan menghitung nilai WCSS (*Within-Cluster Sum of Squares*) untuk variasi jumlah klaster  $k$  dari 1 hingga 10. Script di bawah ini menunjukkan proses yang dilakukan pada tahap tersebut.

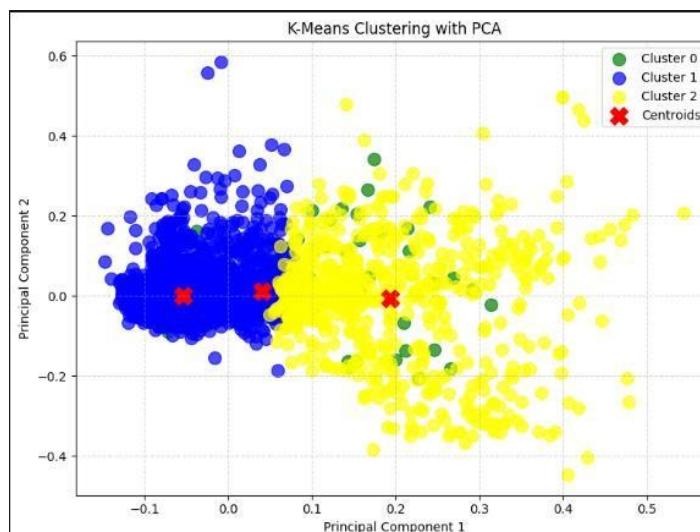
```
from sklearn.cluster import KMeans wcss = []
for i in range(1, 11):
    kmeans = KMeans(n_clusters=i, init='k-means++', max_iter=300, n_init=10,
    random_state=42)
    kmeans.fit(data_tf_idf) wcss.append(kmeans.inertia_)
plt.plot(range(1, 11), wcss, marker='o') plt.title('The Elbow Method')
plt.xlabel('Number of clusters') plt.ylabel('WCSS')
plt.show()
```

Kode ini menerapkan algoritma *K-Means Clustering* dari *sklearn.cluster* untuk mencari jumlah klaster optimal pada data TF-IDF. Nilai WCSS dihitung untuk jumlah klaster 1 hingga 10. Objek K-Means diinisialisasi dengan parameter *init* = '*k - means ++*', *max\_iter* = 300, *n\_init* = 10, dan *random\_state* = 42 untuk hasil yang konsisten. Hasil WCSS kemudian divisualisasikan menggunakan metode Elbow dengan *matplotlib.pyplot* untuk menentukan jumlah klaster terbaik.



Gambar 7. Hasil Visualisasi Metode Elbow.

Dari hasil evaluasi WCSS, terlihat bahwa penurunan nilai yang cukup tajam terjadi pada jumlah klaster 1 hingga 3. Setelah klaster ketiga, penurunan WCSS cenderung melandai, yang mengindikasikan bahwa penambahan klaster tidak lagi memberikan peningkatan kualitas pengelompokan yang signifikan. Dengan mempertimbangkan hal tersebut, penelitian ini menetapkan tiga klaster sebagai jumlah yang paling optimal. Pada penerapan K-Means, penomoran klaster dimulai dari nol, sehingga klaster yang terbentuk diberi label Cluster 0, Cluster 1, dan Cluster 2. Penamaan ini murni bersifat teknis dan tidak memengaruhi interpretasi hasil, karena perhatian utama terletak pada distribusi serta karakteristik data dalam tiap klaster.



**Gambar 8.** Hasil Visualisasi Algoritma PCA.

Hasil pengelompokan menggunakan algoritma *K-Means* yang direduksi ke dua dimensi dengan *Principal Component Analysis* (PCA) menunjukkan bahwa data terbagi menjadi tiga klaster, yaitu *Cluster 0* (hijau), *Cluster 1* (biru), dan *Cluster 2* (kuning). Setiap titik pada plot merepresentasikan satu data, sementara simbol silang merah menunjukkan posisi centroid klaster. *Cluster 1* dan *Cluster 2* mendominasi jumlah data dengan area sebaran yang luas dan posisi yang terpisah jelas, sedangkan *Cluster 0* berisi lebih sedikit data dan berada di area yang lebih dekat dengan dua klaster lainnya. Kondisi ini mengindikasikan bahwa *Cluster 0* memiliki karakteristik campuran atau transisi antar klaster. Jarak centroid yang cukup berjauhan menegaskan adanya perbedaan fitur antar klaster, meskipun terdapat beberapa area dengan titik tumpang tindih yang menunjukkan kemiripan data antar klaster.

Melalui penerapan algoritma K-Means, opini publik mengenai program MBG dapat dikelompokkan ke dalam tiga tema utama yang tercermin dari frekuensi kemunculan kata tertentu di setiap klaster. Visualisasi word cloud digunakan sebagai metode eksplorasi untuk menampilkan kata kunci dominan pada tiap klaster, sehingga memudahkan peneliti dalam

memahami fokus topik serta arah pembahasan masyarakat.



Gambar 9. Hasil Visualisasi Wordcloud Cluster 0.

Berdasarkan Gambar 20, visualisasi *wordcloud* pada *Cluster 0* menekankan pentingnya ketersediaan pangan bergizi sebagai inti dari Program MBG. Isu yang muncul tidak hanya terkait penyediaan makanan, tetapi juga pemenuhan gizi seimbang guna mendukung pertumbuhan anak, mencegah stunting, serta menjaga kesehatan kelompok prioritas seperti balita, anak sekolah, dan ibu hamil. Lebih jauh, pembahasan mengenai sekolah, pesantren, daerah, dan nasional menunjukkan bahwa program ini dipandang sebagai kebijakan yang harus menjangkau berbagai lapisan masyarakat, baik di lembaga pendidikan formal maupun non-formal, serta dilaksanakan secara merata di seluruh wilayah Indonesia. Dengan demikian, cluster ini merefleksikan perhatian publik terhadap kualitas gizi sekaligus pemerataan distribusi pangan, agar manfaat program dapat dirasakan secara adil tanpa menimbulkan kesenjangan antarwilayah maupun antarkelompok sosial.



Gambar 10. Hasil Visualisasi Wordcloud Cluster 1.

Berdasarkan Gambar 10, visualisasi *wordcloud* pada *Cluster 1* menggambarkan adanya dukungan masyarakat terhadap pelaksanaan MBG, khususnya dalam kaitannya dengan agenda pencegahan stunting dan peningkatan kesehatan anak. Dukungan tersebut disertai optimisme bahwa program ini dapat menjadi strategi nyata pemerintah dalam meningkatkan kualitas generasi muda.

Selain itu, keterkaitan dengan kepemimpinan nasional memperlihatkan bahwa MBG dipandang bukan sekadar kebijakan teknis, melainkan juga simbol komitmen politik pemerintah dalam menjawab isu gizi dan pendidikan. Hal ini sekaligus menunjukkan bahwa program tersebut mampu memperkuat legitimasi kebijakan dan citra pemerintah di mata publik. Kehadiran narasi mengenai keterlibatan masyarakat juga menandakan bahwa MBG dipersepsikan sebagai bentuk pelayanan sosial yang mendekatkan pemerintah dengan rakyat. Dengan demikian, cluster ini menegaskan bahwa MBG memperoleh dukungan sosial yang luas serta berfungsi sebagai instrumen penting dalam membangun legitimasi politik.



**Gambar 11.** Hasil Visualisasi *Wordcloud Cluster 2*.

Berdasarkan Gambar 11, visualisasi *wordcloud* pada *Cluster 2* berfokus pada isu pendanaan MBG dan dampaknya terhadap kondisi ekonomi nasional. Kekhawatiran publik muncul karena program ini dianggap berpotensi menambah beban fiskal melalui peningkatan alokasi APBN maupun kewajiban pajak masyarakat. Isu tersebut juga dipandang memiliki keterkaitan dengan persoalan sosial lain, seperti lapangan kerja, daya beli masyarakat, dan prioritas pembiayaan untuk sektor berbeda. Dengan demikian, kritik yang muncul tidak hanya berorientasi pada penolakan, tetapi lebih pada pertanyaan mengenai keberlanjutan program dalam jangka panjang dan efektivitas pengelolaan anggarannya. Secara keseluruhan, cluster ini menyoroti tantangan utama MBG terkait mekanisme pembiayaan, potensi tekanan terhadap APBN, serta implikasi ekonomi yang dapat dirasakan masyarakat.

## 5. KESIMPULAN DAN SARAN

### Kesimpulan

Penelitian ini menunjukkan bahwa respons publik terhadap Program Makan Bergizi Gratis (MBG) di platform X bersifat berimbang antara dukungan dan kritik. Sentimen positif sedikit mendominasi, mencerminkan harapan masyarakat terhadap peningkatan gizi, pencegahan stunting, dan kualitas sumber daya manusia, sementara sentimen negatif menyoroti kekhawatiran terkait pendanaan, beban pajak, serta keberlanjutan fiskal program.

Analisis klustering K-Means mengungkap tiga fokus utama diskursus publik, yaitu pemenuhan gizi bagi kelompok prioritas, legitimasi sosial dan politik program, serta implikasi ekonomi dan anggaran negara. Dari sisi metodologi, algoritma SVM mampu mengklasifikasikan sentimen positif dan negatif dengan cukup baik, namun kurang optimal pada sentimen netral. Penerapan SMOTE terbukti meningkatkan kinerja klasifikasi secara signifikan dengan menyeimbangkan distribusi kelas. Secara keseluruhan, hasil penelitian menegaskan bahwa keberhasilan Program MBG tidak hanya ditentukan oleh tujuan sosialnya, tetapi juga oleh transparansi pembiayaan, efektivitas implementasi, dan pengelolaan komunikasi publik yang responsif.

### Saran

Berdasarkan temuan penelitian, pemerintah disarankan untuk mempertahankan dan meningkatkan dukungan publik terhadap Program Makan Bergizi Gratis dengan memastikan kualitas pelaksanaan, penyediaan menu bergizi seimbang, serta pengawasan distribusi pangan yang merata. Kritik publik terkait beban APBN, kewajiban pajak, dan dampak ekonomi perlu direspon melalui transparansi anggaran, efisiensi pembiayaan, serta komunikasi publik yang akuntabel dan responsif. Selain itu, diskursus publik di media sosial dapat dimanfaatkan sebagai sarana sosialisasi dan edukasi kebijakan secara berkelanjutan. Untuk penelitian selanjutnya, disarankan memperluas sumber data ke berbagai platform media sosial agar diperoleh gambaran opini publik yang lebih beragam, menyempurnakan tahap pra-pemrosesan teks dengan teknik lanjutan seperti *lemmatization*, serta mengombinasikan visualisasi *word cloud* dengan metode *topic modeling* seperti LDA atau NMF guna menghasilkan analisis tema yang lebih mendalam dan komprehensif.

## DAFTAR PUSTAKA

- Abdillah, L. A. (2022). *Peranan media sosial modern*. Bening Media Publishing. <https://www.bening-mediapublishing.com>
- Arsi, P., & Waluyo, R. (2021). Analisis sentimen wacana pemindahan ibu kota Indonesia menggunakan algoritma support vector machine (SVM). *Jurnal Teknologi Informasi dan Ilmu Komputer*, 8(1), 147–154. <https://doi.org/10.25126/jtiik.0813944>
- Batubara, R. P., Skawanti, J. R., Saleha, S., Ishak, R. P., Sriwulandari, T., Setiawan, A. S., Sari, D. P., & Prayugo, S. (2024). *Peluang usaha produk kuliner berbasis puyuh* (S. Prayugo, Ed.). Bypass. <https://books.google.co.id/books?id=eucjEQAAQBAJ>
- Dekasari, Y., & Gunawan, T. (2024). Upaya pencegahan stunting pada anak balita di Indonesia: Pengabdian di Kabupaten Pesawaran. *JPKM: Jurnal Pengabdian kepada Masyarakat*, 2024(2), 129–135. <https://journal.danisapublisher.id>

- Efendi, J., & Lutfianingsih, F. F. (2023). *Aspek hukum media sosial: Perdata dan pidana*. Kencana.
- Jollyta, D., Hajjah, A., Haerani, E., & Siddik, M. (2023). *Algoritma klasifikasi untuk pemula: Solusi Python dan RapidMiner*. Deepublish. <https://books.google.co.id/books?id=y84TEQAAQBAJ>
- Mada Sanjaya, W. S. (2024). *Fisika komputasi berbasis machine learning dengan pemrograman Python*. Bolabot.
- Paramitha, D. I., Al Farauqi, M. D. A., Tyas, I. K. D., & Universitas Muhammadiyah Kalimantan Timur. (2023). Literasi digital pengguna internet Indonesia guna mewujudkan budaya damai di ruang mayantara. *Jurnal Kewarganegaraan*, 7(1), 1208–1215. <https://journal.upy.ac.id/index.php/pkn/article/view/5308>
- Permata, A. (2024). Analisis sentimen media sosial: Mengurai opini publik dengan data. *Teknologi Pintar*, 4(3), 1–19.
- Rahmaddeni, S. K. M. K., Wulandari, D. S. K. M. K., Renova, M., Ramadhan, G., & Sari, R. (2024). *Machine learning. Serasi Media Teknologi*. <https://books.google.co.id/books?id=owoOEQAAQBAJ>
- Rianto, I., & Santosa, I. P. I. (2025). *Data preparation untuk machine learning & deep learning* (Y. Fransisca A., Ed.). Penerbit Andi. <https://books.google.co.id/books?id=Y5U9EQAAQBAJ>
- Situmorang, W., & Hayati, R. (2023). Media sosial Instagram sebagai bentuk validasi dan representasi diri. *Jurnal Sosiologi Nusantara*, 9(1), 111–118. <https://doi.org/10.33369/jsn.9.1.111-118>
- Suryana, A. Y. (2024). *Optimalisasi pengelolaan informasi di era digital guna menyuksekan pembangunan nasional*. Jakarta.
- Tarmizi, S. N. (2024). *Membentengi anak dari stunting*. Kementerian Kesehatan Republik Indonesia. <https://sehatnegeriku.kemkes.go.id/baca/mediakom/20240728/4646123>
- Wiliani, N., Chusna, N. L., & Ramadhan, P. B. (2023). *Analisis sentimen terhadap pro kontra aksi unjuk rasa mahasiswa dengan naïve Bayes dan information gain*. Penerbit NEM.