Journal of Applied Informatics and Computing (JAIC)
Vol.10, No.1, February 2026, pp. 865~876
e-ISSN: 2548-6861

Sentiment Analysis of the Free Nutritious Meal Program (MBG) on Social Media X (Twitter) Using K-Nearest Neighbor and Artificial Neural Network

Fernanda Amri Hakim [1*], Ifnu Wisma Dwi Prastya [2], Jauhara Rana Budiani [3*]**
[1, 2] *Teknik Informatika, Universitas Nahdlatul Ulama Sunan Giri Bojonegoro
[3] *Statistika, Universitas Nahdlatul Ulama Sunan Giri Bojonegoro
nachamri98@gmail.com [1], ifnudwi867@gmail.com [2], jbudiani@unugiri.ac.id [3]

Article Info

Article history:

Received 2026-01-08
Revised 2026-01-26
Accepted 2026-02-09

Keyword:

Artificial Neural Network,
Free Nutritious Meal Program,
K-Nearest Neighbor,
Sentiment Analysis,
Twitter.

ABSTRACT

The Free Nutritious Meal Program (Makan Bergizi Gratis/MBG) is a national policy initiated by the Indonesian government to improve public nutritional status, particularly among children and vulnerable groups. Since its implementation, the program has generated extensive public discussion on social media, reflecting diverse opinions, support, and criticism. This study aims to analyze public sentiment toward the MBG program on social media X (Twitter) using machine learning-based text classification methods. A total of 9,038 Indonesian-language tweets were collected and processed through text preprocessing, semi-automatic sentiment labeling with manual validation, and feature extraction using the Term Frequency–Inverse Document Frequency (TF–IDF) method. Sentiments were classified into three categories: positive, neutral, and negative. The performance of K-Nearest Neighbor (KNN), Artificial Neural Network (ANN), and ANN with class balancing using Synthetic Minority Over-Sampling Technique (ANN + SMOTE) was evaluated using accuracy, precision, recall, and F1-score metrics supported by confusion matrix analysis. The results indicate that the ANN + SMOTE model achieved the highest performance with an accuracy of 93.58%, outperforming ANN (92.59%) and KNN (86.28%). The sentiment distribution indicates that public opinion toward the MBG program is predominantly neutral (52.1%), followed by positive (40.0%) and negative (7.9%) sentiments. These findings suggest that while the MBG program is generally well received, negative sentiments provide important feedback related to program implementation and governance.

I. PENDAHULUAN

Program Makan Bergizi Gratis (MBG) mulai diimplementasikan oleh Pemerintah Indonesia pada 6 Januari 2025 sebagai salah satu program prioritas nasional yang bertujuan meningkatkan status gizi masyarakat, khususnya pada kelompok anak-anak dan ibu hamil[1]. Berdasarkan Peraturan Presiden Nomor 83 Tahun 2024, pelaksanaan program ini berada di bawah koordinasi Badan Gizi Nasional (BGN). Program MBG menargetkan hingga 82,9 juta penerima manfaat yang mencakup peserta didik dari jenjang PAUD hingga SMA, balita, serta ibu hamil dan menyusui[2]. Hingga Agustus 2025, sebanyak 5.800 Satuan Pelayanan Pemenuhan Gizi (SPPG) telah beroperasi di 38 provinsi dengan cakupan sekitar 20 juta jiwa penerima manfaat[3]. Jumlah tersebut terus meningkat dan pada Oktober 2025 dilaporkan mencapai lebih dari 38,5 juta penerima manfaat dengan 13.514 SPPG aktif di seluruh Indonesia[4].

Skala implementasi Program Makan Bergizi Gratis yang sangat besar sejalan dengan berbagai kajian global yang menekankan pentingnya intervensi gizi dalam meningkatkan luaran kesehatan jangka panjang, termasuk status gizi, perkembangan kognitif, serta penurunan risiko penyakit di masa depan[5],[6]. Meskipun demikian, pelaksanaan program berskala nasional ini juga menghadapi berbagai tantangan, seperti kesiapan infrastruktur, distribusi bahan pangan,

keamanan makanan, potensi penyalahgunaan anggaran, serta efektivitas tata kelola program. Pada tahun pertama implementasinya, anggaran MBG dilaporkan mencapai lebih dari Rp70 triliun dan diproyeksikan meningkat hingga Rp450 triliun pada tahun 2029[7], yang menimbulkan perhatian publik terhadap transparansi, akuntabilitas, dan keberlanjutan kebijakan tersebut.

Berbagai respons terhadap program Makan Bergizi Gratis muncul dari lembaga internasional maupun masyarakat luas. Organisasi internasional seperti UNICEF dan *World Food Programme* (WFP) menyatakan dukungannya melalui pelatihan serta bantuan teknis guna menjaga kualitas gizi dan keamanan pangan. Di sisi lain, di ruang digital, program MBG menjadi topik perbincangan yang intens di media sosial, khususnya pada platform Twitter (X). Media sosial memungkinkan masyarakat untuk mengekspresikan pandangan, dukungan, kritik, serta pengalaman langsung terkait kebijakan publik yang sedang berjalan. Sejumlah penelitian menunjukkan bahwa percakapan daring mengenai program makan bergizi memuat pola sentimen yang beragam, mulai dari dukungan hingga kekhawatiran terhadap kualitas implementasi dan tata kelola kebijakan[8],[9].

Analisis sentimen berbasis media sosial telah banyak diterapkan untuk memahami persepsi publik terhadap kebijakan pemerintah. Penelitian terdahulu telah menerapkan berbagai algoritma pembelajaran mesin untuk menganalisis sentimen publik terhadap kebijakan gizi dan program sosial, seperti *Naïve Bayes*, *Support Vector Machine* (SVM), *K-Nearest Neighbor* (KNN), serta metode ensemble[10],[11]. Hasil penelitian sebelumnya menunjukkan bahwa metode KNN cenderung menghasilkan performa yang lebih rendah pada data teks berdimensi tinggi berbasis TF–IDF. Salah satu studi melaporkan bahwa KNN hanya mencapai akurasi sebesar 60%, dengan precision 62%, recall 60%, dan F1-score 59% dalam analisis sentimen kebijakan publik di media sosial X[12]. Temuan ini mengindikasikan keterbatasan KNN dalam menangkap kompleksitas bahasa dan ketidakseimbangan distribusi kelas pada data sentimen publik.

Sebaliknya, pendekatan berbasis jaringan saraf tiruan dan pembelajaran mendalam menunjukkan performa yang lebih kompetitif dalam berbagai studi analisis sentimen. Metode seperti *Bi-LSTM*, *regresi linier*, *random forest*, serta *Artificial Neural Network* (ANN) dilaporkan mampu menghasilkan tingkat akurasi yang lebih tinggi dalam memetakan opini publik terhadap kebijakan gizi dan program sosial[13],[14]. Secara khusus, ANN dilaporkan mampu mencapai tingkat akurasi di atas 90% pada berbagai tugas klasifikasi teks, sehingga menunjukkan kemampuannya dalam menangkap pola nonlinier pada data kompleks[15]. Selain itu, kombinasi beberapa algoritma seperti KNN, regresi logistik, dan SVM juga telah digunakan untuk mengkaji sentimen publik terhadap kebijakan gizi dan program sosial dalam berbagai konteks[16],[17]. Hal ini menunjukkan bahwa isu Program Makan Bergizi Gratis (MBG) bersifat multidimensional dan tidak hanya berkaitan dengan aspek kesehatan, tetapi juga melibatkan dimensi sosial, ekonomi, dan kebijakan publik.

Berdasarkan latar belakang tersebut, penelitian ini berfokus pada analisis sentimen publik terhadap Program Makan Bergizi Gratis (MBG) di media sosial Twitter (X) dengan menggunakan algoritma *K-Nearest Neighbor* (KNN) dan *Artificial Neural Network* (ANN). Algoritma KNN dipilih sebagai model pembanding karena kesederhanaannya dalam mengklasifikasikan teks berdasarkan tingkat kemiripan fitur serta penggunaannya yang luas dalam penelitian analisis sentimen media sosial [18]. Sementara itu, ANN digunakan karena kemampuannya dalam memodelkan pola nonlinier yang kompleks pada data teks, sehingga diharapkan dapat menghasilkan performa klasifikasi yang lebih baik[19]. Mengingat data sentimen publik umumnya memiliki distribusi kelas yang tidak seimbang, penelitian ini juga menerapkan teknik *Synthetic Minority Over-Sampling Technique* (SMOTE) untuk meningkatkan kinerja ANN dalam mengenali kelas minoritas[20]. Dengan membandingkan kinerja KNN, ANN, dan ANN dengan penyeimbangan kelas, penelitian ini diharapkan dapat memberikan gambaran yang lebih komprehensif mengenai persepsi publik terhadap MBG serta mengevaluasi efektivitas pendekatan klasifikasi yang digunakan.

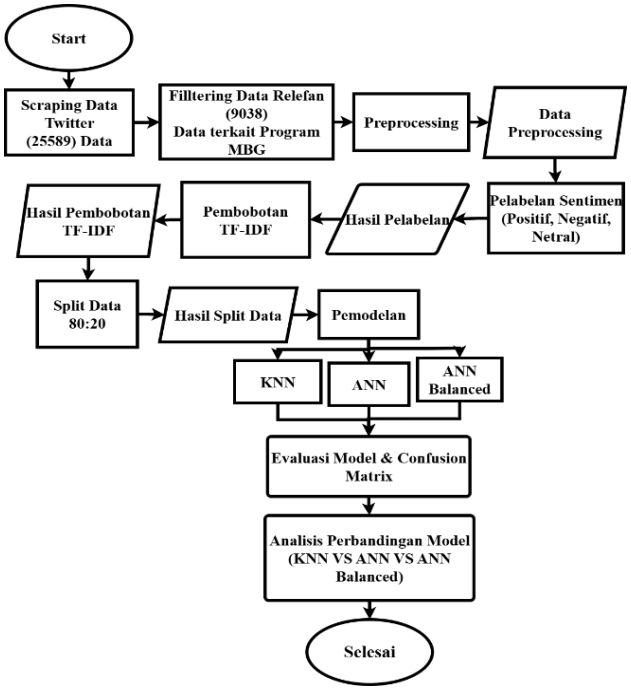
II. METODE

Penelitian ini menggunakan pendekatan analisis sentimen berbasis pembelajaran mesin untuk mengklasifikasikan opini publik terhadap Program Makan Bergizi Gratis (MBG) yang disampaikan melalui media sosial *Twitter* (X). Tahapan penelitian meliputi pengumpulan data, prapemrosesan teks, pelabelan sentimen, pembentukan fitur, pembagian dataset, pelatihan model klasifikasi, serta evaluasi kinerja model. Alur penelitian disusun mengikuti kerangka kerja analisis sentimen yang umum digunakan dalam penelitian opini publik berbasis media sosial[21],[22]. Alur lengkap tahapan penelitian ditunjukkan pada Gambar 1, yang menggambarkan proses mulai dari pengumpulan data hingga evaluasi kinerja model klasifikasi.

A. Dataset

Dataset penelitian diperoleh dari media sosial *Twitter* (X) menggunakan Twitter API v2 dengan autentikasi Bearer Token resmi. Pengambilan data dilakukan pada rentang waktu 1 Januari 2023 hingga 10 September 2025 dengan menggunakan kata kunci utama “Program Makan Bergizi Gratis”, “MBG”, dan “Program Gizi Gratis”, serta beberapa kata pendukung seperti “anak”, “gizi”, “anggaran”, “menu makan”, dan “sekolah”. Proses pengambilan data dibatasi pada tweet berbahasa Indonesia dan tidak menyertakan retweet, sehingga data yang dikumpulkan merepresentasikan opini asli pengguna[23],[24]. Twitter (X) dipilih sebagai sumber data karena platform ini menjadi salah satu ruang utama diskursus publik terkait kebijakan nasional dan

memungkinkan analisis opini masyarakat secara real-time dan terbuka.



Gambar 1. Alur Penelitian

B. Preprocessing Data

Tahap *preprocessing* bertujuan untuk meningkatkan kualitas data teks agar dapat diolah secara optimal oleh model klasifikasi. Proses *preprocessing* dilakukan secara otomatis dan bertahap dengan mengikuti praktik umum pada analisis sentimen Bahasa Indonesia [25]. Tahapan *preprocessing* yang diterapkan meliputi:

1. *Cleaning:* menghapus URL, mention, hashtag, angka, simbol, dan emoji, serta mengonversi seluruh huruf menjadi huruf kecil dan menghilangkan spasi ganda[26].
2. *Case Folding:* menyeragamkan penulisan teks dengan mengubah seluruh huruf menjadi huruf kecil[27].
3. *Tokenization:* memecah teks menjadi unit kata (token) berdasarkan spasi dan tanda baca[28].
4. *Stopword Removal:* menghapus kata umum yang tidak memiliki kontribusi signifikan terhadap makna sentimen dengan menggunakan daftar stopword Bahasa Indonesia[29].
5. *Stemming:* mengubah kata ke bentuk dasar menggunakan algoritma Nazief–Adriani melalui pustaka Sastrawi[30].
6. *Token Join:* menggabungkan kembali token hasil stemming menjadi teks siap analisis[31].

Seluruh tahapan prapemrosesan dilakukan secara konsisten pada seluruh dataset untuk memastikan keseragaman representasi teks. Proses ini bertujuan untuk mengurangi noise, menyederhanakan variasi linguistik, serta meningkatkan kualitas fitur yang digunakan dalam proses pembelajaran model klasifikasi.

C. Pelabelan

Pelabelan sentimen dalam penelitian ini dilakukan menggunakan pendekatan semi-otomatis, yang mengombinasikan metode berbasis leksikon dan validasi manual. Kamus leksikon Bahasa Indonesia digunakan sebagai acuan awal dan diperluas dengan istilah kontekstual yang relevan dengan isu gizi dan kebijakan publik, seperti *gizi*, *anak, bantuan*, *efektif*, dan *gagal* [32]. Pendekatan ini dipilih untuk menyesuaikan karakteristik bahasa informal dan konteks kebijakan publik yang umum ditemukan pada media sosial.

Sentimen diklasifikasikan ke dalam tiga kelas, yaitu positif (1), netral (0), dan negatif (−1). Skor sentimen dihitung untuk setiap *tweet* dengan menjumlahkan bobot kata positif dan negatif yang muncul dalam teks. Skor sentimen dihitung berdasarkan jumlah kemunculan kata positif dan negatif, dengan mempertimbangkan bobot kata *intensifier* yang dirumuskan sebagai berikut[33]:

$$S = \sum_{i=1}^{n} w_i (p_i - n_i) \quad (1)$$

Dengan p_i=1 jika kata ke-i termasuk kata *positif*, n_i= 1 jika kata ke-I termasuk kata *negatif*, dan w_i bobot kata *intensifier* (misalnya "sangat" = 1,5). Berdasarkan nilai skor sentimen yang diperoleh, penentuan label akhir dilakukan menggunakan aturan berikut[34].

$$Label = \begin{cases} 1, & S > 0 \\ 0, & S = 0 \\ -1, & S < 0 \end{cases}$$

Untuk meningkatkan keandalan hasil pelabelan, label yang dihasilkan secara otomatis selanjutnya ditinjau dan divalidasi secara manual oleh dua orang anotator yang memiliki latar belakang akademik di bidang sistem informasi dan statistika. Masing-masing anotator melakukan peninjauan secara independen terhadap hasil pelabelan otomatis guna memastikan kesesuaian konteks sentimen, khususnya pada tweet yang mengandung ambiguitas makna, ironi, sarkasme, atau penggunaan bahasa informal. Apabila terdapat perbedaan penilaian antara anotator atau antara hasil pelabelan otomatis dan hasil peninjauan manual, maka dilakukan diskusi hingga tercapai kesepakatan bersama untuk menentukan label sentimen akhir. Tahapan validasi ini bertujuan untuk meningkatkan keandalan label serta memperkuat kredibilitas hasil klasifikasi sentimen.

Setelah proses pelabelan sentimen selesai, dilakukan analisis distribusi kelas yang menunjukkan dominasi kelas sentimen netral. Ketidakseimbangan distribusi kelas ini berpotensi memengaruhi kinerja model klasifikasi. Oleh karena itu, pada tahap pelatihan model diterapkan teknik Synthetic Minority Over-Sampling Technique (SMOTE) untuk mengatasi ketidakseimbangan kelas, khususnya pada kelas minoritas. Proses pembentukan sampel sintetis dengan SMOTE dirumuskan sebagai berikut[35]:

$$x_{new} = x_i + \delta(x_{zi} - x_i), \delta \in [0,1] \quad (2)$$

Dengan x_i merupakan data asli dan x_{zi} merupakan salah satu tetangga terdekatnya.

D. Pembentukan Fitur TF-IDF

Setelah proses pelabelan, data teks dikonversi ke dalam representasi numerik menggunakan metode Term Frequency–Inverse Document Frequency (TF–IDF). Metode ini dipilih karena efektif dalam merepresentasikan teks pada berbagai penelitian analisis sentimen [36]. Secara umum, rumus TF–IDF yang digunakan[37]:

$$TF(t,d) = \frac{Ft,d}{\max(Fd)} \quad (3)$$

$$IDF(t) = log\left(\frac{N}{dFt}\right) \quad (4)$$

$$Wt,d = TF(t,d)\ x\ IDF(t) \quad (5)$$

Konfigurasi TF–IDF yang diterapkan meliputi *ngram_range* = (1,2), *max_features* = 20.000, *min_df* = 2, *sublinear_tf* = *True*, dan *normalisasi* L2.

E. Pembagian Dataset

Dataset yang telah direpresentasikan dalam bentuk fitur TF–IDF selanjutnya dibagi menjadi data latih dan data uji menggunakan metode train–test split. Pembagian dilakukan dengan rasio 80% data latih dan 20% data uji. Untuk menjaga proporsi masing-masing kelas sentimen (positif, netral, dan negatif) pada kedua subset data, proses pembagian dataset dilakukan menggunakan teknik stratified sampling. Selain itu, nilai random state ditetapkan sebesar 42 untuk memastikan bahwa proses pembagian data bersifat reprodusibel.

Dengan total 9.038 data, proses ini menghasilkan 7.230 data latih dan 1.808 data uji. Data latih digunakan untuk proses pelatihan dan validasi model, sedangkan data uji digunakan secara terpisah untuk mengevaluasi kemampuan generalisasi model terhadap data yang belum pernah dilihat sebelumnya.

Selain evaluasi menggunakan data uji, kestabilan performa model juga dievaluasi melalui metode K-Fold Cross Validation dengan nilai K = 3 yang diterapkan pada data latih. Pada pendekatan ini, data latih dibagi menjadi tiga subset, di mana dua subset digunakan untuk pelatihan dan satu subset digunakan untuk validasi secara bergantian. Metode ini bertujuan untuk mengurangi bias akibat pembagian data tunggal serta memastikan bahwa performa model bersifat konsisten dan dapat digeneralisasi dengan baik. Nilai K = 3 dipilih dengan mempertimbangkan keseimbangan antara kestabilan evaluasi dan efisiensi komputasi, mengingat ukuran dataset dan kompleksitas model yang digunakan.

F. Model Klasifikasi Sentiment

Penelitian ini menerapkan tiga model klasifikasi, yaitu *K-Nearest Neighbor* (KNN), *Artificial Neural Network* (ANN), dan ANN dengan penyeimbangan kelas (ANN + SMOTE). Seluruh model menggunakan representasi fitur TF–IDF sebagai masukan.

1. Model K-Nearest Neighbor (KNN):

Model KNN mengukur kemiripan antar data menggunakan *cosine similarity* pada ruang fitur TF–IDF, yang efektif untuk data teks berdimensi tinggi[38].

$$Sim(A,B) = \frac{\sum_{i=1}^{n} A_i\, B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\ x\ \sqrt{\sum_{i=1}^{n} B_i^2}} \quad (6)$$

Parameter yang digunakan pada model KNN adalah jumlah tetangga (*n_neighbors*) = 5, bobot jarak = *distance*, dan metrik jarak = *cosine*. Evaluasi kestabilan model dilakukan menggunakan *K-Fold Cross Validation* dengan K = 3, dan nilai akurasi rata-rata dihitung berdasarkan hasil setiap fold[39]*:*

$$Acc_{avg} = \frac{1}{K}\sum_{i=1}^{K} Acc_i \quad (7)$$

2. Model Artificial Neural Network (ANN):

Model ANN dibangun dengan arsitektur jaringan *fully connected* yang terdiri dari lapisan input sesuai dimensi TF–IDF, dua lapisan tersembunyi dengan 256 dan 128 *neuron* menggunakan fungsi aktivasi ReLU, serta lapisan *output* dengan tiga *neuron* menggunakan fungsi aktivasi *softmax*. Fungsi aktivasi ReLU didefinisikan sebagai[40]:

$$f(x) = \max(0, x) \quad (8)$$

Sedangkan fungsi *softmax* digunakan untuk menghasilkan probabilitas tiap kelas sentimen[41].

$$\sigma(z_i) = \frac{e^{z_i}}{\sum_{j=1}^{C} e^{z_j}} \quad (9)$$

Dengan C adalah jumlah kelas. Model dilatih menggunakan optimizer Adam dengan *learning rate* 0,0003 dan fungsi *loss Sparse Categorical Crossentropy*. Proses pelatihan dilakukan hingga 50 *epoch* dengan batch size 64 serta mekanisme *early stopping* untuk mencegah *overfitting*.

3. Model ANN Balanced (ANN+SMOTE):

Untuk mengatasi permasalahan ketidakseimbangan kelas, penelitian ini menerapkan teknik *Synthetic Minority Over-sampling Technique* (SMOTE) pada data latih. SMOTE menghasilkan sampel sintetis untuk kelas minoritas berdasarkan persamaan[42].

$$x_{new} = x_i + \delta(x_{zi} - x_i) \quad (10)$$

Dengan δ merupakan bilangan acak dalam rentang [0,1]. Model ANN Balanced menggunakan arsitektur yang lebih dalam dan dilengkapi dengan teknik regularisasi, yaitu: Beberapa lapisan *dense* (512, 256, dan 128 *neuron*) dengan aktivasi ReLU, *Dropout* bertingkat (0,35 – 0,25 – 0,15), *Batch Normalization* pada setiap lapisan tersembunyi, L2 regularization dengan λ=0.001, *ReduceLROnPlateau* untuk penyesuaian *learning rate* secara adaptif.

G. Evaluasi Model

Evaluasi model dilakukan pada data uji menggunakan *confusion matrix* dengan empat matriks utama, yaitu *accuracy*, *precision*, *recall*, dan *F1-score*, yang umum digunakan pada penelitian analisis sentimen teks multikelas [43],[44]:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (11)$$

$$Precision = \frac{TP}{TP + FP} \quad (12)$$

$$Recall = \frac{TP}{TP + FN} \quad (13)$$

$$F1 - Score = \frac{2\ x\ Precision\ x\ Recall}{Precision + Recall} \quad (14)$$

III. HASIL DAN PEMBAHASAN

A. Dataset

Data penelitian diperoleh dari media sosial Twitter (X) menggunakan Twitter API v2 dengan autentikasi Bearer Token resmi pada rentang waktu 1 Januari 2023 hingga 10 September 2025. Proses crawling awal menghasilkan sebanyak 25.589 tweet berbahasa Indonesia yang berkaitan dengan Program Makan Bergizi Gratis (MBG). Selanjutnya, dilakukan proses penyaringan untuk menghapus tweet promosi, spam, retweet berulang, serta konten yang tidak relevan dengan kebijakan MBG. Setelah tahap filtrasi, diperoleh sebanyak 9.038 tweet yang digunakan sebagai dataset utama dalam penelitian ini.

TABEL I
KUMPULAN DATASET

No	Full_text
1	"@Urrangawak Setuju pak... Saya juga ga mau anak keracunan MBG... Udah bener-bener dijaga makannya sama ortunya.. eh malah mau dikorbanin buat cari cuan beberapa pihak aja.."
2	"@Urrangawak Yg bikin keracunan MBG itu kebanyakan karena makanan basi, bayangin daerah gue masak jam 10 malam, packing jam 2 pagi, jam 10 pagi baru dimakan tanpa alat khusus buat menjaga makanan."
3	"@westhernsea Jaman digital sudah terbuka begini maunya dirahasiakan. Yg keracunan MBG dibawa ke RS aja mesti ngaku keracunan MBG itu bisa masuk berita. Ini orang tolol banget."
…	…
9038	"@Urrangawak Hindari keracunan MBG? Ubah wadahnya, ganti pakai kertas box jangan besi. Paman saya bisnis nasi box bertahun-tahun gak pernah ada racun."

Setiap tweet disimpan dalam format CSV dengan beberapa atribut, antara lain *created_at*, *username*, *favorite_count*, *retweet_count*, *reply_count*, *full_text*, *tweet_url*, dan *lang*. Contoh sebagian data mentah disajikan pada Tabel I. Berdasarkan analisis awal terhadap isi *tweet*, mayoritas percakapan publik menyoroti isu keamanan pangan, pengelolaan dapur umum melalui Satuan Pelayanan Pemenuhan Gizi (SPPG), serta risiko keracunan makanan dalam pelaksanaan MBG. Hal ini menunjukkan bahwa perhatian publik tidak hanya tertuju pada tujuan program, tetapi juga pada aspek teknis dan operasional di lapangan.

B. Preprocessing

Tahap prapemrosesan dilakukan untuk menyiapkan teks mentah agar dapat diolah secara optimal oleh model pembelajaran mesin. Proses ini dilaksanakan secara otomatis menggunakan bahasa pemrograman Python dengan bantuan pustaka NLTK, Sastrawi, serta *regular expression*. Tahapan prapemrosesan meliputi *cleaning*, *case folding*, *tokenization*, *stopword removal*, *stemming*, dan *token join*.

1. Cleaning

Tahapan awal ini bertujuan untuk menghapus elemen-elemen yang tidak relevan seperti URL, hashtag, mention pengguna, angka, emoji, serta simbol khusus. Teks kemudian di konversi ke format standar dengan menghapus spasi ganda agar konsisten.

TABEL II
CONTOH HASIL CLEANING

Sebelum	Sesudah
"@Urrangawak Setuju pak... Saya juga ga mau anak keracunan MBG... Udah bener-bener dijaga makannya sama ortunya.. eh malah mau dikorbanin buat cari cuan beberapa pihak aja.."	"Setuju pak Saya juga ga mau anak keracunan MBG Udah bener dijaga makannya sama ortunya eh malah mau dikorbanin buat cari cuan beberapa pihak aja"

2. Case Folding

Dalam tahap ini semua huruf diubah menjadi huruf kecil (*lowercase*) untuk menghindari perbedaan arti akibat kapitalisasi serta mengurangi kompleksitas data.

TABEL III

CONTOH HASIL CASE FOLDING

Sebelum	Sesudah
“Setuju pak saya juga ga mau anak keracunan MBG Udah bener di jaga makan nya sama ortunya eh malah mau dikorbanin buat cari cuan beberapa pihak aja”	“setuju pak saya juga ga mau anak keracunan mbg udah bener di jaga makannya sama ortunya eh malah mau di korbanin buat cari cuan beberapa pihak aja”

3. Tokenization

Proses memecah teks menjadi potongan kata kecil (*token*) berdasarkan spasi dan tanda baca. Tahap ini memungkinkan sistem mengenali setiap kata sebagai unit makna yang berdiri sendiri.

TABEL IV
CONTOH HASIL TOKENIZATION

Sebelum	Sesudah
“setuju pak saya juga ga mau anak keracunan mbg udah bener di jaga makannya sama ortunya eh malah mau di korbanin buat cari cuan beberapa pihak aja”	[“setuju”, “pak”, “saya”, “juga”, “gak”, “mau”, “anak”, “keracunan”, “mbg”, “udah”, “bener”, “dijaga”, “makannya”, “sama”, “ortunya”, “eh”, “malah”, “mau”, “dikorbanin”, “buat”, “cari”, “cuan”, “beberapa”, “pihak”, “aja”]

4. Stopword Removal

Stopword merupakan kata umum yang sering muncul tetapi kurang berkontribusi pada makna, seperti “yang”, “dan”, “di”, “ke”, atau “itu”. Dalam penelitian ini, penghapusan *stopword* dilakukan menggunakan daftar stopword Bahasa Indonesia dari *pustaka Sastrawi* yang diperluas dengan *custom stopword list* sesuai konteks percakapan di media sosial.

TABEL V
CONTOH HASIL STOPWORD REMOVAL

Sebelum	Sesudah
[“setuju”, “pak”, “saya”, “juga”, “ga”, “mau”, “anak”, “keracunan”, “mbg”, “udah”, “bener”, “dijaga”, “makannya”, “sama”, “ortunya”, “eh”, “malah”, “mau”, “dikorbanin”, “buat”, “cari”, “cuan”, “beberapa”, “pihak”, “aja”]	[“setuju”, “anak”, “keracunan”, “mbg”, “bener”, “jaga”, “makan”, “korban”, “cuan”, “pihak”]

5. Stemming

Tahap ini mengubah setiap kata ke bentuk dasarnya (*root word*) menggunakan algoritma *Nazief-Adriani*, sehingga variasi *morfologi* seperti “menyenangkan” menjai “senang” dan “berjalan” menjadi “jalan”. Tujuannya adalah menyatukan kata-kata yang memiliki makna yang sama ke dalam bentuk tunggal.

TABEL VI
CONTOH HASIL STEMMING

Sebelum	Sesudah
[“setuju”, “anak”, “keracunan”, “mbg”, “bener”, “jaga”, “makan”, “korban”, “cuan”, “pihak”]	[“setuju”, “anak”, “racun”, “mbg”, “benar”, “jaga”, “makan”, “korban”, “cuan”, “pihak”]

6. Token Join

Tahap akhir ini adalah penggabungan kembali token hasil *stemming* menjadi satu kalimat utuh. Teks hasil tahap ini sudah bersih, ringkas, dan siap diolah dalam pembentukan fitur menggunakan metode seperti *TF-IDF*.

TABEL VII
CONTOH HASIL TOKEN JOIN

Sebelum	Sesudah
[“setuju”, “anak”, “racun”, “mbg”, “benar”, “jaga”, “makan”, “korban”, “cuan”, “pihak”]	[“setuju anak racun mbg benar jaga makan korban cuan pihak”]

Hasil prapemrosesan menunjukkan bahwa teks tweet menjadi lebih ringkas dan representatif, sehingga mampu mengurangi noise serta meningkatkan kualitas fitur yang digunakan dalam proses klasifikasi. Contoh hasil prapemrosesan pada setiap tahap ditunjukkan pada tabel terkait. Secara umum, proses ini berperan penting dalam memastikan bahwa model klasifikasi dapat mempelajari pola sentimen secara lebih akurat dan konsisten.

C. Pelabelan Sentimen

Pelabelan sentimen dilakukan secara semiotomatis menggunakan pendekatan berbasis leksikon yang diperluas dengan istilah kontekstual terkait kebijakan gizi. Setiap tweet diklasifikasikan ke dalam tiga kelas sentimen, yaitu positif (1), netral (0), dan negatif (−1). Distribusi akhir ketiga label ini setelah seluruh tahap pemrosesan ditampilkan pada Tabel II.

TABEL VIII
DISTRIBUSI SENTIMEN PUBLIK

Label	Jumlah Tweet	Persentase(%)
Positif	3.620	40,0
Netral	4.707	52,1
Negatif	711	7,9
Total	9.038	100,0

Sentimen netral mendominasi opini publik (52,1%), diikuti sentimen positif (40,0%). Sentimen negatif hanya sebesar (7,9%), namun tetap penting karena mewakili kritik terhadap pelaksanaan Program Makan Bergizi Gratis (MBG).

D. Pembentukan Fitur TF-IDF

Data teks yang telah melalui tahap pelabelan kemudian dikonversi menjadi representasi numerik menggunakan metode TF–IDF dengan konfigurasi *ngram_range* = (1,2) dan *max_features* = 20.000. Proses ini menghasilkan matriks fitur berukuran 9.038 × 20.000, di mana setiap *tweet* direpresentasikan oleh bobot unigram dan bigram.

Analisis bobot TF–IDF menunjukkan bahwa kata-kata seperti "gratis", "gizi", dan "anak" sering muncul pada sentimen positif dan netral, sedangkan istilah seperti "korupsi", "anggaran", dan "basi" menjadi indikator kuat sentimen negatif. Pola ini membantu model dalam membedakan konteks sentimen secara lebih efektif.

E. Pembagian Dataset

Distribusi kelas pada data latih dan data uji relatif seimbang karena penerapan teknik stratified sampling pada proses pembagian dataset.

TABEL IX
HASIL PEMBAGIAN DATASET

Kelas Sentimen	Data Latih	Data Uji	Total
Positif	2.896	724	3.620
Netral	3.766	941	4.707
Negatif	568	143	711
Total	7.230	1.808	9.038

F. Hasil Klasifikasi Model

Evaluasi performa dilakukan terhadap tiga model klasifikasi, yaitu *K-Nearest Neighbor* (KNN), *Artificial Neural Network* (ANN), dan ANN dengan penyeimbangan kelas (ANN + SMOTE). Evaluasi dilakukan menggunakan K-Fold Cross Validation (K = 3) pada data latih serta pengujian pada data uji untuk menilai kemampuan generalisasi model.

*1. Hasil Eval*uasi *pada Data Latih*

Evaluasi *K-Fold Cross Validation* bertujuan untuk mengukur konsistensi performa model pada data latih serta mengurangi potensi bias akibat pembagian data tunggal. Pada penelitian ini digunakan K = 3, di mana data latih dibagi menjadi tiga subset dan proses pelatihan–validasi dilakukan secara bergantian.

1) K-Nearest Neighbor (KNN)

Model KNN digunakan sebagai model pembanding awal (*baseline*). Parameter yang digunakan pada model ini adalah: *n_neighbors = 5, weights = distance, metric = cosine.*

TABEL X
HASIL K-FOLD KNN MENGGUNAKAN (DATA LATIH)

Fold	Akurasi
Fold 1	0,8755
Fold 2	0,8622
Fold 3	0,8614
Rata-rata akurasi	0,8664

Hasil *K-Fold Cross Validation* menunjukkan bahwa model KNN memiliki performa yang relatif stabil antar fold dengan rata-rata akurasi sebesar 86,64%. Selisih nilai akurasi antar fold tergolong kecil, yang mengindikasikan bahwa model cukup konsisten pada data latih. Namun demikian, berdasarkan laporan klasifikasi, model KNN masih memiliki keterbatasan dalam mengenali kelas minoritas, khususnya kelas negatif, yang merupakan karakteristik umum metode berbasis jarak pada data teks berdimensi tinggi seperti TF–IDF.

2) Artificial Neural Network (ANN)

Model ANN dilatih menggunakan data latih tanpa penyeimbangan kelas. Arsitektur jaringan yang digunakan adalah 20.000–256–128–3 dengan *optimizer* Adam dan mekanisme *early stopping*.

Proses pelatihan model ANN dihentikan secara otomatis menggunakan mekanisme early stopping, dengan bobot terbaik diperoleh pada epoch ke-10 berdasarkan performa validasi.

TABEL XI
RINGKASAN HASIL PELATIHAN MODEL ANN MENGGUNAKAN (DATA LATIH)

Parameter Pelatihan	Nilai
Epoch maksimum	50
Epoch efektif	15
Epoch terbaik	10
Akurasi latih akhir	0,9999
Loss latih akhir	0,0013

Model ANN menunjukkan konvergensi yang sangat cepat, di mana akurasi meningkat signifikan sejak *epoch* awal hingga mencapai nilai mendekati sempurna pada data latih. Hal ini menunjukkan bahwa ANN mampu mempelajari pola data dengan sangat baik. Namun, tingginya akurasi pada data latih juga mengindikasikan potensi bias akibat ketidakseimbangan kelas, sehingga evaluasi pada data uji menjadi sangat penting untuk menilai kemampuan generalisasi model.

3) ANN Balanced (ANN + SMOTE)

Untuk menunjukkan dampak penerapan teknik penyeimbangan kelas, Tabel XII menyajikan distribusi kelas sentimen pada data latih sebelum dan sesudah penerapan *Synthetic Minority Over-Sampling Technique* (SMOTE). Penyajian ini bertujuan untuk memberikan gambaran kuantitatif mengenai tingkat ketidakseimbangan data awal serta perubahan distribusi kelas setelah proses penyeimbangan dilakukan.

TABEL XII
DISTRIBUSI KELAS SENTIMEN PADA DATA LATIH SEBELUM DAN SESUDAH SMOTE

Kelas Sentimen	Sebelum SMOTE	Sesudah SMOTE
Positif	2.896	3.766
Netral	3.766	3.766
Negatif	568	3.763
Total	7.230	11.295

Berdasarkan Tabel XII, terlihat bahwa sebelum penerapan SMOTE, kelas sentimen negatif memiliki jumlah data yang jauh lebih sedikit dibandingkan kelas positif dan netral. Setelah penerapan SMOTE, jumlah data pada masing-masing kelas menjadi relatif seimbang. Kondisi ini diharapkan dapat mengurangi bias model terhadap kelas mayoritas serta meningkatkan kemampuan model dalam mengenali kelas minoritas pada proses pelatihan.

Model ANN Balanced dilatih menggunakan data latih yang telah diseimbangkan dengan teknik SMOTE. Data latih awal berjumlah 7.230 tweet dengan dimensi fitur TF–IDF sebesar 20.000. Setelah penerapan SMOTE, jumlah data latih meningkat menjadi 11.295 data dengan dimensi fitur yang tetap sama.

Arsitektur jaringan pada model ANN Balanced diperluas menjadi 20.000–512–256–128–3 dan dilengkapi dengan beberapa teknik regularisasi, yaitu *Dropout*, *Batch Normalization*, dan *L2 regularization*, serta mekanisme *early stopping* dan penyesuaian learning rate adaptif menggunakan *ReduceLROnPlateau*. Proses pelatihan model dihentikan secara otomatis menggunakan *early stopping*, dengan bobot terbaik diperoleh pada *epoch* ke-6.

TABEL XIII
RINGKASAN HASIL PELATIHAN MODEL ANN BALANCED (ANN + SMOTE) MENGGUNAKAN DATA LATIH

Parameter Pelatihan	Nilai
Epoch maksimum	50
Epoch efektif	12
Epoch terbaik	6
Akurasi latih akhir	0,9999
Loss latih akhir	0,0731

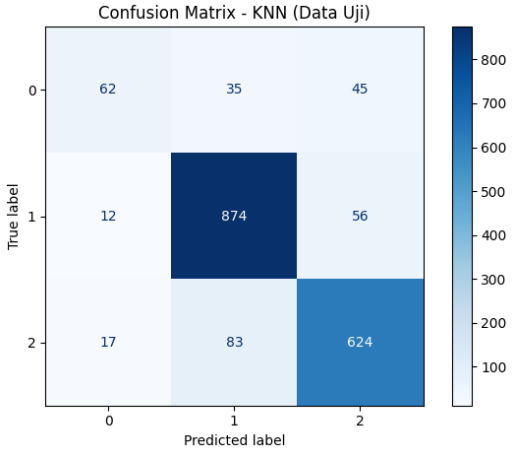
Penerapan SMOTE menghasilkan distribusi data latih yang lebih seimbang sehingga model ANN Balanced mampu mempelajari pola dari seluruh kelas sentimen secara lebih merata. Proses pelatihan berlangsung stabil dengan konvergensi yang konsisten dan tidak menunjukkan indikasi *overfitting* yang berlebihan selama pelatihan.

2. Analisis Confusion Matrix dan Perbandingan Model

Untuk menilai performa keseluruhan, dilakukan perbandingan berbasis *akurasi*, *presisi*, *recall*, dan *F1-score* pada data uji (1.808 *tweet*). Visualisasi dan interpretasi *confusion matrix* membantu memahami pola kesalahan klasifikasi pada tiap kelas sentimen.

1. Interpretasi Hasil Confusion Matrix

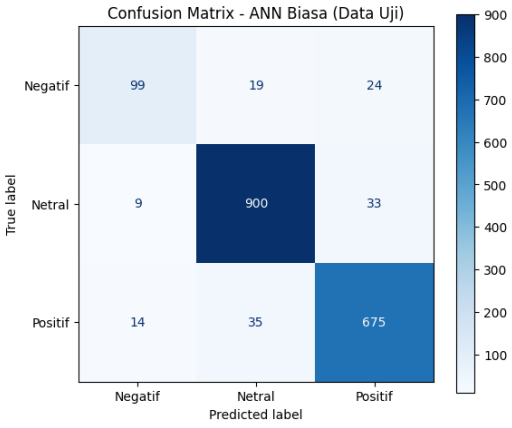
a. Model KNN



Gambar 2. Confusion Matrix KNN

Berdasarkan *confusion matrix*, model KNN menghasilkan: (62) prediksi benar untuk kelas negatif, (874) prediksi benar untuk kelas netral, (624) prediksi benar untuk kelas positif, dengan akurasi total sebesar (86,28%). Kesalahan klasifikasi paling dominan terjadi pada kelas negatif, di mana (35) *tweet* negatif diprediksi sebagai netral dan (45) sebagai positif. Kondisi ini menyebabkan recall kelas negatif hanya sebesar (43,66%), yang menunjukkan bahwa model KNN kurang sensitif terhadap kelas minoritas. Hal ini mencerminkan keterbatasan metode berbasis jarak dalam membedakan nuansa sentimen pada data teks berdimensi tinggi dan tidak seimbang.

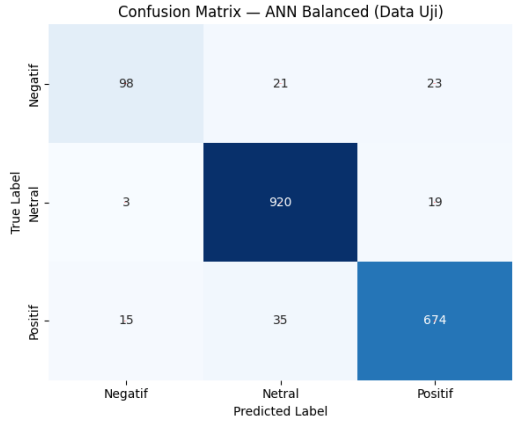
b. Model ANN



Gambar 3. Confusion matrix ANN

Model ANN tanpa penyeimbangan kelas menghasilkan: (99) prediksi benar untuk kelas negatif, (900) prediksi benar untuk kelas netral, (675) prediksi benar untuk kelas positif, dengan akurasi total sebesar (92,59%). Dibandingkan KNN, kemampuan ANN dalam mengenali kelas negatif meningkat signifikan, dengan recall kelas negatif sebesar (69,72%). Meskipun demikian, masih terdapat kesalahan silang antara kelas netral dan positif, yang menunjukkan bahwa ketidakseimbangan distribusi kelas masih memengaruhi proses pembelajaran model.

c. Model ANN Balanced



Gambar 4. Confusion matrix ANN Balanced

Berdasarkan confusion matrix, model ANN Balanced (ANN + SMOTE) menunjukkan performa klasifikasi yang paling seimbang dibandingkan model lainnya. Model ini berhasil mengklasifikasikan dengan benar sebanyak 98 tweet pada kelas negatif, 920 tweet pada kelas netral, dan 674 tweet pada kelas positif, dengan tingkat akurasi keseluruhan sebesar 93,58%.

Penerapan SMOTE berkontribusi signifikan dalam meningkatkan sensitivitas model terhadap kelas minoritas. Hal ini terlihat dari nilai recall kelas negatif yang mencapai 69,01% dan precision sebesar 84,48%, yang menunjukkan bahwa sebagian besar tweet negatif dapat dikenali dengan baik dan kesalahan prediksi terhadap kelas lain relatif rendah.

Pola diagonal yang dominan pada confusion matrix mengindikasikan bahwa sebagian besar prediksi berada pada kelas yang sesuai. Selain itu, jumlah kesalahan klasifikasi silang antar kelas relatif kecil, terutama pada kesalahan prediksi kelas netral menjadi negatif, yang hanya terjadi pada beberapa kasus. Temuan ini menunjukkan bahwa model ANN Balanced mampu menghasilkan prediksi yang lebih stabil dan proporsional antar kelas sentimen.

*2. Perbandingan Hasil Eva*luasi *Model pada data Uji*

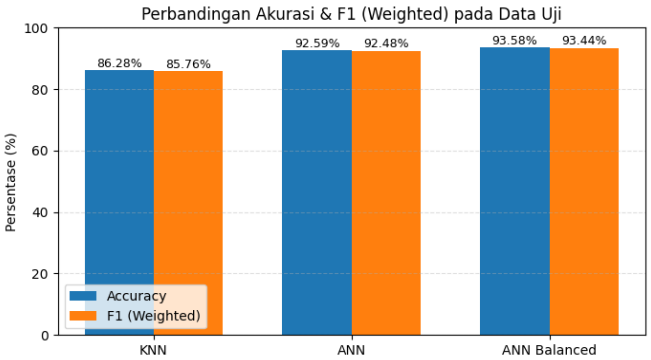
Evaluasi akhir dilakukan pada 1.808 data uji yang tidak terlibat dalam proses pelatihan. Evaluasi ini bertujuan untuk membandingkan performa ketiga model klasifikasi secara langsung berdasarkan kemampuan generalisasi terhadap data baru.

TABEL XIV
PERBANDINGAN PERFORMA MODEL PADA DATA UJI

Model	Akurasi (%)	Precision	Recall	F1-score
KNN	86,28	85,72	86,28	85,76
ANN	92,59	92,45	92,59	92,48
ANN Balanced	**93,58**	**93,44**	**93,58**	**93,44**

Berdasarkan hasil evaluasi pada Tabel XIII, model ANN dan ANN Balanced menunjukkan peningkatan performa yang signifikan dibandingkan KNN pada seluruh metrik evaluasi. Model KNN memperoleh akurasi terendah dan menunjukkan keterbatasan dalam mengenali kelas minoritas, khususnya kelas sentimen negatif, yang merupakan karakteristik umum metode berbasis jarak pada data teks berdimensi tinggi.

Model ANN tanpa penyeimbangan kelas telah mencapai akurasi yang tinggi sebesar 92,59%. Namun, performa model ini masih dipengaruhi oleh distribusi kelas yang tidak seimbang. Setelah diterapkan penyeimbangan kelas menggunakan SMOTE, model ANN Balanced (ANN + SMOTE) menunjukkan peningkatan performa menjadi 93,58% serta nilai precision, recall, dan F1-score yang lebih tinggi. Hal ini mengindikasikan bahwa penyeimbangan data latih memungkinkan model mempelajari pola dari seluruh kelas sentimen secara lebih merata dan mengurangi bias terhadap kelas mayoritas.



Gambar 5. Perbandingan akurasi model.

Secara kuantitatif, selisih performa antar model pada data uji menunjukkan bahwa model ANN memiliki peningkatan akurasi sebesar 6,31% dibandingkan KNN.

Sementara itu, penerapan SMOTE pada ANN menghasilkan peningkatan akurasi tambahan sebesar 0,99% dibandingkan ANN tanpa penyeimbangan. Jika dibandingkan langsung dengan KNN, model ANN Balanced menunjukkan selisih akurasi sebesar 7,30%.

Hasil ini menunjukkan bahwa pendekatan berbasis jaringan saraf secara konsisten mengungguli metode berbasis jarak (KNN) dalam analisis sentimen data media sosial. Selain itu, penerapan SMOTE pada model ANN Balanced memberikan peningkatan performa tambahan dengan memperbaiki keseimbangan prediksi antar kelas tanpa menurunkan kemampuan generalisasi model terhadap data uji.

G. Hasil Analisis Sentimen Publik terhadap Program Makan Bergizi Gratis (MBG)

Hasil klasifikasi akhir dari 9.038 *tweet* yang membahas tentang kebijakan Program Makan Bergizi Gratis (MBG) menghasilkan distribusi sentimen sebagai berikut.

TABEL XV
HASIL KLASIFIKASI SENTIMEN

Sentimen	Jumlah Tweet	Persentase	Karakteristik Utama
Netral	4.707	52,1%	Tweet bersifat informatif dari akun media, instansi, dan sumber resmi; berisi jadwal kegiatan atau pernyataan kebijakan tanpa ekspresi emosional.
Positif	3.620	40.0%	Tweet berisi dukungan dan apresiasi terhadap program, sering memuat kata “gizi”, “anak”, dan “sehat” yang menggambarkan manfaat sosial MBG.
Negatif	711	7,9%	Tweet mengandung kritik terkait pengelolaan, anggaran, potensi korupsi, dan efektivitas program; mencerminkan kekhawatiran terhadap transparansi dan akuntabilitas kebijakan.

Secara keseluruhan hasil menunjukkan bahwa mayoritas masyarakat bersikap netral ataupun positif (92,1%) terhadap program MBG, menandakan tingkat penerimaan publik yang tinggi. Namun meskipun sentimen negatif hanya sebesar (7,9%), kelompok ini tetap penting karena mencerminkan suara kritis terhadap aspek tata kelola dan efisiensi.

H. Pembahasan

Hasil penelitian menunjukkan bahwa model *Artificial Neural Network* (ANN), khususnya ANN dengan penyeimbangan kelas menggunakan *Synthetic Minority Over-Sampling Technique* (ANN + SMOTE), memberikan performa terbaik dalam analisis sentimen Program Makan Bergizi Gratis (MBG) di media sosial X. Model ANN Balanced mencapai akurasi tertinggi dibandingkan ANN tanpa penyeimbangan dan *K-Nearest Neighbor* (KNN), yang mengindikasikan bahwa pendekatan berbasis jaringan saraf lebih efektif dalam memodelkan kompleksitas data teks berdimensi tinggi yang direpresentasikan menggunakan TF–IDF.

Perbandingan antara ANN dan ANN Balanced menunjukkan bahwa penerapan SMOTE memberikan peningkatan performa yang konsisten, khususnya dalam meningkatkan kemampuan model mengenali kelas minoritas. Berdasarkan analisis *confusion matrix*, ANN Balanced menunjukkan distribusi prediksi yang lebih proporsional antar kelas sentimen serta peningkatan sensitivitas terhadap sentimen negatif. Temuan ini mengonfirmasi bahwa ketidakseimbangan kelas pada data sentimen publik dapat memengaruhi kinerja model, dan penyeimbangan data latih merupakan langkah penting untuk mengurangi bias terhadap kelas mayoritas.

Hasil ini sejalan dengan penelitian terdahulu yang melaporkan bahwa metode berbasis jarak seperti KNN memiliki keterbatasan dalam menangani data teks berdimensi tinggi dan tidak seimbang, sementara ANN dan pendekatan pembelajaran mendalam mampu menangkap pola nonlinier secara lebih efektif. Selain itu, efektivitas SMOTE dalam meningkatkan kinerja ANN pada kelas minoritas memperkuat temuan sebelumnya bahwa teknik penyeimbangan kelas berperan penting dalam analisis sentimen multikelas, terutama ketika proporsi sentimen negatif relatif kecil namun memiliki signifikansi substantif.

Dari sisi substansi kebijakan, distribusi sentimen menunjukkan bahwa mayoritas opini publik terhadap MBG bersifat netral dan positif, yang mencerminkan tingkat penerimaan masyarakat yang relatif tinggi terhadap tujuan program. Namun demikian, keberadaan sentimen negatif tetap memiliki arti penting karena merepresentasikan kritik publik terhadap aspek implementasi, seperti keamanan pangan, pengelolaan Satuan Pelayanan Pemenuhan Gizi (SPPG), serta transparansi dan akuntabilitas anggaran. Oleh karena itu, analisis sentimen berbasis media sosial dapat dimanfaatkan sebagai alat evaluasi berbasis data untuk mengidentifikasi isu-isu krusial secara lebih dini, sekaligus mendukung pengambilan keputusan dalam penyempurnaan pelaksanaan Program Makan Bergizi Gratis di masa mendatang.

IV. KESIMPULAN

Penelitian ini berhasil menganalisis sentimen publik terhadap Program Makan Bergizi Gratis (MBG) di media sosial Twitter (X) menggunakan algoritma *K-Nearest Neighbor* (KNN) dan *Artificial Neural Network* (ANN). Hasil analisis menunjukkan bahwa sentimen publik didominasi oleh sentimen netral (52,1%) dan positif (40,0%), sementara sentimen negatif relatif kecil (7,9%), yang mengindikasikan bahwa program MBG secara umum diterima dengan baik oleh masyarakat meskipun masih terdapat kritik yang berfokus pada aspek implementasi, seperti keamanan pangan dan pengelolaan program. Dari sisi metode, model ANN dengan penyeimbangan kelas menggunakan *Synthetic Minority Over-Sampling Technique* (ANN + SMOTE) menunjukkan performa terbaik dengan akurasi sebesar 93,58%, melampaui ANN tanpa penyeimbangan dan KNN, sehingga menegaskan pentingnya penanganan ketidakseimbangan kelas dalam analisis sentimen berbasis data media sosial. Untuk penelitian selanjutnya, disarankan untuk mengembangkan pendekatan dengan memanfaatkan model *deep learning* berbasis transformer seperti BERT atau IndoBERT, mengombinasikan fitur leksikal dengan fitur semantik, serta memperluas sumber data ke berbagai platform media sosial guna memperoleh gambaran sentimen publik yang lebih komprehensif dan mendalam terhadap kebijakan publik.

DAFTAR PUSTAKA

[1] V. Issue, "Pancasila : Jurnal Keindonesiaan," vol. 05, no. 1, 2025.

[2] Kementerian Sekretariat Negara and R. Indonesia, "Presiden Prabowo: Target 82 Juta Penerima Manfaat Makan Bergizi Gratis akan Terwujud Bertahap," 2025. [Online]. Available: https://www.setneg.go.id/baca/index/presiden_prabowo_target_82_juta_penerima_manfaat_makan_bergizi_gratis_akan_terwujud_bertahap

[3] mediakeuangan kemenkeu, "Pemerintah Salurkan Makan Bergizi Gratis (MBG), Ini Sasaran Utama Penerimanya Artikel ini telah tayang di situs Media Keuangan | MK+ dengan judul 'Pemerintah Salurkan Makan Bergizi Gratis (MBG), Ini Sasaran Utama Penerimanya - Media Keuangan' Lihat seleng," 2025. [Online]. Available: https://mediakeuangan.kemenkeu.go.id/article/show/pemerintah-salurkan-makan-bergizi-gratis-mbg-ini-sasaran-utama-penerimanya

[4] Antara News, "Makan Bergizi Gratis capai 38,5 juta penerima jelang akhir 2025", [Online]. Available: https://www.antaranews.com/berita/5207037/bgn-makan-bergizi-gratis-capai-385-juta-penerima-jelang-akhir-2025

[5] D. González-Fernández, O. Muralidharan, P. A. Neves, and Z. A. Bhutta, "Associations of Maternal Nutritional Status and Supplementation with Fetal, Newborn, and Infant Outcomes in Low-Income and Middle-Income Settings: An Overview of Reviews," Nov. 01, 2024, *Multidisciplinary Digital Publishing Institute (MDPI)*. doi: 10.3390/nu16213725.

[6] E. Kristjansson *et al.*, "School feeding programs for improving the physical and psychological health of school children experiencing socioeconomic disadvantage," *Cochrane Database Syst. Rev.*, vol. 2022, no. 8, Aug. 2022, doi: 10.1002/14651858.CD014794.

[7] M. W. Arif and K. Kustiyono, "Analisis Sentimen Kebijakan Makan Bergizi Gratis di Media Sosial Menggunakan Natural Language Processing Berbasis Python TextBlob di Indonesia," *J. Pendidik. dan Teknol. Indones.*, vol. 5, no. 9, pp. 2463–2471, Sep. 2025, doi: 10.52436/1.jpti.931.

[8] A. Ramadhani, I. Permana, M. Afdal, and M. Fronita, "Analisis Sentimen Tanggapan Publik di Twitter Terkait Program Kerja Makan Siang Gratis Prabowo – Gibran Menggunakan Algoritma Naïve Bayes Classifier dan Support Vector Machine," *Build. Informatics, Technol. Sci.*, vol. 6, no. 3, pp. 1509–1516, Dec. 2024, doi: 10.47065/bits.v6i3.6188.

[9] E. Triningsih, M. Afdal, I. Permana, and N. Evrilyan Rozanda, "Analisis Sentimen Terhadap Program Makan Bergizi Gratis Menggunakan Algoritma Machine Learning Pada Sosial Media X," *Technol. Sci.*, vol. 6, no. 4, pp. 2240–2250, 2025, doi: 10.47065/bits.v6i4.6534.

[10] G. R. Ati and P. T. Prasetyaningrum, "Analysis of Community Sentiment Towards Free Nutrition Meal Programs on Twitter Using Naïve Bayes, Support Vector Machine, K-Nearest Neighbors, and Ensemble Methods," *J. Inf. Syst. Informatics*, vol. 7, no. 2, pp. 1443–1460, Jul. 2025, doi: 10.51519/journalisi.v7i2.1098.

[11] Y. Zaen Vebrian, "A Sentiment Analysis Of Free Meal Plans On Social Media Using Naïve Bayes Algorithms Analisis Sentimen Terhadap Rencana Makan Gratis Di Sosial Media X Menggunakan Alogaritma Naive Bayes," vol. 10, no. 1, p. 2025.

[12] Y. R. Dewi, N. W. S. Saraswati, M. O. E. Monny, I. B. G. Sarasvananda, and I. G. Andika, "Sentiment Analysis of the Relocation of the National Capital on Social Media X," *Sinkron*, vol. 9, no. 2, pp. 625–636, 2025, doi: 10.33395/sinkron.v9i2.14622.

[13] D. T. Attaulah, D. Soyusiawaty, V. No, D. T. Attaulah, and D. Soyusiawaty, "Edumatic : Jurnal Pendidikan Informatika Analisis Sentimen Program Makan Siang Gratis di Twitter / X menggunakan Metode BI-LSTM," *Edumatic J. Pendidik. Inform.*, vol. 9, no. 1, pp. 294–303, Apr. 2025, doi: 10.29408/edumatic.v9i1.29725.

[14] I. Hermansyah and M. S. Hasibuan, "Sentiment Analysis of Free Nutritious Meal Programme on Social Media X using Linear Regression and Random Forest Algorithms," vol. 13, no. 225, pp. 49–54, 2026, doi: 10.33558/piksel.v13i1.10633.

[15] Satrio Junaidi, R. Valicia Anggela, and D. Kariman, "Klasifikasi Metode Data Mining untuk Prediksi Kelulusan Tepat Waktu Mahasiswa dengan Algoritma Naïve Bayes, Random Forest, Support Vector Machine (SVM) dan Artificial Neural Nerwork (ANN)," *J. Appl. Comput. Sci. Technol.*, vol. 5, no. 1, pp. 109–119, 2024, doi: 10.52158/jacost.v5i1.489.

[16] Mutiara Sintia Dewi and A. H. Hasugian, "Penerapan Support Vector Machine Untuk Analisis Sentimen Pada Tanggapan Masyarakat Di Media Sosial Terhadap Program Makan Siang Gratis," *Rabit J. Teknol. dan Sist. Inf. Univrab*, vol. 10, no. 2, pp. 911–922, Jul. 2025, doi: 10.36341/rabit.v10i2.6425.

[17] A. Mardiana, Z. Abidin, A. Mardiana, and Z. Abidin, "The Influence of User Sentiment Analysis X on the Free Nutritious Food Program Using the K-Nearest Neighbors Method and Logistic Regression Citation: Article History," *Res. Horiz.*, vol. 5, no. 4, pp. 1655–1668.

[18] J. Muliawan and E. Dazki, "Sentiment Analysis Of Indonesia ' S Capital City Relocation Using Three Algorithms : Naïve Bayes , Knn , And Random Forest Analisis Sentimen Pemindahan Ibu Kota Negara Indonesia Menggunakan Tiga Algoritma : Naïve Bayes , KNN , Dan Random," vol. 4, no. 5, pp. 1227–1236, 2023.

[19] T. Astuti and I. Pratika, "Product Review Sentiment Analysis by Artificial Neural Network Algorithm," vol. 2, no. 2, pp. 61–66, 2019.

[20] H. H. Limbong, "Optimasi Analisis Sentimen Ulasan Aplikasi Amikom One Menggunakan SMOTE pada Algoritma Artificial Neural Network Optimization of Sentiment Analysis for Amikom

One Application Reviews Using SMOTE with Artificial Neural Network Algorithm," vol. 13, pp. 2048–2059, 2024.
[21] R. Afandi, M. Afdal, and R. Novita, "Analisis Sentimen Masyarakat Terhadap Pinjaman Online di Twitter Menggunakan Algoritma Naïve Bayes Classifier dan K-Nearest N eighbor," vol. 6, no. 2, pp. 596–605, 2024, doi: 10.47065/bits.v6i2.5300.
[22] N. Hendrastuty, A. Rahman Isnain, and A. Yanti Rahmadhani, "Analisis Sentimen Masyarakat Terhadap Program Kartu Prakerja Pada Twitter Dengan Metode Support Vector Machine," vol. 6, no. 3, 2021, [Online]. Available: http://situs.com
[23] M. K. Anam, B. N. Pikir, M. B. Firdaus, and S. Erlinda, "Penerapan Na ¨ ıve Bayes Classifier , K-Nearest Neighbor (KNN) dan Decision Tree untuk Menganalisis Sentimen pada Interaksi Netizen dan Pemeritah Applications of Na ¨ ıve Bayes Classifier , K-Nearest Neighbor and Decision Tree to Analyze Sentiment on Ne," vol. 21, no. 1, 2021, doi: 10.30812/matrik.v21i1.1092.
[24] D. Haliza and M. Ikhsan, "Sentiment Analysis on Public Perception of the Nusantara Capital on Social Media X Using Support Vector Machine (SVM) and K-Nearest Neighbor (K-NN) Methods," 2025. [Online]. Available: http://jurnal.polibatam.ac.id/index.php/JAIC
[25] U. Brawijaya, "Pembentukan Daftar Stopword Goffman Transition Point Dengan Pembobotan Emoji Pada Analisis Sentimen Di Twitter Establishment Of Goffman Transition Point Stopword List With Emoji Weighting On Sentiment Analysis In Twitter," vol. 9, no. 4, pp. 1101–1108, 2022, doi: 10.25126/jtiik.202294706.
[26] M. Bilal, G. Ali, M. W. Iqbal, M. Anwar, M. Sheraz, and A. Malik, "Auto-Prep : Efficient and Automated Data Preprocessing Pipeline," *IEEE Access*, vol. 10, no. June, pp. 107764–107784, 2022, doi: 10.1109/ACCESS.2022.3198662.
[27] M. U. Albab, Y. K. P, and M. N. Fawaiq, "Optimization of the Stemming Technique on Text preprocessing President 3 Periods Topic," vol. 20, no. 2, pp. 1–10, 2023.
[28] A. L. I. Erkan and T. Güngör, "Analysis of Deep Learning Model Combinations and Tokenization Approaches in Sentiment Classification," no. November, pp. 134951–134968, 2023.
[29] A. Bijaksana and P. Negara, "The Influence Of Applying Stopword Removal And Smote On Indonesian Sentiment Classification," vol. 14, no. 3, pp. 172–185, 2023.
[30] Z. Abidin, "Text Stemming and Lemmatization of Regional Languages in Indonesia : A Systematic Literature Review," vol. 10, no. 2, pp. 217–231, 2024.
[31] S. K. Narayanasamy, Y. Hu, and S. M. Qaisar, "Effective Preprocessing and Normalization Techniques for COVID-19 Twitter Streams with POS Tagging via Lightweight Hidden Markov Model," vol. 2022, 2022, doi: 10.1155/2022/1222692.
[32] N. Eldha Oktaviana and Y. Arum Sari, "Analisis Sentimen Terhadap Kebijakan Kuliah Daring Selama Pandemi Menggunakan Pendekatan Lexicon Based Features Dan Support Vector Machine", doi: 10.25126/jtiik.202295625.
[33] A. C. Najib, A. Irsyad, G. A. Qandi, and N. Aini, "Perbandingan Metode Lexicon-based dan SVM untuk Analisis Sentimen Berbasis Ontologi pada Kampanye Pilpres Indonesia Tahun 2019 di Twitter Abstrak," vol. 4, no. 2, 2019.
[34] Y. Asri, W. N. Suliyanti, D. Kuswardani, and M. Fajri, "Pelabelan Otomatis Lexicon Vader dan Klasifikasi Naive Bayes dalam menganalisis sentimen data ulasan PLN Mobile," vol. 15, no. 2, pp. 264–275, 2022.
[35] D. Pateman, T. F. Prasetyo, H. Sujadi, and U. Majalengka, "Sentiment Analysis Of Government On Tiktok And X," vol. 10, no. 4, pp. 900–908, 2025, doi: 10.33480/jitk.v10i4.6645.(SVM).
[36] R. Sistem, "Studi Komparatif Metode Ekstraksi Fitur pada Analisis Sentimen," vol. 1, no. 10, pp. 9–12, 2021.
[37] D. E. Cahyani and I. Patasik, "Performance comparison of TF-IDF and Word2Vec models for emotion text classification," vol. 10, no. 5, pp. 2780–2788, 2021, doi: 10.11591/eei.v10i5.3157.
[38] A. A. Huda, R. Fajarudin, and A. Hadinegoro, "Sistem Rekomendasi Content-based Filtering Menggunakan TF-IDF Vector Similarity Untuk Rekomendasi Artikel Berita," vol. 4, no. 3, pp. 1679–1686, 2022, doi: 10.47065/bits.v4i3.2511.
[39] Z. R. Tembusai, H. Mawengkang, M. Zarlis, A. Info, and A. H. Process, "K-Nearest Neighbor with K-Fold Cross Validation and Analytic Hierarchy Process on Data Classification," vol. 2, no. 1, pp. 1–8, 2021, doi: 10.25008/ijadis.v2i1.1204.
[40] B. H. Hayadi, I. G. I. Sudipa, and A. P. Windarto, "Model peramalan pada peserta KB aktif jalur pemerintahan menggunakan Artificial Neural Network Back-propagation Artificial Neural Network Back-propagation models for active family planning participants in the government pathway," vol. 21, no. 1, 2021, doi: 10.30812/matrik.v21i1.1273.
[41] S. H. Gulo, A. H. Lubis, T. Informatika, F. Teknik, and U. M. Area, "Penerapan Multi-Layer Perceptron untuk Mengklasifikasi Penduduk Kurang Mampu," vol. 4, no. 2, pp. 51–59, 2024.
[42] E. Saputra and E. R. Susanto, "Implementation of Deep Learning with Multilayer Perceptron (MLP) for Heart Disease Prediction Using the SMOTE-ENN Technique," vol. 9, no. 3, pp. 1034–1041, 2025.
[43] K. Neighbor *et al.*, "Analisis Sentimen Komentar Konsumen Industri Jamu di Media Sosial menggunakan Artificial Neural Network dan," *J. Sist. Inf. Bisnis*, vol. 03, no. 3, pp. 210–223, Aug. 2024, doi: 10.21456/vol14iss3pp210-223.
[44] A. Y. Pratama, G. A. Sanjaya, N. K. Lubis, and M. R. Aditya, "Analisis Sentimen Publik Terkait Danantara Menggunakan Algoritma IndoBERT pada Platform Media Sosial," 2025, doi: 10.47002/metik.v9i1.1055.