Jurnal Elektronika dan Teknik Informatika Terapan (JENTIK)
Volume 4 Nomor 3 SEPTEMBER 2026
e-ISSN : 2988-0874; p-ISSN : 2988-0866, Hal 384-401
DOI: https://doi.org/10.59061/jentik.v4i3.1582
Tersedia: https://e-journal.poltek-kampar.ac.id/index.php/JENTIK

ANALISIS SENTIMEN TERHADAP PROGRAM (MBG) PADA *TWITTER* MENGGUNAKAN ALGORITMA *K-NEAREST NEIGHBORS*

Muhammad Arilda Aprilianto[1*], Wahyu Hadikristanto[2], Ahmad Turmudi Zy[3]

[1,2,3]Teknik Informatika, Universitas Pelita Bangsa, Indonesia
[1]*rilda.araa@gmail.com

Alamat: Jl. Inspeksi Kalimalang No.9, Cibatu, Cikarang Selatan, Bekasi, Indonesia
Korespondensi penulis: rilda.araa@gmail.com

Abstract. *The Free Nutritious Meals (MBG) program has generated extensive discussion among users of Twitter (X), producing a substantial volume of opinion-based text. This study examines public responses to MBG by classifying Twitter posts with the K-Nearest Neighbors (KNN) method. The resulting model reached an accuracy of 88%. For negative sentiment, precision and recall were 0.90 and 0.97, respectively, whereas the positive class obtained a precision of 0.61 and a recall of 0.30. The unequal performance across classes is associated with the imbalanced dataset, in which negative samples were considerably more numerous than positive samples. These findings indicate that KNN can be applied to Twitter-based sentiment classification of the MBG program, although its recognition performance is stronger for negative sentiment than for positive sentiment.*

Keywords: *Sentiment Analysis, Program Makan Bergizi Gratis (MBG), Twitter (X), K-Nearest Neighbors (KNN), TF-IDF*

Abstrak. Program Makan Bergizi Gratis (MBG) memunculkan banyak tanggapan dari pengguna *Twitter* (X) sehingga tersedia data opini berbentuk teks dalam jumlah besar. Penelitian ini mengkaji kecenderungan sentimen publik terhadap MBG dengan melakukan klasifikasi menggunakan algoritma *K-Nearest Neighbors* (KNN). Hasil pengujian menghasilkan akurasi 88%. Pada kelas negatif, *precision* mencapai 0,90 dengan *recall* 0,97, sedangkan kelas positif memperoleh *precision* 0,61 dan *recall* 0,30. Perbedaan capaian kedua kelas berkaitan dengan komposisi data yang tidak seimbang karena jumlah sampel negatif jauh lebih dominan daripada sampel positif. Temuan tersebut memperlihatkan bahwa KNN dapat digunakan untuk analisis sentimen MBG di *Twitter*, tetapi kemampuan model dalam mengenali sentimen positif masih lebih rendah dibandingkan sentimen negatif.

Kata kunci: Analisis Sentimen, Program Makan Bergizi Gratis (MBG), Twitter (X), K-Nearest Neighbors (KNN), TF-IDF

1. LATAR BELAKANG

Pendahuluan

Kualitas sumber daya manusia menjadi salah satu prasyarat penting untuk mendukung pencapaian visi Indonesia Emas 2045. Salah satu persoalan yang masih perlu ditangani adalah stunting karena dampaknya berhubungan dengan kualitas generasi pada masa mendatang. Survei Status Gizi Indonesia (SSGI) 2024 mencatat bahwa prevalensi stunting nasional turun dari 21,5% pada 2023 menjadi 19,8% pada 2024(Kementerian Kesehatan Republik Indonesia, 2025). Walaupun menunjukkan perbaikan, capaian

Naskah Masuk: 15 September 2026; Revisi: 21 September 2026; Diterima: 22 September 2026; Terbit: 30 September 2026

tersebut belum mencapai sasaran pemerintah sebesar 18% pada 2025 dan masih berada di sekitar batas maksimal 20% yang digunakan World Health Organization (WHO)(Putriyekti et al., 2025).

Dampak stunting tidak hanya berkaitan dengan kondisi kesehatan anak, tetapi juga dapat memengaruhi pembangunan nasional dalam jangka panjang(Rendy, Hamdi, Muhammadiyah Jember, & Pemerintahan, 2025). Anak yang mengalami stunting berisiko menghadapi hambatan pertumbuhan fisik dan gangguan kognitif, sehingga kualitas sumber daya manusia pada masa depan dapat ikut terdampak(Martony, 2023). Untuk merespons persoalan tersebut, pemerintah menjalankan berbagai kebijakan, salah satunya Program Makan Bergizi Gratis (MBG) yang ditujukan bagi pelajar, balita, ibu hamil, serta ibu menyusui(Putu Alya Noviyanti et al., 2025).

Pelaksanaan MBG secara serentak dimulai pada 6 Januari 2025 dengan sasaran penerima hingga 82,9 juta orang(Ilona Estherina, 2025). Tahap awal program mencakup 26 provinsi dan direncanakan berkembang hingga menjangkau seluruh 38 provinsi. Rantai pengadaan dan penyaluran makanan juga melibatkan UMKM lokal, sehingga program ini diharapkan memberi efek tambahan bagi perekonomian daerah sekaligus mendukung ketahanan pangan(Alif Bais Khoiriyah, 2025). Dari sisi tujuan, MBG diarahkan bukan hanya untuk memperbaiki pemenuhan gizi, tetapi juga untuk mendukung pengurangan kemiskinan, peningkatan produktivitas generasi muda, serta pembentukan sumber daya manusia yang sehat dan berdaya saing(Eddy Cahyono Sugiarto, 2024).

Penerapan Program Makan Bergizi Gratis (MBG) memunculkan tanggapan yang beragam di tengah masyarakat, termasuk di *Twitter*. Sebagian pengguna menilai program tersebut bermanfaat bagi pemenuhan gizi siswa dan kelompok rentan, sementara pengguna lain mempertanyakan efektivitas pelaksanaan, kecukupan anggaran, dan mekanisme distribusinya(Fatkhurrohman, Nugroho, & Fadillah, 2025). Keragaman respons tersebut membentuk perbedaan kecenderungan opini, sehingga analisis sentimen diperlukan untuk memetakan bagaimana masyarakat merespons kebijakan MBG.

Twitter menyediakan ruang bagi pengguna untuk menyampaikan pendapat mengenai isu yang sedang berkembang. Dalam konteks MBG, platform ini memuat beragam ekspresi, penilaian, dan respons masyarakat yang dapat dimanfaatkan sebagai sumber data opini. Analisis sentimen kemudian digunakan untuk merangkum kecenderungan respons

tersebut dan membantu membaca perubahan pola opini publik(Putri Ajie & Wardhana, 2025).

Ringkasan penelitian terdahulu disajikan pada tabel berikut berdasarkan sumber penelitian, data masukan, metode yang digunakan, dan hasil yang diperoleh:

Sumber	Input	Metode	Output
(Isnain, Supriyanto, & Kharisma, 2021)	Penelitian menggunakan 1.825 *tweet* berbahasa Indonesia yang membahas kebijakan pembelajaran daring. Data dihimpun melalui *Tweepy* pada periode 1 Februari-30 September 2020, kemudian direpresentasikan dengan *TF-IDF* dan dikelompokkan ke dalam sentimen positif serta negatif.	TF-IDF + KNN	KNN digunakan untuk mengklasifikasikan sentimen mengenai pembelajaran daring. Dari beberapa nilai K yang diuji, K = 10 memberikan hasil tertinggi dengan *Accuracy* 84,65%, *Precision* 87, *Recall* 86%, dan *F-measure* 87%.
(Ariyanto & Saefurrohman, 2025)	Sebanyak 1.337 unggahan X mengenai Pemilihan Gubernur Jawa Tengah 2024 menjadi data penelitian dan diberi kelas positif atau negatif. Pengumpulan dilakukan melalui *RapidMiner*, sedangkan pemrosesan analitik menggunakan *Python* di *Google Colab*.	KNN	Pada KNN, skenario pembagian 90:10 memberikan performa terbaik: *Accuracy* 81,00%, *Precision* 78,36%, *Recall* 64,78%, dan *F1-score* 67,21%.
(Siregar, Dermawan, & Hisyam, 2025)	Penelitian memanfaatkan 1.453 *tweet* atau komentar X tentang Danantara. Data dikumpulkan dengan *crawling* lalu	KNN	Performa KNN tertinggi diperoleh ketika *preprocessing* terdiri atas *Cleaning*, *Case Folding*, *Tokenizing*, dan

	disusun menjadi *dataset* CSV untuk proses analisis.		*Stemming*, dengan *Accuracy* 68,89%.

Tiga penelitian terdahulu memperlihatkan bahwa KNN telah diterapkan untuk mengelompokkan sentimen pada data dari *Twitter*/X. Nilai akurasi yang dilaporkan tidak sama, yakni 84,65%, 81,00%, dan 68,89%, karena setiap penelitian menggunakan karakteristik data, objek kajian, tahapan *preprocessing*, jumlah kelas, serta skema pembagian data yang berbeda. Penelitian ini tetap menggunakan *Twitter*/X dan KNN, tetapi berfokus pada Program Makan Bergizi Gratis (MBG). *Dataset* yang dianalisis terdiri atas 4.968 *tweet* yang dikelompokkan ke dalam sentimen positif dan negatif, kemudian diproses melalui *preprocessing* serta pembobotan *TF-IDF*. Pengujian menghasilkan akurasi 88%, sehingga penerapan KNN pada data penelitian ini mampu menghasilkan klasifikasi sentimen MBG dengan tingkat ketepatan yang cukup tinggi.

2. KAJIAN TEORITIS

Data Mining

Data mining membahas cara memperoleh pola atau pengetahuan yang bernilai dari kumpulan data dalam jumlah besar. Proses tersebut dapat menggabungkan pendekatan statistika, kecerdasan buatan, *machine learning*, dan teknologi basis data untuk menemukan informasi yang tidak langsung terlihat pada data mentah. Informasi hasil penggalian selanjutnya dapat dimanfaatkan sebagai bahan pendukung dalam pengambilan keputusan(Sirichanya & Kraisak, 2021).

Program Makan Bergizi Gratis (MBG)

Program Makan Bergizi Gratis (MBG) merupakan kebijakan pemerintah yang diarahkan untuk memperbaiki pemenuhan gizi masyarakat sekaligus mendukung pembangunan sumber daya manusia. Penerima program mencakup peserta didik pada pendidikan dasar dan menengah, balita, ibu hamil, serta ibu menyusui yang memperoleh makanan bergizi tanpa biaya(Suprapto, Praditya, Dewi, & Adiyoso, 2025). MBG ditempatkan sebagai program prioritas nasional dan dihubungkan dengan agenda jangka panjang peningkatan kualitas generasi muda. Gagasan penyediaan makanan bagi peserta didik juga memiliki

kemiripan dengan *school feeding program* yang telah dijalankan di sejumlah negara, antara lain Brasil, India, Finlandia, serta berbagai negara berkembang lainnya.

Analisis Sentimen

Analisis sentimen merupakan bidang yang memanfaatkan *Natural Language Processing* (NLP), *text mining*, dan *machine learning* untuk membaca kecenderungan opini atau emosi dari data teks. Melalui proses ini, suatu pernyataan dapat dikelompokkan, misalnya, sebagai sentimen positif, negatif, atau netral(Mao, Liu, & Zhang, 2024). Pada tahap awal perkembangannya, analisis sentimen banyak digunakan untuk klasifikasi dua polaritas. Seiring perkembangan *artificial intelligence*, *machine learning*, dan *deep learning*, cakupannya meluas hingga analisis emosi, *aspect-based sentiment analysis*, tingkat kekuatan sentimen, serta konteks bahasa yang lebih rumit seperti ironi, sarkasme, dan ambiguitas(Jim et al., 2024).

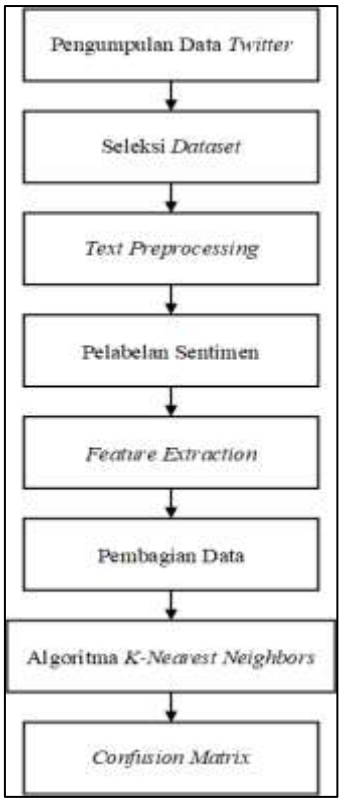
K-Nearest Neighbors

K-Nearest Neighbors (KNN) adalah metode *supervised learning* yang menentukan kelas sebuah data berdasarkan kedekatannya dengan data latih yang telah memiliki label. Ketika data baru akan diklasifikasikan, KNN membandingkannya dengan sejumlah tetangga terdekat dan menggunakan informasi kelas pada tetangga tersebut sebagai dasar penentuan label. Metode ini relatif sederhana untuk diterapkan, sedangkan kualitas hasilnya sangat bergantung pada karakteristik data dan parameter yang dipilih(Halder, Uddin, Uddin, Aryal, & Khraisat, 2024).

Confusion Matrix

Confusion Matrix digunakan untuk merangkum hubungan antara label aktual dan hasil prediksi pada masalah klasifikasi. Untuk klasifikasi dua kelas, tabel ini memuat empat kemungkinan hasil, yaitu *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN). TN menunjukkan data negatif yang diprediksi negatif, sedangkan FP terjadi ketika data negatif diprediksi sebagai positif. TP merupakan data positif yang dikenali secara tepat sebagai positif, sementara FN menunjukkan data positif yang justru diprediksi sebagai negatif(Hovi, Id Hadiana, & Rakhmat Umbara, 2022).

3. METODE PENELITIAN



Gambar 1 Metode Penelitian

Alur penelitian pada Gambar 1 dijabarkan melalui tahapan berikut:

1. Pengumpulan Data

   Data penelitian dikumpulkan dari *Twitter*/X melalui proses *crawling* dengan kata kunci "MBG". Proses tersebut menghasilkan 4.968 *tweet* yang memuat opini masyarakat terkait Program Makan Bergizi Gratis.

2. Seleksi Data

   Kumpulan *tweet* hasil *crawling* diperiksa terlebih dahulu. Data yang tidak berkaitan dengan topik, tercatat lebih dari sekali, kosong, atau tidak layak diolah dikeluarkan dari *dataset* sebelum analisis dilanjutkan.

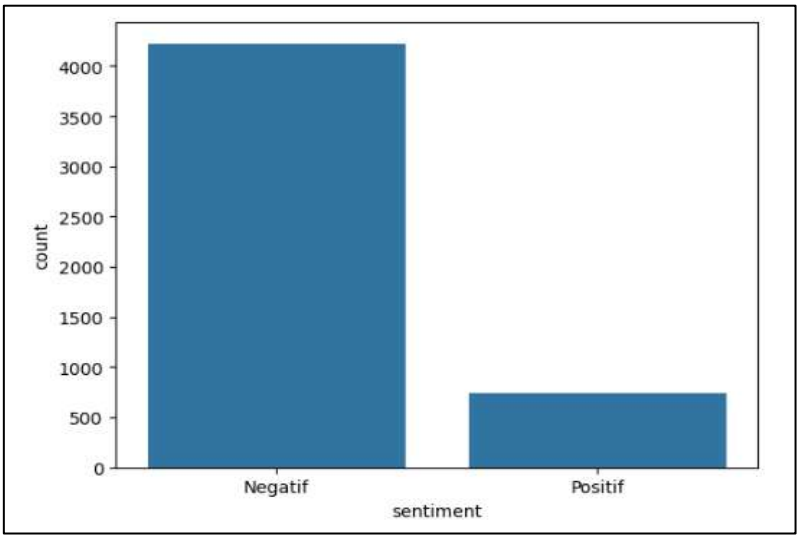
3. *Text Preprocessing*

   Sebelum masuk ke tahap pemodelan, teks distandarkan agar bentuk datanya lebih konsisten. Proses ini mencakup *Case Folding*, *Cleaning*, *Tokenizing*, *Stopword Removal*, dan *Stemming*.

   Rangkaian *preprocessing* mengurangi unsur teks yang tidak diperlukan sehingga data yang diteruskan ke tahap klasifikasi menjadi lebih teratur dan relevan.

4. Pelabelan Sentimen

   Sesudah *preprocessing*, setiap *tweet* diberi label positif atau negatif secara manual. Label tersebut berperan sebagai target yang akan dipelajari pada proses klasifikasi *supervised learning*.



Gambar 2 Visualisasi Jumlah *Class*

5. *Feature Extraction*

   Teks hasil preprocessing selanjutnya dikonversi ke representasi numerik dengan TF-IDF (*Term Frequency-Inverse Document Frequency*). Bobot tiap kata ditentukan berdasarkan frekuensinya di dokumen dan tingkat kekhasannya di seluruh kumpulan dokumen.

6. Pembagian Data

   Data dipisahkan menjadi dua kelompok, yaitu *training data* untuk membentuk model dan *testing data* untuk menilai kemampuan model pada data yang tidak digunakan saat pelatihan.

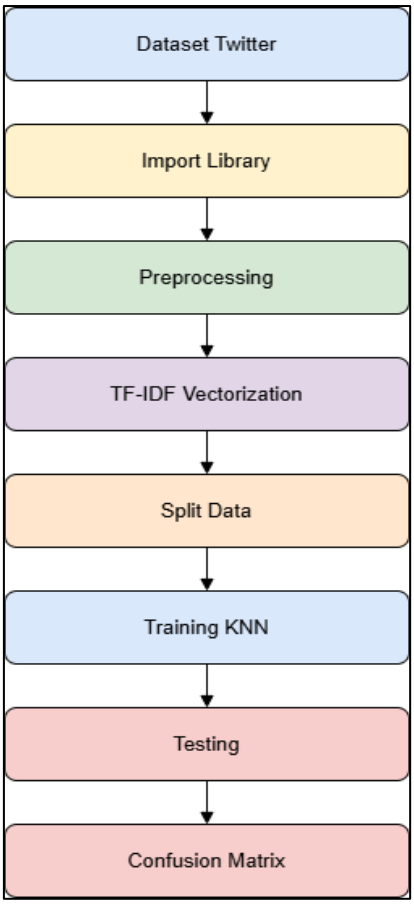
7. Klasifikasi Menggunakan Algoritma *K-Nearest Neighbors*

   Model *K-Nearest Neighbors* dibangun menggunakan *training data*. Untuk menentukan label data baru, algoritma melihat kelas yang paling banyak muncul di antara sejumlah tetangga terdekat sesuai nilai K yang ditetapkan.

8. Evaluasi Model

   Pada tahap akhir, hasil prediksi dinilai melalui *Confusion Matrix*. Ukuran *Accuracy*, *Precision*, dan *Recall* kemudian digunakan untuk menggambarkan kinerja model klasifikasi.

Pengujian



Gambar 3 Model Pengujian

Rangkaian pada model pengujian tersebut dapat dijelaskan sebagai berikut.

1. *Import Dataset*

   Dataset penelitian disiapkan terlebih dahulu sebagai sumber data yang akan diproses dalam pengujian.

2. *Import Library*

   Library yang dibutuhkan dimuat sebelum analisis dimulai agar seluruh fungsi untuk *text mining* dan *machine learning* dapat digunakan selama proses pengolahan.

3. *Text Preprocessing*

   Preprocessing berfungsi menyiapkan data teks dengan mengurangi ketidakteraturan pada data mentah. Hasil dari tahap ini adalah teks yang lebih bersih dan konsisten untuk digunakan dalam klasifikasi.

4. *Feature Extraction*

   Karena algoritma *machine learning* memerlukan masukan numerik, teks yang telah melalui *preprocessing* harus diubah terlebih dahulu. Konversi tersebut dilakukan pada tahap *Feature Extraction* menggunakan *Term Frequency-Inverse Document Frequency* (TF-IDF).

5. Pembagian *Dataset*

   Setelah fitur numerik terbentuk, *dataset* dipisahkan menjadi *training data* dan *testing data* untuk kebutuhan pembentukan serta pengujian model.

6. Penerapan Algoritma *K-Nearest Neighbors*

   K-Nearest Neighbors (KNN) kemudian diterapkan untuk membangun model yang akan mengelompokkan sentimen.

7. Evaluasi Model

   Kinerja model diperiksa pada tahap terakhir dengan memanfaatkan *Confusion Matrix* dan *Classification Report* sebagai dasar evaluasi hasil klasifikasi.

4. HASIL DAN PEMBAHASAN

Hasil Pengujian

Eksperimen dijalankan menggunakan *Python* melalui *Google Colaboratory*. Proses dimulai dengan memuat *library* dan *dataset*, dilanjutkan *preprocessing*, pembentukan fitur TF-IDF, pemisahan *training-testing*, pelatihan KNN, serta pengukuran hasil melalui *Confusion Matrix* dan *Classification Report*.

Import Library

Sejumlah *library* digunakan untuk menjalankan pengolahan data dan membangun model analisis sentimen. Masing-masing menyediakan fungsi yang berbeda, mulai dari manipulasi data, *preprocessing* teks, visualisasi, pemodelan *machine learning*, sampai evaluasi klasifikasi. *Library* utama yang dipakai meliputi *Pandas*, *NumPy*, *Matplotlib*, *Seaborn*, NLTK, *Sastrawi*, dan *Scikit-Learn*.

```
import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
import nltk
import re
from nltk.corpus import stopwords
from Sastrawi.Stemmer.StemmerFactory import StemmerFactory
from sklearn.feature_extraction.text import CountVectorizer

nltk.download('stopwords')

stop_words = set(stopwords.words('indonesian'))

factory = StemmerFactory()
stemmer = factory.create_stemmer()
```

```
[nltk_data] Downloading package stopwords to /root/nltk_data...
[nltk_data]   Package stopwords is already up-to-date!
```

Gambar 4 *Import Library*

Import Dataset

Dataset yang berhasil dimuat berjumlah 4.968 *tweet*. Data tersebut mempunyai dua atribut utama: *clean_text* yang menyimpan teks setelah *preprocessing* dan Sentimen yang menunjukkan label kelas. Susunan atribut *dataset* dirangkum pada tabel berikut.

Tabel 1 Struktur *Dataset*

No	Nama Atribut	Tipe Data	Keterangan
1	*clean_text*	Text	Memuat teks tweet setelah melalui preprocessing.
2	Sentimen	Kategori	Menyimpan kategori sentimen positif atau negatif.

Text Preprocessing

Text preprocessing dilakukan untuk menata dan menormalkan teks sebelum fitur dibentuk dan klasifikasi *K-Nearest Neighbors* (KNN) dijalankan.

1. *Cleaning*

   Cleaning menghilangkan unsur yang tidak diperlukan untuk analisis, seperti URL, *username*, *hashtag*, angka, tanda baca, emoji, karakter khusus, serta spasi berlebih. Penghapusan dilakukan secara otomatis menggunakan pola *Regular Expression* (Regex).

Tabel 2 Hasil Proses *Cleaning*

Sebelum *Cleaning*	**Setelah *Cleaning***
@user Program MBG sangat bagus!! 😊 https://twitter.com	Program MBG sangat bagus

2. *Case Folding*

Sesudah *cleaning*, huruf pada teks diseragamkan melalui *case folding* dengan mengubah karakter alfabet menjadi *lowercase*. Langkah ini mencegah kata yang sama dianggap berbeda hanya karena variasi penggunaan huruf kapital.

Tabel 3 Hasil *Case Folding*

Sebelum	**Sesudah**
Program MBG Sangat Bagus	program mbg sangat bagus

3. *Tokenizing*

Tokenizing memisahkan kalimat menjadi unit-unit kata atau *token*. Token yang terbentuk kemudian digunakan sebagai unsur dasar ketika bobot fitur dihitung dengan *TF-IDF*.

Tabel 4 Proses *Tokenizing*

['program', 'mbg', 'sangat', 'membantu', 'masyarakat']

4. *Stopword Removal*

Stopword removal menyaring kata umum yang frekuensinya tinggi tetapi kontribusinya terhadap pembedaan sentimen relatif kecil. Daftar *stopword* Bahasa Indonesia pada penelitian ini berasal dari NLTK dan disesuaikan kembali dengan kebutuhan data yang dianalisis.

Tabel Hasil 5 *Stopword Removal*

Sebelum	**Sesudah**
program mbg sangat membantu masyarakat	program mbg membantu masyarakat

5. *Stemming*

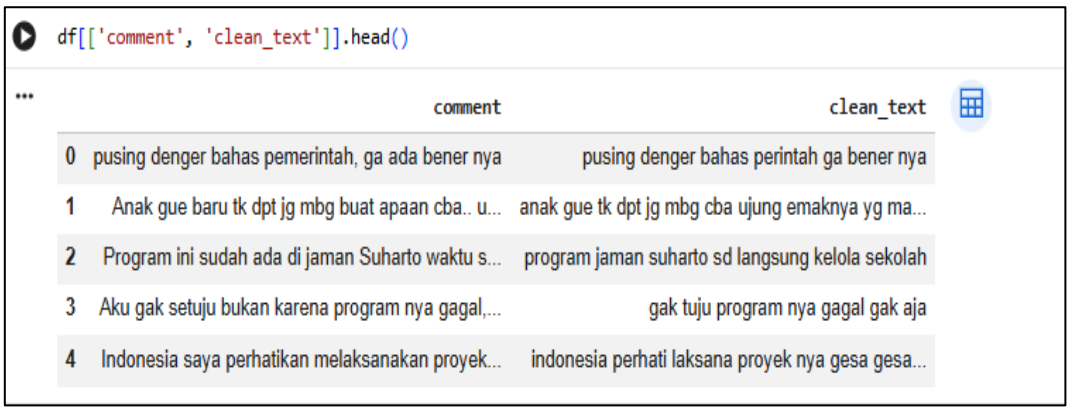
Stemming menjadi langkah penutup pada rangkaian *preprocessing*. Pada tahap ini, kata berimbuhan dikembalikan ke bentuk dasarnya agar variasi kata yang berasal dari akar yang sama dapat direpresentasikan secara lebih konsisten. Proses

tersebut menggunakan *library* Sastrawi yang mendukung pengolahan Bahasa Indonesia.

Tabel 6 Hasil *Stemming*

Sebelum	Sesudah
membantu	bantu
diberikan	beri
makanan	makan
bergizi	gizi

Dengan *stemming*, keragaman bentuk kata di dalam *dataset* dapat dikurangi. Representasi fitur menjadi lebih ringkas dan KNN lebih mudah membandingkan pola kata yang berkaitan dengan masing-masing sentimen.



Gambar 5 Hasil Akhir *Preprocessing*

Hasil akhir *text preprocessing* adalah data teks yang lebih teratur dibandingkan bentuk awalnya. Unsur yang tidak relevan dihapus, kapitalisasi dinormalkan, kalimat dipecah menjadi token, *stopword* disaring, dan kata dikembalikan ke bentuk dasarnya. Rangkaian ini menyiapkan representasi teks yang lebih sesuai untuk pembentukan fitur dan klasifikasi *K-Nearest Neighbors*.

Pembagian Data

Dataset dibagi dengan komposisi 80% sebagai *training data* dan 20% sebagai *testing data*. Porsi yang lebih besar untuk pelatihan diberikan agar model memperoleh cukup contoh dalam mempelajari pola sentimen, sementara bagian pengujian tetap tersedia untuk mengukur kemampuan generalisasi. Jumlah data pada masing-masing bagian disajikan pada tabel berikut.

Tabel 7 Pembagian *Dataset*

Jenis Data	**Jumlah**
Data Training	3.974
Data Testing	994
Total	**4.968**

Pembagian tersebut menghasilkan 3.974 *tweet* untuk pelatihan dan 994 *tweet* untuk pengujian. Proses pemisahan dilakukan secara acak dengan *random_state* = 42, sehingga susunan data dapat diperoleh kembali secara konsisten ketika prosedur yang sama dijalankan ulang.

Penerapan Algoritma *K-Nearest Neighbors* (KNN)

Sesudah pemisahan *training* dan *testing* selesai, model klasifikasi dibentuk dengan algoritma *K-Nearest Neighbors* (KNN). Implementasinya menggunakan *KNeighborsClassifier* yang tersedia pada *library Scikit-Learn*.

```
# KNN
from sklearn.neighbors import KNeighborsClassifier

knn = KNeighborsClassifier(
    n_neighbors=5,
    metric='cosine'
)

knn.fit(X_train, y_train)
```

```
KNeighborsClassifier
KNeighborsClassifier(metric='cosine')
```

Gambar 6 Pelatihan Model KNN

Parameter K ditetapkan sebesar 5 agar keputusan klasifikasi mempertimbangkan sejumlah kecil tetangga terdekat dan tetap mempertahankan informasi lokal pada data teks. Karena *tweet* telah direpresentasikan melalui TF-IDF, ukuran *cosine* digunakan untuk membandingkan arah antarvektor teks. Berdasarkan ukuran kemiripan tersebut, KNN mencari lima sampel terdekat sebagai dasar penentuan kelas.

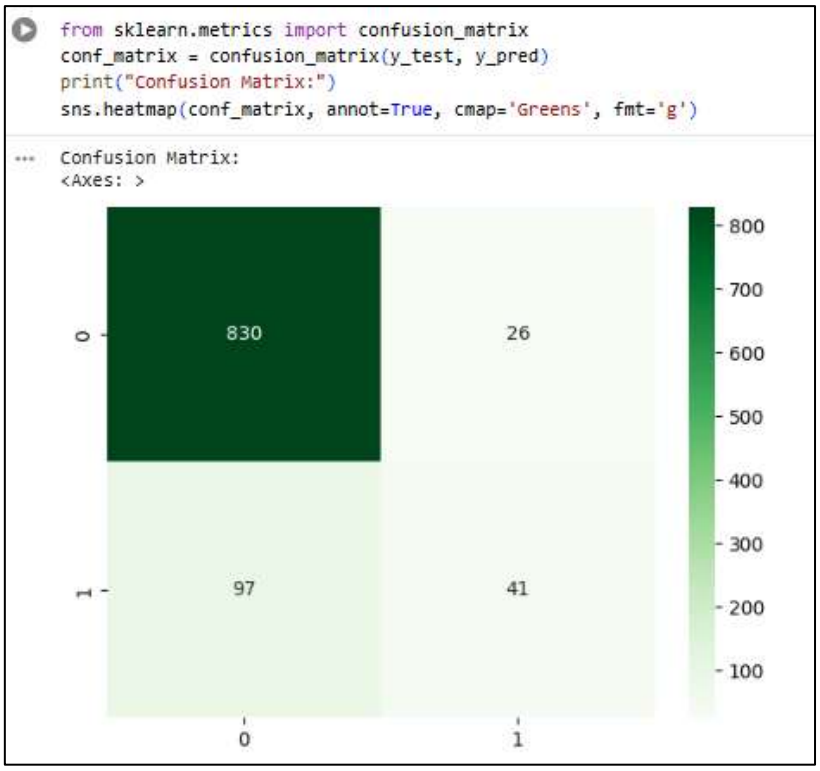
Model yang telah selesai dilatih kemudian diterapkan pada *testing data*. Kelas sentimen setiap sampel diperkirakan melalui fungsi *predict()*.

```
# Prediksi
y_pred = knn.predict(X_test)
```

Gambar 7 Proses Prediksi Menggunakan KNN

Confusion Matrix

Confusion Matrix merinci jumlah prediksi yang sesuai dan tidak sesuai dengan label sebenarnya pada setiap kelas. Informasi ini memperlihatkan pola kesalahan model dalam membedakan sentimen positif dan negatif, sehingga evaluasinya lebih terperinci daripada hanya menggunakan akurasi. Hasil *Confusion Matrix K-Nearest Neighbors* disajikan pada gambar berikut.



Gambar 8 Hasil *Confusion Matrix*

Dari 994 data uji, prediksi yang tepat terdiri atas 830 sentimen negatif dan 41 sentimen positif. Kesalahan klasifikasi mencakup 26 data negatif yang diprediksi positif serta 97 data positif yang diprediksi negatif.

Komposisi hasil tersebut memperlihatkan bahwa pengenalan kelas negatif jauh lebih kuat daripada kelas positif. Salah satu faktor yang berkaitan dengan perbedaan ini adalah ketidakseimbangan *dataset*, karena sampel negatif berjumlah lebih banyak. Secara total, 871 dari 994 *testing data* berhasil diberi label secara benar oleh model.

Kesalahan pada kelas negatif hanya berjumlah 26 data. Hal ini menunjukkan bahwa KNN pada pengujian ini relatif konsisten ketika mengenali *tweet* yang termasuk sentimen negatif terhadap Program Makan Bergizi Gratis (MBG).

Evaluasi Model

Setelah pola benar-salah prediksi dianalisis melalui *Confusion Matrix*, evaluasi dilanjutkan dengan *Classification Report* untuk melihat ukuran kinerja pada masing-masing kelas. Ringkasan hasil evaluasinya ditampilkan pada gambar berikut.

```
# Hasil Valuasi
acc_knn = round(knn.score(X_train, y_train) * 100, 2)
print('KNN:'+str(acc_knn))
print(classification_report(y_test, y_pred))
```

```
KNN:89.33
              precision    recall  f1-score   support

     Negatif       0.90      0.97      0.93       856
     Positif       0.61      0.30      0.40       138

    accuracy                           0.88       994
   macro avg       0.75      0.63      0.67       994
weighted avg       0.86      0.88      0.86       994
```

Gambar 9 Hasil Evaluasi Model KNN

Hasil pada Gambar 9 menunjukkan akurasi *K-Nearest Neighbors* sebesar 88%. Artinya, sekitar delapan puluh delapan persen dari seluruh *testing data* berhasil diklasifikasikan sesuai labelnya.

5. KESIMPULAN DAN SARAN

KESIMPULAN

Berdasarkan rangkaian pengolahan dan pengujian yang telah dilakukan pada analisis sentimen Program Makan Bergizi Gratis (MBG) di *Twitter* dengan *K-Nearest Neighbors*, kesimpulan penelitian dirangkum sebagai berikut.

1. Algoritma *K-Nearest Neighbors* (KNN) telah diterapkan untuk membedakan sentimen positif dan negatif terkait Program Makan Bergizi Gratis (MBG) pada data *Twitter*. Sebelum klasifikasi, teks melalui *cleaning*, *case folding*, *tokenizing*, *stopword removal*, dan *stemming*. Hasil *preprocessing* kemudian diubah menjadi fitur numerik dengan *Term Frequency-Inverse Document Frequency* (TF-IDF) agar dapat menjadi masukan bagi KNN.
2. Pengujian pada 994 *testing data* menghasilkan akurasi 88%. Pada kelas negatif, *Classification Report* mencatat *precision* 0,90 dan *recall* 0,97. Kelas positif memiliki *precision* 0,61 serta *recall* 0,30. Nilai tersebut menunjukkan adanya

perbedaan kemampuan model antar kelas, dengan hasil yang lebih kuat pada pengenalan sentimen negatif.

3. *Confusion Matrix* mencatat 830 data negatif dan 41 data positif yang diklasifikasikan secara tepat. Sebaliknya, 26 data negatif diprediksi sebagai positif dan 97 data positif diprediksi sebagai negatif. Ketimpangan jumlah sampel antar kelas berkaitan dengan perbedaan kinerja tersebut karena kelas negatif lebih dominan di dalam *dataset*.
4. Penerapan *K-Nearest Neighbors* yang dipadukan dengan fitur TF-IDF dan pengukuran *Cosine Similarity* menghasilkan akurasi 88% pada klasifikasi sentimen MBG di *Twitter*. Berdasarkan hasil tersebut, proses penelitian telah menghasilkan model yang dapat mengelompokkan opini masyarakat ke dalam dua kelas sentimen sesuai tujuan yang ditetapkan.

SARAN

Berdasarkan temuan dan keterbatasan yang terlihat selama penelitian, beberapa pengembangan berikut dapat dipertimbangkan pada penelitian selanjutnya.

1. Penelitian berikutnya dapat memperluas jumlah data serta mengambil *tweet* dari rentang waktu yang lebih beragam. Dengan cakupan tersebut, pola sentimen yang diperoleh diharapkan lebih mampu mewakili variasi opini masyarakat.
2. Ketidakseimbangan kelas dapat ditangani pada penelitian lanjutan melalui teknik *data balancing*, misalnya SMOTE, *oversampling*, atau *undersampling*. Pendekatan tersebut dapat diuji untuk meningkatkan kemampuan model dalam mempelajari kelas yang jumlah datanya lebih sedikit.

DAFTAR REFERENSI

Alif Bais Khoiriyah. (2025, November 21). Program MBG Nasional Tingkatkan Peran UMKM, Dorong Pertumbuhan Ekonomi Nasional.

Ariyanto, F., & Saefurrohman. (2025). Komparasi Tingkat Akurasi Sentimen Algoritma K-Nearest Neighbor Dan Naïve Bayes Pemilihan Gubernur Jawa Tengah 2024 di Sosial Media X. *JITSI : Jurnal Ilmiah Teknologi Sistem Informasi*, *6*(2), 112–119. https://doi.org/10.62527/jitsi.6.2.332

Eddy Cahyono Sugiarto. (2024, December 18). Makan Bergizi Gratis dan SDM Unggul.

Fatkhurrohman, F., Nugroho, B. I., & Fadillah, N. (2025). Analisis Sentimen Program Makan Bergizi Gratis Pemerintah RI Melalui Twitter Menggunakan Metode SVM. *RIGGS: Journal of Artificial Intelligence and Digital Business*, *4*(3), 3906–3917. https://doi.org/10.31004/riggs.v4i3.2533

Halder, R. K., Uddin, M. N., Uddin, Md. A., Aryal, S., & Khraisat, A. (2024). Enhancing K-nearest neighbor algorithm: a comprehensive review and performance analysis of modifications. *Journal of Big Data*, *11*(1), 113. https://doi.org/10.1186/s40537-024-00973-y

Hovi, H. S. W., Id Hadiana, A., & Rakhmat Umbara, F. (2022). Prediksi Penyakit Diabetes Menggunakan Algoritma Support Vector Machine (SVM). *Informatics and Digital Expert (INDEX)*, *4*(1), 40–45. https://doi.org/10.36423/index.v4i1.895

Ilona Estherina. (2025, August 15). Penerima Makan Bergizi Gratis Ditargetkan 82,9 Juta hingga Akhir 2025.

Isnain, A. R., Supriyanto, J., & Kharisma, M. P. (2021). Implementation of K-Nearest Neighbor (K-NN) Algorithm For Public Sentiment Analysis of Online Learning. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, *15*(2), 121. https://doi.org/10.22146/ijccs.65176

Jim, J. R., Talukder, M. A. R., Malakar, P., Kabir, M. M., Nur, K., & Mridha, M. F. (2024). Recent advancements and challenges of NLP-based sentiment analysis: A state-of-the-art review. *Natural Language Processing Journal*, *6*, 100059. https://doi.org/10.1016/j.nlp.2024.100059

Kementerian Kesehatan Republik Indonesia. (2025, May 26). SSGI 2024: Prevalensi Stunting Nasional Turun Menjadi 19,8%.

Mao, Y., Liu, Q., & Zhang, Y. (2024). Sentiment analysis methods, applications, and challenges: A systematic literature review. *Journal of King Saud University - Computer and Information Sciences*, *36*(4), 102048. https://doi.org/10.1016/j.jksuci.2024.102048

Martony, O. (2023). Stunting di Indonesia: Tantangan dan Solusi di Era Modern. *Journal of Telenursing (JOTING)*, *5*(2), 1734–1745. https://doi.org/10.31539/joting.v5i2.6930

Putri Ajie, Z., & Wardhana, A. (2025). Analisis Sentimen Negatif Publik pada Tagar #KaburAjaDulu di Media Sosial "X." *Jurnal Audiens*, *6*(3), 430–442. https://doi.org/10.18196/jas.v6i3.608

Putriyekti, A., Sulistiawati, D., Khairani, N., Rizky, R., Kusuma, A., Keuangan, P., & Stan, N. (2025). ANALISIS MULTIDIMENSIONAL SENTIMEN

MASYARAKAT TERHADAP PROGRAM MAKAN BERGIZI GRATIS PADA MEDIA SOSIAL X. *Integrative Perspectives of Social and Science Journal*, *2*(1), 1004.

Putu Alya Noviyanti, N., Gusti Ngurah Anom Cahyadi Putra, I., Studi Informatika, P., Matematika dan Ilmu Pengetahuan Alam, F., Raya Kampus UNUD, J., Jimbaran, B., & Selatan, K. (2025). Penerapan Multinomial Naïve Bayes dan Chi-Square pada Analisis Sentimen Makan Bergizi Gratis. *JNATIA*, *3*(4).

Rendy, M., Hamdi, M., Muhammadiyah Jember, U., & Pemerintahan, J. I. (2025). PERAN PEMERINTAH DALAM MENGATASI STUNTING: UPAYA BERSAMA UNTUK PERTUMBUHAN DAN PEMBANGUNAN BERKELANJUTAN THE GOVERNMENT'S ROLE IN TACKLING STUNTING: JOINT EFFORTS FOR SUSTAINABLE GROWTH AND DEVELOPMENT. In *Jurnal Interelasi* (Vol. 1).

Siregar, M. E., Dermawan, S., & Hisyam, A. A. (2025). PERBANDINGAN KINERJA NAIVE BAYES DAN KNN DALAM ANALISIS SENTIMEN KOMENTAR X DENGAN DAN TANPA TEXT PREPROCESSING (STUDI KASUS: DANANTARA). *Jurnal Informatika Dan Teknik Elektro Terapan*, *13*(3). https://doi.org/10.23960/jitet.v13i3.6612

Sirichanya, C., & Kraisak, K. (2021). Semantic data mining in the information age: A systematic review. *International Journal of Intelligent Systems*, *36*(8), 3880–3916. https://doi.org/10.1002/int.22443

Suprapto, F. A., Praditya, E., Dewi, R. M., & Adiyoso, W. (2025). A Policy Implementation Review of the Free Nutritious Meal (MBG) Program. *The Journal of Indonesia Sustainable Development Planning*, *6*(2), 297–312. https://doi.org/10.46456/jisdep.v6i2.798