



Analisis Sentimen Publik Terhadap Program Makan Bergizi Gratis Menggunakan Metode *Random Forest* Pada Platform X

Ghefira Ainur¹, Noor Latifah², Fajar Nugraha³

^{1,2,3} Program Studi Sistem Informasi, Fakultas Teknik, Universitas Muria Kudus

Article Info:

Dikirim: 3 Juni 2025

Direvisi: 5 Juni 2025

Diterima: 20 Juni 2025

Tersedia Online: 30 Juni 2025

Penulis Korespondensi:

Ghefira Ainur

Program Studi Sistem Informasi,
Fakultas Teknik,

Universitas Muria Kudus

Email: 202153048@std.umk.ac.id

Abstrak: Perkembangan media sosial telah membuka peluang baru dalam memahami opini publik secara real-time terhadap kebijakan pemerintah. Penelitian ini bertujuan untuk menganalisis sentimen masyarakat terkait Program Makan Bergizi Gratis yang diluncurkan oleh pemerintah Indonesia dengan memanfaatkan data dari platform X (Twitter). Metode yang digunakan adalah algoritma *Random Forest* untuk klasifikasi sentimen (positif, negatif, dan netral) dengan pendekatan representasi teks menggunakan TF-IDF. Proses penelitian mencakup pengumpulan data tweet melalui crawling, preprocessing teks (cleaning, case folding, tokenizing, stopword removal, dan stemming), pelabelan data secara manual, serta pelatihan dan pengujian model. Dataset terdiri dari 813 tweet, dibagi dengan rasio 80% data latih dan 20% data uji. Hasil menunjukkan bahwa model mampu mengklasifikasikan sentimen positif dengan akurasi tinggi (precision 82% dan recall 78%), namun kurang akurat pada kelas negatif dan netral. Akurasi keseluruhan model pada data uji adalah 82% dengan F1-score sebesar 90%. Aplikasi klasifikasi juga telah dideploy menggunakan Streamlit untuk memungkinkan klasifikasi komentar secara manual maupun batch melalui file CSV. Kesimpulan dari penelitian ini menunjukkan bahwa algoritma *Random Forest* cukup efektif untuk klasifikasi sentimen berbasis teks, meskipun perbaikan masih diperlukan pada aspek ketidakseimbangan kelas dan pemahaman konteks kalimat.

Kata kunci: Analisis Sentimen; Media Sosial; *Random Forest*; Program Makan Bergizi; TF-IDF

Abstract: The rapid growth of social media has opened new opportunities for understanding public opinion on government policies in real time. This study aims to analyze public sentiment regarding the Free Nutritious Meal Program launched by the Indonesian government, utilizing data from platform X (Twitter). The *Random Forest* algorithm was employed for sentiment classification (positive, negative, and neutral), with text representation using the TF-IDF approach. The research process involved collecting tweets through crawling, performing text preprocessing (cleaning, case folding, tokenizing, stopword removal, and stemming), manually labeling the data, and training and testing the model. The dataset consisted of 813 tweets, split into 80% training and 20% testing data. The results indicate that the model achieved high accuracy in classifying positive sentiments (precision of 82% and recall of 78%) but showed lower accuracy for negative and neutral classes. The overall model accuracy on the test set was 82%, with an F1-score of 90%. The classification application was also deployed using Streamlit, enabling both manual and batch classification via CSV file input. In conclusion, the *Random Forest* algorithm proved effective for text-based sentiment classification, although improvements are still required in addressing class imbalance and enhancing contextual understanding.

Keywords: Sentiment Analysis; Social Media; *Random Forest*; Free Nutritious Meal Program; TF-IDF

1. PENDAHULUAN

Perkembangan teknologi informasi dan komunikasi dalam satu dekade terakhir telah memengaruhi cara masyarakat berinteraksi dan menyampaikan opini, terutama melalui media sosial. *Platform X* (sebelumnya dikenal sebagai *Twitter*) menjadi salah satu media publik digital yang paling banyak digunakan masyarakat Indonesia untuk mengekspresikan pandangan mereka secara real time terhadap berbagai isu sosial, politik, ekonomi, maupun kebijakan pemerintah [1]. Salah satu kebijakan yang menjadi sorotan publik belakangan ini adalah program makan bergizi gratis, sebuah inisiatif pemerintah yang bertujuan meningkatkan kualitas gizi anak-anak sekolah dasar dan mengurangi beban ekonomi rumah tangga. Meskipun memiliki tujuan yang positif, implementasi program ini menuai berbagai tanggapan dari masyarakat yang beragam, mulai dari dukungan atas manfaat jangka panjangnya hingga kritik terhadap efektivitas penggunaan anggaran [2].

Fenomena ini menunjukkan pentingnya pengumpulan dan analisis opini publik secara sistematis untuk mengetahui persepsi masyarakat terhadap kebijakan yang sedang dijalankan. Salah satu pendekatan yang efektif untuk memahami kecenderungan opini dalam skala besar adalah analisis sentimen. Analisis sentimen memungkinkan klasifikasi otomatis opini publik ke dalam kategori seperti sentimen positif, negatif, atau netral, berdasarkan data teks yang diambil dari media sosial. Informasi yang dihasilkan dari analisis ini dapat memberikan gambaran umum bagi pengambil kebijakan dalam mengevaluasi aspek-aspek tertentu dari program yang dijalankan. Dengan kata lain, media sosial tidak hanya berfungsi sebagai saluran komunikasi publik, tetapi juga sebagai sumber data strategis dalam mengkaji respons masyarakat terhadap kebijakan negara.

Berbagai studi telah menunjukkan bahwa metode dalam data mining, khususnya yang menggabungkan teknik pemrosesan bahasa alami (*Natural Language Processing/NLP*) dan algoritma pembelajaran mesin, memberikan hasil yang cukup akurat dalam klasifikasi sentimen. Salah satu algoritma yang banyak digunakan adalah *Random Forest*, yang terbukti efektif dalam menangani data teks berskala besar dengan akurasi tinggi. *Random Forest* bekerja dengan membentuk sejumlah pohon keputusan yang digabungkan (*ensemble learning*) untuk menghasilkan prediksi yang lebih stabil dan tahan terhadap *overfitting* [3]. Penggunaan algoritma ini juga relatif fleksibel terhadap variasi data teks dan mampu menangani fitur yang banyak tanpa mengorbankan performa model.

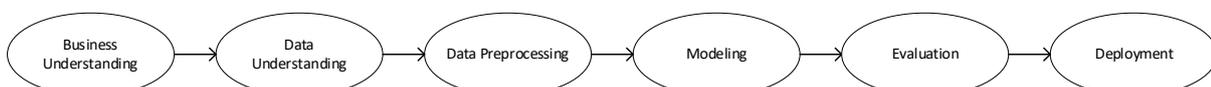
Penelitian ini mengadopsi pendekatan *Random Forest* dalam mengklasifikasikan opini masyarakat terkait program makan bergizi gratis yang dihimpun dari *platform X*. Pemilihan algoritma ini dilatarbelakangi oleh kemampuannya dalam menyeimbangkan kompleksitas pemodelan dengan hasil akurasi yang tinggi, serta efektivitasnya dalam menangani dataset yang tidak seimbang seperti dalam kasus opini publik yang cenderung bias pada kategori tertentu. Selain itu, integrasi teknik TF-IDF (*Term Frequency–Inverse Document Frequency*) dalam proses representasi teks memungkinkan penguatan bobot kata-kata penting yang secara signifikan berkontribusi terhadap sentimen dalam opini [4].

Nilai kebaruan dari penelitian ini terletak pada kombinasi analisis sentimen berbasis media sosial terhadap program kebijakan pemerintah terkini dengan penerapan metode *Random Forest*, serta penggunaan data yang dikumpulkan secara kontekstual dalam kurun waktu tertentu dengan kata kunci relevan. Pendekatan ini diharapkan dapat memperkaya literatur tentang bagaimana algoritma klasifikasi dapat diimplementasikan dalam studi opini publik berbasis media sosial lokal [5]. Selain itu, penelitian ini dapat menjadi model dasar bagi studi-studi berikutnya yang ingin menelusuri keterkaitan antara respons masyarakat dan efektivitas pelaksanaan kebijakan sosial berbasis data.

Dengan demikian, penelitian ini bertujuan untuk menerapkan metode *Random Forest* dalam melakukan klasifikasi sentimen publik terhadap program makan bergizi gratis berdasarkan data dari *platform X*. Penelitian ini juga bertujuan untuk mengevaluasi tingkat akurasi metode *Random Forest* dalam mengelompokkan opini publik ke dalam kategori sentimen positif, negatif, dan netral. Hasil dari penelitian ini diharapkan tidak hanya memperkaya ranah akademik, namun juga dapat memberikan kontribusi praktis bagi pemerintah dalam memahami persepsi masyarakat, sekaligus menjadi dasar pertimbangan dalam perumusan atau perbaikan kebijakan yang lebih responsif dan efektif.

2. METODOLOGI PENELITIAN

Penelitian ini menggunakan pendekatan *machine learning* berbasis kerangka CRISP-DM (*Cross-Industry Standard Process for Data Mining*) dalam melakukan analisis sentimen terhadap opini publik mengenai Program Makan Bergizi Gratis di media sosial X (*Twitter*). CRISP-DM sendiri merupakan standar proses dalam data mining yang dirancang untuk memastikan setiap tahapan pemrosesan data dilakukan secara sistematis, terstruktur, dan efisien, dengan mengarahkan data melalui serangkaian fase yang telah ditentukan dengan jelas [3]. Algoritma yang digunakan dalam klasifikasi sentimen ini adalah *Random Forest*, karena kemampuannya yang baik dalam menangani data teks yang bervariasi serta menghasilkan akurasi yang stabil [6].



Gambar 1. Tahapan Metode *Machine Learning* Berbasis CRISP-DM

Gambar 1 ini menggambarkan alur proses pengolahan data teks (*text mining*) secara sistematis dalam kerangka CRISP-DM. Proses dimulai dari identifikasi masalah dan studi pustaka untuk memahami karakteristik data media sosial dan metode pengumpulan data (*crawling & scraping*). Tahapan ini dilanjutkan dengan proses *preprocessing data*, yang mencakup enam langkah teknis: *cleaning* (menghapus karakter tidak perlu), *case folding* (mengubah huruf menjadi format seragam), *stopword removal* (menghapus kata umum yang tidak bermakna analitik), *tokenizing* (pemecahan kalimat menjadi kata), dan *stemming* (mengubah kata ke bentuk dasar). Data yang telah dibersihkan kemudian direpresentasikan secara numerik menggunakan metode TF-IDF (*Term Frequency–Inverse Document Frequency*) sebelum dimasukkan ke dalam model klasifikasi. Akhir dari proses ini mencakup analisis hasil klasifikasi, pengukuran performa model, serta pelaporan dan dokumentasi proyek. Keseluruhan proses ini mencerminkan pendekatan komprehensif dan terukur dalam mentransformasi data mentah menjadi informasi yang siap digunakan untuk tujuan analisis sentimen atau klasifikasi berbasis data teks.

2.1 Business Understanding

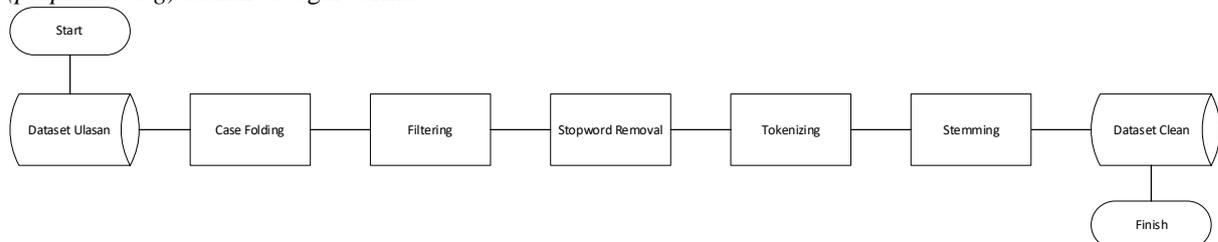
Pada tahap ini dilakukan identifikasi permasalahan penelitian, di mana banyaknya opini publik yang berkembang mengenai Program Makan Bergizi Gratis yang dicanangkan oleh pemerintah menjadi perhatian utama. Opini yang beragam tersebut memerlukan pengolahan dan analisis secara sistematis. Oleh karena itu, dibutuhkan sebuah sistem klasifikasi sentimen yang mampu mengelompokkan opini masyarakat ke dalam kategori positif, negatif, dan netral. Hasil klasifikasi ini diharapkan dapat membantu pemerintah dalam melakukan evaluasi program berdasarkan persepsi publik secara lebih objektif dan terukur.

2.2 Data Understanding

Pada tahap ini data dikumpulkan dari platform media sosial X (*Twitter*) menggunakan teknik *web scraping* berbasis *Python* dengan *library* seperti *snsrape* atau *tweepy*. Data berupa *tweet-tweet* yang berisi opini publik dikumpulkan berdasarkan kata kunci seperti "makan bergizi gratis", "program makan gratis", dan istilah relevan lainnya. Selain itu, dilakukan studi dokumentasi terhadap jurnal-jurnal ilmiah sebelumnya untuk memperdalam pemahaman terkait penerapan analisis sentimen dan algoritma *Random Forest*.

2.3 Data Preprocessing

Tahapan ini bertujuan untuk membersihkan dan menyiapkan data teks agar siap untuk dianalisis [7]. Teks *tweet* akan diubah menjadi bentuk yang lebih terstruktur melalui beberapa tahapan, yaitu *case folding*, *filtering*, *stopword removal*, *stemming*, dan *tokenizing*. Gambar 2 merupakan tahapan pada *data preparation (preprocessing)* adalah sebagai berikut:



Gambar 2. Tahapan *Data Preparation (Preprocessing)*

- Case Folding*, merupakan suatu teknik normalisasi teks yang bertujuan untuk menyeragamkan format penulisan dengan mengubah seluruh huruf dalam dokumen menjadi huruf kecil [8]. Proses *case folding* ditujukan pada gambar 3 berikut ini:

```
[ ] # ----- Case Folding -----
# Using Series.str.lower() function on Pandas
tweet['Text Case Folding'] = tweet['full_text'].str.lower()

print('Case Folding Result : \n')
print(tweet['Text Case Folding'].head(5))
print('\n\n')
```

Gambar 3. Proses *Case Folding*

Seluruh teks dalam *tweet* akan diubah menjadi huruf kecil guna menyeragamkan format dan menghindari perbedaan interpretasi antara huruf kapital dan huruf kecil. Karakter *non-alfabet* seperti simbol, angka, atau tanda baca juga akan dihapus agar data yang digunakan hanya berisi huruf a–z.

b. *Filtering* atau penyaringan adalah proses untuk menghapus kata-kata yang tidak relevan atau jarang muncul dalam konteks analisis, khususnya dalam pencarian informasi [7]. Dalam pembelajaran mesin, *filtering* digunakan untuk menyeleksi kata-kata yang dianggap penting setelah proses tokenisasi. Proses *filtering* ditunjukkan pada gambar 4, dibawah ini:

```
# ----- add stopword from txt file -----  
# read txt stopword using pandas  
txt_stopword = pd.read_csv("stopwordbahasa.txt", names= ["stopwords"], header = None)  
  
# convert stopword string to list & append additional stopword  
list_stopwords.extend(txt_stopword["stopwords"][0].split(' '))  
  
# convert list to dictionary  
list_stopwords = set(list_stopwords)  
#remove stopword pada list token  
def stopwords_removal(words):  
    return [word for word in words if word not in list_stopwords]  
tweet['Text Filtering'] = tweet['Text Stemming'].apply(stopwords_removal)  
print(tweet['Text Filtering'].head())
```

Gambar 4. Proses *Filtering*

Proses ini sering melibatkan penggunaan algoritma seperti *stoplist*, yang bertujuan menghilangkan kata-kata tidak penting, atau *wordlist*, yang mempertahankan kata-kata yang relevan. Kata-kata yang tidak memiliki makna penting dalam analisis seperti kata hubung atau preposisi akan disaring. Hanya kata-kata yang relevan dengan konteks program makan bergizi gratis yang dipertahankan untuk memperkuat kualitas analisis.

c. *Stopword Removal*, yang juga dikenal sebagai *filtering*, adalah proses untuk menyaring kata-kata penting dari hasil tokenisasi [9]. Proses *stopword removal* ditunjukkan pada gambar 5, berikut ini:

```
from nltk.corpus import stopwords  
# ----- get stopword from NLTK stopword -----  
# get stopword indonesia  
list_stopwords = stopwords.words('indonesian')  
# ----- manually add stopword -----  
# append additional stopword  
list_stopwords.extend(['yg', 'dg', 'rt', 'dgn', 'ny', 'klo',  
    'kalo', 'amp', 'bian', 'bikin', 'bilang',  
    'gak', 'ga', 'krn', 'nya', 'nih', 'sih',  
    'si', 'tau', 'tdk', 'tuh', 'utk', 'ya',  
    'jd', 'jgn', 'sdh', 'aja',  
    'nyg', 'hehe', 'pen', 'u', 'nan', 'loh', 'rt',  
    '&amp;', 'yah', 'sdgkan', 'sdg', 'emg', 'sm', 'pls', 'mlu', 'ken',  
    'allah', 'brb', 'btw', 'b/c', 'cod', 'cmiiw', 'fyi',  
    'gg', 'ggwp', 'idk', 'ikr', 'lol', 'ootd', 'lmao', 'oot',  
    'pap', 'otw', 'tfl', 'vc', 'ygy'])
```

Gambar 5. Proses *Stopword Removal*

Proses ini menghapus kata-kata umum yang sering muncul namun tidak berkontribusi signifikan terhadap pemahaman sentimen, seperti "dan", "di", atau "ke". Hal ini berguna untuk mengurangi gangguan dalam pemodelan.

d. *Tokenizing* merupakan proses memecah teks menjadi unit-unit kecil yang disebut token, seperti kata atau frasa, agar analisis teks menjadi lebih efisien [10]. Proses *tokenizing* ditunjukkan pada gambar 6, berikut ini:

```
import string  
import re #regex library  
  
# import word.tokenize  
from nltk.tokenize import word_tokenize  
from nltk.tokenize import wordpunct_tokenize  
  
# ----- Tokenizing -----  
  
def remove_tweet_special(text):  
    if not isinstance(text, str):  
        return "" # atau bisa 'str(text)' kalau ingin paksa jadi string  
    # remove tab, new line, and backslash  
    text = text.replace('\t', " ").replace('\n', " ").replace('\u', " ").replace('\ ', "")  
    # remove non ASCII (emoticon, chinese word, .etc)  
    text = text.encode('ascii', 'replace').decode('ascii')  
    # remove mention, link, hashtag  
    text = ' '.join(re.sub(r'([#][A-Za-z0-9_]+)|(\w+:\w+/\w+)', " ", text).split())  
    # remove incomplete URL  
    return text.replace("http://", " ").replace("https://", " ")
```

Gambar 6. Proses *Tokenizing*

Proses ini biasanya diawali dengan normalisasi, yaitu mengubah seluruh huruf menjadi huruf kecil serta menghapus tanda baca dan simbol seperti “,”, “\$”, atau “*”. Setelah itu, teks dipisahkan berdasarkan spasi sehingga menghasilkan kumpulan kata, misalnya kalimat “program ini sangat membantu” akan diubah menjadi ["program", "ini", "sangat", "membantu"].

e. *Stemming* adalah proses mengubah kata berimbuhan menjadi bentuk dasarnya dengan menghapus awalan, akhiran, atau sisipan [5]. Proses *stemming* ditunjukkan pada gambar 7, berikut ini:

```
# import Sastrawi package
from Sastrawi.Stemmer.StemmerFactory import StemmerFactory
import swifter
# create stemmer
factory = StemmerFactory()
stemmer = factory.create_stemmer()
# stemmed
def stemmed_wrapper(term):
    return stemmer.stem(term)

term_dict = {}
for document in tweet['Text Normalization']:
    for term in document:
        if term not in term_dict:
            term_dict[term] = ''
print(len(term_dict))
print("-----")
for term in term_dict:
    term_dict[term] = stemmed_wrapper(term)
    print(term,":", term_dict[term])

print(term_dict)
print("-----")
# apply stemmed term to dataframe
def get_stemmed_term(document):
    return [term_dict[term] for term in document]
tweet['Text Stemming'] = tweet['Text Normalization'].swifter.apply(get_stemmed_term)
print(tweet['Text Stemming'])
```

Gambar 7. Proses Stemming

Tujuannya adalah mengurangi variasi kata untuk meningkatkan efisiensi pencarian dan analisis. Misalnya, "diberikan", "pemberian", atau "memberi" akan dikembalikan ke kata dasar "beri", guna menyederhanakan representasi kata.

2.4 Pelabelan

Proses pelabelan sentimen dalam penelitian ini dilakukan menggunakan pendekatan berbasis leksikon, di mana setiap kata dalam teks dianalisis berdasarkan daftar kata positif dan negatif yang telah disusun sebelumnya dalam bentuk file *Excel* (*kamus_positive.xlsx* dan *kamus_negative.xlsx*). Kedua file ini dimuat ke dalam *Python* dan dikonversi menjadi *dictionary* (*lexicon_positive_dict* dan *lexicon_negative_dict*) yang berisi pasangan kata dan skor sentimennya. Fungsi *sentiment_analysis_lexicon_indonesia()* kemudian digunakan untuk menghitung total skor sentimen dari setiap teks. Skor ini diperoleh dengan menambahkan nilai dari setiap kata yang ditemukan dalam kamus positif dan mengurangi nilai dari kata-kata yang ditemukan dalam kamus negatif.

```
lexicon_positive = pd.read_excel('kamus_positive.xlsx')
lexicon_positive_dict = {}
for index, row in lexicon_positive.iterrows():
    if row[0] not in lexicon_positive_dict:
        lexicon_positive_dict[row[0]] = row[1]

lexicon_negative = pd.read_excel('kamus_negative.xlsx')
lexicon_negative_dict = {}
for index, row in lexicon_negative.iterrows():
    if row[0] not in lexicon_negative_dict:
        lexicon_negative_dict[row[0]] = row[1]

def sentiment_analysis_lexicon_indonesia(text):
    score = 0
    for word in text:
        if word in lexicon_positive_dict:
            score = score + lexicon_positive_dict[word]
        for word in text:
            if word in lexicon_negative_dict:
                score = score + lexicon_negative_dict[word]
    sentimen=''
    if (score > 0):
        sentimen = 'Positive'
    elif (score < 0):
        sentimen = 'Negative'
    else:
        sentimen = 'Neutral'
    return score, sentimen

results = df['Text Filtering'].apply(sentiment_analysis_lexicon_indonesia)
results = list(zip(results))
df['Polarity Score'] = results[0]
df['Indonesia Sentiment'] = results[1]
#data['sentimen'] = results[1]
#data

df[['Text Filtering', 'Polarity Score', 'Indonesia Sentiment']]
```

Gambar 8. Proses Labeling

Setelah semua kata dalam sebuah teks dianalisis, skor total dievaluasi: jika skor lebih besar dari nol maka teks diberi label positif, jika kurang dari nol diberi label negatif, dan jika sama dengan nol maka dilabeli sebagai netral. Proses ini diterapkan pada kolom teks hasil *preprocessing* (*Text Filtering*) menggunakan fungsi *apply()*, dan hasilnya ditambahkan sebagai kolom baru berupa skor polaritas (*Polarity Score*) dan label sentimen (*Indonesia Sentiment*). Pendekatan ini bersifat efisien dan mudah diterapkan karena tidak memerlukan proses pelabelan manual, namun memiliki keterbatasan dalam menangkap konteks kalimat atau makna implisit seperti ironi atau sarkasme.

2.5 TF-IDF

Salah satu metode populer dalam pemrosesan teks untuk mengukur relevansi sebuah kata dalam suatu dokumen adalah TF-IDF (*Term Frequency – Inverse Document Frequency*) [11]. Teknik ini digunakan untuk memberi bobot pada kata-kata berdasarkan frekuensi kemunculannya di dalam dokumen tertentu dan tingkat kelangkaannya dalam seluruh koleksi dokumen.

```
[40] from nltk.corpus import stopwords
      from sklearn.feature_extraction.text import TfidfVectorizer

      tfidf_vectorizer = TfidfVectorizer (max_features=2500, min_df=7, max_df=0.8, stop_words=stopwords.words('indonesian'))
      response = tfidf_vectorizer.fit_transform(X_train)
```

Gambar 9. Proses TF-IDF

Komponen pertama, yaitu *Term Frequency* (TF), digunakan untuk mengukur seberapa sering sebuah kata muncul dalam satu dokumen [11]. Nilainya diperoleh dengan membandingkan jumlah kemunculan kata tersebut terhadap jumlah maksimum kemunculan kata apa pun dalam dokumen yang sama. Komponen kedua yaitu *Inverse Document Frequency* (IDF), berfungsi untuk menilai tingkat kekhususan suatu kata dalam seluruh dokumen [12]. Jika kata tersebut hanya muncul di sedikit dokumen, maka ia dianggap lebih penting dan bobotnya akan lebih tinggi.

3. HASIL DAN PEMBAHASAN

Bagian menyajikan temuan utama penelitian secara sistematis berdasarkan tahapan metodologi, mulai dari pengumpulan data, *preprocessing*, pelabelan sentimen, hingga evaluasi model. Hasil ditampilkan melalui tabel dan visualisasi seperti *word cloud* dan *confusion matrix* untuk memudahkan pemahaman pembaca terhadap distribusi data dan performa model. Setiap visualisasi dibahas secara komprehensif dengan mengaitkan hasil yang diperoleh pada konteks penelitian dan literatur pendukung. Pembahasan juga menyoroiti kelemahan model, seperti rendahnya akurasi pada kelas netral akibat ketidakseimbangan data dan keterbatasan pendekatan TF-IDF, serta memberikan masukan untuk pengembangan metode yang lebih kontekstual di masa depan [13].

3.1 Data

Data dalam penelitian ini diperoleh melalui proses *crawling* menggunakan *API Tweet Harvest* versi 2.6.1 yang dijalankan di lingkungan *Node.js*. Pengambilan data dilakukan dengan autentikasi token resmi untuk mengakses cuitan publik dari platform X (*Twitter*). Kata kunci pencarian difokuskan pada frasa "makan siang bergizi gratis" dalam rentang waktu 1 Januari hingga 30 Juni 2025, dengan filter bahasa Indonesia, dan dibatasi hingga 1.500 tweet terbaru. Data hasil *crawling* disimpan dalam format CSV untuk analisis selanjutnya.

```
# Import required python package
pip install pandas

# Install Node.js (because tweet-harvest built using Node.js)
sudo apt-get update
sudo apt-get install -y ca-certificates curl gnupg
sudo mkdir -p /etc/apt/keyrings
curl -fsSL https://deb.nodesource.com/gnkey/nodesource-repo.gpg.key | sudo gpg --dearmor -o /etc/apt/keyrings/nodesource.gpg
NODE_MAJOR=20 && echo "deb [signed-by=/etc/apt/keyrings/nodesource.gpg] https://deb.nodesource.com/node_$NODE_MAJOR.x nodistro main" | sudo tee /etc/apt/sources.list.d/nodesource.list
sudo apt-get update
sudo apt-get install nodejs -y
node -v

[ ] # Crawl Data
filename = 'tweet_ulasan_MSBG.csv'
search_keyword = 'makan siang bergizi gratis since:2025-01-01 until:2025-06-30 lang:id'
limit = 1500
Inpx -y tweet-harvest@2.6.1 -o "{filename}" -s "{search_keyword}" --tab "LATEST" -l {limit} --token {twitter_auth_token}

Your tweets saved to: /content/tweets-data/tweet_ulasan_MSBG.csv
Total tweets saved: 1224
-- Scrolling... (1) (2) (3)
```

Gambar 10. Proses Crawling Data Tweet

Tweet Harvest memungkinkan pengambilan data secara sistematis, mencakup informasi seperti teks *tweet*, waktu unggahan, jumlah suka dan *retweet*, serta media yang disisipkan. Proses ini ditampilkan pada Gambar diatas sebagai hasil dokumentasi *crawling*. Dataset ini menjadi dasar penting dalam tahapan analisis sentimen, karena memastikan data yang dikumpulkan bersifat aktual, relevan, dan siap diolah lebih lanjut melalui tahapan *preprocessing* dan klasifikasi.

3.2 Preprocessing

Pada tahap ini, dilakukan proses awal yang mencakup penghapusan elemen-elemen yang tidak relevan seperti URL, *mention* (@username), *hashtag* (#), angka, tanda baca, dan karakter khusus, serta menyeragamkan seluruh teks menjadi huruf kecil (*case folding*) untuk menghindari redundansi [14]. *Tweet* yang terdeteksi sebagai promosi atau iklan juga dihapus guna menjaga kualitas data. Selanjutnya, teks yang telah dibersihkan diproses melalui tokenisasi (pemecahan teks menjadi kata per kata), dilanjutkan dengan penghapusan *stopword*, yaitu kata-kata umum dalam bahasa Indonesia yang tidak memiliki nilai informasi penting. Terakhir, dilakukan proses *stemming* menggunakan *library Sastrawi* untuk mengubah kata berimbuhan menjadi bentuk dasar, sehingga menghasilkan representasi teks yang bersih, relevan, dan siap digunakan dalam tahap klasifikasi sentimen.

a) Case Folding

Pada tahap ini, semua karakter dalam teks diubah menjadi huruf kecil untuk menghindari perbedaan interpretasi antara huruf kapital dan huruf kecil [15].

Tabel 1. Hasil Case Folding

No.	Sebelum	Sesudah
1	MAKAN SIANG GRATIS BERGIZI HARUSNYA BUAT KITA SOALNYA KURANG GIZI	makan siang gratis bergizi harusnya buat kita soalnya kurang gizi

Pada Tabel 1, terlihat bahwa huruf kapital diubah menjadi huruf kecil untuk mempermudah proses analisis berikutnya serta menjaga konsistensi format teks.

b) Filtering

Filtering adalah proses penyaringan kata-kata yang dianggap tidak relevan atau kurang bermakna dalam analisis sentimen [7].

Tabel 2. Hasil Filtering

No.	Sebelum	Sesudah
1	makan siang gratis bergizi harusnya buat kita soalnya kurang gizi	makan, siang, gratis, gizi, gizi

Pada Tabel 2 ditunjukkan bagaimana *filtering* dilakukan dengan menyaring kata-kata yang tidak relevan untuk memperkuat kualitas analisis.

c) Stopword Removal

Stopword removal menghapus kata-kata umum seperti "dan", "di", "ke", "yang", dan lainnya yang sering muncul namun tidak memberikan kontribusi penting terhadap makna kalimat secara keseluruhan [16].

Tabel 3. Hasil Stopword Removal

No.	Sebelum	Sesudah
1	makan, siang, gratis, gizi, gizi	makan, siang, gratis, bergizi, harusnya, buat, kita, soalnya, kurang, gizi

Tabel 3 menunjukkan bagaimana kata-kata umum yang tidak memiliki bobot informasi dihilangkan untuk mengurangi gangguan dalam pemodelan.

d) Tokenizing

Tokenizing memecah kalimat menjadi kata-kata individual (token) berdasarkan spasi, sehingga teks dapat dianalisis secara unit per unit [5]. Proses ini juga membantu dalam pembuatan representasi fitur.

Tabel 4. Hasil Tokenizing

No.	Sebelum	Sesudah
1	makan, siang, gratis, bergizi, harusnya, buat, kita, soalnya, kurang, gizi	makan, siang, gratis, bergizi, harusnya, buat, kita, soalnya, kurang, gizi

Tabel 4 memperlihatkan hasil pemecahan kalimat menjadi token untuk mempermudah proses analisis lanjutan seperti klasifikasi sentimen.

e) Stemming

Stemming adalah proses mengubah kata berimbuhan menjadi bentuk dasar dengan menghapus awalan, akhiran, atau sisipan. Proses ini dilakukan menggunakan *library Sastrawi* untuk Bahasa Indonesia [17].

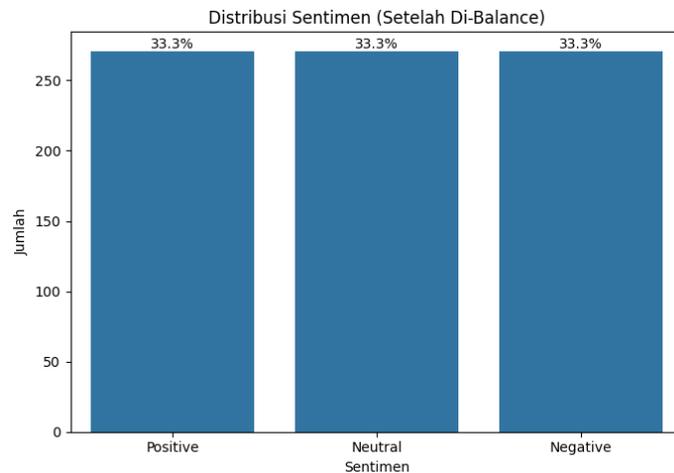
Tabel 5. Hasil Stemming

No.	Sebelum	Sesudah
1	makan, siang, gratis, bergizi, harusnya, buat, kita, soalnya, kurang, gizi	makan, siang, gratis, gizi, harus, buat, kita, soal, kurang, gizi

Pada Tabel 5 terlihat bahwa kata-kata yang memiliki bentuk imbuhan dikembalikan ke bentuk dasarnya untuk menyederhanakan struktur data teks.

3.3 Pelabelan

Proses pelabelan sentimen bertujuan untuk mengelompokkan setiap *tweet* ke dalam kategori positif, negatif, atau netral berdasarkan analisis kata-kata kunci serta konteks keseluruhan kalimat [18]. Metode ini menggabungkan pendekatan berbasis leksikon dengan pemahaman makna dalam teks untuk meningkatkan ketepatan klasifikasi. Hasil pelabelan ini berfungsi sebagai fondasi penting dalam pembangunan model klasifikasi sentimen, sehingga model yang dihasilkan dapat mengenali pola persepsi publik dengan lebih tepat [19]. *Tweet* yang dianalisis berasal dari perbincangan publik mengenai Program Makan Bergizi Gratis yang sedang digalakkan oleh pemerintah.

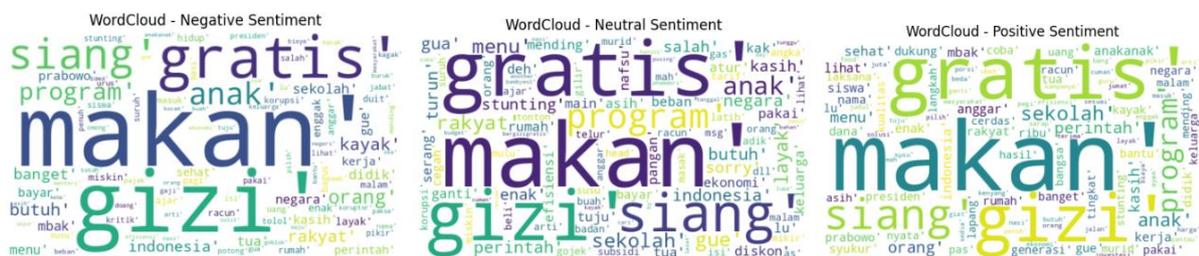


Gambar 11. Hasil Distribusi Sentimen

Dari hasil visualisasi distribusi sentimen, tercatat bahwa dari total 1.516 *tweet* awal, sekitar 75,8% atau 1.149 *tweet* mengandung sentimen positif, menunjukkan bahwa sebagian besar masyarakat menyambut baik program tersebut dan menilainya sebagai langkah yang bermanfaat. Sebaliknya, terdapat 302 *tweet* yang bersentimen negatif, mengindikasikan adanya kritik atau kekhawatiran, sementara 65 *tweet* bersifat netral, cenderung hanya menyampaikan informasi tanpa ekspresi emosional tertentu. Untuk menghindari bias model terhadap salah satu kelas sentimen, data kemudian diseimbangkan melalui proses balancing sehingga setiap kategori *Positive*, *Neutral*, dan *Negative*, memiliki jumlah yang sama, yaitu 271 dari total data. Dengan distribusi yang merata ini, model *machine learning* diharapkan dapat mengenali pola dari ketiga jenis sentimen secara adil dan akurat.

3.4 Visualisasi

Visualisasi *word cloud* pada Gambar 12 menunjukkan representasi linguistik dari persepsi masyarakat terhadap program makan bergizi gratis, berdasarkan tiga kategori sentimen: negatif, positif, dan netral. Dalam sentimen negatif, kata-kata seperti bohong, anggaran, dan korupsi muncul dominan, mencerminkan ketidakpercayaan publik terhadap transparansi dan efektivitas kebijakan tersebut. Hal ini menunjukkan kekhawatiran masyarakat terhadap potensi penyimpangan dana serta pelaksanaan yang tidak tepat sasaran [20]. Selain itu, kemunculan kata-kata seperti janji dan gagal memperkuat narasi bahwa sebagian publik menilai program ini lebih sebagai retorika politik daripada solusi nyata.

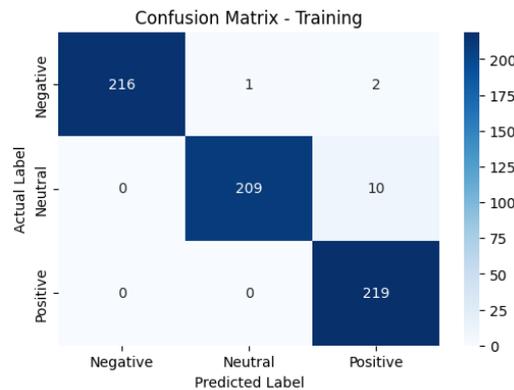


Gambar 12. Visualisai Word Cloud

Sebaliknya, dalam sentimen positif, kata-kata seperti membantu, bergizi, dan anak menggambarkan harapan terhadap manfaat langsung dari program, terutama dalam konteks perbaikan gizi anak dan dukungan terhadap keluarga kurang mampu. Pada sentimen netral, dominasi kata-kata seperti program, makan, dan sekolah menunjukkan kecenderungan narasi yang bersifat informatif tanpa muatan emosi. Namun, minimnya kata seperti adil atau merata di ketiga kategori tersebut menunjukkan bahwa isu pemerataan dan keadilan distribusi belum menjadi fokus utama dalam opini publik terhadap program ini. Temuan ini dapat menjadi pertimbangan penting bagi pemerintah dalam membangun kepercayaan masyarakat dan meningkatkan efektivitas komunikasi kebijakan

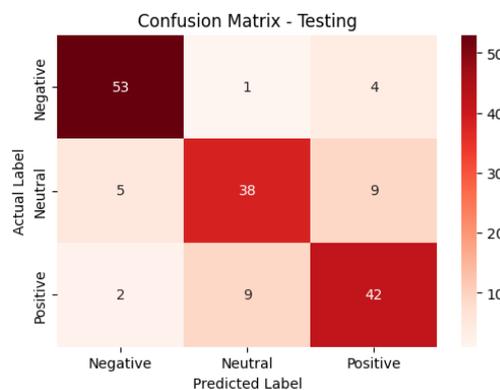
3.5 Modeling Random Forest Classification

Pada tahap klasifikasi menggunakan *Random Forest*, sejumlah pohon keputusan (*decision tree*) dibangun secara acak dari subset data latih dan fitur yang berbeda, lalu hasil prediksi setiap pohon digabungkan melalui mekanisme voting mayoritas untuk menentukan kelas sentimen akhir (Anjani dkk., 2023). Setiap pohon memilih titik pemisahan terbaik berdasarkan ukuran ketidakmurnian seperti *Gini Index*, sehingga *ensemble* ini mampu mengurangi risiko *overfitting* dan menghasilkan prediksi yang lebih stabil. Dataset ulasan pengguna Twitter (X) tentang Program Makan Bergizi Gratis yang telah menjalani *preprocessing* dan transformasi TF-IDF dibagi menjadi data latih 80% dan data uji 20%, di mana data latih digunakan untuk melatih model *Random Forest* dan data uji dipakai untuk mengevaluasi kinerjanya. Untuk hasil dari confusion matrix training data dapat dilihat pada Gambar 13 berikut.



Gambar 13. Confusion Matrix Training Data

Pada data pelatihan gambar 13, model menunjukkan performa yang sangat baik. Sebanyak 216 data pada kelas *Negative* diprediksi benar sebagai *Negative*, sementara hanya 1 data yang salah diprediksi sebagai *Neutral* dan 2 data salah diprediksi sebagai *Positive*. Untuk kelas *Neutral*, sebanyak 209 data diklasifikasikan dengan benar, tidak ada yang diprediksi sebagai *Negative*, namun terdapat 10 data yang salah diprediksi menjadi *Positive*. Sementara itu, pada kelas *Positive*, seluruh 219 data berhasil diprediksi dengan benar tanpa ada kesalahan klasifikasi ke kelas lain. Distribusi ini menunjukkan bahwa model memiliki performa yang sangat baik pada data training, dengan tingkat kesalahan yang sangat kecil, terutama pada perbedaan antara kelas *Neutral* dan *Positive*.



Gambar 14. Hasil Confusion Matrix Testing Data

Sementara itu, pada data pengujian gambar 14, model mengalami penurunan akurasi, khususnya pada kelas *Negative* dan *Neutral*. Dari total data uji, hanya 16 data *Negative* yang berhasil diklasifikasikan dengan benar, sedangkan 38 sisanya justru diklasifikasikan sebagai *Positive*. Kelas *Neutral* juga menunjukkan kelemahan, karena seluruh data *Neutral* gagal diklasifikasikan dengan benar (7 data malah masuk ke kelas *Positive*). Kelas *Positive* sendiri tetap menunjukkan performa yang baik dengan 234 data diklasifikasikan secara akurat. Temuan ini mengindikasikan bahwa model memiliki bias kuat terhadap kelas *Positive*, sehingga perlu dilakukan perbaikan seperti *balancing* dataset, tuning parameter, atau eksplorasi metode lain guna meningkatkan generalisasi model terhadap data uji.

Sementara itu, pada data pengujian pada Gambar 14, model mengalami sedikit penurunan performa dibandingkan data pelatihan. Pada kelas *Negative*, sebanyak 53 data berhasil diprediksi dengan benar, sementara 1 data salah diprediksi sebagai *Neutral* dan 4 data salah diprediksi sebagai *Positive*. Untuk kelas *Neutral*, terdapat 38 data yang diklasifikasikan dengan benar, namun masih ada 5 data yang salah diprediksi sebagai *Negative* dan 9 data salah diprediksi menjadi *Positive*. Pada kelas *Positive*, sebanyak 42 data berhasil diklasifikasikan dengan benar, tetapi terdapat 2 data yang salah diprediksi sebagai *Negative* dan 9 data salah

diprediksi sebagai *Neutral*. Distribusi ini menunjukkan bahwa meskipun model masih mampu mengklasifikasikan sebagian besar data dengan tepat, terdapat kesalahan yang lebih besar dibandingkan pada data pelatihan, terutama pada perbedaan antara kelas *Neutral* dan *Positive* yang terlihat cukup signifikan.

Dalam proses penelitian ini, data komentar dibagi menjadi dua bagian utama dengan rasio 80% sebagai data pelatihan (*training data*) dan 20% sebagai data pengujian (*testing data*). Tujuan dari pembagian ini adalah agar model pembelajaran mesin memiliki cukup data untuk mempelajari pola sentimen, namun tetap menyisakan sebagian data yang belum dilihat sebelumnya sebagai tolak ukur evaluasi kinerja model. Seluruh proses pelatihan dan pengujian dilakukan menggunakan lingkungan berbasis *cloud* yaitu *Google Colaboratory*, yang memungkinkan penggunaan sumber daya GPU/TPU secara gratis untuk mempercepat proses komputasi dan pelatihan model klasifikasi. Evaluasi performa model dilakukan dengan menghitung berbagai metrik utama seperti *accuracy*, *precision*, *recall*, dan *F1-score*, yang semuanya diperoleh melalui analisis *confusion matrix* dan *classification report*. Secara matematis, metrik evaluasi dapat dirumuskan sebagai berikut:

$$\begin{aligned}
 \text{a. Precision} &= \frac{TP}{TP + FP} \times 100\% = \frac{53}{53 + 7} \times 100\% = 88\% \\
 \text{b. Recall} &= \frac{TP}{TP + FN} \times 100\% = \frac{53}{53 + 5} \times 100\% = 91\% \\
 \text{c. Accuracy} &= \frac{TNeg + TNeu + TPos}{Total} \times 100\% = \frac{53 + 38 + 42}{163} \times 100\% = 82\% \\
 \text{d. F1-Score} &= \frac{2 \times Precision \times Recall}{Recall + Precision} \times 100\% = \frac{2 \times 0.8833 \times 0.9138}{0.8833 + 0.9138} \times 100\% = 89\%
 \end{aligned}$$

Berdasarkan hasil *classification report*, model *Random Forest* pada data pengujian menunjukkan bahwa model memiliki akurasi sebesar 82%, yang berarti sebanyak 82% dari seluruh komentar berhasil diklasifikasikan ke dalam label sentimen yang benar. Hasil ini diperoleh berdasarkan *confusion matrix* yang menunjukkan bahwa dari 163 data uji, model mampu mengklasifikasikan 53 komentar positif dengan benar, namun hanya 58 komentar negatif yang terklasifikasi dengan tepat, dan 52 komentar netral yang berhasil dikenali. Berikut merupakan hasil *classification report* yang ditampilkan pada Tabel 6.

Tabel 1. Hasil Classification Report

Kelas	Precision	Recall	F1-Score	Support
Negative	0.88	0.91	0.90	58
Neutral	0.79	0.73	0.76	52
Positive	0.76	0.79	0.78	53
Accuracy			0.82	163
Macro Avg	0.81	0.81	0.81	163
Weighted Avg	0.82	0.82	0.81	163

Secara rinci, nilai *precision* untuk kelas positif adalah 82%, yang berarti dari semua prediksi positif, 82% di antaranya benar. Nilai *recall* mencapai 79%, menunjukkan bahwa sebagian besar komentar positif berhasil dikenali. Namun, kelas negatif dan netral menunjukkan performa yang lemah, khususnya pada kelas netral yang memiliki *recall* 73%, artinya sebagian komentar netral di klasifikasi dengan benar. Metrik *F1-score* sebagai gabungan presisi dan *recall* menghasilkan nilai sebesar 82%, mencerminkan keseimbangan antara kemampuan model dalam mengenali dan memprediksi kelas secara benar. Hasil ini mengindikasikan bahwa meskipun model sangat baik dalam mendeteksi sentimen positif, namun diperlukan perbaikan lebih lanjut dalam mengenali sentimen netral dan negatif untuk meningkatkan generalisasi model.

3.6 Evaluasi

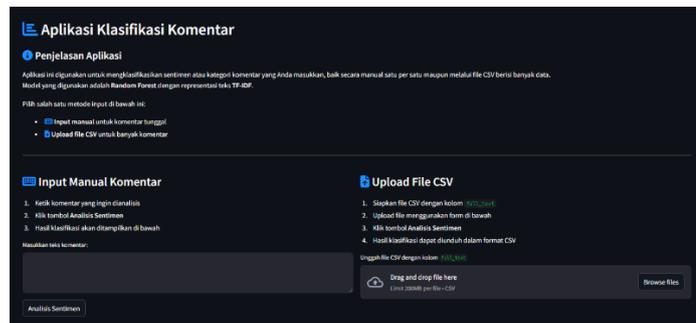
Performa model pada data pengujian menunjukkan kelemahan yang cukup signifikan, terutama dalam mengklasifikasikan komentar dengan sentimen netral dan negatif, yang tercermin dari nilai *recall* yang sangat rendah bahkan nol untuk kelas netral. Meskipun tingkat akurasi keseluruhan mencapai 82%, sebagian besar komentar yang bukan positif gagal dikenali dengan benar. Hal ini dapat dikaitkan dengan tiga penyebab utama: pertama, distribusi data yang tidak seimbang, di mana kelas positif jauh lebih dominan dibandingkan kelas lainnya, sehingga model lebih condong memprediksi kelas mayoritas; kedua, komentar netral umumnya memiliki bahasa yang cenderung kabur dan tidak memiliki kata kunci yang mencolok, sehingga menyulitkan model berbasis TF-IDF untuk mengidentifikasinya; dan ketiga, keterbatasan model *Random Forest* dalam memahami konteks dan urutan kata menyebabkan ketidakmampuannya menangkap makna komentar yang kompleks. Akibatnya, banyak opini netral dan negatif yang tidak terdeteksi, yang dapat berimplikasi pada analisis sentimen yang bias. Untuk memperbaiki performa klasifikasi ke depan, perlu dipertimbangkan

pendekatan seperti *oversampling* (misalnya *SMOTE*) serta penggunaan model berbasis konteks seperti *LSTM* atau *transformer*.

Performa model pada data pengujian menunjukkan kelemahan yang cukup signifikan, terutama dalam mengklasifikasikan komentar dengan sentimen netral dan negatif, yang tercermin dari nilai *recall* yang sangat rendah bahkan nol untuk kelas netral. Meskipun tingkat akurasi keseluruhan mencapai 82%, sebagian besar komentar yang bukan positif gagal dikenali dengan benar. Hal ini dapat dikaitkan dengan tiga penyebab utama: pertama, distribusi data yang tidak seimbang, di mana kelas positif jauh lebih dominan dibandingkan kelas lainnya, sehingga model lebih condong memprediksi kelas mayoritas kedua, komentar netral umumnya memiliki bahasa yang cenderung kabur dan tidak memiliki kata kunci yang mencolok, sehingga menyulitkan model berbasis TF-IDF untuk mengidentifikasinya; dan ketiga, keterbatasan model *Random Forest* dalam memahami konteks dan urutan kata menyebabkan ketidakmampuannya menangkap makna komentar yang kompleks. Akibatnya, banyak opini netral dan negatif yang tidak terdeteksi, yang dapat berimplikasi pada analisis sentimen yang bias. Untuk memperbaiki performa klasifikasi ke depan, perlu dipertimbangkan pendekatan seperti *oversampling* (misalnya *SMOTE*) serta penggunaan model berbasis konteks seperti *LSTM* atau *transformer*.

3.7 Deployment

Tahap *deployment* merupakan proses implementasi model *machine learning* ke dalam sebuah aplikasi interaktif yang dapat diakses oleh pengguna secara langsung. Dalam penelitian ini, proses *deployment* dilakukan menggunakan *framework Streamlit*, yaitu *platform Python* berbasis web yang dirancang untuk membangun antarmuka pengguna secara cepat dan interaktif tanpa perlu keahlian pengembangan *front-end*. Aplikasi dikembangkan dalam file *python*, yang memuat model klasifikasi sentimen berbasis *Random Forest* serta representasi teks TF-IDF yang telah dilatih sebelumnya.



Gambar 15. Deployment Sistem Analisa Sentimen

Aplikasi ini memungkinkan pengguna untuk melakukan klasifikasi komentar secara manual melalui form input teks, maupun secara massal melalui fitur unggah file CSV. Setelah proses klasifikasi, pengguna dapat melihat hasil prediksi sentimen masing-masing komentar, serta visualisasi distribusi sentimen dalam bentuk pie chart yang interaktif. Hasil klasifikasi juga dapat diunduh dalam format CSV untuk keperluan dokumentasi lebih lanjut. Seluruh proses ini dijalankan di *platform Jupyter notebook* atau server lokal, sehingga memudahkan pengujian dan penggunaan aplikasi oleh pengguna non-teknis tanpa perlu instalasi tambahan.

4. KESIMPULAN

Kesimpulan dari penelitian ini menunjukkan bahwa *algoritma Random Forest* yang dikombinasikan dengan representasi teks TF-IDF mampu mengklasifikasikan opini publik terhadap Program Makan Bergizi Gratis dengan akurasi cukup tinggi, terutama dalam mendeteksi sentimen positif. Dengan akurasi sebesar 82% dan F1-score 82% pada data pengujian, model berhasil mengidentifikasi sebagian besar opini positif masyarakat secara efektif. Namun demikian, performa klasifikasi terhadap sentimen negatif dan netral masih rendah, yang menandakan bahwa model belum mampu menangkap kompleksitas bahasa pada dua kategori tersebut secara optimal. Ketidakseimbangan distribusi data, keterbatasan dalam menangkap konteks kalimat, serta kecenderungan model untuk memprediksi kelas mayoritas menjadi tantangan utama yang perlu dibenahi.

Untuk pengembangan lebih lanjut, penerapan teknik *oversampling* seperti *SMOTE* direkomendasikan guna menangani ketimpangan kelas, serta eksplorasi model berbasis konteks seperti *LSTM* atau *transformer* untuk meningkatkan pemahaman terhadap nuansa bahasa alami. Selain itu, sistem klasifikasi yang telah dibangun memiliki prospek aplikasi praktis sebagai alat bantu pemerintah dalam memantau dan mengevaluasi opini masyarakat secara *real-time*, terutama dalam kebijakan publik yang sensitif terhadap persepsi publik. Pendekatan ini tidak hanya mendukung proses pengambilan kebijakan yang lebih responsif, tetapi juga membuka peluang pengembangan sistem analisis sentimen berbasis AI yang adaptif dan akurat untuk isu-isu sosial lainnya.

DAFTAR PUSTAKA

- [1] R. Wijanarko, D. E. Ratnawati, and P. P. Adikara, “Analisis Sentimen Dampak Perkembangan Artificial Intelligence (AI) pada Media Sosial X/Twitter Menggunakan Metode Random Forest,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 1, no. 1, pp. 2548–964, 2017.
- [2] A. Sitanggang, Y. Umaidah, Y. Umaidah, R. I. Adam, and R. I. Adam, “Analisis Sentimen Masyarakat Terhadap Program Makan Siang Gratis Pada Media Sosial X Menggunakan Algoritma Naïve Bayes,” *J. Inform. dan Tek. Elektro Terap.*, vol. 12, no. 3, 2024, doi: 10.23960/jitet.v12i3.4902.
- [3] M. A. Hasanah, S. Soim, and A. S. Handayani, “Implementasi CRISP-DM Model Menggunakan Metode Decision Tree dengan Algoritma CART untuk Prediksi Curah Hujan Berpotensi Banjir,” *J. Appl. Informatics Comput.*, vol. 5, no. 2, pp. 103–108, 2021, doi: 10.30871/jaic.v5i2.3200.
- [4] B. Bayu Baskoro *et al.*, “Analisis Sentimen Pelanggan Hotel di Purwokerto Menggunakan Metode Random Forest dan TF-IDF (Studi Kasus: Ulasan Pelanggan Pada Situs TRIPADVISOR),” *J. Informatics Inf. Syst. Softw. Eng. Appl. (INISTA)*, , vol. Volume 3 N, no. 2, pp. 21–029, 2021, doi: 10.20895/INISTA.V3.
- [5] R. Fachreza and W. T. Handoko, “Analisis Sentimen Masyarakat Terhadap Kinerja Trans Semarang Menggunakan Metode Random Forest,” pp. 724–734, 2024.
- [6] S. Nanda, D. Mualfah, and D. A. Fitri, “Analisis Sentimen Kepuasan Pengguna Terhadap Layanan Streaming Mola Menggunakan Algoritma Random Forest,” *J. Apl. Teknol. Inf. dan Manaj.*, vol. 3, no. 2, pp. 210–219, 2022, doi: 10.31102/jatim.v3i2.1592.
- [7] A. G. Pramita, W. Agus Triyanto, and S. Muzid, “Analisis Sentimen Penggunaan Sunan (Sinau Temenan)E-Learning UMK Sebagai Media PembelajaranMenggunakan Metode Naïve Bayes Classifier,” *J. Pseudocode*, vol. 12, no. 1, pp. 1–6, 2025, doi: 10.33369/pseudocode.12.1.1.
- [8] M. A. A. M. Alamsyah, and M. F. Arif, “Analisis Sentimen Twitter Tentang Pinjaman Online di Indonesia Menggunakan Metode Random Forest,” 2024.
- [9] R. Rahmatulloh, M. I. Ibrahim, and M. R. Handayani, “Analisis Sentimen Publik tentang RUU TNI 2025 dengan Naïve Bayes pada Aplikasi X,” vol. 5, no. 2, pp. 365–379, 2025.
- [10] E. Triningsih, “MACHINE LEARNING PADA SOSIAL MEDIA X,” 2025.
- [11] F. A. Larasati, D. E. Ratnawati, and B. T. Hanggara, “Analisis Sentimen Ulasan Aplikasi Dana dengan Metode Random Forest,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 6, no. 9, pp. 4305–4313, 2022.
- [12] M. Azhari and P. Parjito, “Analisis Sentimen Opini Publik Program Makan Siang Gratis dengan Random Forest Pada Media,” *Build. Informatics, Technol. Sci.*, vol. 6, no. 3, pp. 1932–1942, 2024, doi: 10.47065/bits.v6i3.6423.
- [13] M. Y. Aldean, P. Paradise, and N. A. Setya Nugraha, “Analisis Sentimen Masyarakat Terhadap Vaksinasi Covid-19 di Twitter Menggunakan Metode Random Forest Classifier (Studi Kasus: Vaksin Sinovac),” *J. Informatics, Inf. Syst. Softw. Eng. Appl.*, vol. 4, no. 2, pp. 64–72, 2022, doi: 10.20895/inista.v4i2.575.
- [14] M. R. U. Pulungan, D. E. Ratnawati, and B. Rahayudi, “Analisis Sentimen Ulasan Aplikasi PeduliLindungi dengan Metode Random Forest,” *J. Pengemb. Teknol. Inf. dan Ilmu Komun.*, vol. 6, no. 9, pp. 4378–4385, 2022.
- [15] M. F. Y. Herjanto and C. Carudin, “Analisis Sentimen Ulasan Pengguna Aplikasi Sirekap Pada Play Store Menggunakan Algoritma Random Forest Classifier,” *J. Inform. dan Tek. Elektro Terap.*, vol. 12, no. 2, pp. 1204–1210, 2024, doi: 10.23960/jitet.v12i2.4192.
- [16] W. A. Aziz, *Implementasi metode random forest pada klasifikasi data ulasan konsumen perusahaan (studi kasus: aplikasi kai access)*. 2021.
- [17] V. Fransisco and D. B. Rarasati, “Analisis Sentimen Aplikasi Polri Super App Menggunakan Algoritma Random Forest,” *J. Ilm. Sains dan Teknol.*, vol. 8, no. 2, pp. 183–195, 2024, doi: 10.47080/saintek.v8i2.3383.
- [18] M. Ilham and B. Priambodo, “Analisis Sentimen Publik Terhadap Program Makan Siang Gratis Menggunakan BERT Neural Network Pada Platform X,” vol. 6, no. 2, pp. 1039–1047, 2024.
- [19] A. G. Pramita and F. Nugraha, “Sistem Analisis Sentimen Produk Pada Aplikasi Lazada Menggunakan Metode Naïve Bayes,” *J. Digit*, vol. 14, no. 1, p. 23, 2024, doi: 10.51920/jd.v14i1.362.
- [20] N. Alvia Wirayani *et al.*, “Analisis Sentimen Tanggapan Masyarakat Tentang Garuda IKN Menggunakan Metode Naïve Bayes,” *Decod. J. Pendidik. Teknol. Inf.*, vol. 5, no. 1, pp. 27–40, 2025.
- [21] A. F. Anjani, D. Anggraeni, and I. M. Tirta, “Implementasi Random Forest Menggunakan SMOTE untuk Analisis Sentimen Ulasan Aplikasi Sister for Students UNEJ,” *J. Nas. Teknol. dan Sist. Inf.*, vol. 9, no. 2, pp. 163–172, 2023, doi: 10.25077/teknosi.v9i2.2023.163-172.