ANALISIS PERBANDINGAN KINERJA ALGORITMA NAIVE BAYES DAN SVM DALAM MENGKLASIFIKASI SENTIMEN TERHADAP MBG

Rifki Amelsan, Muntahanah
Teknik Informatika, Fakultas Teknik Universitas Muhammadiyah Bengkulu, Bengkulu
Jalan Bali, Kampung Bali, Kecamatan. Teluk Segara, Kota Bengkulu, Bengkulu
rifkiamelsanbkl@gmail.com

ABSTRAK

Perkembangan platform digital telah mengubah cara masyarakat menyampaikan pendapat, khususnya terhadap kebijakan publik yang berdampak langsung pada kehidupan sehari-hari. Program Makan Bergizi Gratis (MBG) yang digagas pemerintah Indonesia menjadi salah satu topik yang banyak diperbincangkan di media sosial Twitter, menghasilkan volume data teks yang besar namun belum termanfaatkan secara optimal untuk kepentingan evaluasi kebijakan. Kondisi ini mendorong perlunya penerapan metode komputasional yang mampu mengolah data tersebut secara efisien dan objektif. Penelitian ini bertujuan untuk menganalisis dan membandingkan kinerja algoritma Naive Bayes dan Support Vector Machine (SVM) dalam mengklasifikasikan sentimen masyarakat terhadap program MBG. Data yang digunakan merupakan dataset publik dari Kaggle yang berjumlah 3.457 komentar media sosial Twitter. Tahapan penelitian meliputi preprocessing (cleaning, case folding, tokenizing, stopword removal, dan stemming), pelabelan otomatis, pembobotan fitur menggunakan TF-IDF, serta pembagian data sebesar 70% untuk pelatihan dan 30% untuk pengujian. Evaluasi model dilakukan menggunakan confusion matrix dengan metrik accuracy, precision, recall, dan F1-score. Hasil pengujian menunjukkan bahwa Naive Bayes memperoleh akurasi 84,2% dengan F1-score 87,5%, sedangkan SVM menghasilkan akurasi 88,9% dengan F1-score 91,7%. Berdasarkan hasil tersebut, SVM menunjukkan performa yang lebih optimal dibandingkan Naive Bayes dalam mengklasifikasikan sentimen terhadap program MBG.

Kata kunci : *Analisis Sentimen, Kebijakan Publik, MBG, Naive Bayes, Support Vector Machine, TF-IDF*

1. PENDAHULUAN

Implementasi kebijakan Makan Bergizi Gratis (MBG) di Indonesia menjadi salah satu isu strategis yang memperoleh perhatian luas dari masyarakat karena berkaitan langsung dengan peningkatan kualitas gizi nasional serta implikasi sosial dan ekonomi yang menyertainya [1].

Sebagai kebijakan publik yang relatif baru, MBG memunculkan beragam respons yang terekam secara masif melalui media sosial, khususnya platform X (Twitter). Opini publik yang tersebar dalam bentuk teks tersebut mencerminkan persepsi, dukungan, maupun kritik terhadap kebijakan yang diambil pemerintah. Volume data yang besar dan bersifat tidak terstruktur menuntut adanya pendekatan komputasional yang mampu mengekstraksi informasi secara sistematis dan objektif melalui teknik analisis sentimen.

Analisis sentimen merupakan bagian dari text mining yang bertujuan mengidentifikasi serta mengklasifikasikan kecenderungan opini berdasarkan polaritas tertentu, seperti positif dan negatif. Dalam konteks kebijakan publik, pendekatan ini dapat dimanfaatkan sebagai instrumen evaluasi awal untuk memetakan tingkat penerimaan masyarakat secara cepat dan berbasis data [2].

Namun demikian, karakteristik data media sosial yang mengandung singkatan, variasi bahasa informal, simbol, serta tingkat noise yang tinggi menjadi tantangan dalam proses klasifikasi. Oleh karena itu, pemilihan algoritma machine learning yang tepat menjadi faktor krusial dalam menentukan kualitas hasil analisis.

Naive Bayes merupakan algoritma probabilistik yang banyak digunakan dalam klasifikasi teks karena kesederhanaan model serta efisiensi komputasinya. Metode ini bekerja berdasarkan asumsi independensi antar fitur dan sering dijadikan baseline dalam penelitian klasifikasi. Sementara itu, Support Vector Machine (SVM) dikenal sebagai algoritma supervised learning yang efektif dalam menangani data berdimensi tinggi melalui pembentukan hyperplane optimal sebagai pemisah antar kelas. Sejumlah penelitian terdahulu menunjukkan bahwa kedua algoritma tersebut memiliki performa yang kompetitif dalam analisis sentimen, meskipun tingkat akurasinya sangat dipengaruhi oleh karakteristik dataset dan representasi fitur yang digunakan.

Sebagian besar penelitian sebelumnya lebih banyak difokuskan pada domain e-commerce, produk komersial, atau layanan digital yang telah mapan. Kajian terhadap isu kebijakan publik yang bersifat aktual dan dinamis, seperti MBG, masih relatif terbatas. Padahal, domain kebijakan publik memiliki karakteristik linguistik yang berbeda, termasuk penggunaan istilah politik, ekspresi opini kolektif, serta dinamika persepsi masyarakat terhadap pemerintah. Berdasarkan kondisi tersebut, permasalahan yang dikaji dalam penelitian ini adalah bagaimana penerapan algoritma Naive Bayes dan Support Vector Machine dalam mengklasifikasikan sentimen masyarakat terhadap program MBG serta

algoritma mana yang menunjukkan kinerja paling optimal berdasarkan metrik evaluasi yang digunakan.

Penelitian ini bertujuan untuk menganalisis dan membandingkan kinerja algoritma Naive Bayes dan Support Vector Machine dalam klasifikasi sentimen terhadap program MBG dengan menggunakan representasi fitur berbasis Term Frequency–Inverse Document Frequency (TF-IDF). Evaluasi model dilakukan menggunakan metrik accuracy, precision, recall, dan F1-score guna memperoleh gambaran performa yang komprehensif. Hasil penelitian diharapkan dapat memberikan kontribusi ilmiah dalam pengembangan analisis sentimen pada domain kebijakan publik serta menjadi referensi dalam pembangunan sistem pemantauan opini masyarakat berbasis machine learning.

2. TINJAUAN PUSTAKA

2.1. Program Makan Bergizi Gratis (MBG)

Program Makan Bergizi Gratis (MBG) merupakan kebijakan strategis pemerintah yang bertujuan meningkatkan kualitas gizi anak usia sekolah serta menekan angka stunting dan masalah gizi kronis[3]. Kebijakan ini menimbulkan berbagai respons publik di media sosial, baik dukungan maupun kritik, sehingga diperlukan analisis berbasis data untuk memahami persepsi masyarakat secara objektif .

2.2. Analisis Sentimen

Analisis sentimen adalah teknik dalam text mining yang digunakan untuk mengidentifikasi dan mengklasifikasikan opini dalam teks ke dalam kategori positif, negatif. Dalam konteks kebijakan publik, metode ini berfungsi sebagai instrumen evaluasi untuk mengukur respons masyarakat secara cepat dan sistematis.

2.3. Prapemrosesan dan Representasi Teks

Prapemrosesan bertujuan membersihkan dan menyeragamkan data teks agar lebih terstruktur dan siap dianalisis. Tahapan yang umum dilakukan meliputi cleaning, case folding, tokenizing, stopword removal, dan stemming. Setelah itu, teks ditransformasikan ke bentuk numerik menggunakan metode pembobotan seperti TF-IDF untuk merepresentasikan kata sebagai fitur dalam proses klasifikasi.

2.4. Algoritma Naive Bayes

Naive Bayes merupakan algoritma klasifikasi probabilistik berbasis Teorema Bayes dengan asumsi independensi antar fitur. Algoritma ini efisien secara komputasi dan banyak digunakan dalam klasifikasi teks karena performanya yang stabil pada data berdimensi tinggi[4].

2.5. Support Vector Machine (SVM)

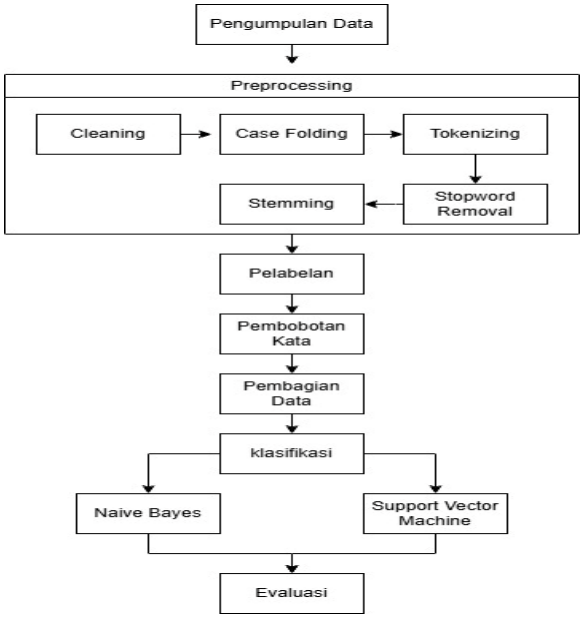
Support Vector Machine (SVM) adalah algoritma supervised learning yang membangun hyperplane optimal untuk memisahkan data antar kelas. Metode ini memiliki kemampuan generalisasi yang baik dan efektif dalam menangani data teks hasil representasi numerik.

2.6. Confusion Matrix

Confusion Matrix digunakan untuk mengevaluasi kinerja model dengan membandingkan hasil prediksi dan label aktual. Berdasarkan matriks ini dihitung metrik seperti accuracy, precision, recall, dan F1-score untuk memberikan gambaran performa model secara komprehensif.

3. METODE PENELITIAN

Penelitian ini menggunakan pendekatan *machine learning* untuk menganalisis sentimen masyarakat terhadap program Makan Bergizi Gratis (MBG). Algoritma Naive Bayes dan Support Vector Machine (SVM) diterapkan untuk membandingkan kinerja klasifikasi sentimen positif dan negatif. Proses penelitian mencakup pengumpulan data, prapemrosesan teks, representasi data numerik, pelatihan model, serta evaluasi kinerja menggunakan metrik yang relevan guna menentukan algoritma dengan performa terbaik.



Gambar 1. Metode Penelitian

3.1. Pengumpulan Data

Data pada penelitian ini diperoleh dari platform Kaggle.com [5], dengan judul "Sentimen Publik Terhadap Makan Bergizi Gratis", berisi komentar masyarakat di Twitter terkait program MBG yang dikumpulkan pada tahun 2025. Dataset berjumlah 3.457 data dengan 2 variabel yaitu username dan Comment. Dataset tersebut digunakan sebagai data sekunder karena bersifat terbuka, relevan dengan topik penelitian, serta memiliki jumlah data yang memadai untuk mendukung proses analisis sentimen. Seluruh data yang diperoleh selanjutnya digunakan sebagai bahan penelitian dan diproses pada tahap berikutnya.

3.2. Preprocessing

Data teks yang diperoleh masih bersifat mentah sehingga memerlukan tahap prapemrosesan sebelum digunakan dalam analisis. Proses ini bertujuan untuk membersihkan, menyeragamkan, dan menyiapkan data agar lebih terstruktur serta relevan untuk klasifikasi sentimen[6].

Tahapan yang dilakukan meliputi:

a. cleaning (menghapus URL, emoji, simbol, angka, dan karakter khusus),
b. case folding (mengubah seluruh teks menjadi huruf kecil),
c. tokenizing (memecah teks menjadi unit kata),
d. stopword removal (menghilangkan kata umum yang tidak informatif), dan
e. stemming (mengubah kata ke bentuk dasar). Rangkaian proses ini meningkatkan konsistensi representasi fitur dan mendukung kinerja model klasifikasi.

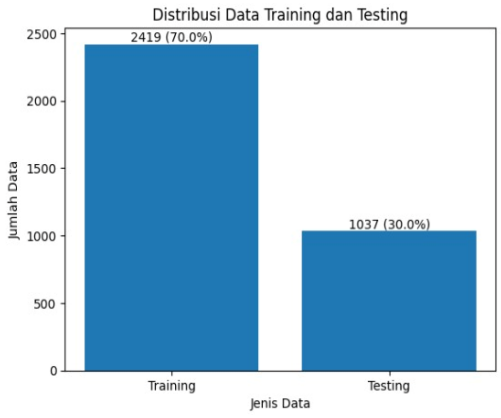
3.3. Pelabelan

Pelabelan data dilakukan secara otomatis menggunakan pendekatan komputasional pada platform Google Colab dengan memanfaatkan bahasa pemrograman Python. Proses pelabelan bertujuan untuk menentukan kategori sentimen setiap komentar ke dalam kelas positif dan negatif berdasarkan hasil analisis teks yang telah melalui tahap prapemrosesan. Label sentimen yang dihasilkan selanjutnya digunakan sebagai variabel target dalam proses pelatihan dan pengujian algoritma Naive Bayes dan Support Vector Machine (SVM).

3.4. Pembobotan Kata

Pembobotan kata dilakukan menggunakan metode **TF-IDF** untuk mengubah data teks menjadi representasi numerik. Metode ini memberikan bobot berdasarkan frekuensi kemunculan kata dan tingkat kepentingannya dalam seluruh dokumen, sehingga menghasilkan fitur yang informatif untuk pelatihan model Naive Bayes dan SVM[7].

3.5. Pembagian Data



Gambar 2. Distribusi Data Training dan Testing

Gambar 2 menunjukkan pembagian dataset menjadi data pelatihan dan data pengujian. Sebanyak 70% data (2.419 data) digunakan untuk melatih model, sedangkan 30% data (1.037 data) digunakan untuk menguji kinerja model. Pembagian ini bertujuan untuk memastikan model dapat dilatih secara optimal serta dievaluasi secara objektif.

3.6. Klasifikasi

Pada tahap klasifikasi, data sentimen yang telah direpresentasikan dalam bentuk numerik digunakan sebagai masukan untuk algoritma Naive Bayes dan Support Vector Machine **(**SVM**)**. Naive Bayes mengklasifikasikan data berdasarkan pendekatan probabilistik, sedangkan SVM bekerja dengan menentukan batas pemisah optimal antar kelas sentimen. Kedua algoritma diterapkan untuk mengklasifikasikan komentar ke dalam kelas positif dan negatif, sehingga memungkinkan dilakukan perbandingan kinerja model dalam analisis sentimen program Makan Bergizi Gratis (MBG).

3.7. Evaluasi

Pada tahap akhir penelitian, evaluasi kinerja dilakukan terhadap algoritma Naive Bayes dan Support Vector Machine **(**SVM**)** dengan menggunakan pendekatan confusion matrix. Metode ini digunakan untuk menilai performa model klasifikasi melalui perbandingan antara hasil prediksi dan label aktual pada data uji. Berdasarkan confusion matrix, kinerja masing-masing algoritma diukur menggunakan metrik accuracy**,** precision**,** recall, dan F1-score. Accuracy merepresentasikan tingkat ketepatan klasifikasi secara keseluruhan, sedangkan precision dan recall menggambarkan kualitas prediksi model terhadap kelas sentimen yang dianalisis. Seluruh metrik evaluasi dihitung berdasarkan formulasi confusion matrix. Berikut merupakan rumus dari hasil dari *confusion matrix* [8].

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \quad (1)$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad (2)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (3)$$

$$\text{F1-Score} = 2 \times \frac{Precision \times Recall}{Precision+Recall} \quad (4)$$

4. HASIL DAN PEMBAHASAN

4.1. Pengumpulan Data

Dataset yang digunakan dalam penelitian ini diperoleh dari platform Kaggle yang memuat opini masyarakat terhadap program Makan Bergizi Gratis (MBG) yang bersumber dari media sosial Twitter. Dataset terdiri atas 3.457 data teks yang telah melalui tahapan preprocessing dan pelabelan sehingga terbagi ke dalam dua kategori sentimen, yaitu positif dan negatif. Data tersebut selanjutnya digunakan sebagai dasar dalam proses pelatihan dan pengujian model klasifikasi menggunakan algoritma Naive Bayes dan Support Vector Machine (SVM).

4.2. Pre-Processing

```
# Fungsi pre-processing
def cleaning(text):
    text = re.sub(r'http\S+|www\S+|https\S+', '', str(text))  # hapus URL
    text = re.sub(r'@\w+|#\w+', '', text)  # hapus mention & hashtag
    text = re.sub(r'[^A-Za-z\s]', '', text)  # hapus angka & tanda baca
    text = re.sub(r'\s+', ' ', text).strip()  # hapus spasi ganda
    return text

def case_folding(text):
    return str(text).lower()

def tokenizing(text):
    return word_tokenize(str(text))

def stopword_removal(tokens):
    return [w for w in tokens if w not in stop_words]

def stemming(tokens):
    return [stemmer.stem(w) for w in tokens]

# Terapkan tahapannya
import nltk
nltk.download('punkt')
nltk.download('punkt_tab')
nltk.download('stopwords')
df['Hasil Cleaning'] = df['full_text'].apply(cleaning)
df['Hasil Case folding'] = df['Hasil Cleaning'].apply(case_folding)
df['Hasil Tokenizing'] = df['Hasil Case folding'].apply(tokenizing)
df['Hasil Stopword'] = df['Hasil Tokenizing'].apply(stopword_removal)
df['Hasil Stemming'] = df['Hasil Stopword'].apply(stemming)
```

Gambar 3. Library Python Preprocessing

Pre-processing merupakan tahap awal dalam analisis sentimen yang bertujuan menyiapkan data teks sebelum proses klasifikasi. Pada tahap ini, data mentah dibersihkan dan disederhanakan untuk menghilangkan elemen yang tidak relevan serta mengurangi ketidakteraturan bahasa. Proses ini menghasilkan data yang lebih konsisten dan representatif sehingga mendukung peningkatan kinerja model klasifikasi. Hasil dari proses ini disajikan pada Gambar 3.

4.3. Cleaning

Cleaning data merupakan proses pembersihan data dengan menghapus karakter yang tidak relevan seperti emoji, URL, hashtag, spasi berlebih, dan angka sehingga data siap untuk tahap preprocessing selanjutnya[9]. Hasil proses pembersihan data disajikan pada Tabel 1.

Tabel 1. Hasil Cleaning

No	Full_Text	Cleaning
1	Makan Siang Bergizi Gratis https://t.co/f27aIthAdX	Makan Siang Bergizi Gratis
2	Momen Prabowo Tanda Tangani Sepatu Siswa di Bogor saat Tinjau Makan Bergizi Gratis https://t.co/2ZlA35BStZ	Momen Prabowo Tanda Tangani Sepatu Siswa di Bogor saat Tinjau Makan Bergizi Gratis
…	…	…
3456	@FOODFESS2 @gibran_tweet bisa ga menu makan siang gratis utk anak sekolah seperti ini??? Hanya 10 rb pasti kenyanggg murid senang ortu senang	bisa ga menu makan siang gratis utk anak sekolah seperti ini Hanya rb pasti kenyanggg murid senang ortu senang
3457	@kurawa Tanya'in suuuu.. AnakÂ²e @gibran_tweet @bobbynasution_ dan cucuÂ² @jokowi lainnya serta anak/cucu menteriÂ² juga makan siang BERGIZI gratis 10rb'an disekolah mereka? Atau barangkali anak/cucumu juga nguntal?? Koe ki lanangan penjilat.! Menjijikan.!	Tanyain suuuu Anake dan cucu lainnya serta anakcucu Menteri juga makan siang BERGIZI gratis rban disekolah mereka Atau barangkali anakcucumu juga nguntal Koe ki lanangan penjilat Menjijikan

4.4. Case Folding

Tahap *case folding* dilakukan dengan mengubah seluruh teks menjadi huruf kecil untuk memastikan konsistensi penulisan dan meningkatkan ketepatan proses klasifikasi[10]. Hasil dari proses ini disajikan pada Tabel 2.

Tabel 2. Hasil Case Folding

No	Cleaning	Case Folding
1	Makan Siang Bergizi Gratis	makan siang bergizi gratis
2	Momen Prabowo Tanda Tangani Sepatu Siswa di Bogor saat Tinjau Makan Bergizi Gratis	momen prabowo tanda tangani sepatu siswa di bogor saat tinjau makan bergizi gratis
…	…	…
3456	bisa ga menu makan siang gratis utk anak sekolah seperti ini Hanya rb pasti kenyanggg murid senang ortu senang	bisa ga menu makan siang gratis utk anak sekolah seperti ini hanya rb pasti kenyanggg murid senang ortu senang
3457	Tanyain suuuu Anake dan cucu lainnya serta anakcucu Menteri juga makan siang BERGIZI gratis rban disekolah mereka Atau barangkali anakcucumu juga nguntal Koe ki lanangan penjilat Menjijikan	tanyain suuuu anake dan cucu lainnya serta anakcucu menteri juga makan siang bergizi gratis rban disekolah mereka atau barangkali anakcucumu juga nguntal koe ki lanangan penjilat menjijikan

4.5. Tokenizing

Tokenizing adalah proses memecah teks menjadi kata-kata (token) sebagai dasar dalam tahap ekstraksi fitur dan klasifikasi sentimen. Hasil proses tokenizing ditampilkan pada Tabel 3.

Tabel 3. Hasil Tokenizing

No	Case Folding	Tokenizing
1	makan siang bergizi gratis	['makan', 'siang', 'bergizi', 'gratis']
2	momen prabowo tanda tangani sepatu siswa di bogor saat tinjau makan bergizi gratis	['momen', 'prabowo', 'tanda', 'tangani', 'sepatu', 'siswa', 'di', 'bogor', 'saat', 'tinjau', 'makan', 'bergizi', 'gratis']
…	…	…

No	Case Folding	Tokenizing
3456	bisa ga menu makan siang gratis utk anak sekolah seperti ini hanya rb pasti kenyanggg murid senang ortu senang	['bisa', 'ga', 'menu', 'makan', 'siang', 'gratis', 'utk', 'anak', 'sekolah', 'seperti', 'ini', 'hanya', 'rb', 'pasti', 'kenyanggg', 'murid', 'senang', 'ortu', 'senang']
3457	tanyain suuuu anake dan cucu lainnya serta anakcucu menteri juga makan siang bergizi gratis rban disekolah mereka atau barangkali anakcucumu juga nguntal koe ki lanangan penjilat menjijikan	['tanyain', 'suuuu', 'anake', 'dan', 'cucu', 'lainnya', 'serta', 'anakcucu', 'menteri', 'juga', 'makan', 'siang', 'bergizi', 'gratis', 'rban', 'disekolah', 'mereka', 'atau', 'barangkali', 'anakcucumu', 'juga', 'nguntal', 'koe', 'ki', 'lanangan', 'penjilat', 'menjijikan']

4.6. Stopword Removal

Stopword removal dilakukan untuk menghapus kata-kata umum yang tidak berkontribusi terhadap makna sentimen, sehingga proses ekstraksi fitur menjadi lebih efektif. Hasil tahap ini disajikan pada Tabel 4.

Tabel 4. Hasil Stopword removal

No	Tokenizing	Stopword removal
1	['makan', 'siang', 'bergizi', 'gratis']	['makan', 'siang', 'bergizi', 'gratis']
2	['momen', 'prabowo', 'tanda', 'tangani', 'sepatu', 'siswa', 'di', 'bogor', 'saat', 'tinjau', 'makan', 'bergizi', 'gratis']	['momen', 'prabowo', 'tanda', 'tangani', 'sepatu', 'siswa', 'bogor', 'tinjau', 'makan', 'bergizi', 'gratis']
…	…	…
3456	['bisa', 'ga', 'menu', 'makan', 'siang', 'gratis', 'utk', 'anak', 'sekolah', 'seperti', 'ini', 'hanya', 'rb', 'pasti', 'kenyanggg', 'murid', 'senang', 'ortu', 'senang']	['ga', 'menu', 'makan', 'siang', 'gratis', 'utk', 'anak', 'sekolah', 'ini', 'rb', 'kenyanggg', 'murid', 'senang', 'ortu', 'senang']
3457	['tanyain', 'suuuu', 'anake', 'dan', 'cucu', 'lainnya', 'serta', 'anakcucu', 'menteri', 'juga', 'makan', 'siang', 'bergizi', 'gratis', 'rban', 'disekolah', 'mereka', 'atau', 'barangkali', 'anakcucumu', 'juga', 'nguntal', 'koe', 'ki', 'lanangan', 'penjilat', 'menjijikan']	['tanyain', 'suuuu', 'anake', 'cucu', 'lainnya', 'anakcucu', 'menteri', 'makan', 'siang', 'bergizi', 'gratis', 'rban', 'disekolah', 'atau', 'barangkali', 'anakcucumu', 'nguntal', 'koe', 'ki', 'lanangan', 'penjilat', 'menjijikan']

4.7. Stemming

Stemming dilakukan untuk mengubah kata ke bentuk dasarnya agar representasi teks menjadi lebih konsisten dan mendukung peningkatan akurasi klasifikasi. Hasil proses ini ditampilkan pada Tabel 5.

Tabel 5. Hasil Stemming

No	Stopword removal	Stemming
1	['makan', 'siang', 'bergizi', 'gratis']]	['makan', 'siang', 'gizi', 'gratis']
2	['momen', 'prabowo', 'tanda', 'tangani', 'sepatu', 'siswa', 'bogor', 'tinjau', 'makan', 'bergizi', 'gratis']	['momen', 'prabowo', 'tanda', 'tangan', 'sepatu', 'siswa', 'bogor', 'tinjau', 'makan', 'gizi', 'gratis']
…	…	…
3456	['ga', 'menu', 'makan', 'siang', 'gratis', 'utk', 'anak', 'sekolah', 'ini', 'rb', 'kenyanggg', 'murid', 'senang', 'ortu', 'senang']	['ga', 'menu', 'makan', 'siang', 'gratis', 'utk', 'anak', 'sekolah', 'ini', 'rb', 'kenyanggg', 'murid', 'senang', 'ortu', 'senang']
3457	['tanyain', 'suuuu', 'anake', 'cucu', 'lainnya', 'anakcucu', 'menteri', 'makan', 'siang', 'bergizi', 'gratis', 'rban', 'disekolah', 'atau', 'barangkali', 'anakcucumu', 'nguntal', 'koe', 'ki', 'lanangan', 'penjilat', 'menjijikan']	['tanyain', 'suuuu', 'anak e', 'cucu', 'lain', 'anakcucu', 'menteri', 'makan', 'siang', 'gizi', 'gratis', 'rban', 'seko', 'atau', 'barangkali', 'anakcucumu', 'nguntal', 'koe', 'ki', 'lanang', 'jilat', 'jijik']

4.8. Pelabelan

Pelabelan dilakukan secara otomatis menggunakan Python di Google Colab tanpa melibatkan annotator manusia, karena jumlah data yang besar yaitu 3.457 data sehingga pelabelan manual dinilai tidak efisien. Proses pelabelan dilakukan dengan membandingkan skor sentimen positif dan negatif pada setiap teks. Jika skor positif lebih tinggi, teks diberi label positif, dan sebaliknya diberi label negatif. Data berlabel ini kemudian digunakan dalam proses pelatihan dan pengujian model, seperti ditunjukkan pada Tabel 6.

Tabel 6. Hasil Label

username	Comment	Label
banten_berita	makan siang gizi gratis	Positif
okezonenews	momen prabowo tanda tangan sepatu siswa bogor tinjau makan gizi gratis	Positif
…	…	…
SamPrd24	ga menu makan siang gratis utk anak sekolah ini rb kenyanggg murid senang ortu senang	Positif
Iam_Nobody_1145	tanyain suuuu anak e cucu lain anakcucu menteri makan siang gizi gratis rban seko atau barangkali anakcucumu nguntal koe ki lanang jilat jijik	Negatif

4.9. Pembobotan kata

Pembobotan kata menggunakan metode TF-IDF untuk mengubah teks menjadi data numerik berdasarkan frekuensi kata dalam dokumen dan keseluruhan korpus. Penelitian ini menggunakan maksimal 5.000 fitur dengan unigram dan bigram. Hasil pembobotan ditampilkan pada Tabel 7.

Tabel 7. Hasil Pembobotan kata

No	abah	abal	makan	…	zulkifli hasan
1	0	0	0.159483	…	0
2	0	0	0.042497		0
…	…	…	…	…	…
3456	0	0	0.042753	…	0
3457	0	0	0.039005	…	0

4.10. Klasifikasi

Tahap ini melakukan klasifikasi sentimen menggunakan algoritma Naive Bayes dan Support Vector Machine (SVM). Naive Bayes memprediksi kelas berdasarkan probabilitas kata, sedangkan SVM menentukan batas pemisah terbaik antar kelas. Hasil prediksi pada data uji kemudian dibandingkan dengan label asli, seperti ditampilkan pada Tabel 8.

Tabel 8. Hasil Klasifikasi ..

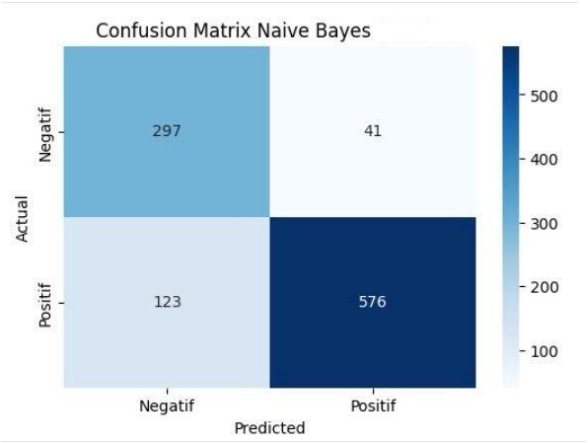
username	Comment	Label	Prediksi NB	Prediksi svm
okezonenews	momen prabowo tanda tangan sepatu siswa bogor tinjau makan gizi gratis	Positif	Positif	Positif
Sofffyannn	moga program makan gizi gratis dri mda tetap lanjut walaupun program sma perintah	Positif	Positif	Positif
…	…	…	…	…
SamPrd24	ga menu makan siang gratis utk anak sekolah ini rb kenyanggg murid senang ortu senang	Positif	Negatif	Negatif
Iam_Nobody_1145	tanyain suuuu anak e cucu lain anakcucu menteri makan siang gizi gratis rban seko atau barangkali anakcucumu nguntal koe ki lanang jilat jijik	Negatif	Negatif	Positif

Berdasarkan Tabel 8, kedua algoritma berhasil memprediksi dengan benar pada komentar yang memiliki konteks sentimen jelas, seperti pada komentar milik okezonenews dan Sofffyannn yang berlabel positif dan diprediksi positif oleh keduanya. Namun terdapat kasus misklasifikasi, seperti komentar milik SamPrd24 yang berlabel positif namun diprediksi negatif oleh kedua algoritma. Hal ini kemungkinan disebabkan penggunaan bahasa informal dan singkatan sehingga model kesulitan menangkap makna sebenarnya. Selain itu, terdapat perbedaan prediksi antar algoritma pada komentar milik Iam_Nobody_1145 yang berlabel negatif, di mana Naive Bayes memprediksi dengan benar sedangkan SVM memprediksi positif, menunjukkan bahwa SVM kurang optimal dalam menangani teks yang mengandung bahasa ambigu.

4.11. Evaluasi

Pada tahap evaluasi, algoritma Naive Bayes diuji menggunakan pembagian data latih dan data uji untuk menilai kemampuan model dalam mengklasifikasikan sentimen. Pengukuran kinerja dilakukan melalui confusion matrix dengan menghitung nilai accuracy, precision, recall, dan F1-score.

Hasil pengujian menunjukkan bahwa model mampu mengelompokkan data ke dalam kategori positif dan negatif dengan tingkat ketepatan yang baik. Distribusi hasil klasifikasi selanjutnya disajikan dalam bentuk Gambar 4 untuk memperjelas perbandingan jumlah masing-masing kelas.

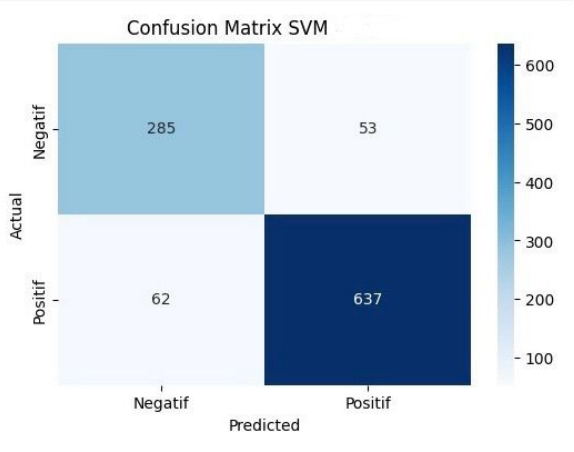



```
Akurasi Naive Bayes : 0.8418514946962391

Classification Report:

              precision    recall  f1-score   support

     Negatif       0.71      0.88      0.78       338
     Positif       0.93      0.82      0.88       699

    accuracy                           0.84      1037
   macro avg       0.82      0.85      0.83      1037
weighted avg       0.86      0.84      0.85      1037
```

Gambar 4. hasil confusion matrix naïve bayes

Pengujian selanjutnya dilakukan pada dataset yang sama dengan menerapkan algoritma Support Vector Machine untuk mengklasifikasikan sentimen ke dalam kategori positif dan negatif. Proses evaluasi dilakukan menggunakan confusion matrix guna memperoleh nilai accuracy, precision, recall, dan F1-score. Hasil pengujian tersebut digunakan untuk

menilai tingkat ketepatan serta kemampuan model dalam mengenali pola sentimen pada data penelitian.



Akurasi SVM : 0.8891031822565092

Classification Report:

	precision	recall	f1-score	support
Negatif	0.82	0.84	0.83	338
Positif	0.92	0.91	0.92	699
accuracy			0.89	1037
macro avg	0.87	0.88	0.87	1037
weighted avg	0.89	0.89	0.89	1037

Gambar 5. hasil confusion matrix SVM

Berdasarkan hasil Gambar 4 dan 5, model Naive Bayes memperoleh akurasi 84,2%, dengan precision 93,3%, recall 82,4%, dan F1-score 87,5%. Nilai ini menunjukkan kemampuan klasifikasi yang cukup baik dalam membedakan sentimen positif dan negatif. Sementara itu, Support Vector Machine mencatat performa lebih tinggi dengan akurasi 88,9%, precision 92,3%, recall 91,1%, serta F1-score 91,7%. Secara umum, SVM memberikan hasil yang lebih optimal dibandingkan Naive Bayes pada dataset penelitian ini.

Keunggulan SVM ini dapat dijelaskan dari beberapa aspek. Pertama, SVM mampu bekerja lebih baik pada data berdimensi tinggi hasil representasi TF-IDF karena mencari hyperplane optimal yang memisahkan antar kelas secara maksimal. Kedua, Naive Bayes memiliki asumsi independensi antar fitur yang dalam kenyataannya tidak selalu terpenuhi pada data teks, sehingga mempengaruhi kualitas prediksinya. Ketiga, nilai recall SVM yang lebih tinggi yaitu 91,1% dibandingkan Naive Bayes 82,4% menunjukkan bahwa SVM lebih mampu mendeteksi sentimen secara menyeluruh, termasuk pada teks yang mengandung bahasa informal dan ambigu.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil evaluasi menggunakan confusion matrix, algoritma Naive Bayes dan Support Vector Machine (SVM) menunjukkan kemampuan yang baik dalam mengklasifikasikan sentimen masyarakat terhadap program Makan Bergizi Gratis (MBG). Naive Bayes memperoleh akurasi sebesar 84,2% dengan F1-score 87,5%, sedangkan SVM mencapai akurasi 88,9% dengan F1-score 91,7%. Hasil tersebut menunjukkan bahwa SVM memiliki performa klasifikasi yang lebih optimal pada dataset penelitian ini, khususnya dalam menangani representasi fitur berbasis TF-IDF pada data teks berdimensi tinggi. Dengan demikian, SVM direkomendasikan sebagai model yang lebih efektif untuk implementasi sistem pemantauan opini publik terhadap kebijakan MBG. Untuk penelitian selanjutnya, disarankan melakukan optimasi parameter model, memperluas ukuran dan variasi dataset, serta menambahkan kategori sentimen netral guna meningkatkan kemampuan generalisasi dan stabilitas model klasifikasi.

DAFTAR PUSTAKA

[1] G. Governance and A. Kualitatif, “Arus Jurnal Sosial dan Humaniora (AJSH) Kebijakan Makan Bergizi Gratis dan Relevansinya terhadap Nilai-nilai,” vol. 5, no. 1, 2025.

[2] S. Jurnalis Pipin and H. Kurniawan, “Analisis Sentimen Kebijakan MBKM Berdasarkan Opini Masyarakat di Twitter Menggunakan LSTM,” *J. SIFO Mikroskil*, vol. 23, no. 2, pp. 197–208, 2022, doi: 10.55601/jsm.v23i2.900.

[3] M. Manurung, B. Pamungkas, K. A. Harahap, and P. Siregar, “Kebijakan Pemerintah dalam Mendukung Program Peningkatan Kualitas SDM Masyarakat Melalui Program Makan Bergizi Gratis (MBG),” *J. Ilm. Dikdaya*, vol. 15, no. 2, p. 691, 2025, doi: 10.33087/dikdaya.v15i2.881.

[4] S. Khoerunnisa, D. F. Shiddieq, and D. Nurhayati, “Penerapan Algoritma Naive Bayes dengan Teknik TF-IDF dan Cross Validation untuk Analisis Sentimen Terhadap Starlink,” *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 5, no. 2, pp. 566–577, 2025, doi: 10.57152/malcom.v5i2.1852.

[5] M. H. (Ambil dari username Hanif281103), “Sentimen Publik Terhadap Makan Bergizi Gratis,” Kaggle Datasets. [Online]. Available: https://www.kaggle.com/datasets/hanif281103/sentimen-publik-terhadap-makan-bergizi-gratis

[6] M. Safrudin and U. Hayati, “Perbandingan kinerja algoritma Naïve Bayes dan Support Vector Machine dalam mengklasifikasikan sentimen pada ulasan game Genshin Impact.,” *J. Mhs. Tek. Inform.*, vol. 8, no. 3, pp. 3182–3188, 2024.

[7] R. Syahputra, G. J. Yanris, and D. Irmayani, “SVM and Naïve Bayes Algorithm Comparison for User Sentiment Analysis on Twitter,” *Sinkron*, vol. 7, no. 2, pp. 671–678, 2022, doi: 10.33395/sinkron.v7i2.11430.

[8] H. Hariyadi, D. Firdo, and M. H. Al Rafi, “Implementasi Algoritma Naïve Bayes dan Support Vector Machine pada Analisis Sentimen Ulasan Aplikasi Canva,” *J. Minfo Polgan*, vol. 13, no. 1, pp. 261–269, 2024, doi: 10.33395/jmp.v13i1.13568.

[9] R. Hayati, M. Marzuki, F. Fachrurazi, A. Karim, S. H. Pratiwi, and R. Dewi, "Penerapan Filsafat Pendidikan Oleh Tenaga Pendidik Di Sekolah Dasar," *Pedagog. J. Ilm. Pendidik. dan Pembelajaran Fak. Tarb. Univ. Muhammadiyah Aceh; Vol 10, No 1, April (2023); 35-48 ; 2622-9005 ; 2337-7364*, vol. 12, no. 1, pp. 5–6, 2023, [Online]. Available: https://ejournal.unmuha.ac.id/index.php/pedagogik/article/view/1702

[10] R. W. Sandra, Y. Vitriani, M. Affandes, and S. Sanjaya, "Cryptocurrency Sentiment Classification Based on Comments On Facebook Using K-Nearest Neighbor," *Int. J. Inf. Syst. Technol. Akreditasi*, vol. 6, no. 158, pp. 259–269, 2022.