

PERBANDINGAN NAÏVE BAYES DAN SVM PADA SENTIMEN KASUS KERACUNAN MAKAN BERGIZI GRATIS DI PLATFORM X

Larasati Mya Mulyono^{1*}, Nina Sulistiyowati², Ahmad Khusaeri³

^{1,2,3}Sistem Informasi; Fakultas Ilmu Komputer; Universitas Singaperbangsa Karawang; Jl. HS. Ronggo Waluyo, Puseurjaya, Telukjambe Timur, Karawang, Jawa Barat 41361

Keywords:

Analisis Sentimen, Naïve Bayes, *Support Vector Machine*, CRISP-DM, Program Makan Bergizi Gratis.

Correspondent Email:

larasati.mulyono20@gmail.com

Abstrak. Program Makan Bergizi Gratis (MBG) merupakan program pemerintah untuk pemenuhan gizi masyarakat. Munculnya kasus keracunan makanan dalam pelaksanaannya memicu beragam tanggapan masyarakat di media sosial, khususnya platform X. Analisis sentimen diperlukan untuk memahami persepsi publik sekaligus mengevaluasi pelaksanaan program. Penelitian ini bertujuan menganalisis sentimen masyarakat terhadap kasus keracunan pada Program MBG serta membandingkan performa algoritma Naïve Bayes dan SVM dalam klasifikasi sentimen. Metodologi yang digunakan adalah CRISP-DM, meliputi *business understanding*, *data understanding*, *data preparation*, *modeling*, *evaluation*, dan *deployment*. Data berupa 1.744 tweet dari platform X pada periode 06 Januari–31 Desember 2025, yang setelah *preprocessing* menjadi 1.533 tweet. Tahap *data preparation* mencakup *cleaning*, *case folding*, *normalisasi*, *tokenizing*, *stopword removal*, dan *stemming*. Pelabelan sentimen menggunakan VADER, ekstraksi fitur menggunakan TF-IDF, dan penyeimbangan kelas menggunakan SMOTE pada data latih. Hasil menunjukkan opini masyarakat didominasi sentimen negatif (1.162 negatif, 366 netral, 5 positif), mengindikasikan persepsi publik yang cenderung negatif terhadap Program MBG. SVM menunjukkan performa lebih baik dibandingkan Naïve Bayes. Pada evaluasi internal, SVM dengan SMOTE mencapai akurasi tertinggi (84,42%), sedangkan pada evaluasi eksternal, SVM tanpa SMOTE unggul dengan 80,13%. Berdasarkan temuan ini, pihak terkait disarankan melakukan evaluasi program, meningkatkan pengawasan keamanan pangan, serta memperbaiki sistem distribusi dan transparansi Program MBG.



Copyright © 2026 The Author(s), Published by Jurnal Informatika dan Teknik Elektro Terapan. This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0).

Abstract. The Makan Bergizi Gratis (MBG) program is a government initiative aimed at fulfilling public nutritional needs. The emergence of food poisoning cases during its implementation triggered various public responses on social media, particularly on platform X. Sentiment analysis is necessary to understand public perception and evaluate program implementation. This study aims to analyze public sentiment toward food poisoning cases in the MBG program and compare the performance of Naïve Bayes and SVM algorithms in sentiment classification. The methodology employed is CRISP-DM, covering *business understanding*, *data understanding*, *data preparation*, *modeling*, *evaluation*, and *deployment*. Data consisted of 1,744 tweets from platform X during the period of January 6–December 31, 2025, which after preprocessing yielded 1,533 tweets. Data preparation included cleaning, case folding, normalization, tokenizing, stopword removal, and stemming. Sentiment labeling used VADER, feature extraction used TF-IDF, and class balancing applied SMOTE on training data.

Results indicate that public opinion was dominated by negative sentiment (1,162 negative, 366 neutral, 5 positive), reflecting an overall negative public perception toward the MBG program. SVM outperformed Naïve Bayes across evaluations. In internal evaluation, SVM with SMOTE achieved the highest accuracy (84.42%), while in external evaluation, SVM without SMOTE led with 80.13%. Based on these findings, relevant stakeholders are advised to conduct program evaluation, strengthen food safety oversight, and improve the distribution system and transparency of the MBG program.

1. PENDAHULUAN

Program Makan Bergizi Gratis (MBG) merupakan program utama pemerintah Prabowo–Gibran yang bertujuan mengatasi permasalahan gizi buruk dan meningkatkan kualitas sumber daya manusia di Indonesia [1]. Data dari Kementerian Kesehatan dan Kemenko Pembangunan Manusia dan Kebudayaan menunjukkan bahwa 41% peserta didik mengalami kekurangan pangan yang berdampak pada penurunan kualitas Pendidikan [2]. Gagasan program ini telah muncul sejak Juli 2006, berangkat dari tingginya angka stunting, dengan sekitar 30% anak usia di bawah lima tahun terdampak akibat ketidakcukupan asupan gizi [3]. Sejak diimplementasikan pada 6 Januari 2025, MBG dilaksanakan melalui kerja sama dengan Badan Gizi Nasional (BGN) untuk memastikan pemenuhan standar kualitas, kebersihan, dan kandungan gizi [4], [5]. Pada tahap awal, program ini menargetkan tiga juta penerima manfaat dari kelompok rentan, meliputi balita, santri, peserta didik dari PAUD hingga SMA, serta ibu hamil dan menyusui [6].

Namun demikian, implementasi Program Makan Bergizi Gratis (MBG) masih menghadapi sejumlah kendala, salah satunya berupa kasus keracunan makanan pada penerima manfaat yang diduga disebabkan oleh penyajian makanan yang tidak memenuhi standar. Pada periode Januari hingga Mei 2025, tercatat puluhan insiden keracunan di lebih dari sepuluh provinsi dengan jumlah korban mencapai ribuan orang [4]. Data Jaringan Pemantau Pendidikan Indonesia (JPPI) menunjukkan bahwa jumlah korban terus meningkat, mencapai 10.482 siswa pada awal Oktober 2025 dan 13.168 anak pada pertengahan Oktober 2025 [7], [8]. Kasus serupa juga dilaporkan di Bandung Barat pada akhir Oktober 2025 dengan total 202 korban dari berbagai kelompok masyarakat [9].

Kondisi ini menimbulkan kekhawatiran publik terhadap standar higienitas dan pengawasan makanan dalam program MBG, yang kemudian mendorong munculnya berbagai opini masyarakat di media sosial, khususnya pada platform X.

Platform X merupakan media sosial yang banyak digunakan masyarakat untuk menyampaikan informasi, opini, dan aspirasi. Pada Juli 2025, jumlah pengguna X secara global mencapai 561 juta, dengan Indonesia menempati peringkat keempat sebagai negara dengan jumlah pengguna terbanyak, yakni sekitar 23,76 juta pengguna [10]. Melalui unggahan teks, pengguna dapat mengekspresikan tanggapan terhadap berbagai isu, baik dalam bentuk sentimen positif maupun negative [11], [12]. Oleh karena itu, analisis terhadap opini di platform X menjadi penting sebagai sumber evaluasi non-formal yang dapat mendukung perbaikan kebijakan dan mitigasi krisis reputasi pemerintah.

Seiring perkembangan teknologi, analisis sentimen (*sentiment analysis*) menjadi pendekatan yang efektif untuk mengidentifikasi pandangan publik terhadap suatu isu melalui data teks di platform digital. Analisis sentimen, atau opinion mining, merupakan proses otomatis untuk mengklasifikasikan opini ke dalam kategori tertentu, seperti positif, negatif, atau netral [13]. Metode ini banyak dimanfaatkan untuk mengukur persepsi masyarakat, memantau citra, serta mendukung evaluasi kebijakan [14]. Dalam konteks kasus keracunan makanan pada program MBG, analisis sentimen dapat memberikan gambaran mengenai respons dan kecenderungan opini publik. Agar proses analisis dilakukan secara sistematis dan menghasilkan model yang valid, penelitian ini menggunakan kerangka kerja CRISP-DM sebagai pendekatan metodologis. Dalam konteks kasus keracunan makanan pada program MBG, analisis sentimen dapat

memberikan gambaran mengenai respons dan kecenderungan opini publik. Agar proses analisis dilakukan secara sistematis dan menghasilkan model yang valid, penelitian ini menggunakan kerangka kerja CRISP-DM sebagai pendekatan metodologis. Pengolahan data teks dalam jumlah besar dan kompleks memerlukan pendekatan yang terstruktur agar proses analisis berjalan sistematis dan terarah. Salah satu metodologi yang umum digunakan adalah CRISP-DM (*Cross-Industry Standard Process for Data Mining*), yang merupakan standar dalam proses data mining dan bersifat fleksibel untuk berbagai permasalahan [15], [16]. CRISP-DM terdiri dari enam tahapan, yaitu *business understanding*, *data understanding*, *data preparation*, *modeling*, *evaluation*, dan *deployment*, dengan alur yang tidak kaku sehingga memungkinkan iterasi antar tahapan [17]. Pada tahap *data preparation*, ketidakseimbangan distribusi antar kelas menjadi permasalahan umum yang dapat menurunkan kinerja model klasifikasi. Salah satu metode penanganannya adalah SMOTE (*Synthetic Minority Over-sampling Technique*), yang menambahkan sampel pada kelas minoritas untuk menyeimbangkan proporsi data [18]. Pada tahap *modeling*, pemilihan algoritma klasifikasi menjadi faktor penting untuk menghasilkan kinerja analisis sentimen yang optimal dan akurat.

Penelitian ini membandingkan kinerja algoritma Naïve Bayes dan *Support Vector Machine* (SVM) dalam menganalisis sentimen terkait kasus keracunan pada program MBG di platform X. Naïve Bayes merupakan metode klasifikasi berbasis probabilitas yang mengacu pada Teorema Bayes dengan asumsi independensi antar fitur [19]. Sementara itu, SVM merupakan metode klasifikasi yang efektif untuk data linear maupun non-linear melalui pemanfaatan fungsi kernel, sehingga mampu menghasilkan akurasi yang tinggi dalam klasifikasi teks [20]. Sejumlah penelitian terdahulu telah membandingkan kinerja algoritma Naïve Bayes dan *Support Vector Machine* (SVM) dalam analisis sentimen. Noviana dan Rasal menunjukkan bahwa SVM memiliki akurasi lebih tinggi (81%) dibandingkan Naïve Bayes (79%) pada data dari platform X, meskipun Naïve Bayes lebih unggul dalam efisiensi proses [12]. Sebaliknya, Firdausy et al. menemukan bahwa Naïve Bayes

memberikan performa yang lebih baik dibandingkan SVM dan Deep Learning pada analisis sentimen e-learning dengan akurasi 82,54% [21]. Penelitian lain oleh Maulana et al. dan Petiwi et al. menunjukkan bahwa SVM cenderung menghasilkan performa yang lebih konsisten dan akurat, terutama pada data dengan variasi kata yang tinggi, sehingga lebih optimal dalam klasifikasi sentimen di media social [22], [23]. Di sisi lain, studi oleh Laia et al. membuktikan bahwa Naïve Bayes juga memiliki kinerja yang stabil dalam menganalisis sentimen terkait program MBG dengan akurasi mencapai 86,46% [24]. Selain itu, Fatkhurrohman et al. melalui pendekatan lexicon-based menemukan bahwa mayoritas sentimen terhadap program MBG bersifat positif, meskipun terdapat kritik terkait aspek pelaksanaan dan keamanan pangan [25]. Temuan-temuan tersebut menunjukkan bahwa kedua algoritma memiliki keunggulan masing-masing, sehingga pemilihan metode sangat memengaruhi hasil analisis sentimen.

Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk menganalisis persepsi masyarakat terhadap insiden keracunan makanan dalam pelaksanaan Program Makan Bergizi Gratis (MBG) di platform X menggunakan pendekatan analisis sentimen berbasis CRISP-DM, serta membandingkan kinerja algoritma Naïve Bayes dan *Support Vector Machine* (SVM) dalam mengategorikan sentiment pada data dari platform X.

2. TINJAUAN PUSTAKA

2.1 *Text Mining*

Text mining merupakan sekumpulan teknik yang digunakan untuk memperoleh informasi bermakna dari dokumen tidak terstruktur maupun semi-terstruktur [26]. Teknik ini dapat dipahami sebagai proses berbasis pengetahuan di mana pengguna berinteraksi dengan kumpulan dokumen menggunakan berbagai alat analisis untuk mengekstraksi pengetahuan yang tidak tersimpan secara eksplisit di dalam teks [27]. Sebagai pengembangan dari data mining, text mining berfokus pada identifikasi pola-pola bermakna dari kumpulan data teks berskala besar dan termasuk salah satu pendekatan yang umum digunakan dalam proses klasifikasi [28]. Tujuan utama *text mining* adalah mengekstraksi

wawasan dan pengetahuan yang relevan dari kumpulan dokumen teks untuk mendukung proses pengambilan keputusan [29]. Dalam konteks media sosial, text mining banyak diterapkan untuk menganalisis sentimen publik, di mana teks pengguna diklasifikasikan ke dalam kategori positif, negatif, atau netral guna memahami pola opini dan tren komunikasi masyarakat terhadap suatu isu [28].

2.2 Analisis Sentimen

Analisis sentimen merupakan teknik pengolahan teks secara otomatis yang bertujuan untuk mengenali dan memahami muatan emosional yang terkandung dalam suatu ekspresi atau pendapat [30]. Teknik ini mencakup bidang kajian yang lebih luas, meliputi analisis opini, emosi, penilaian, serta sikap individu terhadap suatu objek seperti produk, layanan, maupun kebijakan publik [31]. Melalui proses analisis dan pengolahan teks secara otomatis, analisis sentimen menghasilkan informasi mengenai arah atau kecenderungan opini yang terkandung dalam data [32].

2.3 Naïve Bayes

Naïve Bayes adalah algoritma klasifikasi berbasis probabilitas yang bekerja dengan menghitung kemungkinan setiap kelas berdasarkan frekuensi kemunculan data pada tahap pelatihan [27] [33]. Algoritma ini dikembangkan dari konsep Teorema Bayes yang pertama kali dikemukakan oleh Thomas Bayes [14], dengan asumsi bahwa setiap fitur atau atribut dalam data bersifat saling independen [27]. Meskipun memiliki struktur yang sederhana, Naïve Bayes mampu menghasilkan performa klasifikasi yang baik dan mudah diimplementasikan [19].

Teorema Bayes memanfaatkan nilai probabilitas untuk mengubah ketidakpastian data menjadi keputusan yang lebih terukur melalui perbandingan antar kemungkinan kelas [34], [35]. Secara matematis, Teorema Bayes diformulasikan ke dalam persamaan (1):

$$P(C_i|X) = \frac{P(X|C_i) \cdot P(C_i)}{P(X)} \quad (1)$$

Keterangan:

C_i : kelas ke- i yang menjadi target klasifikasi.

X : data atau fitur yang akan diklasifikasikan.

$P(C_i|X)$: probabilitas data X termasuk dalam kelas C_i .

$P(X|C_i)$: probabilitas kemunculan data X di kelas C_i .

$P(C_i)$: probabilitas prior dari kelas C_i .

$P(X)$: probabilitas marginal dari keseluruhan data X .

2.4 Support Vector Machine (SVM)

Support Vector Machine (SVM) adalah algoritma pembelajaran mesin yang dapat digunakan untuk tugas klasifikasi maupun regresi [36]. Prinsip kerja SVM adalah menemukan hyperplane optimal yang memisahkan data ke dalam kelas-kelas berbeda dengan margin terlebar antara titik data dan decision boundary, sehingga menghasilkan klasifikasi yang akurat [37]. Pada penerapannya, SVM Linear digunakan ketika data dapat dipisahkan secara linier. Fungsi kernel linear diformulasikan dalam persamaan (2) berikut:

$$K(x, x_i) = x \cdot x_i \quad (2)$$

Fungsi tersebut menghitung hasil kali dalam (*dot product*) antara dua vektor fitur untuk menentukan pemisahan optimal antar kelas pada ruang berdimensi linier.

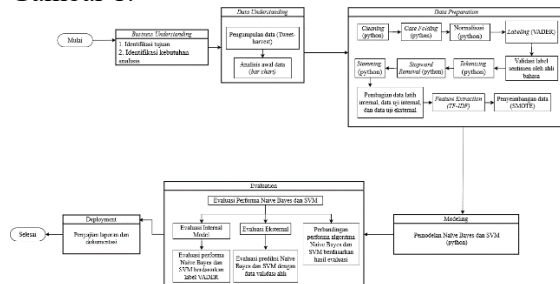
2.5 SMOTE

Synthetic Minority Oversampling Technique (SMOTE) adalah metode penyeimbangan kelas yang bekerja dengan membangkitkan data sintesis pada kelas minoritas hingga proporsinya seimbang dengan kelas mayoritas [18], [38]. Metode ini dipilih karena efektif menangani ketidakseimbangan kelas sekaligus meminimalkan risiko *overfitting* [39]. Keunggulan SMOTE terletak pada kemampuannya mempertahankan seluruh informasi dari setiap kelas karena tidak melibatkan pengurangan data, sehingga representasi kelas minoritas meningkat dan model dapat mempelajari karakteristiknya secara lebih proporsional, yang pada akhirnya berkontribusi pada peningkatan akurasi klasifikasi [40].

3. METODE PENELITIAN

Penelitian ini menerapkan metode CRISP-DM (*Cross-Industry Standard Process for Data Mining*) sebagai kerangka kerja utama karena sifatnya yang sistematis dan iteratif dalam pengolahan data [41]. Metode ini banyak digunakan dalam penelitian *data mining* [42]. termasuk analisis sentimen berbasis teks. Tahapan-tahapan ini mencakup *business understanding*, *data understanding*, *data preparation*, *modeling*, *evaluation*, dan *deployment* [43]. Dataset yang digunakan dalam penelitian ini merupakan data sekunder

yang diperoleh dari platform X pada periode 6 Januari 2025 - 31 Desember 2025 melalui proses crawling menggunakan tools Tweet-Harvest. Dataset terdiri dari 1.744 tweet yang memuat opini masyarakat terkait kasus keracunan dalam Program MBG. Rincian setiap ahaon yang dilakukan ditampilkan pada Gambar 1.



Gambar 1. Rancangan Penelitian CRISP-DM)

3.1. Business Understanding

Pada tahap ini dilakukan analisis untuk memahami konteks permasalahan, yaitu sentimen masyarakat terhadap kasus keracunan pada program MBG di platform X. Tahap ini juga bertujuan membandingkan kinerja algoritma Naïve Bayes dan *Support Vector Machine* (SVM) dalam mengklasifikasikan sentimen ke dalam kategori positif, negatif, dan netral.

3.1 Data Understanding

Tahap ini meliputi pengumpulan data melalui proses *crawling* menggunakan library Tweet-Harvest, yang kemudian dilanjutkan dengan analisis awal (*exploratory data analysis*) untuk memahami karakteristik data, seperti jumlah tweet, distribusi tweet per bulan, serta identifikasi data duplikat dan *noise*.

3.2 Data Preparation

Tahap ini meliputi proses *cleaning*, *case folding*, normalisasi, pelabelan sentimen, validasi oleh ahli bahasa, *tokenizing*, *stopword removal*, *stemming*, pembagian dataset, ekstraksi fitur, serta penyeimbangan data menggunakan SMOTE.

1. **Cleaning:** menghapus elemen yang tidak relevan, seperti URL, tanda baca yang tidak mengandung emosi, *mention*, dan angka, untuk mengurangi *noise* dalam teks. Tahapan ini mereduksi dataset dari 1.744 data menjadi 1.599 data. Perbandingan kondisi data sebelum dan setelah tahapan *cleaning* disajikan pada Gambar 2.

Sebelum Cleaning	Setelah Cleaning
OJK berencana usulkan program Makan Bergizi Gr...	OJK berencana usulkan program Makan Bergizi Gr...
@hxxguu bukan gitulah, dasar aneh. ya tuhan t...	bukan gitulah, dasar aneh. ya tuhan tolong ka...
gwa mending makan hokben tiap hari deh, drpd d...	gwa mending makan hokben tiap hari deh, drpd d...
@sofie_rr @abcdhnty @emgininamakuu Lu gak nyim...	Lu gak nyimak berita woy, di Papua sana mal...

Gambar 2. Perbedaan Sebelum dan Setelah Cleaning

2. **Case Folding:** mengubah seluruh huruf menjadi huruf kecil untuk menjaga konsistensi data. Perbandingan kondisi data sebelum dan setelah tahapan *case folding* disajikan pada Gambar 3.

Sebelum Case Folding	Setelah Case Folding
OJK berencana usulkan program Makan Bergizi Gr...	oik berencana usulkan program makan bergizi gr...
bukan gitulah, dasar aneh. ya tuhan tolong ka...	bukan gitulah, dasar aneh. ya tuhan tolong ka...
gwa mending makan hokben tiap hari deh, drpd d...	gwa mending makan hokben tiap hari deh, drpd d...
Lu gak nyimak berita woy, di Papua sana mal...	lu gak nyimak berita woy, di papua sana mal...
MBG'nversi Prabowo : Makan Bareng Gaes'nGak b...	mbg'nversi prabowo : makan bareng gaes'ngak b...
Program Makan Bergizi Gratis (MBG) Kembali Dis...	program makan bergizi gratis (mbg) kembali dis...

Gambar 3. Perbedaan Sebelum dan Setelah Case Folding

3. **Normalisasi:** mengubah kata tidak baku yang tidak sesuai dengan Kamus Besar Bahasa Indonesia (KBBI) menjadi bentuk baku tanpa mengubah makna. Perbandingan kondisi data sebelum dan setelah tahapan normalisasi disajikan pada Gambar 4.

Sebelum Normalisasi	Setelah Normalisasi
oik berencana usulkan program makan bergizi gr...	oik berencana usulkan program makan bergizi gr...
bukan gitulah, dasar aneh. ya tuhan tolong ka...	bukan gitulah, dasar aneh. ya tuhan tolong kas...
gwa mending makan hokben tiap hari deh, drpd d...	saya mending makan hokben tiap hari deh, darip...
lu gak nyimak berita woy, di papua sana mal...	kamu tidak simak berita woy, di papua sana ma...
mbg'nversi prabowo : makan bareng gaes'ngak b...	mbg versi prabowo : makan bareng gaes tidak be...
program makan bergizi gratis (mbg) kembali dis...	program makan bergizi gratis (mbg) kembali dis...

Gambar 4. Perbedaan Sebelum dan Setelah Normalisasi

4. **Labeling:** Pelabelan sentimen dilakukan menggunakan metode VADER (*Valence Aware Dictionary and sEntiment Reasoner*) untuk mengklasifikasikan teks ke dalam kategori positif, negatif, dan netral setelah proses normalisasi. Pada tahapan ini juga dilakukan penghapusan duplikasi data sehingga dataset berkurang dari 1.599 menjadi 1.533 data yang selanjutnya digunakan dalam analisis.

Hasil dari tahapan ini ditampilkan pada Gambar 5.

	full_text	sentimen
0	ojk berencana usulkan program makan bergizi gr...	netral
1	bukan gitulah dasar aneh ya tuhan tolong kasih...	negatif
2	saya mending makan hokben tiap hari deh daripa...	negatif
3	kamu tidak simak berita woy di papua sana mala...	netral
4	mbg versi prabowo makan bareng guys tidak berg...	negatif
5	program makan bergizi gratis mbg kembali disor...	negatif

Gambar 5. Hasil Labeling VADER

5. **Validasi Label Sentimen:** Validasi label sentimen dilakukan oleh ahli bahasa terhadap 20% sampel dari dataset utama, yaitu sebanyak 317 data dari total 1.533 data, guna memastikan kesesuaian label dengan konteks bahasa dan meningkatkan keakuratan anotasi data. Perbedaan label antara hasil pelabelan otomatis menggunakan VADER dan hasil validasi manual oleh ahli bahasa ditampilkan pada Gambar 6.

full_text	vader_score	vader_score_custom	sentimen	label_manual
makan mbg gratis kalo mau dapet harus setuju d...	0.3400	-0.2600	netral	negatif
badan gizi nasional bgn saat ini masih menyusu...	0.0516	-0.5484	negatif	netral
keracunan mbg lagi mau sampai kapan siswa jadi...	0.0516	-0.5484	negatif	negatif
badan gizi nasional bgn menanggapi insiden ker...	0.0516	-0.5484	negatif	negatif
pemerintah mengambil tindakan tegas menanggapi...	0.0516	-0.5484	negatif	netral
sekolah di kulon progo yang pelajarinya mengala...	0.0516	-0.5484	negatif	negatif

Gambar 6. Perbedaan Pelabelan Sentimen antara VADER dan Validasi Manual

6. **Tokenizing:** memecah kalimat menjadi token kata agar dapat diproses oleh algoritma. Hasil tahapan *tokenizing* disajikan pada Gambar 7.

tokens
[ojk, berencana, usulkan, program, makan, berg...
[bukan, gitulah, dasar, aneh, ya, tuhan, tolon...
[saya, mending, makan, hokben, tiap, hari, deh...
[kamu, tidak, simak, berita, woy, di, papua, s...
[mbg, versi, prabowo, makan, bareng, guys, tid...
[program, makan, bergizi, gratis, mbg, kembali...
[program, makan, bergizi, gratis, mbg, bermasa...

Gambar 7. Hasil Tahapan Tokenizing

7. **Stopword Removal:** proses menghapus kata umum yang tidak memiliki makna signifikan dalam analisis. Perbandingan kondisi data sebelum dan setelah tahapan *stopword removal* disajikan pada Gambar 8.

full_text	tokens
ojk berencana usulkan program makan bergizi gr...	[ojk, berencana, usulkan, program, makan, berg...
bukan gitulah dasar aneh ya tuhan tolong kasih...	[bukan, gitulah, dasar, aneh, ya, tuhan, tolon...
saya mending makan hokben tiap hari deh daripa...	[saya, mending, makan, hokben, tiap, hari, deh...
kamu tidak simak berita woy di papua sana mala...	[kamu, tidak, simak, berita, woy, di, papua, s...

Gambar 8. Perbedaan Sebelum dan Setelah Tahapan Stopword Removal

8. **Stemming:** tahapan dalam pemrosesan teks yang bertujuan untuk mengubah kata berimbuhan menjadi kata dasarnya. Perbandingan kondisi data sebelum dan setelah tahapan *stemming* disajikan pada Gambar 9.

full_text	text_stemmed
ojk berencana usulkan program makan bergizi gr...	ojk rencana usul program makan gizi gratis mbg...
bukan gitulah dasar aneh ya tuhan tolong kasih...	bukan gitulah dasar aneh ya tuhan tolong kasih...
saya mending makan hokben tiap hari deh daripa...	saya mending makan hokben tiap hari deh daripa...
kamu tidak simak berita woy di papua sana mala...	kamu tidak simak berita woy di papua sana mala...
mbg versi prabowo makan bareng guys tidak berg...	mbg versi prabowo makan bareng guys tidak gizi...

Gambar 9. Perbedaan Sebelum dan Setelah Stemming

9. **Pembagian Dataset:** Dataset dibagi menjadi data latih dan data uji internal dengan empat variasi rasio, yaitu 90:10, 80:20, 70:30, dan 60:40, guna mengevaluasi kinerja model secara komprehensif. Sebanyak 317 data hasil validasi manual oleh ahli bahasa dipisahkan dari dataset utama dan digunakan sebagai data uji eksternal untuk mengukur kemampuan generalisasi model terhadap data yang belum pernah dilihat sebelumnya.
10. **Feature Extraction:** Ekstraksi fitur dilakukan menggunakan metode TF-IDF untuk mengubah teks menjadi representasi numerik sebelum pemodelan.
11. **Penyeimbangan Data:** Penyeimbangan data dilakukan menggunakan SMOTE dengan menambahkan sampel sintetis pada kelas minoritas untuk menyeimbangkan distribusi sebelum pemodelan.

3.3 Modeling

Tahap pemodelan dilakukan dengan menerapkan algoritma Naïve Bayes dan

Support Vector Machine (SVM) pada data hasil preproses. Dataset dibagi menjadi data latih dan uji dengan rasio tertentu, kemudian model dilatih dan dievaluasi untuk mengukur kemampuan klasifikasi sentiment.

3.4 Evaluation

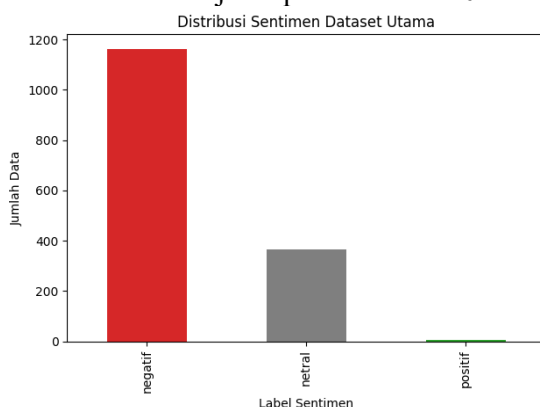
Tahap evaluasi dilakukan untuk menilai performa model melalui evaluasi internal dan eksternal. Evaluasi internal mengukur kinerja model berdasarkan label VADER, sedangkan evaluasi eksternal menggunakan data yang telah divalidasi ahli sebagai *ground truth*. Performa model diukur menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score* untuk menentukan algoritma dengan kinerja terbaik dalam analisis sentimen.

3.5 Deployment

Tahap ini mencakup penyusunan kesimpulan dan pelaporan hasil penelitian, serta pemberian rekomendasi bagi pihak terkait sebagai dasar pengambilan keputusan terkait keberlanjutan program MBG.

4. HASIL DAN PEMBAHASAN

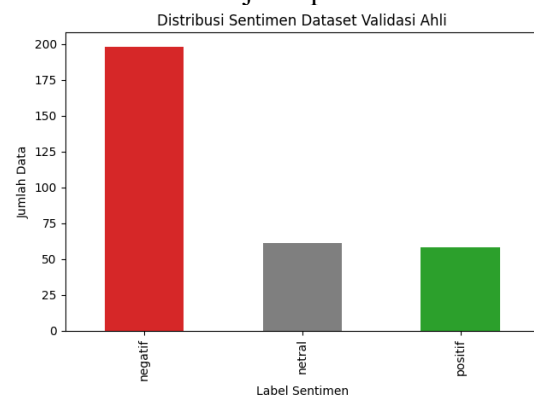
Hasil penelitian ini menyajikan analisis sentimen masyarakat terhadap kasus keracunan Program Makan Bergizi Gratis (MBG) di platform X berdasarkan 1.533 tweet yang telah melalui tahap preprocessing dan diklasifikasikan menggunakan metode VADER. Pembahasan mencakup distribusi sentimen dataset, perbandingan performa algoritma Naïve Bayes dan SVM dengan dan tanpa penerapan SMOTE pada evaluasi internal dan eksternal, serta interpretasi temuan sebagai dasar rekomendasi. Distribusi sentimen pada dataset utama disajikan pada Gambar 10.



Gambar 10. Distribusi Sentimen Dataset Utama

Gambar 10 menunjukkan distribusi sentimen pada dataset utama yang terdiri atas 1.533 tweet. Kelas negatif mendominasi dengan jumlah 1.162 data, diikuti kelas netral sebanyak 366 data, dan kelas positif yang sangat terbatas yaitu hanya 5 data. Ketidakseimbangan distribusi kelas ini mencerminkan bahwa opini masyarakat di platform X cenderung kritis terhadap kasus keracunan pada Program MBG, yang didominasi oleh kekhawatiran, kritik, dan ketidakpercayaan terhadap aspek kualitas makanan, keamanan pangan, distribusi, serta transparansi anggaran program. Kondisi ketidakseimbangan kelas (*class imbalance*) ini menjadi dasar penerapan metode SMOTE pada tahap pemodelan untuk menyeimbangkan distribusi data latih sehingga model dapat mempelajari karakteristik setiap kelas secara lebih proporsional.

Selain dataset utama, distribusi sentimen juga dianalisis pada dataset validasi yang terdiri atas 317 data berlabel manual oleh ahli bahasa dan digunakan sebagai *ground truth* pada evaluasi eksternal. Distribusi sentimen pada dataset validasi disajikan pada Gambar 11.

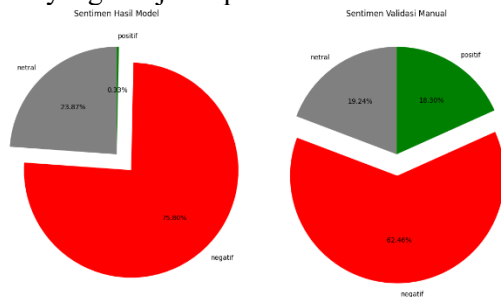


Gambar 11. Distribusi Sentimen Dataset Validasi

Gambar 11 menunjukkan distribusi sentimen pada dataset validasi ahli yang terdiri atas 317 data. Kelas negatif tetap mendominasi dengan jumlah 198 data, diikuti kelas netral sebanyak 62 data, dan kelas positif sebanyak 57 data. Dibandingkan dengan dataset utama, proporsi kelas positif dan netral pada dataset validasi jauh lebih besar, yang mengindikasikan bahwa pelabelan manual oleh ahli bahasa menghasilkan distribusi sentimen yang lebih seimbang dibandingkan pelabelan otomatis menggunakan VADER. Perbedaan distribusi ini menunjukkan adanya keterbatasan VADER dalam menangani konteks bahasa Indonesia

informal, sarkasme, dan ekspresi tidak langsung yang banyak digunakan pada platform X, sehingga sebagian tweet yang sebenarnya bernada positif atau netral cenderung terlabeli sebagai negatif oleh metode otomatis tersebut.

Perbandingan distribusi sentimen antara hasil pelabelan model dan validasi manual dapat dilihat lebih jelas melalui persentase tiap kelas yang disajikan pada Gambar 12.



Gambar 12. Perbandingan Persentase Sentimen Hasil Model dan Validasi Manual

Gambar 12 menyajikan perbandingan persentase sentimen antara hasil pelabelan model menggunakan VADER dan hasil validasi manual oleh ahli bahasa. Pada hasil model, sentimen negatif mendominasi dengan

persentase 75,80%, diikuti netral 23,87%, dan positif hanya 0,33%. Sementara itu, pada hasil validasi manual, sentimen negatif tetap mendominasi namun dengan proporsi yang lebih rendah yaitu 62,46%, diikuti netral 19,24%, dan positif 18,30%. Perbedaan yang paling mencolok terletak pada kelas positif, di mana validasi manual menghasilkan proporsi jauh lebih besar (18,30%) dibandingkan hasil model (0,33%). Hal ini mengonfirmasi bahwa VADER cenderung kurang sensitif dalam mendeteksi sentimen positif pada teks berbahasa Indonesia informal, sehingga banyak tweet bernada positif yang terlabeli secara tidak tepat oleh metode otomatis tersebut.

Selanjutnya, untuk mengetahui performa algoritma dalam mengklasifikasikan sentimen tersebut, dilakukan evaluasi model menggunakan dua tahap pengujian. Evaluasi internal dilakukan menggunakan data uji berlabel VADER dengan empat skenario rasio pembagian data (90:10, 80:20, 70:30, dan 60:40), pada kondisi dengan dan tanpa penerapan SMOTE. Rekapitulasi hasil evaluasi internal disajikan pada Tabel 1.

Tabel 1. Rekapitulasi Evaluasi Internal Performa Naive Bayes dan SVM

Algoritma	SMOTE	Rasio Pembagian Data	Akurasi	Precision	Recall	F1-score
Naïve Bayes	Pakai	90:10	76,23%	75,95%	76,23%	76,08%
		80:20	73,36%	73,82%	73,36%	73,56%
		70:30	75,62%	76,97%	75,61%	76,23%
		60:40	77,00%	77,53%	77,00%	77,26%
SVM		90:10	84,42%	83,73%	84,42%	83,89%
		80:20	81,56%	80,99%	81,56%	81,26%
		70:30	84,11%	83,29%	84,10%	83,32%
		60:40	83,16%	81,97%	83,16%	81,96%
Naïve Bayes	Tidak Pakai	90:10	77,05%	74,78%	77,05%	69,13%
		80:20	76,23%	81,49%	76,23%	66,35%
		70:30	76,16%	81,59%	76,16%	66,13%
		60:40	75,98%	81,35%	75,98%	65,81
SVM		90:10	82,79%	81,83%	82,79%	80,86%
		80:20	79,51%	77,29%	79,51%	77,21%
		70:30	81,92%	80,76%	81,92%	79,52%
		60:40	81,31%	79,94%	81,31%	78,48%

Berdasarkan data yang ada pada Tabel 1 diketahui bahwa pada kondisi dengan penerapan SMOTE, Naïve Bayes mencapai akurasi tertinggi sebesar 77,00% pada skenario 60:40, sedangkan SVM memperoleh performa terbaik pada skenario 90:10 dengan akurasi 84,42%. Pola akurasi Naïve Bayes tidak menunjukkan tren yang konsisten

antarskenario, mengindikasikan bahwa algoritma ini lebih dipengaruhi oleh distribusi data pada masing-masing skenario. Penerapan SMOTE berkontribusi pada peningkatan performa Naïve Bayes, khususnya pada nilai *F1-score* yang lebih seimbang antarkelas dibandingkan kondisi tanpa SMOTE.

Sedangkan, pada kondisi tanpa SMOTE, Naïve Bayes memperoleh akurasi tertinggi 77,05% pada skenario 90:10, namun dengan nilai *F1-score* yang lebih rendah akibat pengaruh ketidakseimbangan kelas. SVM tanpa SMOTE mencapai performa terbaik pada skenario 90:10 dengan akurasi 82,79% dan tetap menunjukkan kinerja yang lebih stabil dibandingkan Naïve Bayes pada seluruh skenario.

Secara keseluruhan, hasil evaluasi internal menunjukkan bahwa SVM memiliki kinerja klasifikasi yang lebih baik dan lebih stabil dibandingkan Naïve Bayes pada seluruh skenario pemodelan.

Selanjutnya, kemampuan generalisasi model diuji melalui evaluasi eksternal menggunakan 317 data berlabel manual oleh ahli bahasa sebagai ground truth yang tidak dilibatkan dalam proses pelatihan. Rekapitulasi hasil evaluasi eksternal disajikan pada Tabel 2.

Tabel 2. Rekapitulasi Evaluasi Eksternal Performa Naive Bayes dan SVM

Algoritma	SMOTE	Skenario Pemodelan	Akurasi	Precision	Recall	F1-score
Naïve Bayes	Pakai	90:10	75,71%	78,95%	75,71%	76,94%
		80:20	76,40%	79,16%	76,34%	77,41%
		70:30	74,76%	77,29%	74,76%	75,79%
		60:40	73,82%	75,26%	73,82%	74,43%
SVM		90:10	78,23%	77,86%	78,23%	78,04%
		80:20	78,55%	77,38%	78,54%	77,87%
		70:30	77,92%	76,07%	77,92%	76,69%
		60:40	77,92%	75,40%	77,92%	75,89%
Naïve Bayes	Tidak Pakai	90:10	76,03%	81,47%	76,03%	65,98%
		80:20	76,03%	81,47%	76,03%	65,98%
		70:30	76,03%	81,47%	76,03%	65,98%
		60:40	75,71%	57,32%	75,71%	65,24%
SVM		90:10	80,13%	78,17%	80,13%	78,36%
		80:20	78,86%	76,34%	78,86%	76,27%
		70:30	79,18%	76,78%	79,18%	75,89%
		60:40	78,55%	76,19%	78,55%	73,86%

Berdasarkan data yang terdapat pada Tabel 2 diketahui bahwa pada kondisi dengan penerapan SMOTE, Naïve Bayes mencapai akurasi tertinggi 76,40% pada skenario 80:20, sedangkan SVM memperoleh akurasi tertinggi 78,55% pada skenario yang sama. Meskipun SVM pada skenario 70:30 dan 60:40 menghasilkan nilai akurasi yang sama sebesar 77,92%, skenario 70:30 memiliki nilai *precision* dan *F1-score* yang lebih tinggi sehingga dianggap lebih optimal di antara keduanya.

Sedangkan, pada kondisi tanpa SMOTE, Naïve Bayes menghasilkan akurasi yang konsisten sebesar 76,03% pada tiga skenario (90:10, 80:20, dan 70:30), mengindikasikan bahwa variasi rasio pembagian data tidak memberikan pengaruh signifikan terhadap kemampuan generalisasi model. SVM tanpa SMOTE memperoleh performa terbaik pada skenario 90:10 dengan akurasi 80,13%, dan

tetap menunjukkan kinerja yang lebih tinggi dan stabil dibandingkan Naïve Bayes pada seluruh skenario.

Perbedaan nilai performa antara evaluasi internal dan eksternal disebabkan oleh perbedaan karakteristik data pengujian. Evaluasi internal menggunakan data yang memiliki distribusi serupa dengan data latih sehingga menghasilkan nilai performa lebih tinggi, sedangkan evaluasi eksternal menggunakan data independen yang merepresentasikan kemampuan generalisasi model terhadap data baru.

Secara keseluruhan, penerapan SMOTE terbukti meningkatkan performa Naïve Bayes pada evaluasi eksternal, sementara SVM menunjukkan performa yang stabil baik dengan maupun tanpa SMOTE. Bahkan pada beberapa skenario, SVM tanpa SMOTE menghasilkan nilai yang lebih tinggi, mengonfirmasi bahwa SVM memiliki kemampuan generalisasi yang

lebih baik dalam menangani data tidak seimbang dibandingkan Naïve Bayes.

Pada tahap *deployment*, hasil analisis sentimen yang diperoleh melalui penerapan algoritma Naïve Bayes dan *Support Vector Machine* digunakan sebagai dasar penyusunan rekomendasi yang disajikan dalam laporan penelitian. Rekomendasi tersebut dapat dijadikan bahan pertimbangan bagi pemangku kepentingan terkait dalam pelaksanaan Program Makan Bergizi Gratis (MBG). Penyusunan rekomendasi didasarkan pada dominasi sentimen dalam dataset, pola kata pada masing-masing kelas sentimen, serta hasil evaluasi performa model klasifikasi.

Hasil analisis menunjukkan bahwa sentimen negatif didominasi oleh kata-kata yang berkaitan dengan isu keamanan pangan, kualitas makanan, pengawasan pelaksanaan program, distribusi MBG pada hari libur sekolah yang dinilai tidak sesuai pedoman, penggunaan anggaran, serta kekhawatiran terhadap potensi penyalahgunaan anggaran dalam pelaksanaan program. Secara keseluruhan, sentimen negatif pada dataset penelitian ini memuat keluhan, kritik, dan kekhawatiran masyarakat terhadap pelaksanaan program, terutama yang berkaitan dengan kualitas dan keamanan makanan, ketepatan distribusi, serta transparansi pengelolaan anggaran. Beberapa contoh data yang terklasifikasi ke dalam kelas sentimen negatif disajikan sebagai berikut.

"program makan bergizi gratis mbg bermasalah temuan makanan tidak layak konsumsi dan kasus keracunan anak". "bukan gitulah dasar aneh ya tuhan tolong kasih paham lah ini orang kalo tidak ada mbg yang makan anggaran berapa triliun itu keracunan ini gaakan terjadi anggaran pendidikan tidak perlu dipangkas kesimpulannya stop mbg liat sudah ada yang stop tuh". "menurutmu program makan bergizi gratis apakah sudah optimal mengingat banyak kasus keracunan dan distribusi mbg tetap dilakukan saat anak sekolah sedang libur, selain itu banyak pejabat yang memiliki dapur mbg yang mengakibatkan dana mbg dikorupsi".

Data tersebut mencerminkan persepsi negatif masyarakat terhadap kualitas makanan, sistem distribusi, kebijakan anggaran, serta transparansi pelaksanaan Program MBG. Temuan ini mengindikasikan bahwa aspek

keamanan pangan, kualitas makanan, ketepatan distribusi, dan transparansi pengelolaan anggaran merupakan faktor yang paling dominan dalam membentuk sentimen negatif pada data penelitian. Dominasi sentimen negatif tersebut menggambarkan bahwa pelaksanaan Program Makan Bergizi Gratis masih menghadapi berbagai kendala yang memerlukan evaluasi dan perbaikan secara menyeluruh.

Berdasarkan temuan penelitian yang menunjukkan dominasi sentimen negatif terhadap pelaksanaan Program Makan Bergizi Gratis, penelitian ini merekomendasikan agar pemangku kepentingan terkait melakukan evaluasi dan peninjauan ulang secara komprehensif terhadap pelaksanaan program. Evaluasi tersebut perlu difokuskan pada aspek keamanan pangan, kualitas makanan, sistem distribusi, serta transparansi pengelolaan anggaran yang teridentifikasi sebagai faktor utama munculnya sentimen negatif masyarakat. Selain itu, pihak pelaksana program disarankan untuk mempertimbangkan penghentian sementara pelaksanaan program di beberapa wilayah hingga permasalahan yang berkaitan dengan kualitas makanan, keamanan pangan, dan sistem pengawasan dapat diselesaikan sesuai dengan standar yang ditetapkan. Penghentian sementara ini bertujuan untuk meminimalkan risiko kejadian keracunan makanan serta mencegah dampak yang lebih luas terhadap penerima manfaat program. Setelah dilakukan perbaikan pada sistem produksi, distribusi, pengawasan, dan pengendalian mutu, program dapat dilaksanakan kembali dengan mekanisme pengawasan yang lebih ketat serta sistem pengendalian kualitas yang lebih terstandarisasi. Perbaikan pada tahap distribusi dapat dilakukan melalui penggunaan wadah makanan yang tertutup dan higienis, penjaminan waktu distribusi yang tidak melampaui batas kelayakan konsumsi, serta pemantauan suhu dan kebersihan selama proses pengiriman berlangsung. Perbaikan pada tahap penyajian dapat dilakukan dengan memastikan makanan disajikan dalam kondisi higienis dan bebas dari kontaminasi, serta dilakukannya pemeriksaan kondisi makanan sebelum dikonsumsi oleh penerima manfaat. Dengan demikian, pelaksanaan Program Makan Bergizi Gratis diharapkan dapat berjalan lebih optimal,

aman, dan sesuai dengan tujuan program yang telah ditetapkan.

Penelitian ini juga merekomendasikan agar pemangku kepentingan terkait meningkatkan sistem pengawasan pelaksanaan program serta transparansi dalam pengelolaan anggaran. Peningkatan sistem pengawasan dapat dilakukan melalui pelaksanaan inspeksi rutin ke Satuan Pelayanan Pemenuhan Gizi (SPPG), penyediaan saluran pengaduan masyarakat sebagai media pelaporan terhadap permasalahan yang ditemukan dalam pelaksanaan program, serta pelaksanaan evaluasi berkala terhadap kinerja pelaksana program. Peningkatan transparansi anggaran dapat diwujudkan melalui publikasi laporan penggunaan anggaran yang dapat diakses secara publik, penyampaian rincian anggaran produksi dan distribusi makanan, serta pelaporan hasil evaluasi program secara berkala kepada masyarakat. Upaya peningkatan pengawasan dan transparansi tersebut diharapkan dapat memperkuat akuntabilitas pelaksanaan program serta mereduksi munculnya persepsi negatif masyarakat terhadap pelaksanaan Program Makan Bergizi Gratis.

5. KESIMPULAN

- a. Analisis sentimen terhadap 1.533 tweet di platform X menunjukkan dominasi sentimen negatif (1.162 data), diikuti netral (366 data), dan positif (5 data), yang mengindikasikan persepsi publik yang cenderung kritis terhadap pelaksanaan Program MBG. Opini negatif didominasi oleh kekhawatiran terhadap kualitas makanan, keamanan pangan, sistem distribusi, dan transparansi anggaran program. Penerapan metodologi CRISP-DM menghasilkan alur analisis yang sistematis dan terstruktur sehingga pola persepsi masyarakat dapat teridentifikasi secara komprehensif.
- b. Perbandingan performa algoritma menunjukkan bahwa SVM memiliki kinerja lebih baik dan lebih stabil dibandingkan Naïve Bayes pada seluruh skenario evaluasi. Pada evaluasi internal, SVM dengan SMOTE mencapai akurasi tertinggi sebesar 84,42%, sedangkan Naïve Bayes dengan SMOTE mencapai 77,00%. Pada evaluasi eksternal, SVM tanpa

SMOTE memperoleh akurasi tertinggi sebesar 80,13%, sementara Naïve Bayes dengan SMOTE mencapai 76,40%. Penerapan SMOTE terbukti lebih signifikan meningkatkan performa Naïve Bayes, sedangkan SVM tetap menunjukkan performa yang konsisten baik dengan maupun tanpa SMOTE. Dengan demikian, SVM direkomendasikan sebagai algoritma yang lebih optimal untuk klasifikasi sentimen pada penelitian ini.

DAFTAR PUSTAKA

- [1] A. Kiftiyah, F. A. Palestina, F. U. Abshar, and K. Rofiah, "Program Makan Bergizi Gratis (MBG) dalam Perspektif Keadilan Sosial dan Dinamika Sosial – Politik," *Pancasila: Jurnal Keindonesiaan*, vol. 5, no. 1, pp. 101–112, Apr. 2025, doi: 10.52738/pjk.v5i1.726.
- [2] A. A. Merlinda and Y. Yusuf, "Analisis Program Makan Gratis Prabowo Subianto Terhadap Strategi Peningkatan Motivasi Belajar Siswa di Sekolah Tinjauan dari Perspektif Sosiologi Pendidikan," *Ranah Research: Journal of Multidisciplinary Research and Development*, vol. 2, no. 7, pp. 1364–1373, Jan. 2025, doi: 10.38035/rj.v7i2.1360.
- [3] M. R. Hidayatullah, "Adik Prabowo Ungkap Asal-usul Ide Makan Bergizi Gratis, Tercetus Sejak 2006," *Bisnis.com*, Oct. 07, 2024.
- [4] R. Fauziah, "Analisis Jaringan Komunikasi (Social Network Analysis) Program MBG Indonesia," *Journal of Science and Social Research*, vol. 8, no. 3, pp. 3556–3561, Aug. 2025, doi: 10.54314/jssr.v8i3.4007.
- [5] M. A. Rojabi, *Mitra BGN: Syarat, Cara Daftar Online, dan Kunci Lolos Verifikasi Program Makan Bergizi Gratis*. Afdan Rojabi Publisher, 2025. doi: <https://books.google.co.id/books?id=cCGKEQAAQBAJ>.
- [6] M. Rayhan and Zulham, "Implementasi Jaminan Halal pada Program Makan Bergizi Gratis (MBG) di Kecamatan Medan Timur," *AL-SULTHANIYAH*, vol. 14, no. 2, pp. 370–380, Aug. 2025, doi: 10.37567/al-sulthaniyah.v14i2.4154.
- [7] N. P. Nugroho, "JPPI: Jumlah Korban Keracunan MBG Kembali Naik, Mencapai 13.168 Orang," *Tempo.co*, Oct. 20, 2025. Accessed: Oct. 29, 2025. [Online]. Available: <https://www.tempo.co/politik/jppi-jumlah-korban-keracunan-mbg-kembali-naik-mencapai-13-168-orang-2081242>
- [8] Redaksi UMJ, "Kasus Keracunan Massal MBG, Pakar UMJ Soroti Pengawasan Gizi dan

- Keamanan Pangan,” Universitas Muhammadiyah Jakarta-rubrik Opini. Accessed: Oct. 29, 2025. [Online]. Available: <https://umj.ac.id/opini-1/kasus-keracunan-massal-mbg-pakar-umj-soroti-pengawasan-gizi-dan-keamanan-pangan/>
- [9] R. Kurniawan, “Kondisi Korban Keracunan MBG di Cibodas Bandung Barat, Mulai Dipulangkan, Tinggal 10 Orang di Posko,” *TribunJabar.id*, Oct. 29, 2025. Accessed: Oct. 29, 2025. [Online]. Available: <https://jabar.tribunnews.com/metro-bandung/1152833/kondisi-korban-keracunan-mbg-di-cibodas-bandung-barat-mulai-dipulangkan-tinggal-10-orang-di-posko>
- [10] A. I. K. Pau, “Indonesia Tempati Peringkat Keempat Pengguna X Terbanyak Dunia,” RRI (Radio Republik Indonesia). Accessed: Dec. 11, 2025. [Online]. Available: <https://rri.co.id/hiburan/1819337/indonesia-tempati-peringkat-keempat-pengguna-x-terbanyak-dunia>
- [11] E. R. Lidinillah, T. Rohana, and A. R. Juwita, “Analisis sentimen twitter terhadap steam menggunakan algoritma logistic regression dan support vector machine,” *TEKNOSAINS: Jurnal Sains, Teknologi dan Informatika*, vol. 10, no. 2, pp. 154–164, Jul. 2023, doi: 10.37373/tekno.v10i2.440.
- [12] R. Noviana and I. Rasal, “Penerapan Algoritma Naive Bayes dan SVM untuk Analisis Sentimen Boy Band BTS pada Media Sosial Twitter,” *JTS (Jurnal Teknik dan Science)*, vol. 2, no. 2, pp. 51–60, Jun. 2023, doi: 10.56127/jts.v2i2.791.
- [13] B. K. Widodo, N. H. Matondang, and D. S. Prasvita, “Penerapan Algoritma Naive Bayes Untuk Analisis Sentimen Penggunaan Aplikasi Jobstreet,” *Techno.COM*, vol. 21, no. 3, pp. 523–533, Aug. 2022, doi: 10.33633/tc.v21i3.6361.
- [14] F. Wulandari, E. Haerani, M. Fikry, and E. Budianita, “Analisis Sentimen Larangan Penggunaan Obat Sirup Menggunakan Algoritma Naive Bayes Classifier,” *Jurnal CoSciTech (Computer Science and Information Technology)*, vol. 4, no. 1, pp. 88–96, Apr. 2023, doi: 10.37859/coscitech.v4i1.4781.
- [15] D. Fernando and R. G. Guntara, “Model Klasifikasi Penyebab Turnover Karyawan Menggunakan Kerangka Kerja CRISP-DM,” *J-INTECH (Journal of Information and Technology)*, vol. 12, no. 2, pp. 383–392, Dec. 2024, doi: 10.32664/j-intech.v12i02.1502.
- [16] D. A. Kusuma and A. D. P. Wicaksono, “Analisis Klastering Dampak Lingkungan Berdasarkan Konsumsi Energi Perusahaan Berbasis Industri 4.0 Menggunakan Metode CRISP-DM,” *Positif: Jurnal Sistem dan Teknologi Informasi*, vol. 9, no. 2, pp. 1–6, Dec. 2023, doi: 10.31961/positif.v9i2.2050.
- [17] R. Winurputra and D. E. Ratnawati, “Peramalan Penjualan Produk Menggunakan Extreme Gradient Boosting (XGBoost) dan Kerangka Kerja CRISP-DM untuk Pengoptimalan Manajemen Persediaan (Studi Kasus: UB Mart),” *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 12, no. 2, pp. 417–428, Apr. 2025, doi: 10.25126/jtiik.2025129451.
- [18] N. N. Sholihah and A. Hermawan, “Implementation of Random Forest and SMOTE Methods for Economic Status Classification in Cirebon City,” *Jurnal Teknik Informatika (Jutif)*, vol. 4, no. 6, pp. 1387–1397, Dec. 2023, doi: 10.52436/1.jutif.2023.4.6.1135.
- [19] A. F. Riany and G. Testiana, “Penerapan Data Mining untuk Klasifikasi Penyakit Stroke Menggunakan Algoritma Naïve Bayes,” *Jurnal SAINTEKOM*, vol. 13, no. 1, pp. 42–54, Mar. 2023, doi: 10.33020/saintekom.v13i1.352.
- [20] M. K. Darmawan and M. Ziveria, “Analisis Kebutuhan UI/UX Mahasiswa Kalbis Institute pada Studi Kasus Learning Management System (LMS) LEAPS Kalbis Institute,” *KALBISIANA: Jurnal Mahasiswa Institut Teknologi dan Bisnis Kalbis*, vol. 9, no. 2, pp. 401–413, Jun. 2023, doi: 2998-1676.
- [21] N. Firdausy, I. Yuadi, and I. Puspitasari, “Analisis Sentimen Evaluasi Reaksi E-Learning Menggunakan Algoritma Naïve Bayes Support Vector Machine dan Deep Learning,” *Techno.COM*, vol. 22, no. 3, pp. 677–689, Aug. 2023, doi: 10.33633/tc.v22i3.8160.
- [22] B. A. Maulana, M. J. Fahmi, A. M. Imran, and N. Hidayati, “Analisis Sentimen Terhadap Aplikasi Pluang Menggunakan Algoritma Naive Bayes dan Support Vector Machine (SVM),” *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, vol. 4, no. 2, pp. 375–384, Feb. 2024, doi: 10.57152/malcom.v4i2.1206.
- [23] M. I. Petiwi, A. Triayudi, and I. D. Sholihati, “Analisis Sentimen Gofood Berdasarkan Twitter Menggunakan Metode Naïve Bayes dan Support Vector Machine,” *JURNAL MEDIA INFORMATIKA BUDIDARMA*, vol. 6, no. 1, p. 542, Jan. 2022, doi: 10.30865/mib.v6i1.3530.
- [24] M. M. Laia, F. N. Hasan, and A. Y. Kuntoro, “Sentimen Program Makan Gratis Pada Platform X Menggunakan Algoritma Naïve Bayes,” *Jurnal Ilmiah Informatika (JIF)*, vol. 13, no. 2, pp. 259–264, Sep. 2025, doi: 10.33884/jif.v13i02.10427.
- [25] F. Fatkhurrohman, B. I. Nugroho, and N. Fadillah, “Analisis Sentimen Program Makan Bergizi Gratis Pemerintah RI Melalui Twitter Menggunakan Metode SVM,” *RIGGS: Journal*

- of Artificial Intelligence and Digital Business*, vol. 4, no. 3, pp. 3906–3917, Aug. 2025, doi: 10.31004/riggs.v4i3.2533.
- [26] A. El Alaoui El Fels, L. Mandi, A. Kammoun, N. Ouazzani, O. Monga, and M. L. Hbid, “Artificial Intelligence and Wastewater Treatment: A Global Scientific Perspective through Text Mining,” *Water (Switzerland)*, vol. 15, no. 19, Oct. 2023, doi: 10.3390/w15193487.
- [27] T. Jo, *Text Mining: Concepts, Implementation, and Big Data Challenge*, 2nd ed. Cham: Springer Cham, 2024. doi: 10.1007/978-3-319-91815-0.
- [28] C. J. E. Munthe, N. A. Hasibuan, and H. Hutabarat, “Penerapan Algoritma Text Mining Dan TF-RF Dalam Menentukan Promo Produk Pada Marketplace,” *RESOLUSI: Rekayasa Teknik Informatika dan Informasi*, vol. 2, no. 3, pp. 110–115, Jan. 2022, doi: 10.30865/resolusi.v2i3.309.
- [29] S. Sukriadi, I. Ismail, and A. M. Andzar, “Penerapan Text Mining Dalam Klasifikasi Judul Skripsi Yang Diusulkan Mahasiswa Menggunakan Metode Naïve Bayes,” *Jurnal Ilmiah Sistem Informasi dan Teknik Informatika (JISTI)*, vol. 6, no. 2, pp. 184–196, Oct. 2023, doi: 10.57093/jisti.v6i2.174.
- [30] A. Safira and F. N. Hasan, “Analisis Sentimen Masyarakat Terhadap Paylater Menggunakan Metode Naive Bayes Classifier,” *ZONASI: Jurnal Sistem Informasi*, vol. 5, no. 1, pp. 59–70, Jan. 2023, doi: 10.31849/zn.v5i1.12856.
- [31] I. P. D. W. Darmawan, G. A. Pradnyana, and I. B. N. Pascima, “Optimasi Parameter *Support Vector Machine* Dengan Algoritma Genetika Untuk Analisis Sentimen Pada Media Sosial Instagram,” *SINTECH JOURNAL (Science and Information Technology)*, vol. 6, no. 1, pp. 58–67, Apr. 2023, doi: 10.31598/sintechjournal.v6i1.1245.
- [32] N. Fitriyah, B. Warsito, and D. A. I. Maruddani, “Analisis Sentimen Gojek pada Media Sosial Twitter dengan Klasifikasi *Support Vector Machine* (SVM),” *JURNAL GAUSSIAN*, vol. 9, no. 3, pp. 376–390, Aug. 2020, doi: 10.14710/j.gauss.9.3.376-390.
- [33] B. Z. Ramadhan, I. Riza, and I. Maulana, “Analisis Sentimen Ulasan Pada Aplikasi E-Commerce Dengan Menggunakan Algoritma Naïve Bayes,” *Journal of Applied Informatics and Computing (JAIC)*, vol. 6, no. 2, pp. 220–225, Dec. 2022, doi: 10.30871/jaic.v6i2.4725.
- [34] D. Gunawan and V. N. Andika, “Implementasi Teorema Bayes Pada Sistem Informasi Posyandu Dalam Mendeteksi Stunting Pada Balita,” *Jurnal Sistem Komputer dan Informatika (JSON)*, vol. 4, no. 4, pp. 692–698, Jun. 2023, doi: 10.30865/json.v4i4.6146.
- [35] M. Zunaidi, U. F. S. S. Pane, and A. H. Nasyuha, “Analisis Teorema Bayes Dalam Mendiagnosa Penyakit Tanaman Pisang,” *Jurnal Media Informatika Budidarma*, vol. 5, no. 4, pp. 1302–1308, Oct. 2021, doi: 10.30865/mib.v5i4.3225.
- [36] H. Eldo, A. Ayuliana, D. Suryadi, G. Chrisnawati, and L. Judijanto, “Penggunaan Algoritma *Support Vector Machine* (SVM) Untuk Deteksi Penipuan pada Transaksi Online,” *Jurnal Minfo Polgan*, vol. 13, no. 2, pp. 1627–1632, Oct. 2024, doi: 10.33395/jmp.v13i2.14186.
- [37] F. Abdusyukur, “Penerapan Algoritma *Support Vector Machine* (SVM) untuk Klasifikasi Pencemaran Nama Baik di Media Sosial Twitter,” *KOMPUTA : Jurnal Ilmiah Komputer dan Informatika*, vol. 12, no. 1, pp. 73–82, Apr. 2023, doi: 10.34010/komputa.v12i1.9418.
- [38] F. M. A. Sofyan, N. Sulistiyowati, and A. Voutama, “Analisis Sentimen Terhadap Respon Perubahan Nama Twitter Menjadi ‘X’ Menggunakan Metode Naive Bayes Classifier,” *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 8, no. 5, pp. 10987–10994, Oct. 2024, doi: doi.org/10.36040/jati.v8i5.10720.
- [39] L. Qadrini, H. Hikmah, and M. Megasari, “Oversampling, Undersampling, Smote SVM dan Random Forest pada Klasifikasi Penerima Bidikmisi Sejava Timur Tahun 2017,” *Journal of Computer System and Informatics (JoSYC)*, vol. 3, no. 4, pp. 386–391, Sep. 2022, doi: 10.47065/josyc.v3i4.2154.
- [40] K. Pramayasa, I. M. D. Maysanjaya, and I. G. A. A. D. Indradewi, “Analisis Sentimen Program Mbkm Pada Media Sosial Twitter Menggunakan KNN Dan SMOTE,” *Science and Information Technology (SINTECH)*, vol. 6, no. 2, pp. 89–98, Aug. 2023, doi: 10.31598/sintechjournal.v6i2.1372.
- [41] F. Rahmadayanti and V. Yolanda, “Penerapan Algoritma C4.5 untuk Prediksi Penerimaan Beasiswa di SMA N.01 Tanjung Sakti Pumu,” *Jurnal Teknologi Informasi Mura*, vol. 16, no. 2, pp. 163–170, Dec. 2024, doi: 10.32767/jti.v16i2.2501.
- [42] D. Kurniawan and M. Yasir, “Optimization Sentiment Analysis Using CRISP-DM and Naive Bayes Methods Implemented on Social Media,” *Cyberspace: Jurnal Pendidikan Teknologi informasi*, vol. 6, no. 2, pp. 74–85, Oct. 2022, doi: 10.22373/cj.v6i2.12793.
- [43] A. E. Noviyanti and S. Juanita, “Rekomendasi Paket Pakaian Berdasarkan Pola Penjualan Menggunakan Algoritma Apriori,” *Jurnal*

SISFOTENIKA, vol. 14, no. 2, pp. 129–139, Jul.
2024, doi: 10.30700/sisfotenika.v14i2.424.