

Perbandingan SVM dan Naive Bayes pada Analisis Sentimen MBG di Platform TikTok dan Instagram

Zahra Prisdian Tiararosa¹, Ibnu Sani Wijaya², Eni Rohaini³

^{1, 2, 3} Fakultas Ilmu Komputer, Program Studi Teknik Informatika, Universitas Dinamika Bangsa, Jambi, Indonesia

Email: ¹zhprisdian04@gmail.com, ²ibnusw17@unama.ac.id, ³enirohaini0104@gmail.com

Email Penulis Korespondensi: zhprisdian04@gmail.com

Artikel Info :

Artikel History :

Submitted : 13-02-2026

Accepted : 15-04-2026

Published : 30-04-2026

Kata Kunci:

Analisis_Sentimen,

Naive_Bayes,

Media_Sosial, MBG,

Machine_Learning..

Abstrak—Perkembangan media sosial yang pesat menjadikan platform seperti TikTok dan Instagram sebagai sarana utama masyarakat dalam menyampaikan opini terhadap berbagai isu publik, termasuk program Makan Bergizi Gratis (MBG). Banyaknya komentar yang muncul menyebabkan analisis opini secara manual menjadi tidak efisien, sehingga diperlukan pendekatan analisis sentimen berbasis machine learning. Penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap program MBG serta membandingkan kinerja metode Support Vector Machine (SVM) dan Naive Bayes (NB) dalam mengklasifikasikan sentimen positif, negatif, dan netral. Data yang digunakan berupa 6001 komentar masyarakat yang dikumpulkan dari platform TikTok dan Instagram melalui proses web scraping menggunakan Apify dan Python. Tahapan penelitian meliputi pengumpulan data, preprocessing teks, pembentukan fitur, pelatihan model, dan evaluasi menggunakan confusion matrix dengan metrik akurasi, presisi, recall, dan F1-score. Hasil penelitian diharapkan dapat menunjukkan perbandingan performa antara metode SVM dan Naive Bayes serta memberikan gambaran umum mengenai kecenderungan opini masyarakat terhadap program MBG di media sosial.

Abstract—The rapid growth of social media has made platforms such as TikTok and Instagram primary channels for the public to express opinions on various public issues, including the Free Nutritious Meal Program (MBG). The large volume of user-generated comments makes manual analysis inefficient, thereby requiring sentiment analysis approaches based on machine learning. This study aims to analyze public sentiment toward the MBG program and to compare the performance of Support Vector Machine (SVM) and Naive Bayes (NB) methods in classifying sentiments into positive, negative, and neutral categories. The dataset consists of 6,001 comments collected from TikTok and Instagram through a web scraping process using Apify and Python. The research stages include data collection, text preprocessing, feature extraction, model training, and evaluation using a confusion matrix with accuracy, precision, recall, and F1-score metrics. The results of this study are expected to provide a comparison of the performance of SVM and Naive Bayes algorithms as well as an overview of public opinion trends regarding the MBG program on social media.

Keywords:

Sentiment_Analysis,

Naive_Bayes,

Social_Media, MBG,

Machine Learning.

1. PENDAHULUAN

Di era digital saat ini, perkembangan teknologi informasi dan komunikasi telah membawa perubahan mendasar dalam berbagai aspek kehidupan masyarakat [1]. Kemajuan ini mendorong semakin banyak individu untuk mengekspresikan pendapat, berbagi informasi, dan berinteraksi secara daring melalui berbagai platform media sosial [2]. Penggunaan media sosial tidak lagi sekadar sarana hiburan, melainkan telah berkembang menjadi ruang publik yang aktif digunakan masyarakat untuk merespons berbagai isu sosial, ekonomi, maupun kebijakan pemerintah, sehingga opini yang muncul di dalamnya menjadi indikator penting dalam mengukur persepsi dan kepuasan publik.

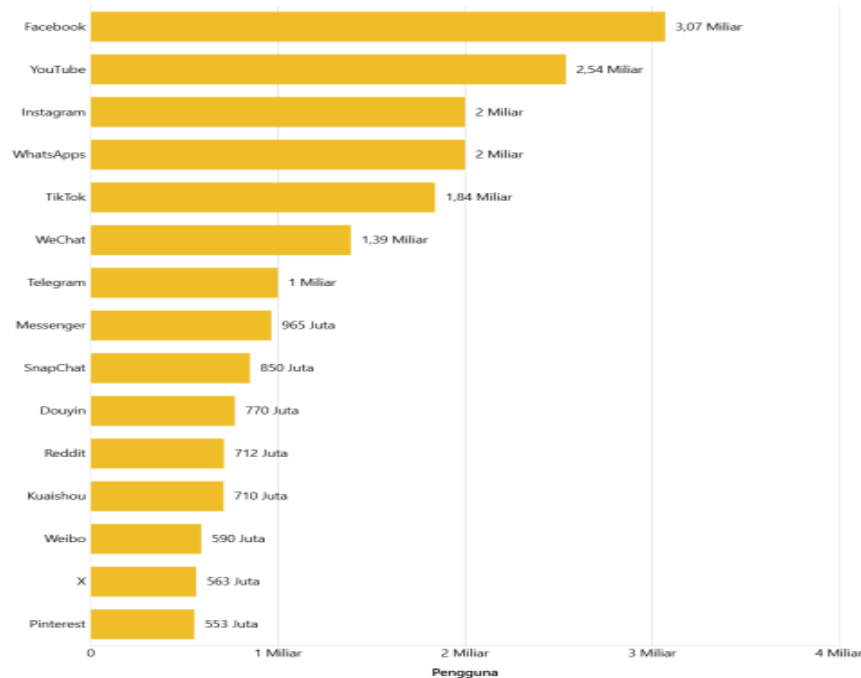
Salah satu dampak nyata dari kemajuan teknologi tersebut adalah meningkatnya jumlah pengguna media sosial di seluruh dunia. Berdasarkan laporan We Are Social, terdapat sekitar 5,31 miliar pengguna media sosial secara global pada April 2025, meningkat sebesar 4,7% dibandingkan tahun sebelumnya [3]. Data ini memperlihatkan betapa signifikannya peran media sosial dalam kehidupan masyarakat modern. Di antara berbagai platform yang ada, TikTok dan Instagram merupakan dua platform yang paling banyak digunakan di Indonesia, sehingga menjadikan kedua platform tersebut sebagai objek kajian utama dalam penelitian ini.

Ini Media Sosial Terpopuler di Dunia pada Pertengahan 2025



Nabilah Muhamad - 2025/05/20, 11.50 AM

15 Media Sosial Terpopuler Global (April 2025)*

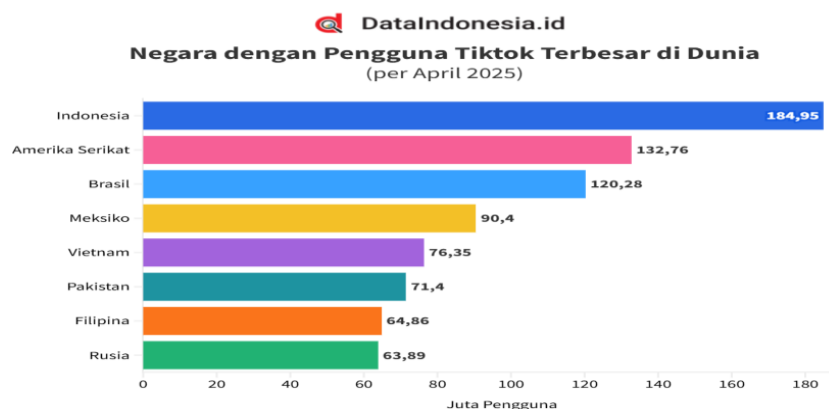


Gambar 1. Media Sosial Terpopuler Di Dunia Tahun 2025 [3]

Berdasarkan laporan tersebut, Facebook menempati posisi teratas dengan 3,07 miliar pengguna aktif, disusul oleh YouTube, Instagram, dan TikTok yang juga mencatatkan jumlah pengguna yang sangat besar [3]. Di Indonesia, TikTok dan Instagram menjadi dua platform media sosial dengan penetrasi tertinggi, menjadikan keduanya sebagai kanal utama penyebaran informasi dan opini publik di kalangan masyarakat Indonesia.

Penelitian ini berfokus pada platform TikTok dan Instagram sebagai objek kajian utama karena keduanya merupakan platform dengan jumlah pengguna terbesar di Indonesia serta memiliki pengaruh yang signifikan terhadap perilaku dan opini masyarakat. TikTok adalah aplikasi berbagi video pendek yang telah berkembang menjadi salah satu platform media sosial paling populer di dunia, digunakan sebagai sarana ekspresi diri, hiburan, dan interaksi sosial oleh ratusan juta pengguna [4].

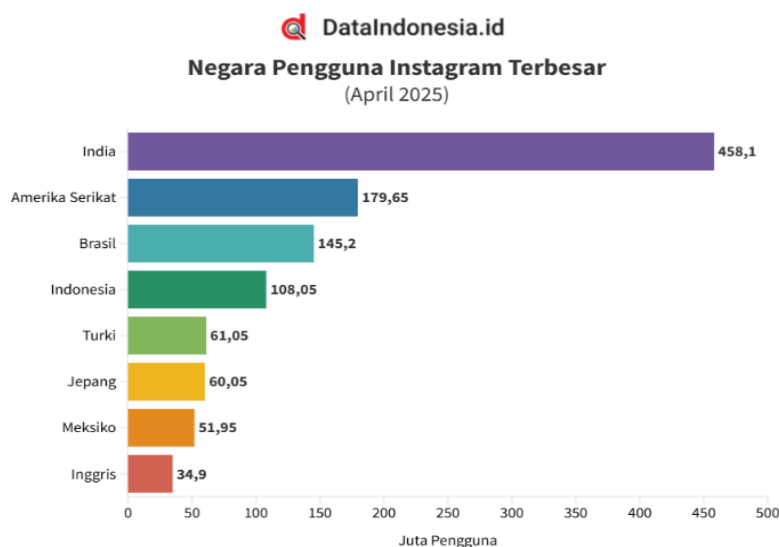
Berdasarkan data dari DataIndonesia.Id, Indonesia menempati posisi pertama sebagai negara dengan jumlah pengguna TikTok terbanyak di dunia, yaitu sebanyak 184,95 juta pengguna aktif per April 2025 [5]. Tingginya penetrasi TikTok di Indonesia menegaskan bahwa platform ini merupakan salah satu ruang digital yang paling strategis untuk menyampaikan dan memantau opini publik di Indonesia.



Gambar 2. Negara Dengan Pengguna TikTok Terbesar Di Dunia [5]

Selain TikTok, Instagram juga menjadi platform media sosial yang sangat populer di Indonesia. Instagram merupakan platform berbagi foto dan video yang digunakan masyarakat untuk merepresentasikan diri, berinteraksi, dan mencari informasi [6]. Platform ini terus diminati karena fitur visualnya yang menarik serta kemampuannya dalam mendukung aktivitas berbagi konten, sehingga menjadikannya sebagai salah satu sarana utama penyebaran opini publik. Berikut ini disajikan data pengguna Instagram terbesar di dunia.

Berdasarkan laporan We Are Social dan Meltwater, jumlah pengguna Instagram di dunia mencapai 1,84 miliar orang pada April 2025, meningkat 5,8% dari periode sebelumnya [7]. Indonesia menempati posisi keempat dengan 108,05 juta pengguna aktif, setelah India, Amerika Serikat, dan Brasil. Angka ini mempertegas bahwa Indonesia merupakan salah satu negara dengan basis pengguna Instagram terbesar di dunia, sekaligus menjadikan Instagram sebagai platform yang sangat relevan untuk menganalisis opini publik masyarakat Indonesia.



Gambar 3. Negara Pengguna Instagram Terbesar Di Dunia [7]

Melihat tingginya aktivitas pengguna di TikTok dan Instagram, khususnya di Indonesia, tidak mengherankan apabila kedua platform ini kerap dimanfaatkan sebagai sarana penyebaran informasi dan opini publik terhadap berbagai isu yang sedang berkembang. Di Indonesia, salah satu topik yang paling ramai diperbincangkan di media sosial belakangan ini adalah program Makan Bergizi Gratis (MBG), yang merupakan program unggulan pemerintahan Presiden Prabowo Subianto melalui Badan Gizi Nasional (BGN) [8]. Program ini dirancang untuk mengatasi permasalahan gizi masyarakat, terutama masalah stunting, sekaligus mendorong pertumbuhan ekonomi daerah dengan memprioritaskan penggunaan bahan pangan lokal [9].

Program MBG menasar sekitar 82,9 juta penerima manfaat, meliputi ibu hamil, balita, serta murid TK, SD, SMP, dan sederajat, dengan tujuan membangun generasi bangsa yang sehat, cerdas, dan produktif demi mewujudkan visi Indonesia Emas 2045 [10]. Namun, sebagaimana program-program besar lainnya, MBG menuai beragam respons dari masyarakat [12]. Banyak pengguna TikTok dan Instagram memberikan komentar dan opini—baik berupa dukungan, kritik, maupun tanggapan yang bersifat netral—terhadap program ini. Volume komentar yang sangat besar di kedua platform tersebut menjadikan analisis opini secara manual menjadi tidak efisien, tidak terstruktur, dan tidak dapat dilakukan secara komprehensif.

Untuk mengatasi tantangan tersebut, diperlukan pendekatan analisis sentimen berbasis machine learning yang mampu mengklasifikasikan opini masyarakat secara otomatis dan akurat [13]. Analisis sentimen adalah proses mengidentifikasi dan mengkategorikan opini yang dinyatakan dalam teks untuk menentukan apakah sikap penulis bersifat positif, negatif, atau netral terhadap suatu topik tertentu [15]. Hasil analisis sentimen terhadap komentar di TikTok dan Instagram dapat memberikan gambaran nyata mengenai persepsi masyarakat terhadap program MBG, sehingga dapat menjadi masukan yang berharga bagi pemerintah dalam mengevaluasi dan meningkatkan pelaksanaan program tersebut.

Beberapa penelitian sebelumnya telah mengkaji analisis sentimen terkait program MBG dan kebijakan publik serupa menggunakan berbagai metode. Rahmatullah et al. [11] menerapkan algoritma Naive Bayes untuk menganalisis sentimen terhadap MBG dan memperoleh hasil klasifikasi yang cukup baik. Noviyanti et al. [13] menggunakan Multinomial Naive Bayes dengan Chi-Square pada data komentar YouTube dan mencapai akurasi 83,48%. Laia et al. [14] menganalisis sentimen MBG pada Platform X menggunakan Naive Bayes dengan IndoBERT untuk pelabelan dan memperoleh akurasi 86,46%. Sementara itu, beberapa penelitian perbandingan

SVM dan Naive Bayes pada konteks lain, seperti analisis sentimen ulasan produk dan komentar politik, secara konsisten menunjukkan bahwa SVM menghasilkan akurasi yang lebih tinggi dibandingkan Naive Bayes [16][17].

Namun demikian, penelitian-penelitian tersebut umumnya hanya menggunakan satu algoritma atau hanya berfokus pada satu platform media sosial, seperti Twitter atau Platform X. Hingga saat ini, belum ada penelitian yang secara khusus membandingkan kinerja metode Support Vector Machine (SVM) dan Naive Bayes secara bersamaan pada dua platform berbeda, yaitu TikTok dan Instagram, khususnya dalam konteks analisis sentimen terhadap program Makan Bergizi Gratis (MBG). Kesenjangan inilah yang menjadi landasan dan motivasi utama penelitian ini, sekaligus menunjukkan kontribusi dan kebaruan yang ditawarkan dibandingkan penelitian sebelumnya.

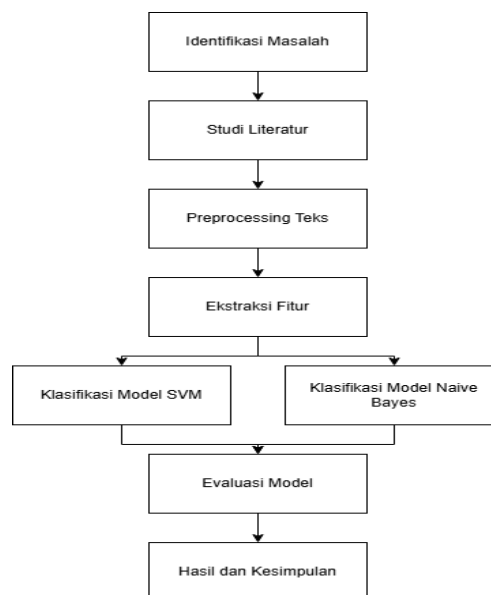
Support Vector Machine (SVM) adalah algoritma machine learning yang bekerja dengan mencari hyperplane optimal untuk memisahkan kelas data secara maksimal dalam ruang fitur berdimensi tinggi, sehingga menghasilkan generalisasi model yang baik [16]. Sementara itu, Naive Bayes adalah teknik klasifikasi probabilistik yang didasarkan pada teorema Bayes dengan asumsi independensi antar fitur. Kelebihan Naive Bayes terletak pada kesederhanaannya dan kemampuannya bekerja dengan baik meskipun hanya menggunakan sedikit data latih, sehingga sangat efisien dalam memproses data teks bervolume besar [17].

Berdasarkan permasalahan yang telah diuraikan, penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap program MBG melalui komentar di platform TikTok dan Instagram, sekaligus membandingkan kinerja metode Support Vector Machine (SVM) dan Naive Bayes dalam mengklasifikasikan sentimen positif, negatif, dan netral. Data yang digunakan berupa 6.001 komentar yang dikumpulkan melalui proses web scraping menggunakan Apify dan Python. Hasil penelitian ini diharapkan dapat memberikan gambaran yang jelas mengenai opini publik terhadap program MBG, sekaligus merekomendasikan metode klasifikasi yang paling efektif untuk analisis sentimen pada data media sosial berbahasa Indonesia.

2. METODOLOGI PENELITIAN

2.1 Kerangka Kerja Penelitian

Untuk melaksanakan penelitian ini secara sistematis, disusun sebuah kerangka kerja yang menggambarkan alur tahapan penelitian secara spesifik. Kerangka kerja ini dirancang khusus untuk menjawab permasalahan analisis sentimen masyarakat terhadap program MBG di platform TikTok dan Instagram dengan membandingkan kinerja dua metode machine learning, yaitu Support Vector Machine (SVM) dan Naive Bayes, menggunakan data komentar sebanyak 6.001 yang dikumpulkan melalui web scraping. Secara keseluruhan, terdapat enam tahap utama yang saling berkaitan dalam penelitian ini, dimulai dari identifikasi masalah, studi literatur, pengumpulan data, analisis menggunakan metode SVM dan Naive Bayes dengan ekstraksi fitur TF-IDF, penyajian hasil dan pembahasan perbandingan kinerja kedua metode, hingga penyusunan laporan. Adapun kerangka kerja penelitian yang digunakan dapat dilihat pada gambar berikut.



Gambar 4. Kerangka Kerja Penelitian

Berdasarkan kerangka kerja di atas maka dapat diuraikan pembahasan masing-masing tahap dalam penelitian adalah sebagai berikut :

1. Identifikasi Masalah
Penelitian ini mengkaji persepsi dan sentimen masyarakat terhadap program Makan Bergizi Gratis (MBG) di media sosial TikTok dan Instagram dengan tujuan menganalisis sentimen serta membandingkan kinerja metode SVM dan Naive Bayes.
2. Studi Literatur
Peneliti menelaah berbagai literatur terkait analisis sentimen, text mining, algoritma SVM dan Naive Bayes, serta konteks sosial program MBG sebagai dasar teoritis dan metodologis penelitian.
3. Preprocessing Teks
Membersihkan dan menormalisasi 6.001 data komentar mentah dari *noise* melalui serangkaian proses: *cleaning*, *case folding*, *tokenizing*, *stopword removal*, dan *stemming*.
4. Ekstraksi Fitur
Mengonversi token kata hasil *preprocessing* menjadi representasi bobot numerik menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF) agar dapat diproses oleh model.
5. Klasifikasi Model SVM dan Naive Bayes
Menerapkan kedua algoritma *machine learning* tersebut secara paralel untuk memetakan dan mengklasifikasikan data ke dalam label sentimen positif, negatif, dan netral.
6. Evaluasi Model
Mengukur kinerja dan tingkat kesalahan kedua model klasifikasi secara kuantitatif menggunakan *confusion matrix* berdasarkan metrik *accuracy*, *precision*, *recall*, dan *f1-score*.
7. Hasil dan Kesimpulan
Membandingkan hasil evaluasi untuk menetapkan algoritma yang paling efektif, serta menyajikan konklusi akhir mengenai kecenderungan opini publik terhadap program MBG.

2.2 Rancangan Eksperimen

Rancangan eksperimen pada penelitian ini bertujuan untuk menjelaskan tahapan-tahapan yang dilakukan dalam proses penerapan algoritma machine learning untuk analisis sentimen. Tahapan eksperimen disusun secara sistematis agar proses penelitian dapat berjalan terarah, mulai dari persiapan data hingga evaluasi hasil model.

Eksperimen ini dimulai dari pengumpulan dan pelabelan data, kemudian data diproses melalui tahap *preprocessing* seperti pembersihan teks, *case folding*, dan tokenisasi. Setelah itu dilakukan ekstraksi fitur menggunakan TF-IDF dan seleksi fitur dengan Chi-Square. Data yang sudah siap diklasifikasikan menggunakan metode SVM dan Naive Bayes. Selanjutnya, kinerja kedua model dievaluasi dengan metrik akurasi, presisi, *recall*, dan *f1-score*, lalu hasilnya dibandingkan untuk mengetahui metode yang paling efektif. Adapun alur eksperimen yang digunakan dalam penelitian ini dapat dijelaskan sebagai berikut :

1. Pengumpulan Data
Data penelitian berupa 6.001 komentar terkait program MBG yang dikumpulkan dari TikTok dan Instagram menggunakan teknik web scraping melalui Apify dan Python.
2. Preprocessing Data
Data komentar dibersihkan dan disiapkan melalui tahapan *cleansing*, *case folding*, tokenisasi, penghapusan *stopword*, dan *stemming* agar teks menjadi seragam dan siap dianalisis.
3. Pelabelan Data
Setiap komentar diberi label sentimen positif, negatif, atau netral secara manual untuk membentuk dataset terlatih yang dapat dipahami oleh algoritma SVM dan Naive Bayes.
4. Ekstraksi Fitur TF-IDF
Metode TF-IDF digunakan untuk mengubah data teks menjadi nilai numerik berdasarkan tingkat kepentingan kata dalam keseluruhan dokumen.
5. Split Data
Dataset dibagi menjadi data training sebesar 80% dan data testing sebesar 20% untuk memastikan proses pelatihan dan evaluasi model berjalan objektif.
6. Proses Klasifikasi
Data yang telah diproses dan diekstraksi fiturnya diklasifikasikan menggunakan algoritma Support Vector Machine dan Naive Bayes.
7. Evaluasi Model
Kinerja kedua algoritma dievaluasi menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score* untuk mengukur tingkat keberhasilan klasifikasi.
8. Perbandingan Analisis
Hasil evaluasi SVM dan Naive Bayes dibandingkan untuk menentukan algoritma yang paling efektif dalam mengklasifikasikan sentimen masyarakat terhadap program MBG.

3. HASIL DAN PEMBAHASAN

Dalam era perkembangan teknologi kecerdasan buatan atau artificial intelligence (AI) yang pesat saat ini, masih sedikit orang yang menyadari bahwa kecerdasan buatan memiliki beberapa cabang, di antaranya adalah machine Learning atau pembelajaran mesin. Machine Learning (ML), salah satu bagian penting dari AI, menarik perhatian karena kemampuannya belajar seperti manusia [21]. Machine Learning (ML) adalah salah satu aplikasi dari Artificial Intilligent (AI) yang fokus kepada pengembangan sebuah sistem yang mampu belajar sendiri tanpa harus diprogram berulang kali. ML membutuhkan sebuah data (data traning) sebagai proses learning sebelum menghasilkan sebuah hasil. Jadi, secara sederhana dapat dijelaskan bahwa Machine Learning adalah pemograman komputer untuk mencapai kriteria/performa tertentu dengan menggunakan sekumpulan data traning atau pengalaman di masa lalu (past experience) [22].

3.1 Dataset Penelitian

Dataset yang digunakan dalam penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap MBG melalui platform media sosial TikTok dan Instagram. Dataset terdiri dari dua sumber data, yaitu dataset TikTok dan dataset Instagram, yang masing-masing memuat data komentar dan interaksi pengguna terkait topik MBG. Data yang dikumpulkan merepresentasikan respons, opini, serta pandangan masyarakat yang disampaikan melalui kedua platform media sosial. Dataset dimanfaatkan dalam proses analisis sentimen untuk membandingkan kinerja metode Support Vector Machine (SVM) dan Naive Bayes dalam mengklasifikasikan sentimen masyarakat. Sebelum dilakukan proses pemodelan, dataset terlebih dahulu melalui tahap eksplorasi dan analisis deskriptif secara umum untuk memahami karakteristik data serta memastikan kesiapan data yang akan digunakan dalam tahapan selanjutnya.

Tabel 1. Struktur *Dataset*

Aspek	Dataset TikTok	Dataset Instagram
Jumlah Baris	2.999	3.002
Jumlah Kolom	9	6
Total Entri	2.999	3.002
Format Data	Pandas DataFrame	Pandas DataFrame
Kolom Bertipe Integer	4 kolom	1 kolom
Kolom Bertipe Object	5 kolom	5 kolom
Penggunaan Memori	± 211 KB	± 140 KB
Kolom Utama Teks Komentar	text	Comments
Sumber Data	Media Sosial TikTok	Media Sosial Instagram

3.2 Pembersihan Data (*Pre-Processing*)

Tahap *preprocessing* data merupakan tahapan awal yang bertujuan untuk mempersiapkan data teks agar layak digunakan dalam proses analisis sentimen. Data yang diperoleh dari media sosial TikTok dan Instagram umumnya masih mengandung berbagai unsur yang tidak relevan, seperti karakter khusus, tautan (*URL*), simbol, serta teks kosong yang dapat memengaruhi kinerja model klasifikasi. Oleh karena itu, diperlukan serangkaian proses pembersihan dan normalisasi data sebelum dilakukan pembobotan kata dan klasifikasi sentimen.

3.2.1 *Cleanning Data*

Tahap *cleaning data* bertujuan untuk membersihkan data teks dari elemen-elemen yang tidak diperlukan dalam analisis sentimen. Proses ini meliputi penghapusan data kosong (*missing value*), penghapusan data duplikat, serta pembersihan teks dari *URL*, *mention*, *hashtag*, angka, tanda baca, dan karakter khusus lainnya. Tahap ini dilakukan agar data teks yang digunakan benar-benar mencerminkan opini atau sentimen pengguna.

```

===== DATASET TIKTOK (BEFORE CLEANING) =====
0 saya kecewa pak, saya lulusan sarjana gizi iku...
1 Anakku juga Lulus murni sarjana manajemen semo...
2 Alhamdulillah anak saya lulus murni.
3 niat bapak sungguh mulia, sangatlah mulia. sem...
4 Sampai sekarang tidak menerima kejelasan hasi...
Name: text, dtype: object

===== DATASET INSTAGRAM (BEFORE CLEANING) =====
0 Inilah fungsi MBG yg tepat sasaran 🍷❤️
1 Sedih banget ya allah 🙏
2
3 sehat sehat yaa adek 🍷❤️
4 Ah ga suka gini². . Nangis banget 🙏 nak jadi p...
Name: Comments, dtype: object

===== DATASET TIKTOK (AFTER CLEANING) =====
0 saya kecewa pak saya lulusan sarjana gizi ikut...
1 Anakku juga Lulus murni sarjana manajemen semo...
2 Alhamdulillah anak saya lulus murni
3 niat bapak sungguh mulia sangatlah mulia semog...
4 Sampai sekarang tidak menerima kejelasan hasi...
Name: text_clean, dtype: object

===== DATASET INSTAGRAM (AFTER CLEANING) =====
0 Inilah fungsi MBG yg tepat sasaran
1 Sedih banget ya allah
2
3 sehat sehat yaa adek
4 Ah ga suka gini² Nangis banget nak jadi profes...
Name: comments_clean, dtype: object

```

Gambar 5. Preprocessing Data

3.2.2 Case Folding

Tahap *case folding* merupakan proses untuk mengubah seluruh huruf pada data teks menjadi huruf kecil (*lowercase*). Tujuan dari proses ini adalah untuk menyeragamkan bentuk kata sehingga kata yang memiliki makna sama tetapi ditulis dengan perbedaan huruf besar dan kecil tidak dianggap sebagai kata yang berbeda oleh sistem. Tahap *case folding* penting dilakukan untuk meningkatkan konsistensi data dan mengurangi kompleksitas pada tahap analisis selanjutnya. Pada penelitian ini, proses *case folding* diterapkan pada hasil teks yang telah melalui tahap *cleaning data*. Dengan demikian, seluruh teks komentar dari dataset TikTok dan Instagram berada dalam format huruf kecil dan siap digunakan pada tahap *tokenizing*, *stopword removal*, dan *stemming*.

```

===== BEFORE CASE FOLDING (TIKTOK) =====
0 saya kecewa pak saya lulusan sarjana gizi ikut...
1 Anakku juga Lulus murni sarjana manajemen semo...
2 Alhamdulillah anak saya lulus murni
3 niat bapak sungguh mulia sangatlah mulia semog...
4 Sampai sekarang tidak menerima kejelasan hasi...
Name: text_clean, dtype: object

===== BEFORE CASE FOLDING (INSTAGRAM) =====
0 Inilah fungsi MBG yg tepat sasaran
1 Sedih banget ya allah
2
3 sehat sehat yaa adek
4 Ah ga suka gini² Nangis banget nak jadi profes...
Name: comments_clean, dtype: object

===== AFTER CASE FOLDING (TIKTOK) =====
0 saya kecewa pak saya lulusan sarjana gizi ikut...
1 anakku juga lulus murni sarjana manajemen semo...
2 alhamdulillah anak saya lulus murni
3 niat bapak sungguh mulia sangatlah mulia semog...
4 sampai sekarang tidak menerima kejelasan hasi...
Name: text_clean, dtype: object

===== AFTER CASE FOLDING (INSTAGRAM) =====
0 inilah fungsi mbg yg tepat sasaran
1 sedih banget ya allah
2
3 sehat sehat yaa adek
4 ah ga suka gini² nangis banget nak jadi profes...
Name: comments_clean, dtype: object

```

Gambar 6. Case Folding

3.2.3 Tokenising

Tahap *tokenizing* merupakan proses pemecahan teks menjadi unit-unit yang lebih kecil yang disebut *token*, umumnya berupa kata. Tujuan dari proses *tokenizing* adalah untuk mengubah data teks berbentuk kalimat menjadi kumpulan kata yang dapat diproses lebih lanjut oleh sistem. Tahap ini sangat penting dalam analisis sentimen karena model klasifikasi bekerja berdasarkan representasi kata, bukan kalimat utuh. Pada penelitian ini, proses *tokenizing* dilakukan setelah tahap *case folding*, sehingga seluruh teks telah berada dalam bentuk huruf kecil. Dengan demikian, hasil *token* yang dihasilkan menjadi lebih konsisten dan siap digunakan pada tahap *stopword removal* dan *stemming*.

```
===== DATASET TIKTOK (BEFORE TOKENIZING) =====
0 saya kecewa pak saya lulusan sarjana gizi ikut...
1 anakku juga lulus murni sarjana manajemen semo...
2 alhamdulillah anak saya lulus murni
3 niat bapak sungguh mulia sangatlah mulia semog...
4 sampai sekarang tidak menerima kejelasan hasi...
Name: text_clean, dtype: object

===== DATASET INSTAGRAM (BEFORE TOKENIZING) =====
0 inilah fungsi mbg yg tepat sasaran
1 sedih banget ya allah
2
3 sehat sehat yaa adek
4 ah ga suka gini2 nangis banget nak jadi profes...
Name: comments_clean, dtype: object

===== DATASET TIKTOK (AFTER TOKENIZING) =====
0 [saya, kecewa, pak, saya, lulusan, sarjana, gi...
1 [anakku, juga, lulus, murni, sarjana, manajeme...
2 [alhamdulillah, anak, saya, lulus, murni]
3 [niat, bapak, sungguh, mulia, sangatlah, mulia...
4 [sampai, sekarang, tidak, menerima, kejelasan...
Name: text_token, dtype: object

===== DATASET INSTAGRAM (AFTER TOKENIZING) =====
0 [inilah, fungsi, mbg, yg, tepat, sasaran]
1 [sedih, banget, ya, allah]
2 []
3 [sehat, sehat, yaa, adek]
4 [ah, ga, suka, gini2, nangis, banget, nak, jad...
Name: comments_token, dtype: object
```

Gambar 7. Tokenizing

3.2.4 Stop Removal

Tahap *stopword removal* merupakan proses lanjutan dalam *preprocessing* teks yang bertujuan untuk menghilangkan kata-kata umum yang sering muncul tetapi tidak memiliki makna penting dalam analisis sentimen. Kata-kata tersebut, seperti kata sambung, kata depan, dan kata ganti, umumnya tidak memberikan kontribusi signifikan terhadap penentuan sentimen suatu teks. Dengan menghapus *stopword*, data teks menjadi lebih ringkas dan fokus pada kata-kata yang mengandung informasi penting, sehingga dapat meningkatkan kinerja model klasifikasi yang digunakan pada tahap selanjutnya. Pada penelitian ini, proses *stopword removal* dilakukan menggunakan daftar *stopword* Bahasa Indonesia yang disediakan oleh pustaka *Natural Language Toolkit* (NLTK). Proses ini diterapkan pada kedua dataset, yaitu dataset TikTok dan dataset Instagram, setelah melalui tahap *case folding* dan *tokenizing*.

```
===== BEFORE STOPWORD REMOVAL (TIKTOK) =====
0 [saya, kecewa, pak, saya, lulusan, sarjana, gi...
1 [anakku, juga, lulus, murni, sarjana, manajeme...
2 [alhamdulillah, anak, saya, lulus, murni]
3 [niat, bapak, sungguh, mulia, sangatlah, mulia...
4 [sampai, sekarang, tidak, menerima, kejelasan...
Name: text_token, dtype: object

===== BEFORE STOPWORD REMOVAL (INSTAGRAM) =====
0 [inilah, fungsi, mbg, yg, tepat, sasaran]
1 [sedih, banget, ya, allah]
2 []
3 [sehat, sehat, yaa, adek]
4 [ah, ga, suka, gini2, nangis, banget, nak, jad...
Name: comments_token, dtype: object

===== AFTER STOPWORD REMOVAL (TIKTOK) =====
0 [kecewa, lulusan, sarjana, gizi, spii, batch, ...
1 [anakku, lulus, murni, sarjana, manajemen, sem...
2 [alhamdulillah, anak, lulus, murni]
3 [niat, sungguh, mulia, mulia, semoga, yg, lolo...
4 [menerima, kejelasan, hasil, pengumuman, web,...
Name: text_token, dtype: object

===== AFTER STOPWORD REMOVAL (INSTAGRAM) =====
0 [fungsi, mbg, yg, sasaran]
1 [sedih, banget, ya, allah]
2 []
3 [sehat, sehat, yaa, adek]
4 [ah, ga, suka, gini2, nangis, banget, nak, pro...
Name: comments_token, dtype: object
```

Gambar 8. Stop Removal

3.2.5 Stemming

Tahap *stemming* merupakan proses *preprocessing* teks yang bertujuan untuk mengubah setiap kata ke dalam bentuk kata dasar. Proses ini dilakukan dengan menghilangkan imbuhan seperti awalan, akhiran, sisipan, maupun gabungan imbuhan yang terdapat dalam suatu kata. Dengan menerapkan *stemming*, variasi kata yang memiliki makna sama dapat diseragamkan sehingga mengurangi kompleksitas fitur pada tahap pembobotan dan klasifikasi. Dalam penelitian ini, proses *stemming* dilakukan menggunakan pustaka *Sastrawi* yang dirancang khusus untuk Bahasa Indonesia. Tahap ini diterapkan setelah proses *stopword removal* sehingga kata-kata yang diproses hanya merupakan kata penting yang telah terbebas dari kata umum. Hasil dari *stemming* diharapkan dapat meningkatkan kualitas representasi teks serta kinerja model klasifikasi sentimen yang digunakan pada tahap selanjutnya.

```

===== BEFORE STEMMING (TIKTOK) =====
0 [kecewa, lulusan, sarjana, gizi, sppi, batch, ...
1 [anakku, lulus, murni, sarjana, manajemen, sem...
2 [alhamdulillah, anak, lulus, murni]
3 [niat, sungguh, mulia, mulia, semoga, yg, lolo...
4 [menerima, kejelasan, hasil, pengumuman, web,...
Name: text_token, dtype: object

===== BEFORE STEMMING (INSTAGRAM) =====
0 [fungsi, mbg, yg, sasaran]
1 [sedih, banget, ya, allah]
2 []
3 [sehat, sehat, yaa, adek]
4 [ah, ga, suka, gini?, nangis, banget, nak, pro...
Name: comments_token, dtype: object

100%|██████████| 2999/2999 [05:47<00:00, 8.64it/s]
100%|██████████| 3002/3002 [03:20<00:00, 14.97it/s]===== AFTER STEMMING (TIKTOK) =====
0 [kecewa, lulus, sarjana, gizi, sppi, batch, tp...
1 [anak, lulus, murni, sarjana, manajemen, moga,...
2 [alhamdulillah, anak, lulus, murni]
3 [niat, sungguh, mulia, mulia, moga, yg, lolos,...
4 [terima, jasa, hasil, umum, web, email]
Name: text_stem, dtype: object

===== AFTER STEMMING (INSTAGRAM) =====
0 [fungsi, mbg, yg, sasar]
1 [sedih, banget, ya, allah]
2 []
3 [sehat, sehat, yaa, adek]
4 [ah, ga, suka, gin, nang, banget, nak, profeso...
Name: comments_stem, dtype: object

```

Gambar 9. Stemming

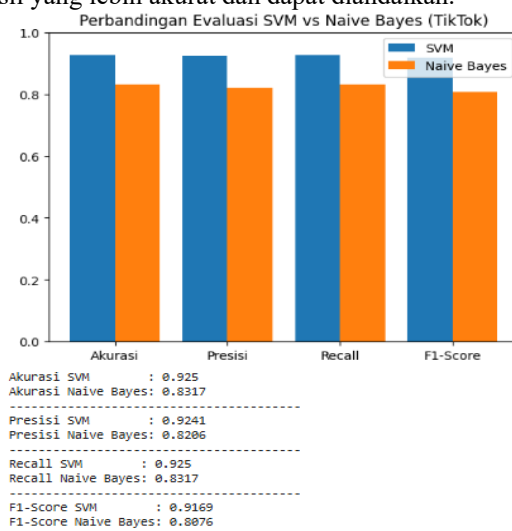
3.3 Evaluasi Model

Berdasarkan hasil evaluasi yang telah dilakukan, dapat disimpulkan bahwa metode Support Vector Machine (SVM) memiliki kinerja yang lebih baik dibandingkan metode Naive Bayes dalam analisis sentimen masyarakat terhadap MBG pada platform TikTok. Hal ini terlihat dari seluruh nilai evaluasi utama, yaitu akurasi, presisi, recall, dan F1-score, di mana SVM selalu memperoleh nilai yang lebih tinggi. Secara umum, SVM mampu mengklasifikasikan sentimen dengan lebih stabil dan seimbang pada setiap kelas sentimen, sedangkan Naive Bayes cenderung mengalami kesulitan terutama dalam mengenali sentimen negatif.

3.3.1 Perbandingan SVM dan Naive Bayes di Platform TikTok

Pada platform TikTok, metode SVM menunjukkan performa yang lebih unggul dibandingkan Naive Bayes. Nilai akurasi SVM sebesar 0,925 lebih tinggi dibandingkan Naive Bayes sebesar 0,8317, yang berarti SVM lebih sering memberikan hasil klasifikasi yang benar. Selain itu, nilai presisi, recall, dan F1-score SVM juga lebih baik, yang menunjukkan bahwa SVM tidak hanya akurat, tetapi juga lebih konsisten dalam mengenali berbagai jenis sentimen. SVM mampu mengklasifikasikan sentimen netral dan positif dengan sangat baik, serta masih lebih efektif dalam mengenali sentimen negatif dibandingkan Naive Bayes.

Sementara itu, metode Naive Bayes pada platform TikTok memiliki keterbatasan yang cukup jelas. Meskipun Naive Bayes masih mampu mengklasifikasikan sentimen netral dan positif dengan cukup baik, metode ini gagal mengenali sentimen negatif, yang terlihat dari nilai presisi, recall, dan F1-score negatif yang bernilai nol. Hal ini menyebabkan performa Naive Bayes menjadi kurang seimbang dan cenderung bias ke sentimen netral. Oleh karena itu, berdasarkan hasil pengujian pada data TikTok, metode SVM lebih direkomendasikan untuk analisis sentimen karena memberikan hasil yang lebih akurat dan dapat diandalkan.



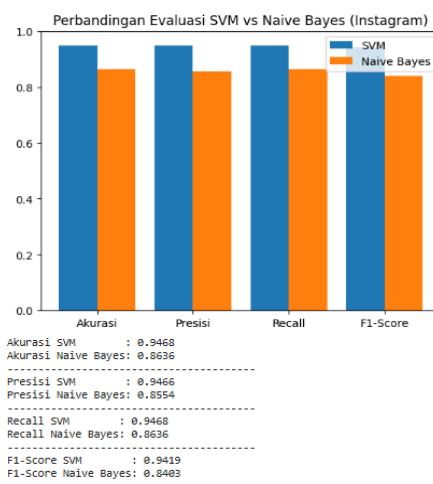
Gambar 10. Perbandingan SVM dan Naive Bayes di Platform TikTok

3.3.2 Perbandingan SVM dan Naive Bayes di Platform Instagram

Berdasarkan hasil pengujian pada platform Instagram, metode Support Vector Machine (SVM) kembali

menunjukkan kinerja yang lebih baik dibandingkan metode Naive Bayes. Hal ini terlihat dari nilai akurasi SVM sebesar 0,9468 yang lebih tinggi dibandingkan Naive Bayes sebesar 0,8636, yang berarti SVM lebih sering memberikan hasil klasifikasi yang benar. Selain itu, nilai presisi, recall, dan F1-score SVM juga lebih tinggi, menunjukkan bahwa SVM lebih konsisten dan stabil dalam mengklasifikasikan sentimen masyarakat terhadap MBG di Instagram.

Sementara itu, metode Naive Bayes pada platform Instagram masih memiliki keterbatasan yang sama seperti pada TikTok, terutama dalam mengenali sentimen negatif. Meskipun Naive Bayes mampu mengklasifikasikan sentimen netral dan positif dengan cukup baik, metode ini cenderung mengelompokkan banyak data ke dalam kelas netral. Hal ini menyebabkan hasil klasifikasi menjadi kurang seimbang. Oleh karena itu, berdasarkan perbandingan hasil evaluasi di platform Instagram, metode SVM dinilai lebih unggul dan lebih direkomendasikan untuk analisis sentimen karena memberikan hasil yang lebih akurat dan dapat diandalkan.



Gambar 11. Perbandingan SVM dan Naive Bayes di Platform Instagram

3.3.3 Perbandingan SVM dan Naive Bayes di Platform TikTok dan Instagram

Dapat disimpulkan bahwa metode SVM secara konsisten menghasilkan nilai evaluasi yang lebih tinggi dibandingkan Naive Bayes pada kedua platform, baik TikTok maupun Instagram. Hal ini menunjukkan bahwa SVM lebih akurat, stabil, dan seimbang dalam mengklasifikasikan sentimen masyarakat terhadap MBG. Sementara itu, Naive Bayes memiliki performa yang cukup baik, namun cenderung lebih rendah terutama dalam mengenali sentimen negatif.

Tabel 2. Perbandingan Keseluruhan Metode SVM dan Naive Bayes

Platform	Metode	Akurasi	Presisi	Recall	F1-Score
TikTok	SVM	0.9250	0.9241	0.9250	0.9169
	Naive Bayes	0.8317	0.8206	0.8317	0.8076
Instagram	SVM	0.9468	0.9466	0.9468	0.9419
	Naive Bayes	0.8636	0.8554	0.8636	0.8403

4. KESIMPULAN

Berdasarkan hasil penelitian, metode *Support Vector Machine* (SVM) terbukti secara konsisten mengungguli kinerja *Naive Bayes* dalam mengklasifikasikan sentimen masyarakat terhadap program Makan Bergizi Gratis (MBG). Pada platform TikTok, SVM mencapai tingkat akurasi 92,50% (berbanding 83,17% pada *Naive Bayes*), dan pada Instagram mencapai 94,68% (berbanding 86,36%). Keunggulan utama SVM terletak pada tingkat stabilitas dan kemampuannya dalam mengenali kelas sentimen minoritas, khususnya sentimen negatif yang sama sekali gagal dideteksi oleh *Naive Bayes* pada dataset TikTok. Selain itu, temuan analisis menunjukkan bahwa opini publik di kedua platform didominasi oleh sentimen netral, mengindikasikan respons masyarakat yang lebih bersifat informatif. Oleh karena itu, algoritma SVM sangat direkomendasikan sebagai metode yang lebih reliabel untuk klasifikasi analisis sentimen pada data media sosial berbahasa Indonesia.

REFERENCES

- [1] G. C. Hermawan And D. F. Suyatno, "Analisis Kepuasan Pelanggan Terhadap Layanan Aplikasi Klik Indomaret Zahra Prisdian Tiararosa, 2026, *JAKAKOM*, Page 1992

- Menggunakan Metode End User Computing Satisfaction (Eucs) Dan Delone And Mclean,” *Progr. Stud. Sist. Inf. Fak. Sains Dan Teknol. Univ. Islam Negeri Syarif Hidayatullah*, Vol. 4, No. 1, Pp. 88–100, 2023, doi: <https://doi.org/10.26740/jeisbi.v4i3.54230>
- [2] N. P. Husain, S. Sukirman, And S. Sajiah, “Analisis Sentimen Ulasan Pengguna TikTok Pada Google Play Store Berbasis Tf-Idf Dan Support Vector Machine,” *J. Syst. Comput. Eng.*, Vol. 5, No. 1, Pp. 91–102, 2024, doi: <https://doi.org/10.61628/jsce.v5i1.1105>.
- [3] N. Muhamad, “Ini Media Sosial Terpopuler Di Dunia Pada Pertengahan 2025.” [Online]. Available: <https://databoks.katadata.co.id/media-sosial/statistik/6891de083c9b7/indonesia-masuk-20-negara-paling-sering-akses-media-sosial-kuartal-i-2025>.
- [4] N. Apriliani *Et Al.*, “Analisis Sentimen Review Penggunaan TikTok Melalui Pendekatan Algoritma Naïve Bayes,” Vol. 7, No. 6, Pp. 3725–3731, 2023.
- [5] M. A. Rizaty, “Kumpulan Data Seputar TikTok Di Dunia Dan Indonesia Hingga April 2025.” [Online]. Available: <https://dataindonesia.id/internet/detail/laporan-kumpulan-data-seputar-tiktok-di-dunia-dan-indonesia-hingga-april-2025>.
- [6] N. Cahyono, S. Informatika, And U. A. Yogyakarta, “Analisis Sentimen Komentar Instagram Pada Program Kampus,” Vol. 8, No. 2, Pp. 2372–2381, 2024.
- [7] M. A. Rizaty, “Negara Pengguna Instagram Terbesar.” [Online]. Available: <https://dataindonesia.id/internet/detail/jajaran-negara-pengguna-instagram-terbanyak-pada-oktober-2025-indonesia-keempat>.
- [8] K. Andreas *Et Al.*, “Tinjauan Kritis Tentang Program Makan Bergizi Gratis Terhadap Produktivitas Belajar Siswa,” Vol. 2, 2025. doi : <https://share.google/zrIMkjFxJPZDY2bCO>
- [9] M. A. Rojabi, *Mitra Bgn: Syarat, Cara Daftar Online, Dan Kunci Lolos Verifikasi Program Makan Bergizi Gratis*. Yogyakarta: Penerbit Andi, 2024.j
- [10] L. Nursingghah, T. Mufizar, And U. Perjuangan, “Analisis Sentimen Pengguna Aplikasi X Terhadap Program Makan Siang Gratis,” Vol. 12, No. 3, 2024.
- [11] B. Rahmatullah, S. A. Saputra, P. Budiono, And D. P. Wigandi, “Sentimen Analisis Makan Bergizi Gratis Menggunakan Algoritma Naive Bayes,” *J. Inf. Technol.*, Vol. 5, No. 1, 2025, doi: <https://doi.org/10.46229/jifotech.v5i1.978>.
- [12] A. Sitanggang, Y. Umaidah, R. I. Adam, U. S. Karawang, And T. Timur, “Analisis Sentimen Masyarakat Terhadap Program Makan Siang Gratis Pada Media,” Vol. 12, No. 3, 2024.
- [13] N. Putu Alya Noviyanti *Et Al.*, “Penerapan Multinomial Naïve Bayes Dan Chi-Square Pada Analisis Sentimen Makan Bergizi Gratis,” *Jnatia*, Vol. 3, No. 4, Pp. 785–794, 2025.
- [14] M. M. Laia *Et Al.*, “Analisis Sentimen Program Makan Gratis Pada Platform X Algoritma Naïve Bayes Menggunakan,” 2025.
- [15] A. Ardiansyah, Suleman, E. A. Pratama, And N. I. Fadlilah, “Analisis Sentimen Pengguna Terhadap Aplikasi Chatgpt Di Google Play Store: Penerapan Algoritma Support Vector Machine,” *J. Tek. Inform. Dan Sist. Inf.*, Vol. 11, No. 2, Pp. 247–254, 2024.
- [16] Sukri And D. Arisandi, “Jurnal Resti,” *Resti*, Vol. 1, No. 1, Pp. 19–25, 2020.
- [17] A. Pebdika, R. Herdiana, And D. Solihudin, “Klasifikasi Menggunakan Metode Naive Bayes Untuk Menentukan Calon Penerima Pip,” *Jati (Jurnal Mhs. Tek. Inform.)*, Vol. 7, No. 1, Pp. 452–458, 2023, doi: <https://doi.org/10.36040/jati.v7i1.6303>.
- [18] R. R. Putri And N. Cahyono, “Publik Pemerintah Dki Jakarta Dengan Algoritma,” Vol. 8, No. 2, Pp. 2363–2371, 2024.
- [19] F. Yahya, D. E. Ratnawati, And B. Rahayudi, “Analisis Sentimen Ulasan Pengguna Dari Google Maps Menggunakan Metode Long Short-Term Memory (Studi Kasus : Rumah Sakit Gatoel),” Vol. 9, No. 4, Pp. 1–11, 2025.
- [20] R. Zaevan, K. Putra, M. Febrian, A. Pradhana, And F. R. Nansyah, “Klasifikasi Sentimen Ulasan Aplikasi Mobile Jkn Menggunakan Naive Bayes Untuk Memenuhi Persyaratan Penelitian Pada Mata Kuliah Proyek Penelitian Program Studi Teknik Informatika,” 2025.
- [21] M. C. Ari Muzakir, S.Kom., M. . Prof.Dr.Kusworo Adi, S.Si., And M. K. Dr.Retno Kusumanigrum, S.Si., *Penerapan Konsep Machine Learning & Deep Learning*. 2024.
- [22] J. Informatika And S. Informasi, “Informasi (Jurnal Informatika Dan Sistem Informasi) Volume 12 No.1 / Mei/ 2020,” Vol. 12, No. 1, Pp. 67–80, 2020.