Building of Informatics, Technology and Science (BITS)
Volume 8, No 1, June 2026 Page: 623−634
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.10032

Perbandingan Kinerja Naive Bayes dan SVM dalam Analisis Sentimen Program Makanan Bergizi Gratis (MBG) sebagai Pendukung Pengambilan Keputusan

Vebi Adeka Putra, Nirwana Hendrastuty*

Fakultas Teknik dan Ilmu Komputer, Sistem Informasi, Universitas Teknokrat Indonesia, Bandar Lampung, Indonesia
Email: [1]vebi_adeka_putra@teknokrat.ac.id, [2,*]nirwanahendrastuty@teknokrat.ac.id
Email Penulis Korespondensi: nirwanahendrastuty@teknokrat.ac.id
Submitted: **22/05/2026**; Accepted: **30/06/2026**; Published**: 30/06/2026**

Abstrak−Program Makan Bergizi Gratis (MBG) merupakan salah satu program pemerintah yang bertujuan meningkatkan kualitas gizi masyarakat dan mengurangi angka stunting di Indonesia. Program tersebut menimbulkan berbagai tanggapan masyarakat yang banyak disampaikan melalui media sosial, salah satunya pada kolom komentar YouTube. Penelitian ini dilakukan untuk mengetahui sentimen masyarakat terhadap Program Makan Bergizi Gratis (MBG) serta membandingkan performa algoritma Naive Bayes dan Support Vector Machine (SVM) dalam klasifikasi sentimen komentar pengguna YouTube. Data penelitian diperoleh melalui proses *scraping* komentar YouTube kemudian diproses menggunakan tahapan *preprocessing* yang meliputi *cleaning*, *case folding*, normalisasi, *tokenization*, *stopword removal*, dan *stemming*. Selanjutnya dilakukan pembobotan fitur menggunakan TF-IDF serta pelabelan data menggunakan metode *lexicon-based*. Proses klasifikasi dilakukan menggunakan algoritma Naive Bayes dan Support Vector Machine (SVM), sedangkan evaluasi model dilakukan menggunakan *confusion matrix*, *accuracy*, *precision*, *precision*, dan *f1-score*. Hasil penelitian menunjukkan bahwa algoritma Support Vector Machine (SVM) memiliki performa lebih baik dibandingkan Naive Bayes. Algoritma SVM memperoleh *accuracy* sebesar 77,4%, *precision* sebesar 78,4%, *precision* sebesar 77,4%, dan *f1-score* sebesar 77,6%, sedangkan Naive Bayes memperoleh *accuracy* sebesar 70,5%, *precision* sebesar 74,4%, *precision* sebesar 70,5%, dan *f1-score* sebesar 67,7%. Kontribusi penelitian ini adalah memberikan analisis komparatif terhadap performa algoritma Naive Bayes dan Support Vector Machine (SVM) pada klasifikasi sentimen komentar YouTube terkait Program Makan Bergizi Gratis (MBG), sehingga dapat menjadi referensi dalam pemilihan metode klasifikasi sentimen yang lebih efektif untuk mendukung analisis opini publik terhadap kebijakan pemerintah berbasis media sosial.

Kata Kunci: Analisis Sentimen; Naive Bayes; Support Vector Machine; YouTube; TF-IDF; MBG

Abstract−The Free Nutritious Meal Program is one of the Indonesian government programs aimed at improving community nutritional quality and reducing stunting rates. The program has generated various publik responses expressed through media social platforms, particularly in YouTube comment sections. This study was conducted to analyze publik sentiment toward the Free Nutritious Meal Program (MBG) and to compare the performance of the Naive Bayes and Support Vector Machine (SVM) algorithms in classifying sentiment from YouTube user comments. The research data were obtained through a YouTube comment scraping process and then processed through several preprocessing stages, including cleaning, case folding, normalization, tokenization, stopword removal, and stemming. Furthermore, feature weighting was performed using the TF-IDF method, and data labeling was carried out using a lexicon-based approach. The sentiment classification process employed the Naive Bayes and Support Vector Machine (SVM) algorithms, while model evaluation was conducted using confusion matrix, accuracy, precision, precision, and f1-score metrics. The results showed that the Support Vector Machine (SVM) algorithm achieved better performance than Naive Bayes. The SVM algorithm obtained an accuracy of 77.4%, precision of 78.4%, precision of 77.4%, and f1-score of 77.6%, whereas the Naive Bayes algorithm achieved an accuracy of 70.5%, precision of 74.4%, precision of 70.5%, and f1-score of 67.7%. The main contribution of this study is the comparative evaluation of Naive Bayes and Support Vector Machine (SVM) for classifying public sentiment from YouTube comments related to the MBG program, providing empirical evidence on the most effective classification approach for supporting social media–based public opinion analysis of government policies.

Keywords: Sentiment Analysis; Naive Bayes; Support Vector Machine; YouTube; TF-IDF; MBG

1. PENDAHULUAN

Program Makan Bergizi Gratis (MBG) merupakan salah satu program pemerintah yang bertujuan untuk meningkatkan kualitas gizi masyarakat, khususnya bagi anak sekolah dan kelompok rentan. Program ini hadir sebagai upaya pemerintah dalam menekan angka stunting dan meningkatkan kualitas sumber daya manusia Indonesia. Berdasarkan hasil Survei Status Gizi Indonesia (SSGI) tahun 2022, prevalensi stunting di Indonesia masih mencapai 21,6%, sehingga pemerintah terus mendorong berbagai program strategis di bidang kesehatan dan pemenuhan gizi masyarakat [1]. Kehadiran Program MBG mendapatkan perhatian luas dari masyarakat karena dianggap sebagai kebijakan yang memiliki dampak besar terhadap kesehatan, pendidikan, sosial, hingga ekonomi masyarakat. Namun demikian, implementasi program ini juga memunculkan berbagai tanggapan, baik berupa dukungan maupun kritik, terutama terkait efektivitas pelaksanaan, kualitas makanan, serta besarnya anggaran yang digunakan. Perkembangan media sosial menjadikan masyarakat lebih mudah menyampaikan opini terhadap kebijakan publik secara terbuka dan *real-time*. Salah satu platform yang aktif digunakan masyarakat untuk memberikan tanggapan adalah YouTube melalui fitur komentar pada video yang diunggah. Karakteristik komentar pada YouTube yang bersifat terbuka menyebabkan opini masyarakat berkembang sangat cepat dan menghasilkan data teks dalam jumlah besar [2]. Berbagai komentar yang muncul menunjukkan adanya sentimen positif, negatif, maupun netral terhadap pelaksanaan Program Makan Bergizi Gratis. Sebagian masyarakat mendukung program tersebut karena dinilai mampu membantu pemenuhan gizi anak dan mengurangi stunting, sedangkan sebagian lainnya memberikan kritik terkait pengelolaan anggaran, distribusi

Building of Informatics, Technology and Science (BITS)
Volume 8, No 1, June 2026 Page: 623−634
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.10032

makanan, hingga kualitas makanan yang diberikan. Besarnya jumlah komentar masyarakat pada YouTube menghasilkan data yang tidak terstruktur sehingga sulit dianalisis secara manual. Oleh karena itu, diperlukan suatu pendekatan berbasis teknologi untuk mengolah dan memahami opini publik secara sistematis. Salah satu pendekatan yang dapat digunakan adalah analisis sentimen. Analisis sentimen merupakan bagian dari *text mining* dan *Natural Language Processing (NLP)* yang digunakan untuk mengidentifikasi, mengelompokkan, dan mengevaluasi opini masyarakat ke dalam kategori sentimen positif, negatif, atau netral [3]. Analisis sentimen menjadi penting karena mampu membantu pemerintah maupun peneliti dalam memahami persepsi masyarakat terhadap suatu kebijakan secara lebih objektif dan terukur.

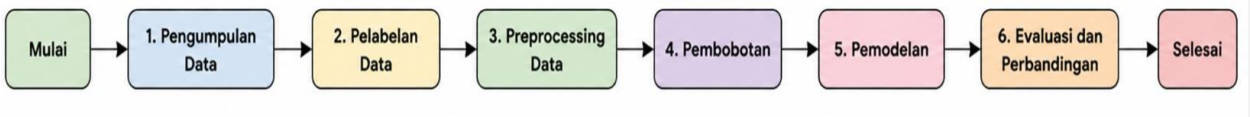
Beberapa penelitian sebelumnya telah melakukan analisis sentimen terhadap Program Makan Bergizi Gratis menggunakan berbagai metode *machine learning*. Penelitian yang dilakukan oleh Putriyekti dkk. menggunakan pendekatan multidimensional sentiment analysis dan menunjukkan bahwa sentimen negatif lebih mendominasi dibandingkan sentimen positif dan netral. Penelitian lain yang dilakukan oleh Hidayat dan Ratnaningsih membandingkan metode Random Forest dan Naive Bayes dalam analisis sentimen Program MBG dan memperoleh akurasi tertinggi sebesar 96,38% pada metode Random Forest [4]. Selain itu, penelitian oleh Ghofur dan Putri membandingkan algoritma Support Vector Machine (SVM), K-Nearest Neighbor (KNN), dan Naive Bayes pada analisis sentimen MBG, di mana metode SVM memperoleh performa terbaik dengan akurasi sebesar 91,7%. Penelitian lain juga menunjukkan bahwa metode klasifikasi yang berbeda dapat menghasilkan tingkat akurasi yang berbeda pada data media sosial berbahasa Indonesia. Meskipun penelitian-penelitian tersebut telah memberikan hasil yang baik, masih terdapat beberapa *research gap*. Pertama, sebagian besar penelitian memanfaatkan data dari platform X (Twitter), sedangkan penggunaan komentar YouTube sebagai sumber data analisis sentimen Program MBG masih relatif terbatas. Padahal, YouTube memiliki jumlah pengguna yang besar serta karakteristik komentar yang lebih panjang dan beragam sehingga berpotensi memberikan gambaran opini publik yang lebih komprehensif. Kedua, sebagian besar penelitian hanya mengevaluasi satu algoritma atau membandingkan beberapa algoritma dengan pendekatan yang berbeda, sehingga belum banyak penelitian yang secara khusus melakukan evaluasi komparatif antara Naive Bayes dan Support Vector Machine (SVM) pada dataset komentar YouTube mengenai Program MBG. Ketiga, penelitian terdahulu umumnya lebih berfokus pada hasil akurasi tanpa membahas secara komprehensif performa klasifikasi berdasarkan beberapa metrik evaluasi seperti *accuracy*, *precision*, *recall*, dan *F1-score*.

Naive Bayes dan Support Vector Machine (SVM) dipilih dalam penelitian ini karena keduanya merupakan algoritma klasifikasi yang banyak digunakan dalam analisis sentimen dan memiliki karakteristik yang berbeda. Naive Bayes merupakan algoritma probabilistik yang bekerja berdasarkan Teorema Bayes dengan asumsi independensi antarfitur, sehingga memiliki proses komputasi yang sederhana, efisien, dan mampu menangani data berdimensi tinggi. Sementara itu, Support Vector Machine (SVM) merupakan algoritma yang membangun *hyperplane* optimal untuk memisahkan kelas sehingga dikenal memiliki kemampuan generalisasi yang baik dan menghasilkan performa tinggi pada klasifikasi data teks. Perbedaan karakteristik tersebut menjadikan kedua algoritma relevan untuk dibandingkan dalam menentukan metode yang paling efektif pada klasifikasi sentimen komentar YouTube mengenai Program MBG.

Berdasarkan *research gap* tersebut, penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap Program Makan Bergizi Gratis berdasarkan komentar pengguna YouTube menggunakan algoritma Naive Bayes dan Support Vector Machine (SVM). Kontribusi utama penelitian ini adalah menyajikan analisis komparatif terhadap performa kedua algoritma pada dataset komentar YouTube yang diproses melalui tahapan *preprocessing*, pembobotan TF-IDF, dan pelabelan *lexicon-based*. Hasil penelitian diharapkan dapat memberikan bukti empiris mengenai algoritma yang lebih efektif dalam klasifikasi sentimen serta menjadi referensi bagi penelitian selanjutnya maupun pemerintah dalam memanfaatkan analisis sentimen berbasis media sosial sebagai bahan evaluasi kebijakan publik.

2. METODOLOGI PENELITIAN

Tahapan penelitian analisis sentimen komentar YouTube terkait Program Makan Bergizi Gratis (MBG) dimulai dari pengumpulan data komentar YouTube, pelabelan sentimen positif, negatif, dan netral, serta *preprocessing* data untuk membersihkan dan menyeragamkan teks. Selanjutnya dilakukan pembobotan fitur menggunakan TF-IDF dan proses klasifikasi menggunakan algoritma Naive Bayes dan Support Vector Machine (SVM). Tahap akhir penelitian dilakukan dengan evaluasi dan perbandingan performa model menggunakan *accuracy*, *precision*, *precision*, dan *f1-score* untuk menentukan algoritma terbaik dalam klasifikasi sentimen komentar masyarakat terhadap Program Makan Bergizi Gratis (MBG). Alur lengkap tahapan penelitian ditunjukkan pada Gambar 1.



Gambar 1. Tahapan Penelitian

Berdasarkan Gambar 1, tahapan penelitian terdiri atas beberapa proses sebagai berikut.

Building of Informatics, Technology and Science (BITS)
Volume 8, No 1, June 2026 Page: 623−634
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.10032

a. Pengumpulan Data

Pengumpulan data pada penelitian ini dilakukan menggunakan teknik *web scraping* terhadap komentar pengguna YouTube pada video yang membahas Program Makan Bergizi Gratis (MBG) dengan tautan https://www.youtube.com/watch?v=D_atWzv8Zy4. Proses *scraping* menghasilkan sebanyak 1.538 komentar yang kemudian disimpan dalam format *.csv* untuk memudahkan proses pengolahan data pada tahap selanjutnya. Teknik *web scraping* dipilih karena mampu mengumpulkan data secara otomatis dalam jumlah besar dan telah banyak digunakan dalam penelitian analisis sentimen berbasis media sosial [5]. Selanjutnya dilakukan proses seleksi data dengan menghapus komentar yang tidak relevan, *spam*, serta komentar yang tidak menggunakan bahasa Indonesia sehingga dataset yang digunakan lebih representatif untuk proses analisis [6], [7].

b. *Preprocessing* Data

Tahap *preprocessing* dilakukan untuk mempersiapkan data komentar sebelum proses klasifikasi sehingga kualitas data menjadi lebih baik [8]. Pada penelitian ini, *preprocessing* meliputi *cleaning, case folding, tokenizing, stopword removal*, dan *stemming*. Tahap *cleaning* dilakukan dengan menghapus karakter yang tidak diperlukan seperti angka, tanda baca, URL, *hashtag*, dan *emoji*. Selanjutnya, *case folding* mengubah seluruh huruf menjadi huruf kecil (*lowercase*) untuk menyeragamkan penulisan kata [9]. Tahap *tokenizing* memecah kalimat menjadi kumpulan token, kemudian *stopword removal* menghapus kata-kata yang tidak memiliki makna penting dalam proses klasifikasi [10]. Tahap terakhir adalah *stemming* menggunakan pustaka Sastrawi untuk mengubah kata berimbuhan menjadi kata dasar. Hasil *preprocessing* selanjutnya digunakan pada proses pembobotan fitur.

c. Pembobotan Fitur

Tahap pembobotan fitur dilakukan menggunakan metode TF-IDF (*Term Frequency–Inverse Document Frequency*) untuk mengubah data teks hasil *preprocessing* menjadi representasi numerik [11]. Implementasi TF-IDF dilakukan menggunakan TfidfVectorizer dari pustaka *scikit-learn* sehingga setiap komentar direpresentasikan dalam bentuk vektor numerik yang dapat diproses oleh algoritma klasifikasi[12]. Hasil pembobotan fitur selanjutnya digunakan sebagai masukan pada algoritma Naive Bayes dan Support Vector Machine (SVM).

d. Klasifikasi Sentimen

Proses klasifikasi sentimen dilakukan menggunakan algoritma Naive Bayes dan Support Vector Machine (SVM) terhadap data hasil pembobotan TF-IDF. Kedua algoritma dipilih karena telah banyak digunakan pada penelitian analisis sentimen berbasis teks dan memiliki karakteristik yang berbeda dalam proses klasifikasi. Selanjutnya, performa kedua model dievaluasi dan dibandingkan menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score* untuk menentukan algoritma yang memberikan hasil klasifikasi terbaik[13], [14].

1. Naive Bayes
   Naive Bayes merupakan algoritma klasifikasi berbasis probabilitas yang menggunakan Teorema Bayes dalam menentukan kategori suatu data. Algoritma ini bekerja dengan menghitung probabilitas kemunculan kata pada masing-masing kelas sentimen sehingga dapat menentukan kategori sentimen yang paling sesuai. Pada penelitian ini, Naive Bayes digunakan untuk mengklasifikasikan komentar YouTube terkait Program Makan Bergizi Gratis (MBG) ke dalam kategori sentimen positif, negatif, dan netral. Naive Bayes dipilih karena memiliki proses komputasi yang sederhana, cepat, serta efektif dalam mengolah data teks dalam jumlah besar [14], [15]. Selain itu, algoritma ini cukup baik digunakan dalam analisis sentimen media sosial karena mampu menghasilkan performa klasifikasi yang stabil dengan waktu pemrosesan yang relatif singkat.
2. Support Vector Machine (SVM)
   Support Vector Machine (SVM) merupakan algoritma *supervised learning* yang digunakan untuk melakukan klasifikasi data dengan membentuk *hyperplane* optimal sebagai pemisah antar kelas. Algoritma ini mampu bekerja dengan baik pada data teks berdimensi tinggi dan memiliki kemampuan klasifikasi yang cukup tinggi dalam analisis sentimen. Pada penelitian ini, algoritma SVM digunakan untuk mengklasifikasikan komentar YouTube terkait Program Makan Bergizi Gratis (MBG) berdasarkan polaritas sentimennya. Penggunaan SVM dipilih karena mampu mengenali pola data teks tidak terstruktur dengan lebih baik serta memiliki tingkat akurasi yang tinggi pada penelitian analisis sentimen media sosial [16], [17]. Penelitian sebelumnya menunjukkan bahwa metode SVM memiliki performa yang lebih baik dibandingkan Naive Bayes dalam klasifikasi sentimen media sosial. Hal tersebut terlihat dari hasil evaluasi akurasi, *precision*, *precision*, dan *F1-score* yang lebih tinggi pada metode SVM dibandingkan Naive Bayes. Oleh karena itu, kedua metode tersebut digunakan pada penelitian ini untuk membandingkan kinerja klasifikasi sentimen terhadap komentar YouTube mengenai Program Makan Bergizi Gratis (MBG).

e. Evaluasi Model

Evaluasi model dilakukan untuk mengetahui performa algoritma Naive Bayes dan Support Vector Machine (SVM) dalam mengklasifikasikan sentimen komentar YouTube terkait Program Makan Bergizi Gratis (MBG). Proses evaluasi dilakukan menggunakan *confusion matrix* untuk membandingkan hasil prediksi model dengan label data sebenarnya. Berdasarkan *confusion matrix*, performa model diukur menggunakan beberapa metrik evaluasi, yaitu *accuracy*, *precision*, *precision*, dan *F1-score*. *Accuracy* digunakan untuk mengukur tingkat ketepatan model secara keseluruhan dalam melakukan klasifikasi sentimen. *Precision* digunakan untuk melihat kemampuan model dalam memprediksi data secara tepat pada kelas tertentu, sedangkan *precision* digunakan untuk mengukur kemampuan

Building of Informatics, Technology and Science (BITS)
Volume 8, No 1, June 2026 Page: 623−634
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.10032

model dalam menemukan data yang sesuai dengan kelas sebenarnya. Sementara itu, *F1-score* digunakan untuk menyeimbangkan nilai *precision* dan *precision* sehingga hasil evaluasi menjadi lebih objektif [7], [18]. Rumus perhitungan metrik evaluasi yang digunakan pada penelitian ini adalah sebagai berikut:

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

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

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

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

Pada Persamaan (1)–(4), True Positive (TP) merupakan jumlah data yang berhasil diprediksi benar sebagai kelas positif, True Negative (TN) merupakan jumlah data yang berhasil diprediksi benar sebagai kelas negatif, False Positive (FP) merupakan jumlah data yang diprediksi sebagai kelas positif tetapi sebenarnya termasuk kelas negatif, sedangkan False Negative (FN) merupakan jumlah data yang diprediksi sebagai kelas negatif tetapi sebenarnya termasuk kelas positif. Keempat komponen tersebut digunakan untuk menghitung nilai *accuracy, precision, recall*, dan *F1-score* sehingga dapat memberikan evaluasi yang komprehensif terhadap performa model klasifikasi.

Hasil evaluasi menggunakan metrik tersebut selanjutnya digunakan untuk membandingkan performa algoritma Naive Bayes dan Support Vector Machine (SVM) serta menentukan algoritma yang memberikan hasil klasifikasi terbaik pada analisis sentimen komentar YouTube terkait Program Makan Bergizi Gratis (MBG).

3. HASIL DAN PEMBAHASAN

3.1 Tahap *Preprocessing* Data

Tahap *preprocessing* merupakan tahapan awal yang dilakukan dalam penelitian ini untuk membersihkan dan menyiapkan data komentar YouTube terkait Program Makan Bergizi Gratis (MBG) sebelum dilakukan proses analisis sentimen. Data hasil *scraping* masih mengandung berbagai bentuk noise seperti URL, emoji, tanda baca, huruf kapital yang tidak konsisten, singkatan kata, dan karakter lain yang tidak memiliki pengaruh terhadap proses klasifikasi sentimen. Oleh karena itu, diperlukan beberapa tahapan *preprocessing* agar data menjadi lebih bersih, terstruktur, dan mudah diproses oleh algoritma *machine learning*. Sebelum dilakukan *preprocessing*, data yang digunakan masih berupa data mentah hasil *scraping* komentar pengguna YouTube [19]. Data tersebut terdiri dari beberapa atribut, yaitu *author*, *comment*, *likes*, dan *published_at*. Atribut *author* menunjukkan nama akun pengguna YouTube, atribut *comment* berisi komentar pengguna terkait Program Makan Bergizi Gratis (MBG), atribut *likes* menunjukkan jumlah suka pada komentar, sedangkan atribut *published_at* menunjukkan waktu publikasi komentar dapat dilihat pada Tabel 1 dibawah ini.

Tabel 1. Contoh Data Mentah Hasil *Scraping* Komentar YouTube

No	*Author*	*Comment*	*Likes*	Published At
1	@pulihrhamwanto4006	https://youtube.com/shorts/_D7ktyKNwns?si=Ec__...	0	2026-05-09T08:47:14Z
2	@khannsoress	dengerin fakta emang bikin gelisah ya wk...	0	2026-05-07T09:49:53Z
3	@kezzyagabriella4331	Mau *speak up* disini, tmn w org china cerita di...	0	2026-05-02T05:23:42Z
4	@ImamAburaihan-s9t3n	Sebagai orang yang pernah mengalami susah maka...	0	2026-05-02T04:02:48Z
5	@MegaMaharani-k1v	bagus banget jdi slalu tau	0	2026-04-24T01:06:17Z
6	@dewilestariAtm	Program MBG jangan hanya dilihat sebagai bagi-...	0	2026-04-13T08:42:19Z
7	@NauraSalsabilaw	Program MBG jangan hanya dilihat sebagai bagi-...	0	2026-04-13T08:41:20Z
8	@AlsaZega	MBG ini sejak awal jadi lahan basah buat para ...	0	2026-04-10T10:45:09Z
9	@aanpambudi9713	bagi2 proyek	0	2026-04-09T03:18:47Z
10	@TheShurifon	kata2 "*free thing is the most expensive*" bisa ...	0	2026-04-06T10:45:40Z

Berdasarkan Tabel 1 terlihat bahwa data komentar masih mengandung URL, singkatan kata, penggunaan bahasa tidak baku, dan variasi penulisan yang belum konsisten. Selain itu, terdapat komentar yang masih terpotong serta mengandung campuran bahasa Indonesia dan bahasa Inggris. Kondisi tersebut menunjukkan bahwa data mentah hasil *scraping* belum dapat langsung digunakan pada proses klasifikasi sentimen karena masih mengandung banyak

Building of Informatics, Technology and Science (BITS)
Volume 8, No 1, June 2026 Page: 623−634
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.10032

noise yang dapat memengaruhi kualitas analisis. Tahap *preprocessing* dilakukan untuk membersihkan dan menyeragamkan data sehingga komentar menjadi lebih mudah dipahami oleh sistem. Proses ini sangat penting dalam penelitian analisis sentimen karena kualitas data *preprocessing* akan memengaruhi hasil pembobotan fitur dan performa algoritma klasifikasi seperti Naive Bayes dan Support Vector Machine (SVM).

3.1.1 Proses *Cleaning*

Tahap *cleaning* dilakukan untuk membersihkan data komentar YouTube dari karakter yang tidak diperlukan seperti URL, simbol, tanda baca, angka, dan karakter khusus lainnya yang dapat mengganggu proses analisis sentimen. Proses ini bertujuan agar data teks menjadi lebih bersih, terstruktur, dan mudah diproses oleh algoritma *machine learning* pada penelitian analisis sentimen Program Makan Bergizi Gratis (MBG). Pada tahap ini dilakukan penghapusan URL, simbol tertentu, serta beberapa bentuk penulisan yang tidak relevan terhadap isi sentimen komentar. Hasil proses *cleaning* ditunjukkan pada Tabel 2.

Tabel 2. Contoh Hasil Proses *Cleaning*

No	*Comment*	*Cleaning*
1	https://youtube.com/shorts/_D7ktyKNwns?si=Ec__...	
2	dengerin fakta emang bikin gelisah ya wk...	dengerin fakta emang bikin gelisah ya wk
3	Mau *speak up* disini, tmn w org china cerita di...	Mau *speak up* disini tmn w org china cerita di
4	Sebagai orang yang pernah mengalami susah maka...	Sebagai orang yang pernah mengalami susah maka
5	bagus banget jdi slalu tau	bagus banget jdi slalu tau
6	Program MBG jangan hanya dilihat sebagai bagi-...	Program MBG jangan hanya dilihat sebagai bagib...
7	Program MBG jangan hanya dilihat sebagai bagi-...	Program MBG jangan hanya dilihat sebagai bagib...
8	MBG ini sejak awal jadi lahan basah buat para ...	MBG ini sejak awal jadi lahan basah buat para ...
9	bagi2 proyek	bagi proyek
10	kata2 "*free thing is the most expensive*" bisa ...	kata *free thing is the most expensive* bisa dia...

Berdasarkan Tabel 2 proses *cleaning* berhasil menghapus URL dan karakter yang tidak diperlukan sehingga komentar menjadi lebih bersih tanpa menghilangkan inti informasi. Tahap ini penting untuk meningkatkan kualitas data sebelum dilakukan proses *case folding*, normalisasi, dan klasifikasi sentimen menggunakan algoritma Naive Bayes dan Support Vector Machine (SVM).

3.1.2 Proses *Case folding* dan Normalisasi

Setelah proses *cleaning* selesai dilakukan, tahap berikutnya adalah *case folding* dan normalisasi data. *Case folding* dilakukan dengan mengubah seluruh huruf pada komentar menjadi huruf kecil (*lowercase*) untuk menyeragamkan penulisan kata. Tahap ini bertujuan untuk menghindari perbedaan representasi kata akibat penggunaan huruf kapital, misalnya kata "Program", "PROGRAM", dan "program" yang sebenarnya memiliki makna yang sama. Selanjutnya dilakukan normalisasi untuk mengubah kata tidak baku, singkatan, atau bahasa informal menjadi bentuk kata baku sesuai kaidah bahasa Indonesia. Proses normalisasi penting dilakukan karena komentar media sosial umumnya menggunakan bahasa tidak formal dan singkatan yang beragam. Hasil proses *case folding* dan normalisasi ditunjukkan pada Tabel 3.

Tabel 3. Contoh *Case folding* dan Normalisasi Data

No	*Comment*	*Case folding*	Normalisasi
1	https://youtube.com/shorts/_D7ktyKNwns?si=Ec__...		
2	dengerin fakta emang bikin gelisah ya wk...	dengerin fakta emang bikin gelisah ya wk	mendengarkan fakta memang bikin gelisah ya waktu
3	Mau *speak up* disini, tmn w org china cerita di...	mau *speak up* disini tmn w org china cerita di...	mau *speak up* disini teman gue orang china cerita di...
4	Sebagai orang yang pernah mengalami susah maka...	sebagai orang yang pernah mengalami susah maka...	sebagai orang yang pernah mengalami susah maka...
5	bagus banget jdi slalu tau	bagus banget jdi slalu tau	bagus banget jadi selalu tau
6	Program MBG jangan hanya dilihat sebagai bagi-...	program mbg jangan hanya dilihat sebagai bagib...	program mbak jangan hanya dilihat sebagai bagi...

Berdasarkan Tabel 3 proses *case folding* berhasil mengubah seluruh huruf kapital menjadi huruf kecil sehingga penulisan kata menjadi lebih seragam. Selain itu, beberapa kata yang sebelumnya masih tidak konsisten menjadi lebih mudah diproses pada tahap berikutnya. Tahap ini penting untuk meningkatkan kualitas data teks sehingga proses *tokenization*, *stopword removal*, *stemming*, dan klasifikasi sentimen menggunakan algoritma Naive Bayes dan Support Vector Machine (SVM) dapat berjalan lebih optimal.

Building of Informatics, Technology and Science (BITS)
Volume 8, No 1, June 2026 Page: 623−634
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.10032

3.1.3 *Tokenization*, *Stopword removal*, dan *Stemming*

Tahap selanjutnya setelah *case folding* dan normalisasi adalah *tokenization*, *stopword removal*, dan *stemming*. Ketiga proses ini dilakukan untuk menyederhanakan data teks agar lebih mudah diproses pada tahap pembobotan fitur dan klasifikasi sentimen. *Tokenization* merupakan proses pemecahan kalimat menjadi kumpulan kata tunggal (*token*). Setelah itu dilakukan *stopword removal* untuk menghapus kata-kata umum yang tidak memiliki pengaruh signifikan terhadap sentimen, seperti kata hubung dan kata depan. Tahap terakhir adalah *stemming*, yaitu proses mengubah kata berimbuhan menjadi bentuk kata dasar agar kata yang memiliki arti sama dapat dikenali sebagai satu fitur yang seragam. Hasil proses *tokenization*, *stopword removal*, dan *stemming* ditunjukkan pada Tabel 4.

Tabel 4. *Tokenization*, *Stopword removal*, dan *Stemming*

No	*Tokenize*	*Stopword removal*	*Stemming*
1	-	-	-
2	[mendengarkan, fakta, memang, bikin, gelisah]	[mendengarkan, fakta, bikin, gelisah, ya]	dengar fakta bikin gelisah ya
3	[mau, speak, up, disini, teman, gue, orang, ch...]	[speak, up, teman, gue, orang, china, cerita]	*speak up* teman gue orang china cerita china mb...
4	[sebagai, orang, yang, pernah, mengalami, susa...]	[orang, mengalami, susah, makan, sehari, setuj...]	orang alami susah makan hari tuju program mbak...
5	[bagus, banget, jadi, selalu, tau]	[bagus, banget, tau]	bagus banget tau
6	[program, mbak, jangan, hanya, dilihat, sebaga...]	[program, mbak, bagibagi, makanan, motor, peng...]	program mbak bagibagi makan motor gerak ekonom...

Berdasarkan Tabel 4 proses *tokenization* berhasil memecah komentar menjadi kumpulan kata sehingga lebih mudah diproses oleh sistem. Selanjutnya, *stopword removal* berhasil menghapus kata-kata yang tidak memiliki pengaruh terhadap analisis sentimen sehingga hanya menyisakan kata-kata penting yang merepresentasikan isi komentar. Tahap *stemming* juga berhasil mengubah kata berimbuhan menjadi bentuk kata dasar, seperti kata "mendengarkan" menjadi "dengar" dan "mengalami" menjadi "alami". Hasil *stemming* membuat data menjadi lebih ringkas dan konsisten sehingga membantu meningkatkan kualitas fitur pada proses pembobotan TF-IDF dan klasifikasi sentimen menggunakan algoritma Naive Bayes dan Support Vector Machine (SVM).

3.2 Pembobotan Fitur Menggunakan TF-IDF

Setelah tahap *preprocessing* selesai dilakukan, data komentar diubah ke dalam bentuk numerik menggunakan metode TF-IDF (*Term Frequency–Inverse Document Frequency*). Metode ini digunakan untuk memberikan bobot pada setiap kata berdasarkan tingkat kemunculan kata pada suatu dokumen dan seluruh dokumen komentar. TF-IDF bertujuan untuk menentukan kata yang paling penting dan paling merepresentasikan isi komentar pengguna terkait Program Makan Bergizi Gratis (MBG). Kata yang sering muncul pada komentar tertentu namun jarang muncul pada komentar lain akan memiliki bobot lebih tinggi. Hasil pembobotan fitur menggunakan metode TF-IDF ditunjukkan pada Tabel 5.

Tabel 5. Kata dengan Bobot TF-IDF Tertinggi

No	Kata	Bobot TF-IDF
1	bergizi	0.084
2	gratis	0.079
3	bantu	0.072
4	anak	0.068
5	sekolah	0.064
6	anggaran	0.059
7	kualitas	0.053
8	makan	0.051
9	program	0.049
10	masyarakat	0.045

Berdasarkan Tabel 5 kata "bergizi" memiliki bobot TF-IDF tertinggi sebesar 0,084, diikuti kata "gratis" sebesar 0,079 dan "bantu" sebesar 0,072. Hasil tersebut menunjukkan bahwa kata-kata tersebut paling dominan dan paling merepresentasikan topik komentar masyarakat terkait Program Makan Bergizi Gratis (MBG). Pembobotan TF-IDF membantu sistem dalam memilih fitur kata yang paling relevan sehingga proses klasifikasi sentimen menggunakan algoritma Naive Bayes dan Support Vector Machine (SVM) dapat menghasilkan performa yang lebih baik.

3.3 *Wordcloud* Setelah *Preprocessing*

Setelah seluruh tahapan *preprocessing* selesai dilakukan, tahap berikutnya adalah visualisasi data menggunakan *wordcloud* dan grafik frekuensi kata. Visualisasi ini bertujuan untuk mengetahui kata-kata yang paling dominan

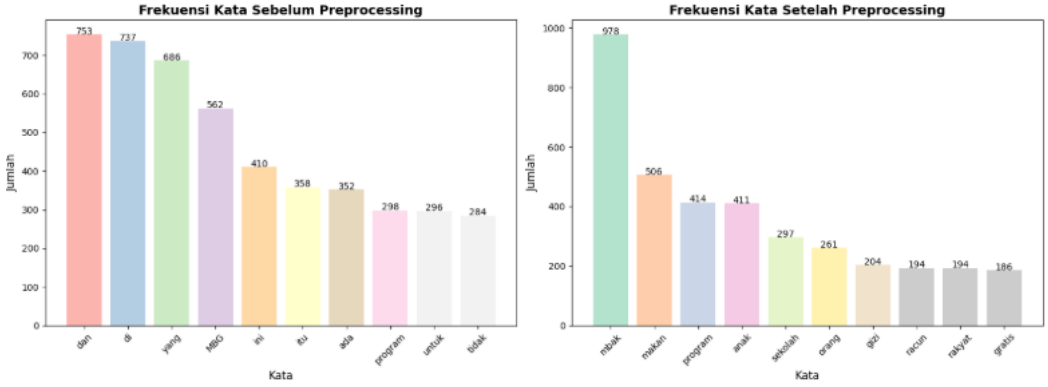
Building of Informatics, Technology and Science (BITS)
Volume 8, No 1, June 2026 Page: 623−634
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.10032

muncul pada komentar pengguna YouTube terkait Program Makan Bergizi Gratis (MBG) sebelum dan sesudah *preprocessing*. *Wordcloud* digunakan untuk memberikan gambaran visual mengenai kata yang paling sering muncul pada *dataset* komentar. Semakin besar ukuran suatu kata pada *wordcloud*, maka semakin tinggi frekuensi kemunculan kata tersebut dalam komentar pengguna. Visualisasi *wordcloud* sebelum dan sesudah proses *preprocessing* ditunjukkan pada Gambar 2.

(a) (b)

Gambar 1 (a) Sebelum *Preprocessing*. (b) Setelah *Preprocessing*

Pada Gambar 2 (a), sebelum *preprocessing* masih terdapat banyak kata umum seperti “dan”, “di”, “yang”, dan “untuk” yang mendominasi data. Setelah *preprocessing* pada Gambar 2 (b), kata-kata yang tidak relevan berhasil dikurangi sehingga kata yang berkaitan dengan topik penelitian seperti “makan”, “program”, “anak”, “gizi”, dan “gratis” menjadi lebih dominan. Selain *wordcloud*, dilakukan juga analisis frekuensi kata untuk melihat jumlah kemunculan kata sebelum dan sesudah *preprocessing*. Frekuensi kata sebelum dan sesudah proses *preprocessing* ditunjukkan pada Gambar 3.



(a) (b)

Gambar 2 (a) Frekuensi Kata sebelum *Preprocessing*. (b) Frekuensi Kata setelah *Preprocessing*

Berdasarkan Gambar 3 (a), sebelum *preprocessing* kata “dan” muncul sebanyak 753 kali, “di” sebanyak 737 kali, dan “yang” sebanyak 686 kali. Setelah *preprocessing* pada Gambar 3.2(b), kata yang paling dominan berubah menjadi “mbak” sebanyak 978 kali, diikuti “makan” sebanyak 506 kali dan “program” sebanyak 414 kali. Hasil visualisasi menunjukkan bahwa *preprocessing* berhasil mengurangi noise dan mempertahankan kata-kata yang lebih relevan dengan topik Program Makan Bergizi Gratis (MBG), sehingga data menjadi lebih baik untuk proses klasifikasi sentimen menggunakan Naive Bayes dan Support Vector Machine (SVM).

3.4 Pelabelan Data Menggunakan Metode *Lexicon-based*

Tahap pelabelan data dilakukan menggunakan metode *lexicon-based*, yaitu metode yang menentukan sentimen komentar berdasarkan skor polaritas kata positif dan negatif yang terdapat pada setiap komentar. Pada metode ini, setiap kata akan dicocokkan dengan kamus sentimen (*lexicon*) untuk mengetahui nilai polaritasnya. Jika jumlah skor bernilai positif maka komentar akan diberi label positif, sedangkan jika skor bernilai negatif maka komentar akan diberi label negatif. Komentar dengan skor netral atau bernilai nol akan diberi label netral. Tahap pelabelan data sangat penting dalam penelitian ini karena hasil label sentimen digunakan sebagai data target pada proses klasifikasi menggunakan algoritma Naive Bayes dan Support Vector Machine (SVM). Dengan adanya proses pelabelan, sistem dapat mempelajari pola sentimen masyarakat terhadap Program Makan Bergizi Gratis (MBG). Hasil pelabelan sentimen menggunakan metode *lexicon-based* ditunjukkan pada Tabel 6.

Tabel 6. Contoh Pelabelan Data

No	*Stemming* Data	Score	Sentiment
1	dengar fakta bikin gelisah ya	0	Netral

Building of Informatics, Technology and Science (BITS)
Volume 8, No 1, June 2026 Page: 623−634
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.10032

No	*Stemming* Data	Score	Sentiment
2	*speak up* teman gue orang china cerita china mb...	1	Positif
3	orang alami susah makan hari tuju program mbak...	-6	Negatif
4	bagus banget tau	0	Netral
5	program mbak bagibagi makan motor gerak ekonom...	1	Positif
6	program mbak bagibagi makan motor gerak ekonom...	1	Positif

Berdasarkan Tabel 6 terlihat bahwa komentar dengan skor positif diberi label sentimen positif, komentar dengan skor negatif diberi label sentimen negatif, sedangkan komentar dengan skor nol diberi label netral. Nilai skor tertinggi terdapat pada komentar "tuju mbak henti cegah stunting sedia pasar pet..." dengan skor 5 yang menunjukkan sentimen positif yang kuat terhadap Program Makan Bergizi Gratis (MBG). Sementara itu, komentar "orang alami susah makan hari tuju program mbak..." memperoleh skor -6 yang menunjukkan sentimen negatif yang cukup kuat. Selain itu, terdapat beberapa komentar dengan skor 0 seperti "dengar fakta bikin gelisah ya" dan "proyek" yang dikategorikan sebagai sentimen netral karena tidak menunjukkan kecenderungan opini positif maupun negatif secara jelas. Hasil pelabelan ini menunjukkan bahwa komentar masyarakat memiliki variasi sentimen yang beragam terhadap Program Makan Bergizi Gratis (MBG). Data yang telah diberi label tersebut selanjutnya digunakan pada tahap klasifikasi sentimen menggunakan algoritma Naive Bayes dan Support Vector Machine (SVM).

3.5 Klasifikasi Sentimen

Pada tahap klasifikasi sentimen, data yang telah melalui proses *preprocessing* dan pembobotan fitur menggunakan TF-IDF kemudian dibagi menjadi data latih (*training data*) dan data uji (*testing data*). Pembagian data dilakukan menggunakan metode *split data* dengan perbandingan 80:20. Data latih digunakan untuk membangun model klasifikasi, sedangkan data uji digunakan untuk mengukur performa model dalam mengklasifikasikan sentimen komentar pengguna YouTube terkait Program Makan Bergizi Gratis (MBG). Pembagian data *training* dan *testing* pada penelitian ini ditunjukkan pada Tabel 7.

Tabel 7. Pembagian Data Training dan Data Testing

Jenis Data	Jumlah Data	Persentase
Data Training	1.219	79,99%
Data Testing	305	20,01%
Total	1.524	100%

Berdasarkan Table 7 jumlah data latih yang digunakan sebanyak 1.219 data atau 79,99% dari total *dataset*, sedangkan jumlah data uji sebanyak 305 data atau 20,01%. Pembagian data tersebut dilakukan agar model dapat mempelajari pola sentimen pada data latih dan kemudian diuji menggunakan data yang belum pernah dipelajari sebelumnya sehingga hasil evaluasi model menjadi lebih objektif.

3.5.1 *Naive Bayes*

Model klasifikasi menggunakan algoritma Naive Bayes diterapkan pada data hasil *preprocessing* dan pembobotan TF-IDF. Hasil *classification report* algoritma Naive Bayes ditunjukkan pada Gambar 4.

Classification Report for Naive Bayes:

	precision	recall	f1-score	support
Negatif	0.717	0.839	0.773	118.000
Netral	0.913	0.280	0.429	75.000
Positif	0.660	0.848	0.742	112.000
accuracy	0.705	0.705	0.705	0.705
macro avg	0.763	0.656	0.648	305.000
weighted avg	0.744	0.705	0.677	305.000

Gambar 4. *Classification report* Naive Bayes

Berdasarkan Gambar 4, algoritma Naive Bayes menghasilkan nilai *accuracy* sebesar 0,705 atau 70,5%. Pada kelas sentimen negatif diperoleh nilai *precision* sebesar 0,717, *precision* sebesar 0,839, dan *f1-score* sebesar 0,773. Pada kelas sentimen netral diperoleh nilai *precision* sebesar 0,913, namun *precision* hanya sebesar 0,280 sehingga menunjukkan bahwa model masih kesulitan mengenali komentar netral. Sementara itu, pada kelas sentimen positif diperoleh nilai *precision* sebesar 0,660, *precision* sebesar 0,848, dan *f1-score* sebesar 0,742. Hasil tersebut menunjukkan bahwa Naive Bayes cukup baik dalam mengenali komentar positif dan negatif, namun performanya masih kurang optimal pada kelas netral.

Building of Informatics, Technology and Science (BITS)
Volume 8, No 1, June 2026 Page: 623−634
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.10032

3.5.2 Support Vector Machine (SVM)

Model klasifikasi menggunakan algoritma Support Vector Machine (SVM) juga diterapkan pada dataset yang sama untuk membandingkan performanya dengan algoritma Naive Bayes. Hasil ***classification report*** algoritma Support Vector Machine (SVM) ditunjukkan pada Gambar 5.

Classification Report for SVM:

	precision	recall	f1-score	support
Negatif	0.864	0.754	0.805	118.000
Netral	0.636	0.747	0.687	75.000
Positif	0.798	0.812	0.805	112.000
accuracy	0.774	0.774	0.774	0.774
macro avg	0.766	0.771	0.766	305.000
weighted avg	0.784	0.774	0.776	305.000

Gambar 5. *Classification report* SVM

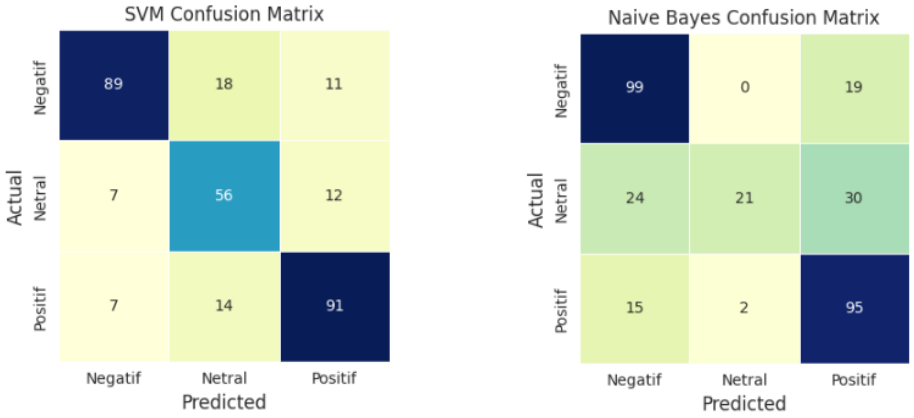
Berdasarkan Gambar 5, algoritma SVM menghasilkan nilai *accuracy* sebesar 0,774 atau 77,4%, lebih tinggi dibandingkan Naive Bayes. Pada kelas sentimen negatif diperoleh nilai *precision* sebesar 0,864, *precision* sebesar 0,754, dan *f1-score* sebesar 0,805. Pada kelas sentimen netral diperoleh nilai *precision* sebesar 0,636, *precision* sebesar 0,747, dan *f1-score* sebesar 0,687. Sementara itu, pada kelas sentimen positif diperoleh nilai *precision* sebesar 0,798, *precision* sebesar 0,812, dan *f1-score* sebesar 0,805. Hasil tersebut menunjukkan bahwa algoritma SVM memiliki performa yang lebih stabil dan lebih baik dibandingkan Naive Bayes dalam melakukan klasifikasi sentimen komentar YouTube terkait Program Makan Bergizi Gratis (MBG).

3.6 Evaluasi Model

Tahap evaluasi model dilakukan untuk mengetahui tingkat performa algoritma dalam mengklasifikasikan sentimen komentar pengguna YouTube terkait Program Makan Bergizi Gratis (MBG). Evaluasi model dilakukan menggunakan *confusion matrix* serta metrik *accuracy*, *precision*, *precision*, dan *f1-score*.

3.6.1 *Confusion matrix*

Confusion matrix digunakan untuk melihat jumlah prediksi benar dan kesalahan klasifikasi yang dilakukan oleh model pada masing-masing kelas sentimen. Melalui *confusion matrix*, dapat diketahui bagaimana kemampuan model dalam membedakan komentar positif, negatif, dan netral. Hasil *confusion matrix* algoritma SVM dan Naive Bayes ditunjukkan pada Gambar 6.



(a) *Confusion matrix SVM* (b) *Confusion matrix Naive Bayes*

Gambar 3 (a) *Confusion matrix* SVM. (b) *Confusion matrix* Naive Bayes

Berdasarkan *confusion matrix* SVM pada Gambar 6 (a), model berhasil memprediksi 89 data negatif dengan benar, 56 data netral dengan benar, dan 91 data positif dengan benar. Namun masih terdapat beberapa kesalahan klasifikasi, seperti 18 data negatif yang diprediksi sebagai netral dan 14 data positif yang diprediksi sebagai netral. Sementara itu, berdasarkan *confusion matrix* Naive Bayes pada Gambar 6 (b), model berhasil memprediksi 99 data negatif dengan benar dan 95 data positif dengan benar. Namun pada kelas netral, model hanya mampu memprediksi 21 data dengan benar, sedangkan sebagian besar data netral diprediksi sebagai negatif sebanyak 24 data dan positif sebanyak 30 data. Hasil *confusion matrix* menunjukkan bahwa algoritma SVM memiliki kemampuan yang lebih baik

Building of Informatics, Technology and Science (BITS)
Volume 8, No 1, June 2026 Page: 623−634
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.10032

dalam mengenali seluruh kelas sentimen secara seimbang dibandingkan Naive Bayes, khususnya pada kelas sentimen netral.

3.6.2 Evaluasi Menggunakan Akurasi, *Precision*, *Precision*, dan *F1-score*

Evaluasi performa model dilakukan menggunakan metrik *accuracy*, *precision*, *precision*, dan *f1-score* untuk mengetahui tingkat kemampuan algoritma dalam mengklasifikasikan sentimen komentar pengguna YouTube terkait Program Makan Bergizi Gratis (MBG). Penggunaan beberapa metrik evaluasi diperlukan agar performa model dapat dianalisis secara lebih menyeluruh, tidak hanya berdasarkan tingkat ketepatan prediksi secara umum, tetapi juga berdasarkan kemampuan model dalam mengenali setiap kelas sentimen. Nilai *accuracy* digunakan untuk mengukur persentase keseluruhan data yang berhasil diklasifikasikan dengan benar oleh model. Nilai *precision* digunakan untuk melihat tingkat ketepatan model dalam memberikan prediksi pada suatu kelas sentimen. Sementara itu, nilai *precision* digunakan untuk mengukur kemampuan model dalam menemukan seluruh data yang termasuk ke dalam kelas tertentu. Adapun *f1-score* merupakan kombinasi harmonis antara *precision* dan *precision* yang digunakan untuk melihat keseimbangan performa model klasifikasi.

Berdasarkan hasil *classification report*, algoritma Support Vector Machine (SVM) memperoleh nilai *accuracy* sebesar 77,4%, *precision* rata-rata sebesar 78,4%, *precision* sebesar 77,4%, dan *f1-score* sebesar 77,6%. Nilai tersebut menunjukkan bahwa sebagian besar komentar berhasil diklasifikasikan dengan benar oleh model SVM. Selain itu, nilai *precision* dan *precision* yang relatif seimbang menunjukkan bahwa algoritma SVM memiliki kemampuan yang cukup baik dalam mengenali komentar positif, negatif, maupun netral secara stabil. Sementara itu, algoritma Naive Bayes memperoleh nilai *accuracy* sebesar 70,5%, *precision* sebesar 74,4%, *precision* sebesar 70,5%, dan *f1-score* sebesar 67,7%. Meskipun Naive Bayes mampu mengenali komentar negatif dan positif dengan cukup baik, performa model masih kurang optimal pada kelas sentimen netral. Hal tersebut terlihat dari nilai *f1-score* yang lebih rendah dibandingkan algoritma SVM.

Perbedaan hasil evaluasi tersebut menunjukkan bahwa algoritma SVM memiliki performa klasifikasi yang lebih baik dibandingkan Naive Bayes dalam penelitian ini. Nilai *accuracy* dan *f1-score* yang lebih tinggi menunjukkan bahwa SVM lebih efektif dalam mengenali pola sentimen masyarakat terhadap Program Makan Bergizi Gratis (MBG). Selain itu, SVM juga mampu menghasilkan performa yang lebih stabil pada seluruh kategori sentimen sehingga tingkat kesalahan klasifikasi menjadi lebih rendah. Hasil evaluasi ini memberikan kontribusi penting dalam penelitian karena menunjukkan bahwa pemilihan algoritma sangat memengaruhi kualitas hasil analisis sentimen. Penggunaan algoritma SVM pada data komentar YouTube terbukti lebih sesuai untuk menangani data teks dengan variasi bahasa yang beragam dan jumlah fitur yang besar dibandingkan algoritma Naive Bayes. Dengan demikian, penelitian ini menunjukkan bahwa algoritma Support Vector Machine (SVM) merupakan metode yang lebih efektif untuk digunakan dalam analisis sentimen komentar YouTube terkait Program Makan Bergizi Gratis (MBG). Hasil penelitian ini diharapkan dapat menjadi referensi bagi penelitian selanjutnya dalam pengembangan analisis sentimen berbasis media sosial, khususnya pada data komentar masyarakat terhadap kebijakan atau program pemerintah.

3.7 Pembahasan

Berdasarkan hasil evaluasi model, algoritma Support Vector Machine (SVM) memperoleh nilai accuracy sebesar 77,4%, sedangkan algoritma Naive Bayes memperoleh nilai accuracy sebesar 70,5%**.** Selain itu, nilai *F1-score* SVM mencapai 77,6%, lebih tinggi dibandingkan Naive Bayes yang memperoleh 67,7%. Hasil tersebut menunjukkan bahwa SVM memiliki kemampuan yang lebih baik dalam mengklasifikasikan sentimen komentar YouTube terkait Program Makan Bergizi Gratis (MBG). Keunggulan SVM juga terlihat pada confusion matrix, di mana model mampu mengenali ketiga kelas sentimen, yaitu positif, negatif, dan netral secara lebih seimbang dibandingkan Naive Bayes. Sebaliknya, Naive Bayes masih mengalami kesulitan dalam mengklasifikasikan komentar dengan sentimen netral sehingga menghasilkan nilai recall dan F1-score yang lebih rendah pada kelas tersebut.

Hasil penelitian ini sejalan dengan penelitian Ghofur dan Putri [5**]**, yang menyatakan bahwa algoritma Support Vector Machine menghasilkan performa klasifikasi yang lebih baik dibandingkan Naive Bayes dan K-Nearest Neighbor (KNN). Pada penelitian tersebut, SVM mampu memperoleh tingkat akurasi sebesar 91,7%, sehingga menunjukkan bahwa algoritma ini memiliki kemampuan yang baik dalam menangani data teks berdimensi tinggi. Kesamaan hasil tersebut menunjukkan bahwa SVM mampu membentuk batas pemisah (hyperplane) yang optimal sehingga proses klasifikasi menjadi lebih akurat, terutama ketika data telah melalui tahapan preprocessing dan pembobotan fitur menggunakan TF-IDF.

Temuan penelitian ini juga didukung oleh penelitian-penelitian terdahulu yang menyatakan bahwa kombinasi preprocessing yang baik dengan pembobotan fitur TF-IDF mampu meningkatkan kualitas representasi data teks sebelum proses klasifikasi dilakukan. Pada penelitian ini, tahapan cleaning, case folding, normalisasi, tokenization, stopword removal, dan stemming berhasil mengurangi noise pada data komentar YouTube. Setelah melalui proses tersebut, kata-kata yang memiliki makna penting menjadi lebih dominan sehingga model memperoleh fitur yang lebih representatif. Kondisi ini berkontribusi terhadap peningkatan performa klasifikasi, khususnya pada algoritma SVM yang mampu memanfaatkan representasi fitur berdimensi tinggi secara lebih efektif dibandingkan Naive Bayes.

Di sisi lain, hasil penelitian ini berbeda dengan penelitian Hidayat dan Ratnaningsih [4], yang melaporkan bahwa algoritma Random Forest memperoleh tingkat akurasi tertinggi sebesar 96,38%. Perbedaan hasil tersebut dapat dipengaruhi oleh beberapa faktor, antara lain karakteristik dataset, jumlah data yang digunakan, metode pelabelan

Building of Informatics, Technology and Science (BITS)
Volume 8, No 1, June 2026 Page: 623−634
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.10032

sentimen, teknik preprocessing, serta algoritma klasifikasi yang diterapkan. Selain itu, penelitian ini menggunakan komentar YouTube sebagai sumber data, sedangkan sebagian besar penelitian terdahulu menggunakan data dari platform X (Twitter). Komentar pada YouTube umumnya memiliki panjang kalimat yang lebih beragam, mengandung bahasa tidak baku, singkatan, emoji, serta campuran bahasa Indonesia dan bahasa asing. Karakteristik tersebut menyebabkan proses klasifikasi menjadi lebih kompleks sehingga nilai akurasi yang diperoleh cenderung lebih rendah dibandingkan penelitian yang menggunakan dataset dengan karakteristik lebih homogen.

Selain dipengaruhi oleh karakteristik data, perbedaan performa antara SVM dan Naive Bayes juga dipengaruhi oleh mekanisme kerja masing-masing algoritma. Naive Bayes menggunakan asumsi bahwa setiap fitur bersifat independen, sedangkan pada data teks sering kali terdapat hubungan antar kata yang saling memengaruhi. Asumsi tersebut menyebabkan performa Naive Bayes menurun ketika menghadapi data komentar yang memiliki variasi bahasa dan konteks yang kompleks. Sebaliknya, SVM mampu menemukan hyperplane optimal yang memisahkan setiap kelas sentimen sehingga lebih efektif dalam menangani data berdimensi tinggi seperti hasil pembobotan TF-IDF. Oleh karena itu, SVM mampu menghasilkan nilai accuracy dan F1-score yang lebih tinggi dibandingkan Naive Bayes pada penelitian ini.

Berdasarkan keseluruhan hasil penelitian, dapat disimpulkan bahwa algoritma Support Vector Machine merupakan metode yang lebih efektif dibandingkan Naive Bayes dalam mengklasifikasikan sentimen komentar YouTube mengenai Program Makan Bergizi Gratis (MBG). Penelitian ini memberikan kontribusi berupa bukti empiris bahwa kombinasi tahapan preprocessing, pembobotan fitur TF-IDF, serta algoritma SVM mampu menghasilkan performa klasifikasi yang lebih baik pada data komentar media sosial. Temuan ini diharapkan dapat menjadi referensi bagi penelitian selanjutnya dalam memilih algoritma klasifikasi sentimen, khususnya pada penelitian yang memanfaatkan data media sosial sebagai sumber informasi untuk mengevaluasi opini publik terhadap kebijakan atau program pemerintah.

4. KESIMPULAN

Penelitian ini berhasil melakukan analisis sentimen terhadap komentar pengguna YouTube terkait Program Makan Bergizi Gratis (MBG) menggunakan algoritma Naive Bayes dan Support Vector Machine (SVM). Penelitian dilakukan untuk mengetahui sentimen masyarakat terhadap Program Makan Bergizi Gratis (MBG) serta membandingkan performa kedua algoritma dalam mengklasifikasikan komentar ke dalam kategori sentimen positif, negatif, dan netral. Berdasarkan hasil pengujian dan evaluasi model, algoritma Support Vector Machine (SVM) memperoleh performa yang lebih baik dibandingkan Naive Bayes. Algoritma SVM menghasilkan nilai *accuracy* sebesar 77,4%, *precision* sebesar 78,4%, *recall* sebesar 77,4%, dan *f1-score* sebesar 77,6%. Sementara itu, algoritma Naive Bayes memperoleh nilai *accuracy* sebesar 70,5%, *precision* sebesar 74,4%, *recall* sebesar 70,5%, dan *f1-score* sebesar 67,7%. Hasil tersebut menunjukkan bahwa algoritma SVM lebih efektif dalam mengenali pola sentimen pada komentar pengguna YouTube terkait Program Makan Bergizi Gratis (MBG) dibandingkan algoritma Naive Bayes. Penelitian ini memberikan kontribusi dalam penerapan analisis sentimen berbasis *machine learning* pada komentar YouTube terkait kebijakan pemerintah, khususnya Program Makan Bergizi Gratis (MBG). Selain itu, penelitian ini juga memberikan perbandingan performa algoritma Naive Bayes dan Support Vector Machine (SVM) sehingga dapat menjadi referensi dalam pemilihan metode klasifikasi sentimen pada data berbahasa Indonesia. Hasil penelitian ini diharapkan dapat membantu pemerintah dalam memahami opini masyarakat terhadap Program Makan Bergizi Gratis (MBG). Namun, penelitian ini masih memiliki keterbatasan, yaitu jumlah data yang digunakan relatif terbatas dan proses pelabelan sentimen masih menggunakan metode *lexicon-based* sehingga masih berpotensi terjadi kesalahan klasifikasi. Oleh karena itu, penelitian selanjutnya disarankan menggunakan jumlah data yang lebih besar, menerapkan hybrid labeling, serta menggunakan metode deep learning agar hasil klasifikasi menjadi lebih akurat.

REFERENCES

[1] A. Putriyekti *et al.*, “Analisis Multidimensional Sentimen Masyarakat Terhadap Program Makan Bergizi Gratis Pada Media Sosial X,” *Integrative Perspectives of Social and Science Journal*, vol. 2, no. 1, pp. 1004–1024, Feb. 2025, doi: https://ipssj.com/index.php/ojs/article/view/142.

[2] I. Abdul Ghofur, A. Dasa Putri, A. Sentimen, N. Bayes, and P. Makan Bergizi Gratis, “Perbandingan Kinerja Svm, Knn, Dan Naïve Bayes Pada Analisis Sentimen Tweet Mbg Kata Kunci,” *Computer Based Information System Journal*, vol. 14, no. 01, pp. 100–117, Mar. 2026, doi: https://doi.org/10.33884/cbis.v14i1.10994.

[3] R. Hidayat and D. J. Ratnaningsih, “Analisis Sentimen Program Makanan Bergizi Gratis Menggunakan Algoritma Random Forest dan Naive Bayes,” *Journal of Computing and Informatics Research*, vol. 5, no. 1, pp. 395–400, 2025, doi: 10.47065/comforch.v5i1.2355.

[4] K. Yoga Pratama, S. Achmadi, and K. A. Sari, “Analisis Sentimen Terhadap Makan Bergizi Gratis Pada Sosial Media X Menggunakan Metode Decicion Tree Dengan Pembanding Naive Bayes,” *Jurnal Mahasiswa Teknik Informatika)*, vol. 9, no. 6, pp. 11061–11067, Dec. 2025, doi: https://doi.org/10.36040/jati.v9i6.15525.

[5] P. M. Kadafi, H. Z. Riyadi, R. S. G. Raya, Ika Kurniawati, Waeisul Bismi, and Riza Fahlapi, “Penerapan Machine Learning Untuk Analisis Sentimen Program Mbg Pada Platform X Dan Youtube,” *INFOTECH journal*, vol. 12, no. 1, pp. 67–74, Mar. 2026, doi: 10.31949/infotech.v12i1.17472.

Building of Informatics, Technology and Science (BITS)
Volume 8, No 1, June 2026 Page: 623−634
ISSN 2684-8910 (media cetak)
ISSN 2685-3310 (media online)
DOI 10.47065/bits.v8i1.10032

[6] Z. Fatah and R. A. Ningsih, "Analisis Sentimen Komentar YouTube terhadap Tragedi Demo 25 Agustus Menggunakan Pendekatan Lexicon-Based," *JAMASTIKA*, vol. 4, pp. 217–224, Oct. 2025, doi: https://doi.org/10.35473/jamastika.v4i2.4525.

[7] N. D. Firdaus and R. R. Suryono, "Perbandingan Algoritma Naive Bayes dan SVM dalam Analisis Sentimen Pengguna AI di Platform X," *Technology and Science (BITS)*, vol. 6, no. 4, 2025, doi: 10.47065/bits.v6i4.7081.

[8] R. Marta Dinata, E. Rayhana, V. Hadi, and U. Alkaf, "Analisis Komprehensif Kinerja Model Klasifikasi Sentimen: Evaluasi Lintas Metrik Pada Dataset Tweet Film Bahasa Indonesia," *Jurnal Rekayasa Informasi*, vol. 14, no. 1, pp. 38–47, Jul. 2025, doi: https://journal.istn.ac.id/index.php/rekayasainformasi/article/view/2361/1472.

[9] W. Rayyana Muntaza, "Analisis Sentimen Tanggapan Masyarakat Terhadap Program Makan Bergizi Gratis Pada Platform YouTube Menggunakan SVM," *PROSIDING SEMINAR NASIONAL TEKNOLOGI DAN SAINS*, vol. 5, pp. 297–302, Jan. 2026, doi: https://doi.org/10.29407/d2x6hw85.

[10] O. I. Gifari, M. Adha, I. Rifky Hendrawan, F. Freddy, and S. Durrand, "Analisis Sentimen Review Film Menggunakan TF-IDF dan Support Vector Machine," *JIFOTECH (JOURNAL OF INFORMATION TECHNOLOGY*, vol. 2, no. 1, pp. 36–40, Mar. 2022, doi: 10.46229/jifotech.v2i1.330.

[11] T. A. Y. S. N. A. V. Mi'raj Fattah, "Analisis Komparasi TF-IDF dan Word2Vec pada Algoritma SVM Untuk Sentimen Pembangunan IKN di YouTube," *JURNAL BUFFER INFORMATIKA*, vol. 12, no. 1, pp. 25–34, Oct. 2026, doi: https://doi.org/10.25134/buffer.v12i1.551.

[12] R. Yulika Go, H. Sarah Divanda Gaspersz, R. Hendra Kusumawardhana, and N. Aeni Hidayah, "Analisis Sentimen Publik Program Makan Bergizi Gratis (MBG) di Youtube: Perbandingan Kinerja Algoritma Support Vector Machine (SVM) dan Random Forest," *JURNAL ILMU KOMPUTER DAN TEKNOLOGI (IKOMTI)*, vol. 7, no. 1, pp. 24–31, Feb. 2026, doi: 10.35960/ikomti.v7i1.2424.

[13] A. F. dan J. I. Toif Muhayat, "Analisis Sentimen Terhadap Komentar Video Youtube Menggunakan Support Vector Machines," *Progresif: Jurnal Ilmiah Komputer*, vol. 19, pp. 231–240, Feb. 2023, doi: 10.35889/progresif.v19i1.1060.

[14] S. Sofiana and A. Fitrianingsih, "Klasifikasi Sentimen Publik Terhadap Makanan Bergizi Gratis Berbasis Algoritma Decision Tree," *IKN: Jurnal Informatika dan Kesehatan*, vol. 3, no. 1, pp. 13–21, Feb. 2026, doi: https://doi.org/10.35473/ikn.v3i1.4885.

[15] R. Amelia Br Siregar, A. Ramadhan Manik, A. Syahri, M. Budi Akbar, F. Ramadhani, and J. W. Iskandar Psr V Medan Esatate Kab Deli Serdang, "Penerapan Naïve Bayes Untuk Analisis Sentimen Netizen Terhadap Program Makan Siang Gratis Pada Media Social X," *Jurnal Mahasiswa Teknik Informatika)*, vol. 9, no. 4, pp. 5621–5628, 2025, doi: https://doi.org/10.36040/jati.v9i4.13870.

[16] F. C. E. W. Adi Ariyo Munandar, "Sentimen Analisis Aplikasi Belajar Online Menggunakan Klasifikasi SVM," *JOINTECS (Journal of Information Technology and Computer Science)*, vol. 7, no. 2, pp. 77–84, Jul. 2023, doi: JOINTECS(JournalofInformationTechnologyandComputerScience).

[17] Z. Prisdian Tiararosa, I. Sani Wijaya, and E. Rohaini, "Perbandingan SVM dan Naive Bayes pada Analisis Sentimen MBG di Platform TikTok dan Instagram," *Jurnal Informatika Dan Rekayasa Komputer (JAKAKOM)*, vol. 1, p. 1983, Apr. 2026, doi: 10.33998/jakakom.v6i1.

[18] P. Prayesy, A. Pujakesuma, and M. R. Qisthiano, "Evaluasi Kinerja Random Forest dan Naïve Bayes untuk Prediksi Risiko Kredit Berdasarkan Pekerjaan Debitur," *Explore*, vol. 15, no. 2, pp. 114–120, Jul. 2025, doi: 10.35200/ex.v15i2.175.

[19] B. Hakim, "Analisa Sentimen Data Text Preprocessing Pada Data Mining Dengan Menggunakan Machine Learning," *JBASE - Journal of Business and Audit Information Systems*, vol. 4, no. 2, Aug. 2021, doi: 10.30813/jbase.v4i2.3000.