

Perbandingan Algoritma SVM dan Logistic Regression dalam Klasifikasi Sentimen Masyarakat terhadap Program MBG

Ulfi Muzayyanah Fadil*, Aidil Halim Lubis

Sains dan Teknologi, Ilmu Komputer, Universitas Islam negeri Sumatera Utara, Medan, Indonesia

INFORMASI ARTIKEL

Diterima Redaksi: 13 Agustus 2026
Revisi Akhir: 27 September 2026
Diterbitkan Online: 28 September 2026

KATA KUNCI

Analisis Sentimen
Klasifikasi Teks
Logistic Regression
MBG
SVM

KORESPONDENSI (*)

E-mail: ulfimuzayyanah826@gmail.com

A B S T R A K

Penelitian ini bertujuan untuk membandingkan kinerja algoritma *Support Vector Machine* (SVM) dan *Logistic Regression* dalam klasifikasi sentimen masyarakat terhadap Program Makan Bergizi Gratis (MBG). Mengingat besarnya volume data teks tidak terstruktur pada media digital, evaluasi sentimen secara konvensional menjadi tidak efektif. Penelitian ini mengimplementasikan metode *supervised learning* untuk memetakan polaritas opini publik secara otomatis. Data penelitian diperoleh melalui teknik *web crawling* pada media sosial X (Twitter) dengan total 1.019 data komentar. Setelah melalui tahapan *preprocessing* yang meliputi *cleaning*, *case folding*, *tokenizing*, *filtering*, *normalisasi*, *stopword removal*, dan *stemming*, diperoleh 799 data bersih yang kemudian diekstraksi fiturnya menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF). Berdasarkan hasil pengujian menggunakan *confusion matrix*, algoritma SVM menunjukkan kinerja yang lebih unggul dengan nilai akurasi 82,50%, presisi 82,69%, *recall* 99,23%, dan *F1-score* 90,21%. Sementara itu, algoritma *Logistic Regression* menghasilkan nilai akurasi sebesar 81,25%. Hasil penelitian ini menyimpulkan bahwa SVM lebih efektif dalam mengklasifikasikan sentimen masyarakat terhadap Program MBG, yang dapat menjadi rekomendasi bagi pemangku kebijakan dalam memantau opini publik.

PENDAHULUAN

Efisiensi serta ketepatan pemilihan model klasifikasi komputasi menjadi faktor krusial dalam implementasi analisis sentimen berbasis teks terhadap isu kebijakan publik, seperti halnya program Makan Bergizi Gratis (MBG) yang marak dipublikasikan oleh media massa. Mengingat besarnya volume data teks tidak terstruktur yang tersebar pada artikel berita dan kolom diskusi media digital resmi, proses evaluasi sentimen secara konvensional atau manual dinilai tidak efektif serta rentan terhadap subjektivitas[1]

Dalam ranah pembelajaran mesin (machine learning), algoritma Support Vector Machine (SVM) dan Logistic Regression muncul sebagai dua metode klasifikasi terbimbing (supervised learning) yang sangat kompetitif untuk menangani karakteristik data teks. Walaupun demikian, tantangan metodologis yang mendasar muncul ketika harus menguji secara empiris performa kedua algoritma tersebut pada karakteristik unik data portal berita resmi yang memiliki tingkat ambiguitas tinggi, variasi gaya bahasa jurnalistik, serta perpaduan sentimen yang kompleks dalam satu kesatuan artikel [2]. Oleh karena itu, studi komparatif langsung antara kedua metode ini diperlukan guna membuktikan apakah pendekatan geometris berbasis margin pemisah maksimum ataukah pendekatan statistik berbasis probabilitas yang memiliki adaptabilitas lebih tinggi dalam memetakan polaritas opini publik secara akurat.

Fokus analisis komparatif ini didasari oleh dilema antara batas linearitas data dan kompleksitas ruang dimensi fitur teks. Proses ekstraksi fitur menggunakan metode Term Frequency-Inverse Document Frequency (TF-IDF) pada dokumen berita akan menghasilkan matriks data bernilai numerik dengan dimensi yang sangat tinggi. Algoritma Logistic

Regression bekerja dengan asumsi dasar bahwa hubungan antar fitur kata bersifat linear. Namun pada kenyataannya, narasi opini masyarakat di media resmi kerap memiliki batas kelas sentimen yang non-linear akibat kerumitan bahasa sosiologis dan konteks kalimat [3]. Melalui penelitian komparatif ini, akan dibuktikan secara empiris apakah penyederhanaan linear milik Logistic Regression cukup tangguh mereduksi noise, ataukah superioritas algoritma SVM diperlukan untuk memisahkan batas kelas yang rumit secara optimal dengan memaksimalkan margin antara hyperplane dan support vectors terdekat[4].

Di samping karakteristik geometris data, urgensi membandingkan kedua algoritma ini dipicu oleh adanya pertukaran timbal-balik (trade-off) antara tingkat akurasi generalisasi model dan nilai interpretabilitasnya bagi pemangku kepentingan. Pihak penentu kebijakan publik tidak sekadar membutuhkan sistem cerdas yang mampu memprediksi sentimen masyarakat dengan akurasi tinggi, melainkan juga memerlukan keterbukaan informasi model yang dapat ditafsirkan guna mendukung evaluasi dan perbaikan program MBG secara praktis di lapangan[5]. Logistic Regression menawarkan keunggulan berupa nilai interpretabilitas fitur yang kuat melalui nilai koefisiennya, memudahkan identifikasi kata kunci yang paling berdampak secara signifikan terhadap pembentukan sentimen positif maupun negatif di media sosial [6]. Sebaliknya, meskipun SVM dikenal memiliki tingkat keakuratan yang tinggi pada dataset teks kompleks, mekanisme internal algoritma ini cenderung bersifat kotak hitam (black-box) yang sulit diterjemahkan secara intuitif bagi pengambil kebijakan[7].

Kondisi ini kian dipertegas dengan fakta bahwa sebagian besar penelitian komparatif sebelumnya umumnya didominasi oleh domain komersial seperti ulasan retail atau e-commerce yang memiliki polaritas kata kunci seragam. Sebaliknya, data teks mengenai kebijakan publik seperti program MBG melibatkan ekspresi emosional terkait isu kesehatan nasional dan alokasi anggaran negara yang strukturnya jauh lebih rumit. Keterbatasan studi komparatif pada ranah evaluasi kebijakan berbasis data berita resmi ini menyebabkan belum adanya rekomendasi model klasifikasi yang optimal [1]. Oleh karena itu, penelitian ini dirancang untuk membandingkan performa kedua algoritma tersebut secara komprehensif menggunakan metrik accuracy, precision, recall, dan F1-score. Hasil penelitian ini diharapkan memberikan kontribusi metodologis berupa rekomendasi ilmiah mengenai model klasifikasi yang paling seimbang antara performa akurasi dan nilai interpretasi dalam mendukung sistem pemantauan opini publik secara berkelanjutan[5].

TINJAUAN PUSTAKA

Text Mining

Text mining merupakan teknik untuk mengolah dan menganalisis data teks yang tidak terstruktur maupun semi-terstruktur dalam jumlah besar guna menemukan informasi dan pola tertentu [8]. Tahapan *text mining* meliputi normalisasi teks untuk mengubah kata tidak baku menjadi bentuk standar, *text cleaning* untuk menghapus karakter yang tidak relevan, *case folding* untuk mengubah huruf menjadi *lowercase*, tokenisasi untuk memecah teks menjadi kata, *stopword removal* untuk menghapus kata yang kurang bermakna, serta *stemming* untuk mengubah kata berimbuhan menjadi kata dasar.

Klasifikasi Teks

Klasifikasi teks (*text classification*) merupakan pilar utama dalam *Natural Language Processing* (NLP) yang bertujuan untuk mengelompokkan dokumen ke dalam kategori tertentu secara otomatis[9]. Secara teknis, proses klasifikasi teks bekerja melalui tahapan yang sistematis, dimulai dari pra-pengolahan data (*text preprocessing*) untuk membersihkan teks dari gangguan, seperti menghapus kata sambung atau mengubah kata ke bentuk dasarnya [10]. Melalui optimasi algoritma yang tepat, klasifikasi teks mampu meningkatkan efisiensi operasional dan membantu pengambilan keputusan berbasis data tekstual secara lebih akurat[11].

Ekstraksi Fitur TF-IDF

Metode *Term Frequency-Inverse Document Frequency* (TF-IDF) merupakan teknik ekstraksi fitur yang sangat populer dalam analisis teks. TF-IDF memberikan bobot pada setiap kata berdasarkan frekuensi kemunculannya dalam dokumen (TF) dikalikan dengan logaritma kebalikan frekuensi dokumen yang memuat kata tersebut (IDF). Hal ini bertujuan untuk mengurangi bobot kata yang terlalu umum dan menonjolkan kata kunci yang spesifik dan informatif dalam dokumen[6].

Support Vector Machine (SVM)

Support Vector Machine (SVM) adalah metode klasifikasi yang bekerja dengan menemukan *hyperplane* terbaik sebagai pemisah dua kelas. Prinsip utamanya adalah memaksimalkan margin antara *support vectors* terdekat dari masing-masing

kelas [12]. SVM memiliki ketangguhan dalam menangani data teks berdimensi tinggi (*sparse matrix*) dan sangat efektif dalam meminimalkan risiko *overfitting* melalui prinsip *maximal margin classifier*. Penggunaan kernel linear pada SVM sering diterapkan karena efisiensi komputasi yang lebih baik pada dataset teks dibandingkan dengan kernel non-linear[13].

Logistic Regression

Logistic Regression merupakan model statistik probabilistik yang digunakan untuk klasifikasi biner. Algoritma ini memetakan kombinasi linier dari fitur input ke dalam rentang 0 hingga 1 menggunakan fungsi aktivasi sigmoid. Keunggulan utama *Logistic Regression* dibandingkan dengan model *black-box* adalah tingkat interpretabilitas koefisiennya yang tinggi, sehingga memudahkan peneliti dalam mengidentifikasi bobot atau pengaruh setiap fitur kata terhadap kategori sentimen. Meskipun model ini berbasis hubungan linier, efisiensinya sangat tinggi pada dataset teks yang besar.

Program Makan Bergizi Gratis (MBG)

Program Makan Bergizi Gratis (MBG) adalah kebijakan strategis pemerintah yang bertujuan untuk meningkatkan kualitas sumber daya manusia (SDM) dan kesehatan anak usia sekolah[14]. Merujuk praktik internasional dan studi global (WFP, African Union, World Bank), pemberian makanan bergizi di sekolah terbukti berpartisipasi terhadap kenaikan partisipasi pendidikan, konsentrasi belajar, dan pemerataan kesejahteraan sosial[15]. Opini masyarakat terkait kebijakan ini sering mencerminkan harapan, kritik, maupun apresiasi yang kompleks di media sosial, sehingga diperlukan analisis sentimen untuk memetakan dinamika persepsi tersebut [16].

Evaluasi Model

Evaluasi model merupakan tahap krusial untuk mengukur kinerja algoritma klasifikasi[17]. *Confusion matrix* sangat berguna untuk mengevaluasi kinerja model klasifikasi biner, karena memberikan gambaran yang lebih rinci tentang kinerja model daripada hanya melihat akurasi saja. Dengan memeriksa jumlah *true positive*, *true negative*, *false positive*, dan *false negative*, dapat dihitung metrik *precision*, *recall*, dan *F1-score*. Metrik ini dapat membantu untuk lebih memahami kekuatan dan kelemahan model *Machine Learning* yang dihasilkan. Untuk menghitung akurasi, presisi, *recall*, dan *F1-score* dari *Confusion matrix*, kita perlu mengetahui terlebih dahulu definisi dari masing-masing metrik tersebut. *Confusion matrix* adalah tabel yang digunakan untuk mengevaluasi performa suatu model klasifikasi dengan membandingkan hasil prediksi model dengan nilai sebenarnya dari data yang diuji[18].

Terdapat empat komponen utama dalam *Confusion matrix* yang harus dihitung. *Confusion matrix* digunakan sebagai instrumen untuk membandingkan hasil prediksi dengan label aktual, yang menghasilkan metrik seperti *accuracy* (ketepatan keseluruhan), *precision* (ketepatan prediksi positif), *recall* (kemampuan menangkap kelas positif), dan *F1-score* (rata-rata harmonik antara presisi dan *recall*) [19].

Analisis Sentimen

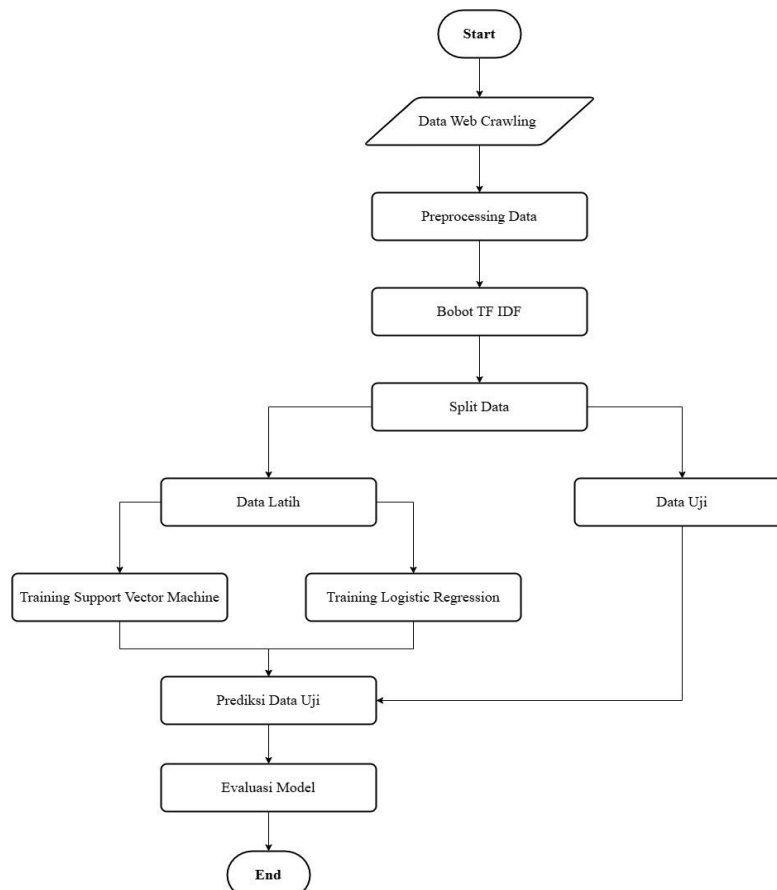
Analisis sentimen merupakan salah satu teknik dalam *Natural Language Processing* (NLP) yang bertujuan untuk mengidentifikasi, mengekstraksi, dan mengklasifikasikan opini, pandangan, atau emosi yang terkandung dalam teks ke dalam kategori tertentu[20]. Dalam prosesnya, NLP berperan sebagai tahapan awal untuk melakukan pengolahan data teks, meliputi pembersihan data, normalisasi, serta representasi teks agar dapat diproses lebih lanjut oleh algoritma pembelajaran mesin analisis sentimen memungkinkan pemetaan kecenderungan sikap masyarakat terhadap suatu topik atau kebijakan secara sistematis, objektif, dan berbasis data [8]. Klasifikasi sentimen umumnya dibagi menjadi tiga kategori utama:

1. Sentimen positif: Merepresentasikan opini yang menunjukkan dukungan, persetujuan, kepuasan, atau apresiasi terhadap suatu objek atau kebijakan[21]. Identifikasi sentimen positif penting untuk menilai aspek kebijakan yang berjalan efektif dan mendapat respons baik dari masyarakat.
2. Sentimen negatif: Menggambarkan opini, sikap, atau ekspresi ketidakpuasan, kritik, maupun kekecewaan. Sentimen ini sering muncul dalam bentuk respons emosional yang mencerminkan ketidakselarasan antara harapan masyarakat dan realitas kebijakan.
3. Sentimen netral: Kategori di mana teks tidak menunjukkan kecenderungan emosional yang jelas. Teks bersifat informatif atau deskriptif, menyampaikan fakta tanpa ekspresi dukungan atau kritik yang signifikan[22].

METODOLOGI

Pelabelan data dilakukan menggunakan metode *lexicon-based* dengan dua kamus sentimen, yaitu *positive lexicon* dan *negative lexicon* dari file *positive.tsv* dan *negative.tsv*. Skor sentimen dihitung dari selisih jumlah kata positif dan negatif. Jika skor ≥ 0 , data diberi label Positif, sedangkan jika skor < 0 , data diberi label Negatif.

Penelitian ini dapat dilihat pada gambar 1 mengkaji analisis klasifikasi teks melalui pendekatan komparatif dua algoritma pembelajaran mesin yang diawali dengan tahap akuisisi dan standardisasi data. Tahapan pertama diinisiasi dengan proses pengumpulan data mentah menggunakan metode *web crawling*, yang kemudian dialirkan menuju fasa *pre-processing data* guna mereduksi *noise*.



Gambar 1 Flowchart klasifikasi SVM dan Logistic Regression

Melalui teknik seperti *cleaning*, *tokenizing*, dan *stemming*. Setelah diperoleh korpus data yang bersih, dilakukan proses *split dataset* dengan rasio 80:20, di mana data latih sebesar 80% dialokasikan untuk fase pembelajaran model, sementara data uji sebesar 20% dicadangkan untuk pengujian. Sebelum memasuki tahap pemodelan, kedua subset data tersebut wajib melewati proses ekstraksi fitur menggunakan metode *TF-IDF* (*Term Frequency - Inverse Document Frequency*) untuk mentransformasikan teks kontekstual menjadi nilai bobot vektor numerik yang siap diproses oleh algoritma.

Memasuki fasa inti pemodelan, data latih yang telah merepresentasikan nilai TF-IDF diimplementasikan secara paralel ke dalam dua algoritma klasifikasi, yaitu *Support Vector Machine* (SVM) dan *Logistic Regression*. Setelah kedua model berhasil dibangun, tahapan berikutnya adalah melakukan pengujian data uji secara objektif menggunakan subset data uji 20% yang telah dipersiapkan sebelumnya untuk mengukur kemampuan generalisasi masing-masing algoritma.

HASIL DAN PEMBAHASAN

Dalam penelitian ini, data diperoleh dari media sosial X (*Twitter*) melalui proses *crawling* data menggunakan kata kunci atau *hashtag* #MakanBergiziGratis. Data yang dikumpulkan berupa komentar masyarakat yang berisi opini, tanggapan, serta pandangan terhadap pelaksanaan Program Makan Bergizi Gratis (MBG). Setelah proses pengumpulan data selesai,

diperoleh sebanyak 1.019 komentar yang selanjutnya digunakan sebagai *dataset* dalam penelitian ini. *Dataset* tersebut menjadi dasar dalam proses klasifikasi sentimen untuk mengetahui kecenderungan opini masyarakat terhadap Program Makan Bergizi Gratis (MBG). Dapat dilihat pada table berikut.

Table 1 Data Sentimen Masyarakat terhadap Program MBG

No	created_at	full_text
1	2026-08-05	Demi mutu makanan anak Indonesia, SPPG yang tak memenuhi standar bakal langsung ditindak tegas. Keselamatan dan gizi anak-anak kita nomor satu! #jagaindonesia #MakanBergiziGratis #KesehatanAnak #GiziSeimbang #IndonesiaSehat https://t.co/5A8GZ6H5CK
2	2026-08-05	Program MBG optimalkan bahan pangan lokal dari petani dan peternak sekitar. Ekonomi daerah makin muter, harga pangan pun tetap stabil terjaga! #jagaindonesia #MakanBergiziGratis #EkonomiLokal #PetaniMaju #PanganAman https://t.co/0yAQgUkBjS
3	2026-08-05	Program Makan Bergizi Gratis (MBG) di Provinsi Bali terus memberikan manfaat bagi ratusan ribu penerima melalui Satuan Pelayanan Pemenuhan Gizi (SPPG) yang tersebar di seluruh kabupaten/kota. #ProgramMBG #MakanBergiziGratis #MBGBali #BaliSehat #GenerasiSehat #IndonesiaEmas2045 https://t.co/PKUVeREHGk
4	2026-08-05	Program MBG mendukung tumbuh kembang anak sekaligus memberi dampak positif bagi perekonomian lokal. #MakanBergiziGratis #EkonomiLokal #ManfaatMBG #DukungMBG https://t.co/0KArdtyGpY
5	2026-08-05	MBG dan KDMP adalah bukti nyata perhatian untuk masa depan bangsa. Ketika fisik anak terawat dan ekonomi warga terangkat, Indonesia Emas bukan lagi sekadar impian. Yuk, dukung dan awasi pelaksanaannya agar tepat sasaran! #DukungProgramRakyat #MakanBergiziGratis https://t.co/OASNucAzqG
6	2026-08-05	Gizi anak terpenuhi, ekonomi lokal ikut bergerak. MBG memberi manfaat bagi masyarakat di berbagai daerah. #MakanBergiziGratis #EkonomiLokal #ManfaatMBG #DukungMBG https://t.co/R9g5lytYeO
7	2026-08-05	MBG tidak hanya mendukung pemenuhan gizi anak, tetapi juga menggerakkan ekonomi lokal dan membuka peluang kerja. #MakanBergiziGratis #EkonomiLokal #ManfaatMBG #DukungMBG https://t.co/MtLNCO8XNk
...	...	...
...	...	...
...	...	...
1017	2025-12-12	Program Makan Bergizi Gratis bukan cuma bikin anak makin sehat, tapi juga nge-boost semangat belajar dan bantu ekonomi keluarga. Dampaknya nyata banget buat masa depan! Kita dukung terus biar makin luas manfaatnya ðŸ™€ðŸ†•ðŸ†•ðŸ†•" #MakanBergiziGratis #DukungProgramPresiden #IndonesiaMaju https://t.co/6fs1mRhmUu
1018	2025-12-12	Program Makan Bergizi Gratis bukan cuma bikin anak makin sehat, tapi juga nge-boost semangat belajar dan bantu ekonomi keluarga. Dampaknya nyata banget buat masa depan! Kita dukung terus biar makin luas manfaatnya ðŸ™€ðŸ†•ðŸ†•ðŸ†•"

		#MakanBergiziGratis #DukungProgramPresiden #IndonesiaMaju https://t.co/dbluBEb09z
1019	2025-12-12	Makan Bergizi Gratis menjadi langkah progresif yang memperkuat ketahanan generasi penerus Indonesia #MBG #MakanBergiziGratis https://t.co/O0L7E9vMNM

Tabel berikut menyajikan hasil akhir dari proses *preprocessing* data yang telah melalui tahapan *cleaning data*, *case folding*, *tokenizing*, *filtering*, *normalisasi*, *stopword removal*, dan *stemming*. Hasil *preprocessing* ini berupa data teks yang telah bersih, terstruktur, dan siap digunakan pada tahap pembobotan TF-IDF serta proses klasifikasi menggunakan algoritma *Support Vector Machine* (SVM) dan *Logistic Regression*.

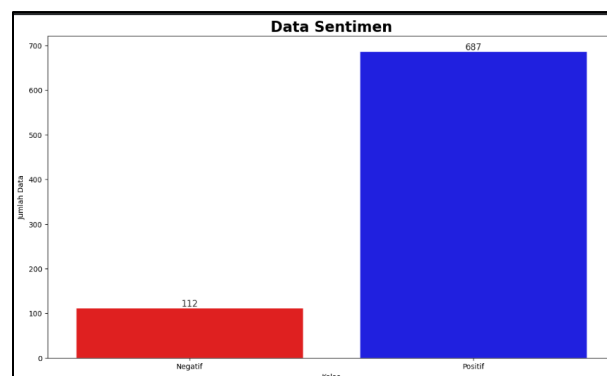
Table 2 Hasil Akhir *Preprocessing Data*

No	Full Text
1	['mutu', 'makan', 'anak', 'penuh', 'standar', 'tindak', 'selamat', 'gizi', 'anak']
2	['dukung', 'program', 'makan', 'gizi', 'gratis']
3	['makan', 'gizi', 'gratis', 'generasi', 'cerdas', 'papua']
4	['anak', 'sehat', 'kuat']
5	['manfaat', 'makan', 'gizi', 'gratis']
6	['badan', 'gizi', 'nasional', 'tingkat', 'standar', 'aman', 'distribusi', 'makan', 'gizi', 'gratis']
7	['makan', 'gizi', 'gratis', 'langkah', 'progresif', 'kuat', 'tahan', 'generasi', 'terus']

Pada tahap ini disajikan hasil implementasi algoritma *Support Vector Machine* (SVM) dan *Logistic Regression* menggunakan bahasa pemrograman Python pada platform *Google Colaboratory* (*Google Colab*). Implementasi dilakukan untuk mengolah data yang telah melalui tahapan *preprocessing*, pembobotan TF-IDF, pelatihan model, hingga proses klasifikasi sentimen. Hasil implementasi ini digunakan untuk mengevaluasi kinerja kedua algoritma serta membandingkan performanya dalam mengklasifikasikan sentimen masyarakat terhadap Program Makan Bergizi Gratis (MBG).

1. Pelabelan Data

Gambar berikut menunjukkan proses pelabelan data menggunakan metode *lexicon-based*. Berdasarkan hasil pelabelan, diperoleh sebanyak 687 data dengan sentimen positif dan 112 data dengan sentimen negatif. Hasil pelabelan ini selanjutnya digunakan sebagai acuan dalam proses pelatihan dan pengujian model klasifikasi sentimen.



Gambar 2 Label Sentimen

2. Wordcloud

Gambar berikut menampilkan visualisasi *word cloud* untuk sentimen positif dan sentimen negatif. Visualisasi ini menunjukkan kata-kata yang paling sering muncul pada masing-masing kategori sentimen, di mana ukuran kata merepresentasikan tingkat frekuensi kemunculannya dalam dataset.



Gambar 3 Wordcloud sentimen positif



Gambar 4 Wordcloud sentimen negatif

3. Pembobotan TF - IDF

Gambar berikut menunjukkan hasil pembobotan TF-IDF pada data yang telah melalui proses *preprocessing*. Hasil pembobotan ini mengubah data teks menjadi representasi numerik berdasarkan tingkat kepentingan setiap kata, sehingga dapat digunakan sebagai masukan pada algoritma *Support Vector Machine (SVM)* dan *Logistic Regression*.

Table 3 Hasil TF - IDF semua data

No.	Dokumen	Kata	Nilai TF-IDF
0	0	anak	0.3780544221912934
1	0	gizi	0.08996883041082297
2	0	makan	0.09520970326686062
3	0	mutu	0.4753733413557513
4	0	penuh	0.268853212580698
5	0	selamat	0.503245487904054
6	0	standar	0.337545010309542
7	0	tindak	0.415955528848044
8	0	bahan	0.3108083723757516
9	0	ekonomi	0.18362814116898918
...	...	...	...
10330	798	kuat	0.27674122520459976
10331	798	langkah	0.29244766113927456
10332	798	makan	0.10022149374758373
10333	798	progresif	0.5138382079993868
10334	798	tahan	0.500396752892332
10335	798	terus	0.4784829241030049

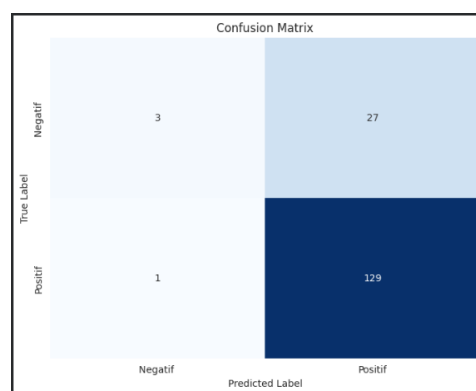
4. *Split Data*

Data yang telah melalui tahap preprocessing selanjutnya dibagi menjadi data latih (*training data*) dan data uji (*testing data*) dengan perbandingan 80:20. Dari total 799 data, sebanyak 639 data digunakan sebagai data latih untuk membangun dan melatih model klasifikasi, sedangkan 160 data digunakan sebagai data uji untuk mengevaluasi kinerja model terhadap data yang belum pernah digunakan selama proses pelatihan.

5. Klasifikasi Algoritma *Support Vector Machine*

Pada tahap klasifikasi, algoritma *Support Vector Machine* (SVM) digunakan untuk membangun model berdasarkan data latih yang telah melalui proses *preprocessing* dan pembobotan TF-IDF. Model yang telah dilatih kemudian digunakan untuk memprediksi kelas sentimen pada data uji. Hasil prediksi tersebut selanjutnya dibandingkan dengan label sebenarnya untuk mengetahui dan mengevaluasi kinerja model SVM dalam melakukan klasifikasi sentimen.

Gambar berikut merupakan *confusion matrix* hasil klasifikasi menggunakan algoritma *Support Vector Machine* (SVM). *Confusion matrix* digunakan untuk mengevaluasi kinerja model dengan membandingkan hasil prediksi terhadap label sebenarnya, sehingga dapat diketahui jumlah prediksi yang benar maupun yang salah pada setiap kelas sentimen.



Gambar 5 Confusion Matriks Algoritma SVM

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100\%$$

$$Accuracy = \frac{129 + 3}{129 + 3 + 27 + 1} \times 100\%$$

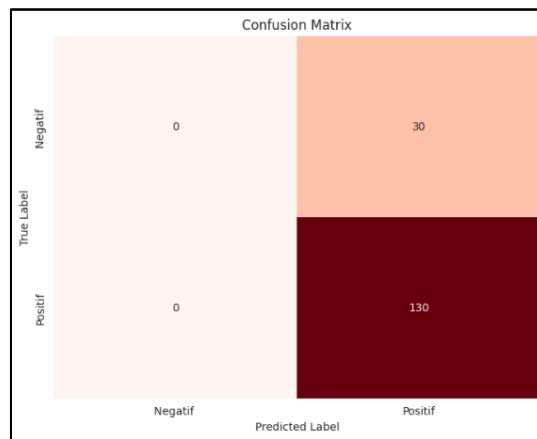
$$Accuracy = 82,50\%$$

Berdasarkan confusion matrix yang diperoleh dari hasil pengujian, algoritma *Support Vector Machine* (SVM) menghasilkan nilai akurasi sebesar 82,50%. Nilai tersebut menunjukkan bahwa model mampu mengklasifikasikan 82,50% data uji dengan benar, sedangkan sisanya merupakan data yang mengalami kesalahan klasifikasi.

6. Klasifikasi Algoritma Logistic Regression

Pada tahap klasifikasi, algoritma *Logistic Regression* digunakan untuk membangun model berdasarkan data latih yang telah melalui proses *preprocessing* dan pembobotan TF-IDF. Model yang telah dilatih kemudian digunakan untuk memprediksi kelas sentimen pada data uji. Hasil prediksi selanjutnya dibandingkan dengan label sebenarnya untuk menghitung nilai akurasi dan mengevaluasi kinerja algoritma *Logistic Regression* dalam melakukan klasifikasi sentimen.

Gambar berikut merupakan *confusion matrix* hasil klasifikasi menggunakan algoritma *Logistic Regression*. *Confusion matrix* digunakan untuk mengevaluasi kinerja model dengan membandingkan hasil prediksi terhadap label sebenarnya, sehingga dapat diketahui jumlah prediksi yang benar maupun yang salah pada setiap kelas sentimen.



Gambar 6 Confusion Matriks Algoritma LR

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100\%$$

$$Accuracy = \frac{130 + 0}{130 + 0 + 30 + 0} \times 100\%$$

$$Accuracy = 81,25\%$$

Berdasarkan hasil *confusion matrix* pada data uji, algoritma *Logistic Regression* memperoleh nilai akurasi sebesar 81,25%. Nilai tersebut menunjukkan bahwa model mampu mengklasifikasikan 81,25% data uji dengan benar, sedangkan 18,75% lainnya mengalami kesalahan klasifikasi. Hasil tersebut digunakan sebagai dasar untuk mengevaluasi kemampuan algoritma *Logistic Regression* dalam mengklasifikasikan sentimen positif dan negatif.

7. Evaluasi Model

Pada tahap evaluasi model, disajikan perhitungan manual dengan membandingkan label sebenarnya pada data uji dan hasil prediksi yang diperoleh menggunakan algoritma *Support Vector Machine* (SVM) dan *Logistic Regression*. Perbandingan ini bertujuan untuk mengevaluasi kinerja kedua algoritma dalam mengklasifikasikan sentimen serta menjadi dasar dalam perhitungan *confusion matrix* dan metrik evaluasi model.

Algoritma SVM

Tabel berikut menyajikan perbandingan antara label sebenarnya (*actual label*) dan hasil prediksi yang diperoleh menggunakan algoritma *Support Vector Machine* (SVM). Perbandingan tersebut digunakan sebagai dasar dalam menentukan nilai *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN) pada proses evaluasi model.

Table 4. Perbandingan Label Aktual dan Hasil Prediksi Algoritma SVM

	Prediksi Negatif	Prediksi Positif
Aktual Negatif	3	27
Aktual Positif	1	129

Algoritma LR

Tabel berikut menyajikan perbandingan antara label sebenarnya (*actual label*) dan hasil prediksi yang diperoleh menggunakan algoritma *Logistic Regression* (LR). Perbandingan tersebut digunakan sebagai dasar dalam menentukan nilai *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN) pada proses evaluasi model.

Table 5. Perbandingan Label Aktual dan Hasil Prediksi Algoritma LR

	Prediksi Negatif	Prediksi Positif
Aktual Negatif	3	27
Aktual Positif	1	129

KESIMPULAN DAN SARAN

Kesimpulan

Berdasarkan hasil penelitian, data yang berhasil dikumpulkan melalui proses *crawling* media sosial X (Twitter) menggunakan hashtag #MakanBergiziGratis berjumlah 1.019 komentar. Setelah melalui tahapan *preprocessing*, sebanyak 220 data duplikat dihapus sehingga diperoleh 799 data yang digunakan dalam proses klasifikasi sentimen. Hasil perbandingan kinerja menunjukkan bahwa algoritma *Support Vector Machine* (SVM) memperoleh nilai *accuracy* sebesar 82,50%, *precision* 82,69%, *recall* 99,23%, dan *F1-score* 90,21%. Sementara itu, algoritma *Logistic Regression* memperoleh nilai *accuracy* sebesar 81,25%, *precision* 81,25%, *recall* 100%, dan *F1-score* 89,66%. Berdasarkan nilai evaluasi tersebut, SVM memiliki akurasi keseluruhan yang lebih tinggi dibandingkan *Logistic Regression*, sehingga SVM lebih tepat digunakan dalam penelitian ini. Namun, kedua algoritma masih mengalami kesulitan dalam mengklasifikasikan kelas sentimen negatif sebagai kelas minoritas akibat ketidakseimbangan distribusi data.

Hasil klasifikasi menggunakan algoritma *Support Vector Machine* (SVM) dan *Logistic Regression* dapat mengelompokkan opini masyarakat ke dalam kategori sentimen positif dan negatif. Hasil tersebut memberikan gambaran mengenai persepsi masyarakat terhadap Program Makan Bergizi Gratis (MBG) dan dapat menjadi salah satu bahan evaluasi bagi pemangku kepentingan dalam memahami opini masyarakat terhadap pelaksanaan program. Meskipun demikian, hasil klasifikasi perlu diinterpretasikan dengan mempertimbangkan ketidakseimbangan dataset yang dapat memengaruhi kemampuan model dalam mengenali kelas minoritas.

Saran

Penelitian selanjutnya disarankan untuk menggunakan dataset dengan jumlah yang lebih besar serta mengambil data dari berbagai platform media sosial, seperti Instagram, Facebook, TikTok, dan YouTube. Dengan demikian, data yang diperoleh akan lebih beragam sehingga hasil analisis sentimen dapat memberikan gambaran yang lebih representatif mengenai persepsi masyarakat terhadap Program Makan Bergizi Gratis (MBG).

Penelitian selanjutnya juga disarankan untuk membandingkan algoritma *Support Vector Machine* (SVM) dan *Logistic Regression* dengan algoritma klasifikasi lainnya, seperti *Naïve Bayes*, *Random Forest*, *Decision Tree*, maupun metode *Deep Learning*. Perbandingan tersebut diharapkan dapat menghasilkan model klasifikasi dengan performa dan tingkat akurasi yang lebih baik.

Selain menggunakan metode *lexicon-based*, penelitian selanjutnya disarankan untuk menerapkan metode pelabelan data lainnya, seperti *TextBlob*, sehingga hasil pelabelan sentimen dapat dibandingkan. Penggunaan metode pelabelan yang berbeda diharapkan mampu meningkatkan kualitas label dan memberikan hasil klasifikasi sentimen yang lebih akurat.

DAFTAR PUSTAKA

- [1] Anindita, "Logistic Regression with Min-Max Scaling and TF-IDF for App Classification and Recommendation on Google Play Store," vol. 5, no. 1, pp. 79–90, 2025.
- [2] Budianto, "Perbandingan Performa Algoritma Support Vector Machine (SVM) dan Logistic Regression untuk Analisis Sentimen Pengguna Aplikasi Retail di Android," vol. 10, no. November, pp. 1–10, 2024.
- [3] S. Mazya, P. Tyas, R. Sarno, and B. S. Rintyarna, "Analisis Perbandingan Metode Klasifikasi Sentimen Berita Saham : Pendekatan Machine Learning , Deep Learning , Transfer Learning , dan Graf," vol. 9, no. 1, pp. 58–64, 2024.
- [4] I. S. Wibowo *et al.*, "Keyword Extraction Judul Berita Online Di Indonesia Menggunakan Metode TF-IDF 1,2," vol. 11, no. 1, 2024.
- [5] N. A. Fahrurrozi, S. Hadiani, and M. Kom, "ANALISIS SENTIMENT ULASAN PRODUK E-COMMERCE

- MENGGUNAKAN METODE LOGISTIC REGRESSION,” vol. XX, no. X, pp. 1–6, 2025.
- [6] A. Davina, M. Putri, N. Sulistianingsih, and R. Rismayati, “Pengaruh Teknik Representasi Teks Bag-of-Words dan TF-IDF,” vol. 7, no. 4, pp. 675–688, 2025.
 - [7] Dometian, “Analisis Sentimen Ulasan Google Play Store : Studi Komparatif Algoritma SVM , Naïve Bayes , dan Logistic Regression,” vol. 15, no. 3, pp. 610–620, 2025.
 - [8] N. N. Azizah, I. Purnamasari, and S. Prangga, “Pengelompokan Judul Laporan Skripsi Berbasis Text Mining dengan Metode Fuzzy K-Means,” 2024.
 - [9] Alfando, “Klasifikasi Teks Berita Berbahasa Indonesia Menggunakan Machine Learning Dan Deep Learning: Studi Literatur,” vol. 7, no. 1, pp. 681–686, 2023.
 - [10] M. M. Davina Felisha Putri, “Sentimen Analisis Masyarakat Terhadap Inflasi Dan Kenaikan Harga Sembako Berdasarkan Platform X,” vol. 6, no. 1, pp. 12057–12062, 2026.
 - [11] Sunarti, “Klasifikasi Pengaduan Pelayanan Fakultas Teknik Universitas Muhammadiyah Makassar menggunakan Natural Language Processing,” vol. 2, no. 2, 2024.
 - [12] I. S. Aisah, B. Irawan, and T. Suprapti, “ALGORITMA SUPPORT VECTOR MACHINE (SVM) UNTUK ANALISIS SENTIMEN ULASAN APLIKASI AL QUR ’ AN DIGITAL,” vol. 7, no. 6, pp. 3759–3765, 2023.
 - [13] R. K. Dinata, N. Hasdyna, and R. M. Sari, *SUPERVISED LEARNING: Strategi Prediksi dan Klasifikasi Data*. Serasi Media Teknologi, 2025.
 - [14] M. T. Rambe, R. A. Ritonga, T. Harahap, and R. Sitio, “Pemetaan Awal Implementasi Program Makan Bergizi Gratis (MBG) di Kabupaten Labuhanbatu : Perspektif Manajemen SDM,” vol. 4, no. 4, pp. 5177–5184, 2025.
 - [15] U. Agustini and S. Mulyani, “EFEKTIVITAS DAN TANTANGAN KEBIJAKAN PROGRAM MAKAN BERGIZI GRATIS SEBAGAI INTERVENSI PENDIDIKAN DI INDONESIA EFFECTIVENESS AND CHALLENGES OF FREE NUTRITIOUS MEAL PROGRAM POLICY AS AN EDUCATIONAL,” vol. 4, pp. 362–368, 2025.
 - [16] M. W. Arif, “Analisis Sentimen Kebijakan Makan Bergizi Gratis di Media Sosial Menggunakan Natural Language Processing Berbasis Python TextBlob di Indonesia Teknik Informatika , Universitas Ngudi Waluyo , Indonesia Sentiment Analysis of Free Nutritious Meal Policy on Social Media Using Python TextBlob-Based Natural Language Processing in Indonesia,” vol. 5, no. 9, pp. 2463–2471, 2025.
 - [17] I. Daqiqil, *MACHINE LEARNING : Teori, Studi Kasus dan Implementasi Menggunakan Python*. Unri Press, 2021.
 - [18] R. Rustiyana *et al.*, *Data Mining: Algoritma dan Penerapannya*. PT. Sonpedia Publishing Indonesia, 2025.
 - [19] S. J. Hari Suparwito *et al.*, *Pengantar Pembelajaran Mesin Menggunakan Bahasa Pemrograman Python*. Sanata Dharma University Press, 2023.
 - [20] T. Widayanti *et al.*, *Buku Ajar Matematika Dasar*. PT. Sonpedia Publishing Indonesia, 2025.
 - [21] F. Aftab, S. U. Bazai, S. Marjan, L. Baloch, and S. Aslam, “A Comprehensive Survey on Sentiment Analysis Techniques,” vol. 14, no. September, pp. 1288–1298, 2023.
 - [22] F. S. Maharani Imran, Venessa Agusta Gogali, “Media dan Komunikasi Kebijakan Publik : Analisis Frekuensi dan Sentimen Pemberitaan Program Makan Bergizi Gratis (MBG),” vol. 4, no. 5, pp. 3444–3452, 2026.