

Analisis Sentimen Terhadap Program Makan Siang Gratis pada Media Sosial X Menggunakan Logistic Regression dan SVM

Taufik Abdul Aziz, Iqbal Ismayadi, Budiman

Informatika, Fakultas Teknologi dan Informatika, Universitas Informatika dan Bisnis Indonesia
Email: taufik.aa22@student.unibi.ac.id; iqbal.i22@student.unibi.ac.id; budiman@unibi.ac.id;

Abstrak

Program makan siang gratis bertujuan meningkatkan gizi anak dan konsentrasi belajar, namun pelaksanaannya mendapat beragam respons publik. Penelitian ini menganalisis sentimen masyarakat terhadap program tersebut di media sosial X menggunakan *Support Vector Machine* (SVM) dan *Logistic Regression*. Dari 6.346 komentar yang dikumpulkan, sentimen positif mendominasi 4.323 komentar, diikuti oleh negatif 1.121 dan netral 902. *Logistic Regression* menunjukkan akurasi lebih tinggi 82,00% dibandingkan SVM 77,00%, menandakan kemampuannya yang lebih baik dalam membedakan variasi sentimen. Hasil ini memberikan wawasan bagi pembuat kebijakan dalam mengevaluasi opini publik terhadap program. Studi lanjutan disarankan untuk mengeksplorasi model deep learning guna meningkatkan akurasi serta mempertimbangkan aspek temporal dan kontekstual dalam analisis sentimen.

Kata Kunci: Analisis Sentimen, Makan Siang Gratis, Logistic Regression, SVM, X

Abstract

The free lunch program aims to improve children's nutrition and learning concentration, but its implementation has received mixed public responses. This study analyses public sentiment towards the programme on social media X using a Support Vector Machine (SVM) and Logistic Regression. Of the 6,346 comments collected, positive sentiment dominated 4,323 comments, followed by negative 1,121 and neutral (902). Logistic Regression showed higher accuracy 82.00% than SVM 77.00%, signaling its better ability to distinguish sentiment variations. These results provide insights for policymakers in evaluating public opinion toward the program. Further studies are recommended to explore deep learning models to improve accuracy and consider temporal and contextual aspects in sentiment analysis.

Keywords: Sentiment Analysis, Free Lunch, Logistic Regression, SVM, X

1 PENDAHULUAN

Program makan siang gratis merupakan salah satu kebijakan sosial yang telah dilaksanakan di berbagai negara, termasuk Indonesia, yang bertujuan meningkatkan status gizi anak-anak dan konsentrasi belajar siswa di sekolah (Rimbawan, Nurdiani, Rachman, Kawamata, & Nozawa, 2023). Masyarakat Indonesia masih menghadapi tantangan serius terkait kesehatan akibat kekurangan nutrisi. pada tahun 2022, angka *stunting* di Indonesia tercatat sebesar 21,6%, menurun dibandingkan tahun sebelumnya yang mencapai 24,4% pada tahun 2021. Meski demikian angka tersebut masih belum memenuhi target pemerintah yang menargetkan dibawah 20% pada tahun 2024 (Kementerian

Kesehatan Republik Indonesia, 2024). Pemerintah serta berbagai institusi terkait telah menyediakan dana yang besar untuk menjamin keberlangsungan dan efisiensi program ini. Namun, seperti kebijakan publik lainnya, pelaksanaan program makan siang gratis juga menghadapi sejumlah tantangan dan reaksi dari masyarakat.

Di era digital saat ini, media sosial menjadi platform utama bagi masyarakat untuk menyampaikan pandangan mereka terhadap kebijakan pemerintah, termasuk program makan siang gratis. Platform seperti X (sebelumnya dikenal sebagai Twitter) memungkinkan pengguna mengungkapkan pendapat secara langsung dan dalam waktu nyata. Akibatnya, muncul berbagai data opini dalam jumlah besar, mencakup dukungan positif, kritik, hingga saran konstruktif. Namun, keterbukaan ini juga berpotensi memicu penyebaran informasi yang keliru serta pandangan negatif yang dapat memengaruhi pelaksanaan kebijakan secara keseluruhan (Chatterjee, Rana, & Dwivedi, 2020).

Pendekatan analisis sentimen yang didasarkan pada *machine learning* telah terbukti efektif dalam menganalisis serta memahami pandangan publik di media sosial (Mallick, Mishra, Giri, & Paikaray, 2023). Berbagai algoritma *machine learning* telah terbukti efektif dalam mengklasifikasikan sentimen dengan tingkat akurasi yang tinggi (Sihombing, Halmi Dar, & Aini Nasution, 2024). Banyak penelitian yang telah dilakukan mengenai analisis sentimen, tetapi sedikit penelitian yang secara khusus meneliti bagaimana persepsi masyarakat berubah seiring waktu dan bagaimana berbagai faktor seperti demografi, tempat, dan isu sosial lainnya memengaruhi pandangan publik mengenai program ini. Studi ini bertujuan untuk mengatasi kekurangan tersebut melalui penggabungan beberapa kata kunci yang terkait dengan program makan bergizi gratis sehingga dapat mengidentifikasi pola dan tren pandangan masyarakat yang muncul di platform media sosial X. Temuan dari penelitian ini diharapkan bisa memberikan saran berharga bagi pihak pemerintah dalam meningkatkan efektivitas pelaksanaan program makan siang gratis di Indonesia.

2 KAJIAN PUSTAKA

Analisis sentimen merupakan proses penggunaan teknik pemrosesan bahasa alami (NLP) untuk mengidentifikasi dan mengklasifikasikan opini atau perasaan dalam teks sebagai positif, negatif, atau netral. Analisis sentimen dapat dilakukan menggunakan pendekatan *lexicon based*, *machine learning*, atau gabungan dari keduanya. Pendekatan ini memungkinkan pengambilan keputusan yang lebih informatif dan berbasis data dalam berbagai sektor industri (Liu, 2012). Pada penelitian ini akan menerapkan dua algoritma *machine learning* yaitu *Logistic Regression* dan *Super Vector Machine* (SVM).

Support Vector Machine (SVM) adalah algoritma pembelajaran mesin yang digunakan untuk tugas klasifikasi dan regresi. Algoritma ini bekerja dengan mencari *hyperplane* optimal yang memisahkan kelas-kelas dalam data dengan margin terbesar. Jika data tidak dapat dipisahkan secara linear, SVM menggunakan fungsi kernel untuk memetakan data ke dimensi yang lebih tinggi agar dapat dipisahkan. Dalam SVM, pemilihan kernel sangat penting karena menentukan bagaimana model menangani data yang kompleks dan tidak terpisah secara linear. Berikut adalah tiga jenis kernel yang umum digunakan dalam SVM (Rabbani, Safitri, Rahmadhani, Sani, & Anam, 2023):

1. Linear Kernel

Linear kernel digunakan ketika data dapat dipisahkan secara linear tanpa perlu transformasi ke dimensi yang lebih tinggi. Rumusnya adalah:

$$K(x_i, x_j) = x_i \cdot x_j \quad (1)$$

Kernel ini cocok untuk data yang memiliki hubungan linier antar fitur.

2. Polinomial Kernel

Kernel ini digunakan untuk menangani hubungan non-linear antara fitur dalam data. Didefinisikan sebagai:

$$K(x_i, x_j) = (x_i \cdot x_j + c)^d \quad (2)$$

di mana c adalah konstanta dan d adalah derajat polinomial. Kernel ini memungkinkan SVM untuk menangkap pola kompleks dalam data.

3. Radial Basis Function (RBF) Kernel

Kernel RBF sering digunakan untuk menangani data yang memiliki pola non-linear kompleks. Rumusnya adalah:

$$K(x_i, x_j) = \exp \exp \left(-\gamma \|x_i - x_j\|^2 \right) \quad (3)$$

di mana γ adalah parameter yang mengontrol seberapa jauh pengaruh satu titik data terhadap titik lainnya. Kernel ini efektif untuk menangani data dengan distribusi yang tidak dapat dipisahkan secara linear.

Logistic Regression adalah algoritma klasifikasi yang digunakan untuk memprediksi probabilitas suatu sampel termasuk dalam kategori tertentu. Model ini bekerja dengan mengubah keluaran linier dari regresi menjadi nilai probabilitas menggunakan fungsi sigmoid atau logistic function (Maulana et al., 2023). Fungsi dasar dari Logistic Regression dinyatakan sebagai:

$$P(X) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n)}} \quad (4)$$

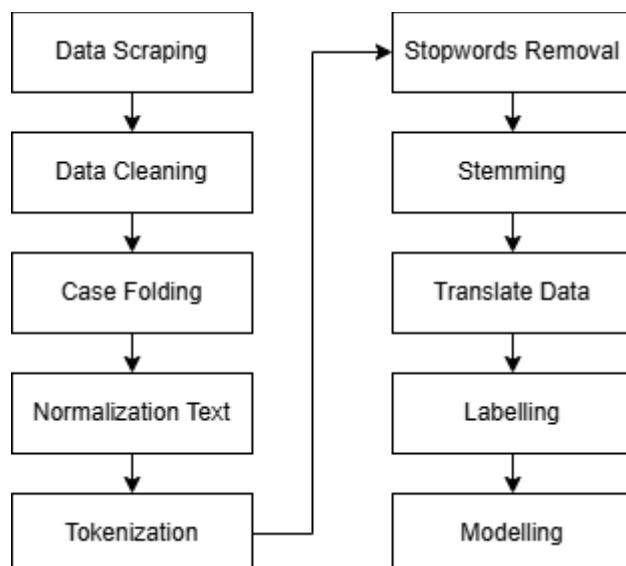
Parameter β_0 berperan sebagai bias atau intercept yang mengatur titik awal keputusan model, sementara $\beta_1, \beta_2, \dots, \beta_n$ merupakan koefisien regresi yang menentukan seberapa besar pengaruh setiap fitur terhadap hasil prediksi. Fitur-fitur input itu sendiri dilambangkan sebagai X_1, X_2, \dots, X_n yang mewakili karakteristik data yang digunakan dalam analisis. Fungsi eksponensial dalam persamaan ini melibatkan bilangan *Euler*(e), yang bernilai sekitar 2,718 dan digunakan untuk membentuk kurva sigmoid yang mengubah output menjadi nilai probabilitas.

Sejumlah penelitian telah dilakukan dalam analisis sentimen dengan menggunakan algoritma machine learning terhadap berbagai isu di media sosial. Analisis sentimen opini publik mengenai Bank BSI dengan memanfaatkan data *Twitter* melalui metode *Support Vector Machine* (SVM), *Naïve Bayes*, dan *Logistic Regression*. Hasilnya menunjukkan bahwa SVM memiliki kinerja tertinggi dengan akurasi 88% (Husen, Astuti, Marlia, Rahmaddeni, & Efrizoni, 2023), diikuti oleh *Logistic Regression* (86%) dan *Naïve Bayes* (76%). Studi analisis sentimen mengenai topik *childfree* di *Twitter* melalui SVM, dengan tingkat akurasi 69,69%, *precision* 51,56%, *recall* 45,60%, dan *F1-score* 46,61% (Ade Dwi Dayani, Yuhandri, & Widi Nurcahyo, 2024). Analisis sentimen terhadap penilaian Pembelajaran Tatap Muka (PTM) 100% di Twitter dengan menggunakan *Logistic Regression*, yang menunjukkan hasil yang baik dengan *precision* 76,92%, *recall* 83,3%, *F1-score* 80%, dan akurasi 78,57% (Assaidi & Amin, n.d.). Penelitian mengenai analisis sentimen di platform *Steam* dan menemukan bahwa SVM menghasilkan kinerja terbaik dengan akurasi 0,81, *precision* 0,85, dan *recall* 0,7 (Lidinillah, Rohana, & Juwita, 2023).

3 METODE PENELITIAN

3.1 Tahapan Analisis

Untuk menjalankan analisis sentimen, ada beberapa langkah yang harus diikuti agar hasil pengujian yang diperoleh optimal. Analisis ini menjadi krusial karena dapat memberikan pemahaman mendalam tentang pandangan yang terdapat dalam teks. Di bawah ini adalah langkah-langkah yang terjadi selama proses berjalan.



Gambar 1. Alur Penelitian

Gambar 1 menggambarkan tahapan proses analisis sentimen, yang dimulai dengan pengumpulan data melalui teknik *data scraping*. Setelah data terkumpul, tahap selanjutnya adalah pembersihan data untuk menghapus elemen yang tidak relevan. Selanjutnya, dilaksanakan *case folding* untuk menyamakan format huruf, diikuti dengan normalisasi teks. Tahapan selanjutnya adalah tokenisasi untuk memisahkan teks menjadi kata-kata, diikuti dengan *stopwords removal* guna mengeliminasi kata-kata yang tidak memiliki makna penting. Langkah berikutnya mencakup *stemming* untuk mengembalikan kata ke bentuk asalnya, kemudian diikuti dengan analisis visual seperti *Wordcloud*. Data kemudian diberi label pada setiap teks, dan akhirnya menggunakan pemodelan untuk meramalkan sentimen. Alur ini dibuat untuk menjamin hasil analisis yang tepat dan terorganisir.

3.2 Scraping Data

Pada fase awal pengambilan data, langkah yang dilakukan adalah mengumpulkan data dari media sosial X sebanyak 3.600 data. Proses ini dimulai dengan melakukan pemeriksaan untuk memperoleh token yang sesuai dengan kata kunci yang dicari. Setelah token didapat, proses pengumpulan data dilaksanakan dengan memanfaatkan alat bantu *Tweet Harvest*. Metode ini memfasilitasi pengumpulan data dengan cara yang efisien dan teratur, sehingga dapat dijadikan landasan untuk analisis sentimen berikutnya.

3.3 Cleaning Data

Proses pembersihan data dilakukan untuk memastikan data teks bersih tanpa elemen yang tidak relevan. Tahap awal adalah menghapus URL yang mengenali dan menghilangkan tautan web dari data. Selanjutnya, menghapus username yang dimulai dengan tanda @. Proses dilanjutkan

dengan menghapus elemen-elemen HTML memakai sehingga yang tersisa hanya teks utama, lalu menghilangkan emoji yang dapat mengganggu analisis dengan menghapus karakter emotikon dan simbol non-verbal. Tahap selanjutnya dengan menghapus teks dari simbol-simbol yang tidak perlu untuk menghilangkan angka dalam teks, sehingga data teks hanya mencakup kata-kata yang bersih dan relevan. Proses ini juga melibatkan penghilangan data ganda dan pengecekan data yang kosong untuk menjaga integritas dataset.

3.4 Case Folding

Proses *case folding* bertujuan untuk mengubah semua huruf dalam teks menjadi huruf kecil agar data menjadi konsisten. Langkah ini krusial untuk mencegah perbedaan akibat penggunaan huruf besar dan kecil, sehingga data menjadi konsisten dan lebih mudah untuk dianalisis. Hasil dari proses ini akan dilanjutkan untuk tahapan berikutnya.

3.5 Text Normalization

Proses penggantian kata tidak baku dimulai dengan memecah teks menjadi kata-kata terpisah untuk diperiksa kesesuaiannya dengan kamus kata tidak baku yang didapat melalui *kaggle* (Fornigulo, n.d.). Jika kata yang ditemukan sesuai dengan entri kamus, kata tersebut akan diganti dengan padanan kata baku yang telah ditentukan. Selama proses ini, kata baku yang dihasilkan, kata yang diganti, serta hash dari kata tidak baku dicatat untuk keperluan pelacakan atau analisis lebih lanjut. Hasil akhirnya adalah teks yang telah dimodifikasi, berupa kalimat baru yang menggabungkan kata-kata yang telah diganti maupun yang tetap, sehingga menghasilkan teks yang lebih terstruktur dan sesuai dengan kaidah bahasa.

3.6 Tokenization

Proses tokenisasi dilakukan untuk membagi kalimat menjadi unit-unit kecil yang disebut token, umumnya berupa kata-kata tunggal. Tindakan ini bertujuan untuk mempersiapkan teks agar lebih gampang dianalisis di tahap berikutnya, seperti analisis frekuensi kata atau pemodelan. Setiap teks yang telah menjalani proses normalisasi diproses untuk dibagi menjadi sekumpulan kata berdasarkan spasi, menghasilkan daftar token yang selanjutnya disimpan dalam kolom baru.

3.7 Stopwords Removal

Stopwords removal adalah langkah dalam preprocessing yang berfokus pada penghilangan kata-kata umum yang tidak memberikan kontribusi penting terhadap makna atau konteks utama dari sebuah teks. Dalam pelaksanaannya, studi ini memanfaatkan daftar stopwords dalam bahasa Indonesia yang telah disesuaikan, meliputi kata-kata seperti "yang", "di", "ke", "dari", serta kata penghubung lainnya yang sering muncul tetapi tidak memberikan informasi krusial untuk analisis sentimen. Proses ini dilakukan dengan cara mencocokkan setiap token hasil tokenisasi dengan daftar *stopwords* yang sudah ditentukan, lalu menghapus token yang ada dalam daftar itu. Penghapusan *stopwords* memainkan peran krusial dalam meningkatkan efisiensi komputasi dan menurunkan kebisingan data, karena penghilangan kata-kata yang tidak relevan memungkinkan sistem untuk lebih fokus pada kata-kata yang lebih signifikan dan berpengaruh dalam penentuan sentimen atau klasifikasi teks.

3.8 Stemming

Proses *stemming* bertujuan untuk mengubah setiap kata dalam teks ke bentuk dasarnya, sehingga dapat mengurangi variasi kata yang memiliki arti serupa. Teks yang sudah melewati tahap *stopwords removal* diproses lebih lanjut untuk diuraikan menjadi kata-kata terpisah. Setiap kata itu

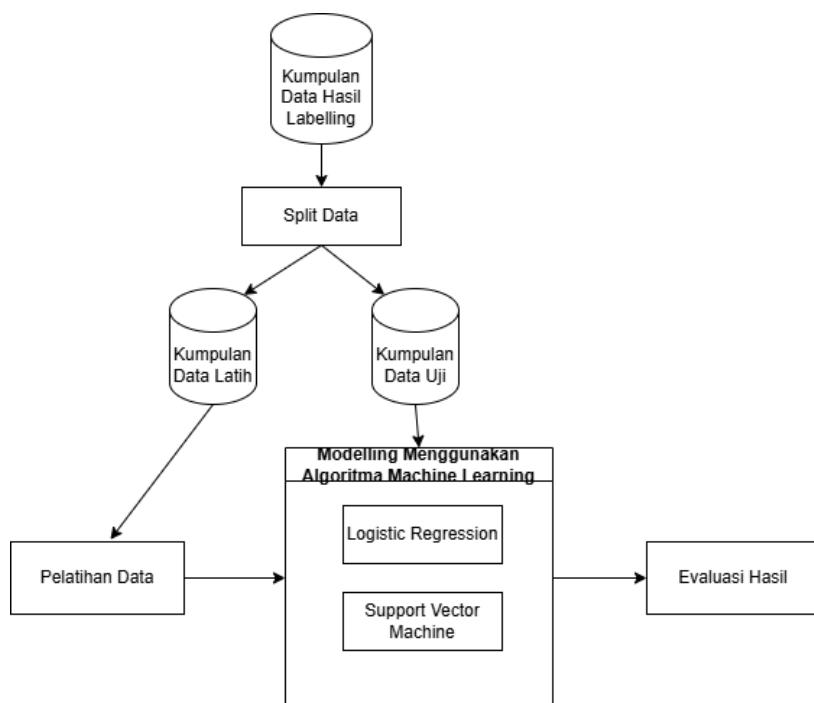
kemudian diproses dengan menggunakan *stemmer*, yang secara otomatis mengenali dan mengubahnya menjadi bentuk dasarnya.

3.9 Labelling

Proses analisis sentimen dilakukan untuk mengidentifikasi polaritas emosi dalam sebuah teks, apakah sifatnya positif, negatif, atau netral. Pada tahap ini, setiap teks yang telah menjalani proses translasi dan stemming dianalisis dengan metode *lexikon based* untuk penandaan emosi. Menyertakan berkas TSV yang berisi istilah positif dan negatif yang diambil dari *Kaggle* (Fajri & Rahamaningtyas, 2017). Metode ini didasarkan pada daftar kata yang telah ditentukan sebelumnya, di mana setiap kata memiliki nilai emosi yang spesifik. Nilai polaritas tersebut ditentukan berdasarkan nilai yang telah ditentukan untuk masing-masing kata, di mana nilai positif mencerminkan sentimen positif, nilai negatif mencerminkan sentimen negatif, dan nilai nol mencerminkan sentimen netral.

3.10 Modelling

Sebelum memasuki tahap pemodelan, data yang sudah dilabeli dipersiapkan untuk proses pelatihan dan pengujian. Data ini terbagi menjadi dua segmen, yaitu 80% untuk pelatihan dan 20% untuk pengujian. Pembagian ini bertujuan agar model bisa belajar dari sebagian besar data, sambil tetap menyimpan data yang belum pernah dilihat untuk menguji kinerja model. Proses ini bertujuan untuk mencegah *overfitting* dan menjamin generalisasi model pada data yang belum pernah dilihat. Gambar 2 menunjukkan alur pemodelan data untuk menghasilkan prediksi yang akurat.



Gambar 2. Alur pada saat menggunakan *machine learning* untuk pemodelan

Setelah pembagian data selesai, dilakukan pemodelan menggunakan algoritma *Logistic Regression* dan *Support Vector Machine*. Data latih digunakan untuk melatih model agar mampu mengenali pola, sedangkan data uji digunakan untuk mengevaluasi performa model. Hasil evaluasi ini akan menentukan sejauh mana model dapat mengklasifikasikan data dengan akurasi yang baik dan sesuai dengan tujuan analisis. Setelah tahap pemodelan selesai, performa model akan dianalisis lebih lanjut menggunakan *confusion matrix* untuk mengidentifikasi jumlah prediksi yang benar dan

salah di setiap kelas. Matriks kebingungan berperan dalam menghitung metrik evaluasi seperti akurasi, *presisi*, *recall*, dan *F1-score*, yang memberikan informasi lebih mendalam tentang efisiensi model dalam mengklasifikasikan data.

4 HASIL DAN PEMBAHASAN

Pada fase ini, dilakukan pencarian dan ekstraksi informasi dari media sosial X dengan menggunakan berbagai kata kunci yang sesuai dengan program makan siang gratis. Informasi yang diperoleh mencakup berbagai postingan, komentar, dan reaksi pengguna terhadap program tersebut, kemudian setiap data dari kata kunci digabung menjadi satu. Setelah proses penghapusan entri duplikat, jumlah data yang dapat dipakai untuk analisis sebanyak 6.346. Sehingga menjamin mutu data yang dipakai dalam penelitian, agar hasil analisis sentimen dapat merefleksikan pandangan publik dengan lebih tepat.

4.1 Preprocessing Data

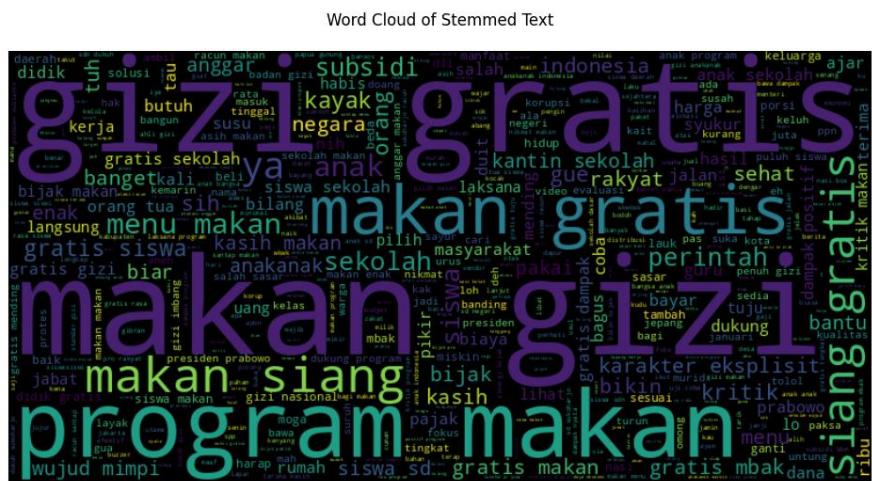
Tahap *preprocessing* dilakukan untuk merapikan dan menyiapkan data sebelum dianalisis secara mendalam. Proses ini meliputi pembersihan untuk menyingkirkan karakter khusus, angka, dan tautan yang tidak penting, dilanjutkan dengan case folding yang merubah seluruh teks menjadi huruf kecil agar format data lebih konsisten. Selanjutnya, normalisasi digunakan untuk mengganti istilah nonstandar atau singkatan dengan bentuk yang lebih umum. Setelahnya, tokenisasi dilakukan untuk memecah teks menjadi bagian-bagian kata yang lebih kecil, dilanjutkan dengan *stopwords removal* untuk menyingkirkan kata-kata umum yang tidak memiliki arti penting dalam analisis sentimen. Lalu, *stemming* digunakan untuk mengubah kata menjadi bentuk dasar, sehingga analisis menjadi lebih konsisten dan tepat. Tabel 1 menunjukkan hasil dari setiap tahapan pra-pemrosesan pada salah satu komentar.

Tabel 1. Hasil setiap tahapan *pre-processing* pada satu komentar

No .	Tahapan	Hasil
1	Full_text	@barengwarga Makan siang Gratis tujuannya memang bukan untuk nutrisi bergizi tapi untuk menambah rasa syukur siswa
2	Cleaning	Makan siang Gratis tujuannya memang bukan untuk nutrisi bergizi tapi untuk menambah rasa syukur siswa
3	Case_folding	makan siang gratis tujuannya memang bukan untuk nutrisi bergizi tapi untuk menambah rasa syukur siswa
4	Normalisasi	makan siang gratis tujuannya memang bukan untuk nutrisi bergizi tapi untuk menambah rasa syukur siswa
5	Tokenize	['makan', 'siang', 'gratis', 'tujuannya', 'memang', 'bukan', 'untuk', 'nutrisi', 'bergizi', 'tapi', 'untuk', 'menambah', 'rasa', 'syukur', 'siswa']
6	Stopwords_removal	['makan', 'siang', 'gratis', 'tujuannya', 'nutrisi', 'bergizi', 'menambah', 'syukur', 'siswa']
7	Stemming	makan siang gratis tuju nutrisi gizi tambah syukur siswa

4.2 *WordCloud*

WordCloud memberikan gambaran visual tentang distribusi kata-kata yang sering muncul dalam dataset. *WordCloud* adalah teknik visualisasi teks yang menampilkan kata-kata dengan ukuran berbeda berdasarkan frekuensi kemunculannya, di mana kata-kata dengan frekuensi lebih tinggi ditampilkan dalam ukuran lebih besar. Visualisasi ini membantu mengidentifikasi kata-kata dominan yang sering muncul dalam dataset setelah melalui proses *preprocessing* seperti *stemming*. Berikut adalah hasil visualisasi *Word Cloud* yang menggambarkan distribusi kata-kata dalam dataset yang telah diproses.



Gambar 3. Wordclouds setelah *preprocessing* data

4.3 *Labeling Data*

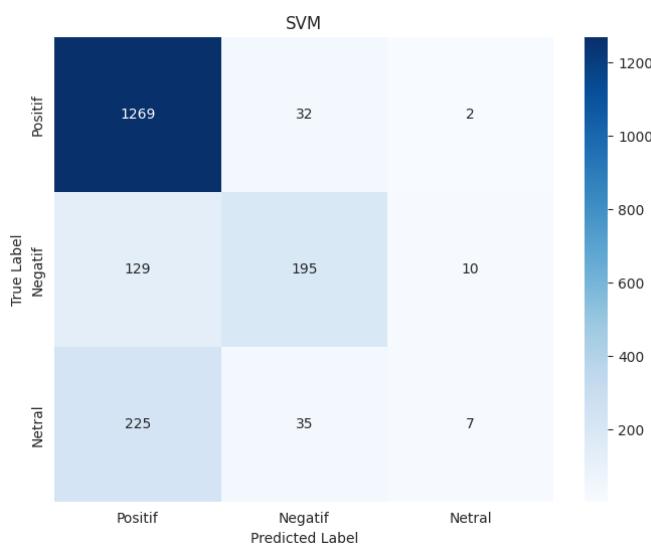
Dalam proses pelabelan data, *lexicon based* digunakan untuk mengidentifikasi sentimen dari setiap komentar berdasarkan kata-kata yang memiliki nilai polaritas dalam kamus sentimen. Setiap komentar dievaluasi dan diberikan label sentimen positif, negatif, atau netral berdasarkan skor polaritas yang telah dihitung. Apabila skor total menunjukkan angka positif, komentar akan digolongkan sebagai sentimen positif; jika angka tersebut negatif, maka termasuk dalam sentimen negatif; sedangkan jika skor hampir nol, komentar tersebut dianggap netral. Dari hasil pelabelan terhadap seluruh data, didapat sebanyak 4.323 komentar tergolong dalam sentimen positif, 1.121 komentar bersentimen negatif, dan 902 komentar bersifat netral. Tabel di bawah ini menyajikan lima komentar beserta klasifikasinya menggunakan metode *lexicon based*.

Tabel 2. Hasil sentimen lima komentar menggunakan Lexicon Based

Text	Skor	Sentimen
program makan gizi gratis siswa sekolah indonesia jalan pekan hasil laksana implementasi agam lapang evaluasi respon istana ketidakseragaman hasil laksana	7	Positif
sabar ya guru pulau nias perintah sibuk asih makan gratis siswa sekolah pulau jawa bisa anggar coba perhati oh ya anak sekolah nias harap makan gratis ya plus bunuh usaha kantin gara siswa dah makan gratis	9	Positif
puluhan siswa sd negeri dukuh kecamatan kabupaten sukoharjo jawa duga racun santap menu makan program makan gizi gratis mbak keluh pusing mual muntah catat	-1	Negatif
ibuk takut i buk makan siang gratis takut kena racun berita siswa racun makan gizi ya racun anak kelas sd lihat berita	-1	Negatif

4.4 Pemodelan dan Evaluasi

Pada tahap ini, dua algoritma *machine learning*, yaitu SVM dan *Logistic Regression*, digunakan untuk mengklasifikasikan sentimen tentang program makan siang gratis. Setiap model dalam analisis memanfaatkan *confusion matrix* untuk memberikan gambaran mengenai jumlah prediksi yang benar dan salah dalam setiap kategori sentimen (positif, negatif, dan netral). Dengan menganalisis distribusi *true positive*, *false positive*, *true negative*, dan *false negative*, pola kesalahan dalam pengklasifikasian sentimen bisa diungkap. Tabel di bawah ini menunjukkan confusion matrix dari kedua model tersebut.



Gambar 4. *Confusion Matrix SVM*



Gambar 5. *Confusion matrix Logistic Regression*

Berdasarkan confusion matrix yang disajikan, SVM menunjukkan jumlah *true positive* yang besar untuk kelas positif, namun memiliki jumlah *false negative* yang cukup signifikan, khususnya pada kelas netral yang sering salah dikategorikan sebagai positif. Sementara itu, *Logistic Regression* menunjukkan kinerja yang baik dalam mengidentifikasi kelas positif, tetapi lebih seimbang dalam

mengklasifikasikan kelas negatif dan netral daripada SVM. *Logistic Regression* terlihat lebih handal dalam menangkap variasi sentimen dengan lebih tepat, khususnya dalam membedakan antara kelas negatif dan netral, sementara SVM cenderung lebih sering memprediksi komentar sebagai sentimen positif.

Berdasarkan hasil tersebut, dua algoritma machine learning di analisa menggunakan metrik evaluasi seperti *precision*, *recall*, *f1-score*, dan akurasi. Hasil penilaian dari kedua model kemudian dihitung rata-ratanya untuk memperoleh gambaran umum kinerja secara keseluruhan. Tabel di bawah ini menyajikan hasil perbandingan antara dua model berdasarkan metrik evaluasi yang telah dihitung.

Tabel 3. Perbandingan SVM dan Logistic Regression

Model	Precision	Recall	F1 – Score	Accuracy
SVM	63.00%	52.60%	52.30%	77.00%
Logistic Regression	78.67%	64.00%	65.30%	82.00%

Berdasarkan hasil evaluasi yang terdapat dalam tabel, model *Logistic Regression* menunjukkan kinerja yang lebih unggul dibanding SVM pada semua metrik evaluasi. Logistic Regression menghasilkan *precision* 78.67%, *recall* 64.00%, dan *f1-score* 65.30%, serta tingkat akurasi yang mencapai 82.00%. Sementara itu, SVM mencatat *precision* 63.00%, *recall* 52.60%, dan *f1-score* 52.30%, serta akurasi 77.00%. Perbedaan ini mengindikasikan bahwa *Logistic Regression* lebih baik dalam mengklasifikasikan pandangan publik mengenai program makan siang gratis dibandingkan SVM.

5 SIMPULAN

Penelitian ini menganalisis sentimen publik terhadap program makan siang gratis menggunakan machine learning berbasis *Support Vector Machine* (SVM) dan *Logistic Regression*. Dari 6.346 data komentar yang dikumpulkan dan diproses, hasil pelabelan menunjukkan dominasi sentimen positif sebesar 4.323 komentar, sementara 1.121 komentar bersentimen negatif dan 902 bersifat netral. *Logistic Regression* menunjukkan kinerja yang lebih unggul dibandingkan SVM dengan akurasi 82.00%, *precision* 78.67%, *recall* 64.00%, dan *f1-score* 65.30%, sedangkan SVM hanya mencapai akurasi 77.00% dengan *precision* 63.00%, *recall* 52.60%, dan *f1-score* 52.30%. Performa yang lebih baik dari *Logistic Regression* menandakan kemampuannya yang lebih tinggi dalam membedakan variasi sentimen, terutama dalam mengklasifikasikan komentar negatif dan netral dengan lebih akurat dibandingkan SVM yang cenderung lebih sering menggolongkan komentar sebagai positif.

Temuan ini menunjukkan bahwa metode *machine learning* dapat digunakan secara efektif untuk menganalisis persepsi publik terhadap kebijakan sosial di media sosial. Hasil penelitian ini dapat dimanfaatkan oleh pemerintah atau pemangku kebijakan untuk memahami respons masyarakat dan meningkatkan strategi komunikasi serta implementasi kebijakan. Penelitian selanjutnya dapat mengeksplorasi model *deep learning* guna meningkatkan akurasi klasifikasi sentimen atau memperluas cakupan analisis dengan mempertimbangkan aspek temporal dan kontekstual dari opini publik.

DAFTAR PUSTAKA

- Ade Dwi Dayani, Yuhandri, & Widi Nurcahyo, G. (2024). Analisis Sentimen Terhadap Opini Publik pada Sosial Media Twitter Menggunakan Metode Support Vector Machine. *Jurnal KomtekInfo*, 1–10. doi:10.35134/komtekinfo.v1i1.439

- Assaidi, S. A., & Amin, F. (n.d.). *Analisis Sentimen Evaluasi Pembelajaran Tatap Muka 100 Persen pada Pengguna Twitter menggunakan Metode Logistic Regression*.
- Chatterjee, S., Rana, N. P., & Dwivedi, Y. K. (2020). Social media as a tool of knowledge sharing in academia: an empirical study using valance, instrumentality and expectancy (VIE) approach. *Journal of Knowledge Management*, 24(10), 2531–2552. doi:10.1108/JKM-04-2020-0252
- Fajri, K., & Rahmaningtyas, G. Y. (2017). InSet Lexicon: Evaluation of a word list for Indonesian sentiment analysis in microblogs. GitHub. Retrieved January 23, 2025, from <https://github.com/fajri91/InSet>
- Fornigulo. (n.d.). Kamus Slang. Kaggle. Retrieved January 23, 2025, from <https://www.kaggle.com/datasets/fornigulo/kamus-slag>
- Husen, R. A., Astuti, R., Marlia, L., Rahmaddeni, R., & Efrizoni, L. (2023). Analisis Sentimen Opini Publik pada Twitter Terhadap Bank BSI Menggunakan Algoritma Machine Learning. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 3(2), 211–218. doi:10.57152/malcom.v3i2.901
- Kementerian Kesehatan Republik Indonesia. (2024). Panduan Hari Gizi Nasional ke 64 Tahun 2024. Retrieved January 23, 2025, from <https://ayosehat.kemkes.go.id/panduan-hari-gizi-nasional-ke-64-tahun-2024>
- Lidinillah, E. R., Rohana, T., & Juwita, A. R. (2023). Analisis sentimen twitter terhadap steam menggunakan algoritma logistic regression dan support vector machine. *TEKNOSAINS : Jurnal Sains, Teknologi Dan Informatika*, 10(2), 154–164. doi:10.37373/tekno.v10i2.440
- Liu, B. (2012). *Sentiment Analysis and Opinion Mining*. Morgan & Claypool Publishers.
- Mallick, C., Mishra, S., Giri, P. K., & Paikaray, B. K. (2023). Machine learning approaches to sentiment analysis in online social networks. *International Journal of Work Innovation*, 3(4), 317. doi:10.1504/IJWI.2023.128860
- Maulana, A., Inayah Khasnaputri Afifah, Asghafi Mubarak, Kiagus Rachmat Fauzan, Ardhan Dwintara, & Zen, B. P. (2023). COMPARISON OF LOGISTIC REGRESSION, MULTINOMIALNB, SVM, AND K-NN METHODS ON SENTIMENT ANALYSIS OF GOJEK APP REVIEWS ON THE GOOGLE PLAY STORE. *Jurnal Teknik Informatika (Jutif)*, 4(6), 1487–1494. doi:10.52436/1.jutif.2023.4.6.863
- Rabbani, S., Safitri, D., Rahmadhani, N., Sani, A. A. F., & Anam, M. K. (2023). Perbandingan Evaluasi Kernel SVM untuk Klasifikasi Sentimen dalam Analisis Kenaikan Harga BBM. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 3(2), 153–160. doi:10.57152/malcom.v3i2.897
- Rimbawan, R., Nurdiani, R., Rachman, P. H., Kawamata, Y., & Nozawa, Y. (2023). School Lunch Programs and Nutritional Education Improve Knowledge, Attitudes, and Practices and Reduce the Prevalence of Anemia: A Pre-Post Intervention Study in an Indonesian Islamic Boarding School. *Nutrients*, 15(4), 1055. doi:10.3390/nu15041055
- Sihombing, E., Halmi Dar, M., & Aini Nasution, F. (2024). Comparison of Machine Learning Algorithms in Public Sentiment Analysis of TAPER Policy. *International Journal of Science, Technology & Management*, 5(5), 1089–1098. doi:10.46729/ijstm.v5i5.1164