

Analisis Sentimen Program Makan Bergizi Gratis Menggunakan *Lexicon-Based* dan *Support Vector Machine*

Zulfikri Akbar^{1,2}, Imam Riadi³, Rusydi Umar⁴

¹S3 Informatika, Fakultas Teknologi Industri, Universitas Ahmad Dahlan

²Informatika, Fakultas Sains, Teknologi dan Kesehatan, Universitas Muhammadiyah Jambi

³Sistem Informasi, Fakultas Sains dan Teknologi Terapan, Universitas Ahmad Dahlan

⁴Magister Informatika, Fakultas Teknologi Industri, Universitas Ahmad Dahlan

¹2537083020@webmail.uad.ac.id, ²zulfikriakbar8668@gmail.com*, ³imam.riadi@is.uad.ac.id, ⁴rusydi@mti.uad.ac.id

Abstract

Public policy initiatives often trigger massive shifts in digital public opinion, such as the Free Nutritious Meal Program (MBG), which has garnered extensive attention from the Indonesian public on social media. Sentiment analysis serves as a vital instrument to map public opinion trends, particularly when dealing with large-scale, unstructured, and heterogeneous textual data. This study aims to analyze the distribution of public sentiment toward the MBG Program and evaluate the effectiveness of the lexicon-based method and Support Vector Machine (SVM) algorithm in classifying opinion texts. The dataset was collected from Twitter (X) via the Kaggle platform, comprising 10,524 public comments. The methodology begins with text preprocessing, including cleaning, case folding, tokenization, normalization, stopword removal, and stemming. Sentiment labeling was performed automatically using a lexicon-based approach referring to the InSet Lexicon to categorize data into three classes: positive, negative, and neutral. Subsequently, text representation was conducted using the Term Frequency–Inverse Document Frequency (TF–IDF) method and classified using an SVM model with a nested cross-validation scheme to maintain performance stability. The results indicate that public opinion is dominated by neutral sentiment at 48.1% (5,066 data points), followed by positive sentiment at 30.8%, and negative sentiment at 21.0%. This dominance of neutral sentiment reflects an informative, descriptive, and cautious public stance toward a policy still in its early implementation stages. Evaluation of the SVM model demonstrates highly stable and reliable performance, achieving an accuracy of 89.26%, with precision, recall, and F1-score each at 89%. This study concludes that the combination of lexicon-based automatic labeling and SVM is effective for public policy sentiment analysis, providing insights into public expectations and concerns regarding government programs.

Keywords: sentiment analysis, lexicon-based approach, support vector machine, TF–IDF, public policy

Abstrak

Kebijakan publik sering kali memicu opini yang masif di ruang digital, salah satunya Program Makan Bergizi Gratis (MBG) yang mendapatkan perhatian luas masyarakat Indonesia di media sosial. Analisis sentimen menjadi instrumen untuk memetakan kecenderungan opini publik, terutama dalam menghadapi data teks berskala besar yang tidak terstruktur dan heterogen. Penelitian ini bertujuan untuk menganalisis distribusi sentimen masyarakat terhadap Program MBG serta mengevaluasi efektivitas metode *lexicon-based* dan algoritma *Support Vector Machine* (SVM) dalam mengklasifikasikan teks opini. Data dikumpulkan dari Twitter (X) melalui Kaggle sebanyak 10.524 komentar masyarakat. Metodologi penelitian diawali *text preprocessing* yang meliputi *cleaning*, *case folding*, *tokenization*, *normalization*, *stopword removal*, dan *stemming*. Pelabelan sentimen dilakukan secara otomatis menggunakan pendekatan *lexicon-based* dengan merujuk pada *InSet Lexicon* untuk mengategorikan data ke dalam tiga kelas: positif, negatif, dan netral. Selanjutnya, teks direpresentasikan melalui metode *Term Frequency–Inverse Document Frequency* (TF–IDF) dan diklasifikasikan menggunakan model SVM dengan skema *nested cross-validation* untuk menjaga stabilitas kinerja model. Hasil penelitian menunjukkan bahwa opini masyarakat didominasi oleh sentimen netral sebesar 48,1% (5.066 data), diikuti sentimen positif 30,8%, dan negatif 21,0%. Dominasi sentimen netral ini merefleksikan sikap masyarakat yang cenderung informatif, deskriptif, dan hati-hati dalam menilai kebijakan yang masih berada pada tahap awal implementasi. Evaluasi model SVM menunjukkan kinerja yang sangat stabil dan handal dengan nilai *accuracy* sebesar 89,26%, serta *precision*, *recall*, dan *F1-score* masing-masing sebesar 89%. Kesimpulan dari penelitian ini adalah kombinasi pelabelan otomatis berbasis *leksikon* dan SVM terbukti efektif untuk analisis sentimen kebijakan publik, memberikan gambaran mengenai ekspektasi serta kekhawatiran publik terhadap program pemerintah.

Kata kunci: analisis sentimen, *lexicon-based*, *support vector machine*, TF–IDF, kebijakan publik

©This work is licensed under a Creative Commons Attribution - ShareAlike 4.0 International License

1. Pendahuluan

Opini masyarakat terhadap kebijakan publik kini banyak diungkapkan melalui media sosial seperti *Twitter* (X), di mana pengguna aktif membagikan komentar dan reaksi mereka terhadap berbagai isu aktual[1]. Analisis sentimen menjadi salah satu metode utama untuk memahami respons publik secara

kuantitatif, terutama dalam menangkap kecenderungan emosional dari teks opini masyarakat yang bersifat besar, tidak terstruktur, dan beragam[2][3]. Melalui pendekatan ini, peneliti dapat mengukur persepsi masyarakat terhadap program atau kebijakan berdasarkan data, sehingga dapat menjadi masukan yang relevan bagi pembuat kebijakan.

Program Makan Bergizi Gratis (MBG) merupakan kebijakan sosial yang mendapatkan respons luas dari masyarakat di Indonesia. Program ini bertujuan untuk meningkatkan akses masyarakat terhadap makanan yang bergizi demi mewujudkan kesehatan dan kesejahteraan publik[4]-[6]. Namun, meskipun penting secara sosial, respon dan opini yang muncul di media sosial terkait pelaksanaan dan kualitas program tersebut belum sepenuhnya dipetakan secara menyeluruh. Hal ini menjadi tantangan karena komentar yang berkembang bersifat heterogen, sering kali ambigu, dan sulit ditafsirkan secara manual karena volumenya yang besar[7].

Permasalahan ilmiah yang muncul adalah bagaimana mengolah dan menganalisis opini publik yang beragam tersebut secara tepat. Analisis manual sering kali tidak efektif karena memakan waktu yang panjang dan rentan terhadap penilaian pribadi. Untuk itu, penelitian ini mengadopsi pendekatan *lexicon-based* dan *Support Vector Machine* (SVM) sebagai metode pendukung untuk melakukan klasifikasi sentimen. Pendekatan *lexicon-based* memanfaatkan kamus kata bersentimen yang telah disesuaikan dengan konteks kebijakan untuk pelabelan awal, sedangkan SVM digunakan sebagai model pembelajaran mesin yang lebih terstruktur untuk meningkatkan kualitas klasifikasi berdasarkan pelabelan awal tersebut[8].

Kajian literatur menunjukkan bahwa model *lexicon-based* telah digunakan secara efektif dalam berbagai konteks untuk memahami opini publik dan memberikan klasifikasi awal terhadap sentimen teks [9]. Sementara itu, SVM dikenal luas sebagai algoritma yang kuat untuk klasifikasi teks-teks karena kemampuannya dalam menangani data berdimensi tinggi dan menghasilkan konsistensi yang baik pada data teks [10][11]. Dengan demikian, kombinasi kedua pendekatan ini diharapkan dapat menghasilkan analisis sentimen yang efisien dari segi komputasi dan juga bermanfaat secara sosial.

Namun demikian, kajian untuk Program MBG di Indonesia masih sangat minim. Penelitian terdahulu belum banyak menggali pola sentimen masyarakat terhadap kebijakan ini dengan metode yang terintegrasi secara menyeluruh. Dengan demikian, terdapat celah penelitian yang perlu diisi dalam literatur tentang persepsi publik terhadap MBG. Tujuan penelitian ini adalah untuk menganalisis distribusi sentimen masyarakat terhadap MBG menggunakan pendekatan *lexicon-based* serta mengevaluasi performa metode SVM dalam klasifikasi sentimen teks opini publik. Diharapkan penelitian ini bisa memberikan gambaran mengenai kecenderungan emosional masyarakat serta kontribusi metodologis dalam konteks analisis sentimen kebijakan publik.

2. Metode Penelitian

Dalam penelitian ini data yang digunakan adalah opini masyarakat yang diperoleh dari hasil *crawling data* media sosial. *Crawling* adalah suatu teknik

mengumpulkan data dari internet secara otomatis untuk kemudian diindeks atau dianalisis [12]. Data penelitian bersumber dari platform Twitter (X) yang dikumpulkan melalui dataset publik Kaggle berjudul *Program Makan Bergizi Gratis 10K Data*. Dataset tersebut berisi 10.524 komentar masyarakat dalam format CSV yang berkaitan dengan topik Program MBG. Seluruh data bersifat teks tidak terstruktur dan belum memiliki label sentimen, sehingga diperlukan proses pengolahan lebih lanjut sebelum dianalisis.

Tahap selanjutnya adalah *text preprocessing* merupakan proses mentransformasi teks mentah menjadi bentuk yang siap dianalisis lebih lanjut dengan tujuan utamanya adalah membersihkan dan memfilter data teks untuk mengurangi *noise* [13][14]. Tahapan *preprocessing* meliputi *cleaning* untuk menghapus URL, *mention*, *hashtag*, angka, dan simbol non-alfabet; *case folding* digunakan untuk menyeragamkan teks menjadi huruf kecil (*lowercase*); *tokenization* untuk memisahkan teks menjadi potongan-potongan kata; *normalization* untuk mengonversi kata tidak baku atau slang ke bentuk formal; *stopword removal* untuk mengeliminasi kata-kata umum yang tidak mengandung bobot informasi atau makna sentimen; serta *stemming* untuk mengubah kata-kata berimbuhan kembali ke bentuk asalnya. Seluruh tahapan ini bertujuan menghasilkan bentuk teks yang lebih konsisten dan relevan untuk proses analisis sentimen[15].

Setelah tahap *preprocessing*, selanjutnya dilakukan pelabelan sentimen secara otomatis menggunakan pendekatan *lexicon-based*. Pendekatan *lexicon-based* sebagai metode analisis sentimen yang mengandalkan keberadaan kata/frasa tertentu dalam dokumen untuk menyimpulkan sentimen, dengan memakai kamus berisi kata dan nilai sentimennya[16]. Pendekatan ini memanfaatkan kamus kata bersentimen yang terdiri dari kata bermakna positif dan negatif, yang disesuaikan dengan konteks kebijakan publik serta merujuk pada *InSet lexicon* sebagai acuan bahasa Indonesia. Setiap komentar diberi label sentimen positif, sentimen negatif, atau sentimen netral berdasarkan gabungan skor kata yang ditransformasikan ke dalam bentuk numerik, dengan skema negatif $\rightarrow 0$, netral $\rightarrow 1$, dan positif $\rightarrow 2$. Pendekatan *lexicon-based* dipilih karena bersifat transparan, dan sesuai untuk pelabelan awal pada dataset berukuran besar [17] dan Penggunaan skema multikelas dipilih karena mampu menggambarkan kompleksitas opini publik secara lebih realistis dibandingkan klasifikasi biner. Dengan pendekatan ini, model diharapkan dapat membedakan sentimen ekstrem dan non-ekstrem, sekaligus meningkatkan kemampuan keakuratan pada data kebijakan publik yang bersifat heterogen [18].

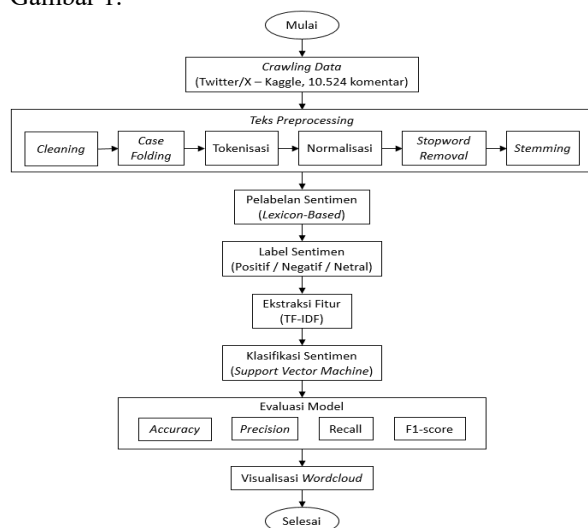
Sebagai validasi awal, sebagian kecil data diperiksa secara manual untuk memastikan kesesuaian polaritas sentiment dengan konteks kebijakan. Label sentimen hasil pendekatan leksikal selanjutnya akan digunakan

sebagai data latih untuk proses klasifikasi menggunakan algoritma *Support Vector Machine* (SVM). SVM dipilih karena memiliki performa yang baik dalam klasifikasi teks dan mampu menghasilkan pemisahan kelas yang optimal [19].

Sebelum proses klasifikasi, teks direpresentasikan ke bentuk numerik dengan metode *Term Frequency–Inverse Document Frequency* (TF-IDF), yang bertujuan mengukur seberapa penting suatu kata dalam satu dokumen relatif terhadap seluruh koleksi dokumen. Bobot tinggi berarti kata itu khas untuk dokumen tersebut dan membantu membedakannya dari dokumen lain[20].

Tahap selanjutnya adalah evaluasi model, dilakukan untuk mengukur kinerja model klasifikasi SVM menggunakan metrik *accuracy*, *precision*, *recall*, dan *F1-score*. Nilai evaluasi ini digunakan untuk menilai sejauh mana model mampu mengklasifikasikan sentimen secara tepat berdasarkan label hasil pendekatan leksikal[21].

Tahap akhir yang dilakukan adalah membuat visualisasi *Wordcloud*. *Wordcloud* merupakan teknik visualisasi teks yang menampilkan kata-kata dari suatu dokumen sehingga kata yang lebih sering muncul digambar lebih besar atau lebih menonjol. Tujuannya untuk memberi gambaran cepat tentang topik utama atau kata-kata penting dalam kumpulan teks, tanpa perlu membaca semuanya[22]. Visualisasi *word cloud* digunakan sebagai ringkasan cepat isi teks, untuk menemukan tema kunci, dan untuk mengkomunikasikan hasil secara visual[23]. Alur metodologi penelitian secara umum ditunjukkan pada Gambar 1.



Gambar 1. Alur Metodologi Penelitian

Gambar 1. menunjukkan tahapan penelitian yang dimulai dari *crawling data* opini masyarakat dari media sosial Twitter (X), dilanjutkan dengan proses *text preprocessing*, pelabelan sentimen dilakukan dengan menggunakan pendekatan *lexicon-based*, klasifikasi sentimen menggunakan SVM, evaluasi performa model, serta visualisasi *wordcloud* untuk

menggambarkan kata dominan pada setiap kategori sentimen.

3. Hasil dan Pembahasan

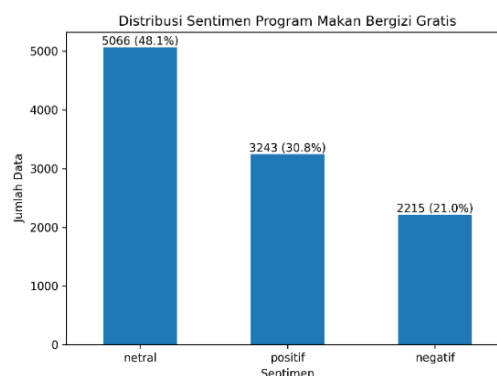
3.1. Distribusi Sentimen Masyarakat

Analisis distribusi sentimen terhadap Program MBG menunjukkan variasi sikap dari masyarakat yang terwakili melalui komentar di media sosial. Berdasarkan pelabelan sentimen dilakukan dengan menggunakan pendekatan *lexicon-based* yang dilakukan pada dataset sebanyak 10.524 komentar *Twitter* (X), diperoleh tiga kategori sentimen yaitu sentimen positif, sentimen negatif, dan sentimen netral, data teks contoh komentar masing-masing kategori sentimen dapat dilihat pada Tabel 1.

Tabel 1. Contoh Komentar Masing-Masing Kategori

Tweet	Sentimen
program makan siang gratis bantu siswa penuhi gizi	Positif
makan siang gratis dukung sehat siswa	Positif
makan siang gratis tidak gizi proyek	Negatif
efek makan siang gratis usaha sepi	Negatif
program makan siang gratis jalan sekolah	Netral
siswa terima makan siang gratis	Netral

Hasil distribusi sentimen divisualisasikan dalam bentuk grafik batang pada Gambar 2 yang menunjukkan jumlah dan persentase masing-masing kategori.



Gambar 2. Grafik Batang Distribusi Sentimen

Gambar 2 yang ditampilkan menunjukk hasil sentimen netral sebanyak 5.066 (48,1%) mendominasi mayoritas opini masyarakat terhadap Program MBG, diikuti oleh sentimen positif 3.243 (30,8%), dan kategori negatif 2.215 (21,0%) memiliki jumlah paling sedikit. Dominasi sentimen netral ini mencerminkan bahwa banyak masyarakat cenderung menyampaikan tanggapan yang bersifat informatif, deskriptif atau berupa penyampaian pendapat tanpa ekspresi emosional yang kuat mengenai kebijakan MBG. Selain itu, netral yang dominan mencerminkan sikap menunggu dan melihat dampak jangka panjang, sementara positif/negatif muncul pada isu efisiensi vs kekhawatiran dan kehilangan pekerjaan[24].

Proporsi antara sentimen positif dan negatif juga memberikan gambaran bahwa meskipun terdapat dukungan terhadap program, beberapa respons negatif muncul tetapi dalam jumlah yang lebih kecil. Pola seperti ini konsisten dengan karakteristik opini publik terhadap isu kebijakan sosial yang luas, dimana respons bersifat seimbang dan tidak terpolarisasi secara berlebihan[25]. Pola distribusi semacam ini sejalan dengan temuan penelitian sentimen kebijakan sosial lainnya, yang menggambarkan opini publik sebagai beragam namun tidak terpolarisasi secara ekstrem, dengan perbedaan pandangan lebih banyak menyangkut aspek implementasi ketimbang penolakan total terhadap tujuan kebijakan[26][27][2].

Penggunaan pendekatan *lexicon-based* dalam pelabelan awal memungkinkan pemetaan distribusi sentimen yang relevan dengan konteks sosial kebijakan publik yang tengah dikaji[28]. Oleh karena itu, dominasi sentimen netral dapat dipahami sebagai gambaran dari kecenderungan masyarakat untuk menyampaikan komentar yang lebih netral dan realistis terhadap kebijakan yang berdampak luas, dibandingkan komentar yang secara eksplisit menunjukkan dukungan atau penolakan.

3.2. Distribusi Sentimen

Dominasi sentimen netral dalam opini masyarakat terhadap Program MBG tidak dapat dimaknai sebagai bentuk ketidakpedulian publik terhadap kebijakan tersebut. Dalam konteks analisis sentimen kebijakan publik, kategori netral justru sering menggambarkan respons yang bersifat informatif, deskriptif, atau evaluatif bersyarat, seperti penyampaian informasi, terlepas dari implementasi program, maupun opini yang belum secara eksplisit menyatakan dukungan atau penolakan. Hal ini sejalan dengan temuan hasil penelitian sebelumnya yang menyebutkan sentimen netral dalam pembahasan kebijakan publik kerap mencerminkan proses cara pandang masyarakat dalam memahami kebijakan, bukan ketiadaan perhatian publik terhadap Program MBG [2][29][30][31].

Karakter sentimen netral pada penelitian ini juga tidak dapat dilepaskan dari sifat Program MBG sebagai kebijakan nasional berskala besar yang masih berada pada tahap awal implementasi. Pada fase awal kebijakan publik, masyarakat cenderung merespons dengan sikap hati-hati dan menunggu bukti nyata terkait dampak kebijakan tersebut, sehingga ekspresi opini yang muncul lebih bersifat netral dan informatif dibandingkan penilaian yang emosional. Kondisi ini diperkuat oleh fakta bahwa tidak seluruh masyarakat telah merasakan manfaat program secara langsung, yang menyebabkan sebagian opini disampaikan dalam bentuk pertanyaan, laporan situasional, atau penilaian bersyarat terhadap potensi keberhasilan program.

Sentimen netral yang mendominasi ini menunjukkan bahwa pendekatan *lexicon-based* mampu menangkap karakteristik opini publik yang tidak selalu terpolarisasi secara berlebihan. Daripada memaksakan klasifikasi biner positif atau negatif, keberadaan kelas netral justru

memberikan gambaran realistis mengenai dinamika persepsi masyarakat terhadap kebijakan publik nasional. Dengan demikian, analisis sentimen pada penelitian ini bukan hanya berfungsi sebagai alat klasifikasi, tetapi juga sebagai sarana untuk memahami tingkat kesiapan, kehati-hatian, dan ekspektasi publik terhadap implementasi Program MBG.

Sentimen positif yang muncul umumnya berkaitan dengan harapan terhadap manfaat program MBG, seperti peningkatan gizi dan kesejahteraan masyarakat. Sebaliknya, sentimen negatif lebih banyak mencerminkan kekhawatiran terhadap aspek implementasi, seperti distribusi, efektivitas, dan potensi kendala teknis. Pola ini menunjukkan bahwa respons masyarakat terhadap MBG tidak hanya sekedar menilai, tetapi juga mencerminkan terhadap tingkat kesulitan pelaksanaan kebijakan publik.

3.3. Analisis Hasil Pelabelan *Lexicon-Based*

Pelabelan sentimen dilakukan secara otomatis dengan menggunakan pendekatan *lexicon-based*. Pendekatan ini dipilih untuk mengatasi keterbatasan skala data yang besar serta untuk menjaga konsistensi label awal sebelum tahap klasifikasi lanjutan menggunakan SVM[2]. Hasil pelabelan menunjukkan bahwa *lexicon-based* mampu menghasilkan distribusi sentimen yang masuk akal secara konteks kebijakan, ditandai dengan dominasi sentimen netral serta proporsi sentimen positif dan negatif yang relatif seimbang.

Label awal yang diberikan *lexicon-based* dapat diamati dari kesesuaian antara polaritas sentimen dan makna isi komentar. Komentar yang mengandung ekspresi dukungan, harapan, atau apresiasi terhadap program cenderung terklasifikasi sebagai positif, sedangkan komentar yang menyoroti kekhawatiran terhadap implementasi, anggaran, atau efektivitas program lebih sering terklasifikasi sebagai negatif. Sementara itu, komentar yang bersifat informatif, deskriptif, atau evaluatif bersyarat secara konsisten masuk ke dalam kategori sentimen netral.

Hasil analisis menunjukkan pendekatan *lexicon-based* berfungsi sebagai mekanisme pelabelan awal yang terkontrol dan masuk akal, sekaligus menyediakan fondasi yang konsisten untuk proses klasifikasi lanjutan. Dengan demikian, metode yang diusulkan dalam penelitian ini tidak hanya efisien secara teknis, tetapi relevan secara mendalam dalam menganalisis opini publik terhadap kebijakan nasional seperti Program MBG.

3.4. Hasil Klasifikasi SVM

Sentimen diklasifikasi menggunakan metode SVM dengan fitur *Term Frequency-Inverse Document Frequency* (TF-IDF). Model dilatih menggunakan label sentimen hasil pendekatan *lexicon-based* dan dioptimalkan menggunakan *nested cross-validation*, di mana *inner loop* digunakan untuk proses optimasi hiperparameter melalui *GridSearchCV*, sedangkan *outer loop* digunakan untuk mengukur kinerja

generalisasi model secara objektif. Pendekatan ini diterapkan untuk memastikan konsistensi kinerja model sekaligus mencegah terjadinya data *leakage* selama proses pelatihan dan evaluasi.

Hasil evaluasi menunjukkan model SVM mampu mencapai nilai konsisten dan relatif tinggi dengan nilai *accuracy* rata-rata 89,26%, nilai *precision (macro)*, *recall (macro)*, dan *F1-score (macro)* masing-masing berada di kisaran 89% dengan simpangan baku yang relatif kecil, yang menunjukkan konsistensi model dalam mengklasifikasikan sentimen positif, sentimen negatif, dan sentimen netral berdasarkan label hasil pendekatan leksikal.

Tabel 2. Hasil Evaluasi Kinerja Model SVM

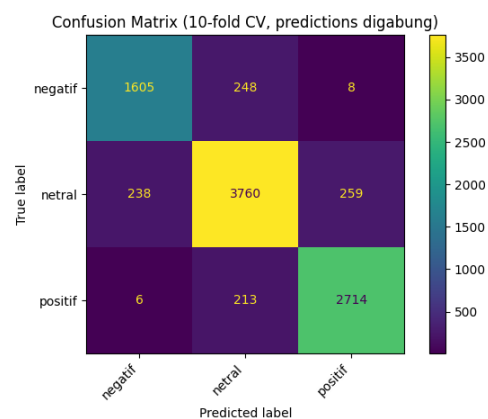
Metrik Evaluasi	Nilai (%)
<i>Accuracy</i>	89,26
<i>Precision</i>	89
<i>Recall</i>	89
<i>F1-Score</i>	89

Konsistensi kinerja model juga tercermin dari hasil *cross-validation* pada skema *nested cross-validation*, di mana variasi performa antar lipatan (*fold*) relatif kecil. Hal ini menunjukkan bahwa model tidak hanya bekerja baik pada *subset* data tertentu, tetapi juga memiliki kemampuan yang memadai terhadap keseluruhan dataset. Temuan ini sejalan dengan penelitian sebelumnya yang menyatakan bahwa SVM merupakan algoritma yang paling efektif dan stabil dalam klasifikasi teks berdimensi tinggi, khususnya ketika dikombinasikan dengan fitur TF-IDF.

Fitur TF-IDF berperan penting dalam menekan pengaruh kata-kata umum dan menonjolkan istilah yang memiliki makna diskriminatif terhadap sentimen. Dengan demikian, ketika label sentimen hasil pendekatan *lexicon-based* digunakan sebagai data latih, SVM dapat mempelajari pola linguistik yang konsisten antara fitur dan label. Hal ini menunjukkan bahwa performa klasifikasi yang diperoleh bukan merupakan kebetulan semata, melainkan hasil dari pengolahan fitur yang tepat dan proses pembelajaran yang terstruktur.

3.5. Visualisasi Confusion Matrix

Analisis *confusion matrix* ditunjukkan pada Gambar 3 menunjukkan dominasi nilai pada diagonal utama, yang menandakan bahwa sebagian besar data berhasil diklasifikasikan secara tepat oleh model SVM. Dominasi diagonal ini menunjukkan bahwa model mampu mengenali pola sentimen utama pada opini masyarakat terhadap Program MBG dan tidak melakukan prediksi secara acak.



Gambar 3. Confusion Matrix hasil klasifikasi SVM

Kesalahan klasik yang muncul terjadi pada kelas sentimen netral. Komentar dengan karakteristik opini bersyarat atau opini ringan cenderung mengalami pergeseran klasifikasi ke kelas positif atau negatif seperti pernyataan yang mengandung unsur harapan sekaligus kekhawatiran seperti yang ditunjukkan pada Tabel 3. Sebaliknya, kesalahan mencolok antara kelas positif dan negatif relatif sangat minim. Pola ini menunjukkan bahwa model SVM mampu memisahkan polaritas sentimen yang berlawanan secara cukup jelas, namun menghadapi tantangan pada opini yang bersifat ambigu atau bersyarat. Hal tersebut selaras dengan karakteristik bahasa opini publik, di mana ekspresi penilaian sering kali tidak disampaikan secara eksplisit sebagai dukungan penuh atau penolakan total. Namun demikian, kesalahan ini relatif dapat diterima pada tahap pelabelan awal dan justru menjadi dasar yang kuat untuk tahap pelatihan model dengan data berlabel menggunakan SVM.

Tabel 3. Contoh Komentar yang Salah Klasifikasi

Tweet	Sentimen	Masalah
program makan gratis niat baik tapi laksana lapang perlu evaluasi	Netral	Mengandung kata positif (pujian), tetapi juga bukan kritik keras
makan gratis jalan tapi belum rasa dampak nyata	Netral	Tidak menyebutkan kata negatif secara eksplisit, tetapi mengandung ketidakpuasan secara implisit.
harap program makan bergizi gratis bantu anak sekolah	Netral	Komentar berbentuk harapan (harap) secara makna mendekati positif, tetapi secara struktur masih netral.
anggaran besar untuk makan gratis semoga tepat guna	Netral	Tidak ada kata negatif, tetapi mengandung kekhawatiran keuangan negara.
ide makan gratis bagus tinggal laksana	Netral	Ada kata positif (bagus), namun diikuti peringatan implementasi.

Temuan tersebut memperkuat argumen bahwa kombinasi pelabelan awal berbasis *lexicon* dengan klasifikasi SVM merupakan pendekatan yang efektif untuk analisis sentimen kebijakan publik. Label *lexicon-based* menyediakan fondasi awal yang terkontrol, sementara SVM berperan dalam mempelajari pola linguistik yang lebih kompleks dari data. Dengan demikian, hasil ini menjawab tujuan

utama penelitian bahwa SVM mampu bekerja secara efektif meskipun dilatih menggunakan label otomatis, selama konteks domain dan *preprocessing* dilakukan secara menyeluruh. Hasil ini juga konsisten dengan studi terkini yang menekankan efektivitas pendekatan *hibrida lexicon-based* dan *machine learning* dalam analisis sentimen berskala besar.

3.5. Visualisasi *Wordcloud* dan Analisis Frekuensi Kata

Untuk memperkuat penjelasan hasil, dilakukan visualisasi *wordcloud* pada setiap kategori sentimen positif, sentimen negatif, dan sentimen netral ditunjukkan pada Gambar 4, 5 dan 6.



Gambar 4. *Wordcloud* Sentimen Positif



Gambar 5. *Wordcloud* Sentimen Negatif



Gambar 6. *Wordcloud* Sentimen Netral

Visualisasi *wordcloud* dan analisis frekuensi kata digunakan sebagai analisis pendukung untuk memberikan gambaran umum mengenai fokus topik dan pola leksikal pada opini masyarakat terkait Program MBG. Hasil visualisasi menunjukkan bahwa kata “makan” dan “gratis” muncul secara dominan pada seluruh kelas sentimen. Dominasi ini bersifat wajar mengingat kedua kata tersebut merupakan inti dari kebijakan yang dibahas, sehingga sering digunakan oleh pengguna media sosial baik dalam konteks evaluatif maupun informatif.

Perbedaan pola kata antar kelas sentimen menunjukkan adanya variasi fokus persepsi publik. Pada sentiment

positif, kata yang sering muncul cenderung berkaitan dengan manfaat dan harapan, seperti “gizi”, “anak”, dan “sehat”, yang mencerminkan ekspektasi masyarakat terhadap dampak kebijakan bagi kesejahteraan generasi muda. Sebaliknya, pada sentimen negatif, kata dominan lebih banyak berkaitan dengan aspek pelaksanaan dan kekhawatiran, seperti “anggaran”, “pelaksanaan”, dan “tidak merata”, yang mengindikasikan perhatian publik terhadap risiko implementasi kebijakan di lapangan.

Kelas sentimen netral didominasi oleh kata-kata informatif seperti “program”, “pemerintah”, dan “sekolah”, yang menunjukkan kecenderungan masyarakat untuk berbagi informasi, bertanya, atau menyampaikan fakta tanpa ekspresi emosional yang kuat. Pola ini memperkuat temuan sebelumnya bahwa kelas netral dalam opini kebijakan publik sering kali berfungsi sebagai ruang diskursif yang bersifat deskriptif dan belum mengarah pada penilaian eksplisit.

4. Kesimpulan

Penelitian ini mengkaji respons publik terhadap Program Makan Bergizi Gratis (MBG) menggunakan pelabelan *lexicon-based*, fitur TF-IDF, dan klasifikasi *Support Vector Machine* (SVM). Hasil analisis menunjukkan opini masyarakat didominasi oleh sentimen netral sebesar 48,1%, diikuti sentimen positif 30,8%, dan negatif 21,0%. Dominasi netral ini merefleksikan sikap publik yang informatif dan hati-hati dalam menilai kebijakan yang baru diimplementasikan. Evaluasi model SVM menunjukkan kinerja yang sangat stabil dan andal dengan pencapaian akurasi sebesar 89,26%, serta nilai *precision*, *recall*, dan *F1-score* yang masing-masing mencapai 89%. Temuan ini membuktikan bahwa kombinasi pelabelan otomatis berbasis leksikon dan SVM efektif sebagai instrumen pemetaan opini publik berskala besar yang efisien namun tetap akurat secara metodologis.

Meskipun memberikan kontribusi signifikan, penelitian ini memiliki keterbatasan pada cakupan data yang hanya bersumber dari satu platform media sosial dan ketergantungan pada kualitas kamus sentimen. Selain itu, ketiadaan metadata waktu membatasi analisis perubahan sentimen secara temporal. Untuk pengembangan ke depan, disarankan penggunaan model *deep learning* atau *transformer-based* sebagai pembanding, serta penerapan *aspect-based sentiment analysis* guna menangkap dimensi evaluasi kebijakan secara lebih spesifik. Penambahan sumber data lintas platform dan informasi berbasis waktu juga sangat potensial untuk memperkaya analisis serta meningkatkan kualitas temuan penelitian mengenai dinamika persepsi masyarakat terhadap kebijakan publik di masa mendatang.

Ucapan Terimakasih

Ucapkan terima kasih penulis sampaikan kepada Bapak Prof. Dr. Imam Riadi, M.Kom dan Bapak Rusydi Umar

S.T., M.T., Ph.D yang merupakan dosen pembimbing yang telah memberikan arahan, masukan, dan bimbingan selama proses penelitian ini. Penulis juga menyampaikan terimakasih kepada pihak Kaggle dan kontributor dataset Program MBG atas ketersediaan data yang digunakan dalam penelitian ini. Selain itu, ucapkan terima kasih penulis sampaikan kepada Universitas Ahmad Dahlan (UAD) serta Universitas Muhammadiyah Jambi (UM Jambi) atas dukungan akademik yang diberikan selama pelaksanaan penelitian ini.

Daftar Rujukan

- [1] D. Latif, M. A. Samad, Rinawulandari, and S. A. Kadir, "Social Media in Shaping Public Opinion Roles and Impact: A Systematic Review," *J. Komun. Malaysian J. Commun.*, vol. 40, no. 2, pp. 205–223, 2024, doi: 10.17576/JKMJC-2024-4002-12.
- [2] R. W. Harwenda, M. David Angelo, I. Budi, A. B. Santoso, and K. Putra, "Sentiment Analysis on Government Public Policies: A Systematic Literature Review," *DIJEMSS*, vol. 6, no. 5, pp. 4192–4211, Jun. 2025, doi: 10.38035/dijemss.v6i5.
- [3] M. Kumar, L. Khan, and H.-T. Chang, "Evolving techniques in sentiment analysis: a comprehensive review," *PeerJ Comput. Sci.*, vol. 11, p. e2592, 2025, doi: 10.7717/peerj-cs.2592.
- [4] M. Dama, A. D. Mayasari, and L. Situmorang, "The Controversy of the Free Meal Programme in the General Election 2024 in Indonesia," *DISCOURSE Indones. J. Soc. Stud. Educ.*, vol. 2, no. 3, pp. 179–191, Jul. 2025, doi: 10.69875/djosse.v2i3.200.
- [5] A. A. Merlinda and Y. Yusuf, "Analisis Program Makan Gratis Prabowo Subianto Terhadap Strategi Peningkatan Motivasi Belajar Siswa di Sekolah Tinjauan dari Perspektif Sosiologi Pendidikan," *Ranah Res. J. Multidiscip. Res. Dev.*, p., 2025, doi: 10.38035/rj.v7i2.1360.
- [6] Z. Akbar, E. Sopandi, B. Badruzzaman, and M. F. Khalik, "The Role of Artificial Intelligence-Based Recommendation Systems in Selection of Courses for Students," *J. Soc. Sci. Util. Technol.*, vol. 1, no. 4, pp. 249–260, Dec. 2023, doi: 10.70177/jssut.v1i4.671.
- [7] N. R. Suaib and D. A. Pratiwi, "Social Network Analysis in the Dissemination of Makan Bergizi Gratis (MBG) Program Information on Social Media X," *J. Sos. Hum. dan Pendidik.*, p., 2025, doi: 10.56127/jushpen.v4i2.2067.
- [8] R. H. Muhammadiyah, T. G. Laksana, and A. B. Arifa, "Combination of Support Vector Machine and Lexicon-Based Algorithm in Twitter Sentiment Analysis," *Khazanah Inform. J. Ilmu Komput. dan Inform.*, p., 2022, doi: 10.23917/khif.v8i1.15213.
- [9] S. Loomba, M. Dave, H. Arolkar, and S. Sharma, "Sentiment Analysis using Dictionary-Based Lexicon Approach: Analysis on the Opinion of Indian Community for the Topic of Cryptocurrency," *Ann. Data Sci.*, vol. 11, pp. 2019–2034, 2023, doi: 10.1007/s40745-023-00496-y.
- [10] L. Luo and L. Li, "Defining and Evaluating Classification Algorithm for High-Dimensional Data Based on Latent Topics," *PLoS One*, vol. 9, p., 2014, doi: 10.1371/journal.pone.0082119.
- [11] B. Hsu, "Comparison of Supervised Classification Models on Textual Data," *Mathematics*, p., 2020, doi: 10.3390/math8050851.
- [12] I. K. Izzah and A. S. Girsang, "Modified TF-Assoc Term Weighting Method for Text Classification on News Dataset from Twitter," *IAENG Int. J. Comput. Sci.*, vol. 48, p. 142, Mar. 2021.
- [13] C. Chai, "Comparison of text preprocessing methods," *Nat. Lang. Eng.*, vol. 29, pp. 509–553, 2022, doi: 10.1017/s1351324922000213.
- [14] M. Braga, G. C. Milanese, and G. Pasi, "Investigating Large Language Models' Linguistic Abilities for Text Preprocessing," *ArXiv*, vol. abs/2510.11482, p., 2025, doi: 10.48550/arxiv.2510.11482.
- [15] A. Rajesh and T. Hiwarkar, "Exploring Preprocessing Techniques for Natural LanguageText: A Comprehensive Study Using Python Code," *Int. J. Eng. Technol. Manag. Sci.*, p., 2023, doi: 10.46647/ijetms.2023.v07i05.047.
- [16] T. Islam *et al.*, "Lexicon and Deep Learning-Based Approaches in Sentiment Analysis on Short Texts," *J. Comput. Commun.*, p., 2024, doi: 10.4236/jcc.2024.121002.
- [17] M. Rizinski, H. Peshov, K. Mishev, M. Jovanovik, and D. Trajanov, "Sentiment Analysis in Finance: From Transformers Back to eXplainable Lexicons (XLex)," *IEEE Access*, vol. 12, pp. 7170–7198, 2023, doi: 10.1109/access.2024.3349970.
- [18] C. Hill, J. Fresneda, and M. Anandarajan, "The wisdom of the lexicon crowds: leveraging on decades of lexicon-based sentiment analysis for improved results," *J. Big Data*, vol. 12, p., 2025, doi: 10.1186/s40537-025-01186-7.
- [19] K. G., "Margin Maximization of Text Classification based on Support Vector Machine," *Int. J. Res. Appl. Sci. Eng. Technol.*, p., 2023, doi: 10.22214/ijraset.2023.57420.
- [20] M. Das, S. Kamalanathan, and P. Alphonse, "A Comparative Study on TF-IDF Feature Weighting Method and Its Analysis Using Unstructured Dataset," pp. 98–107, 2023, doi: 10.48550/arxiv.2308.04037.
- [21] S. Syahrudin, F. A. Artanto, A. R. Maulana, and F. Filsafat, "Metode Support Vector Machine (SVM) dan Lexicon-Based dalam Analisis Sentiment Ulasan Pengguna Aplikasi Wink," *JUMINTAL J. Manaj. Inform. dan Bisnis Digit.*, p., 2025, doi: 10.55123/jumintal.v4i1.5236.
- [22] I. K. Izzah and A. S. Girsang, "Association on Supervised Term Weighting Method for Classification on Data Twitter," *Int. J. Recent Technol. Eng.*, vol. 8, no. 6, pp. 859–863, Mar. 2020, doi: 10.35940/ijrte.f6975.038620.
- [23] M. Skeppstedt, M. Ahltop, K. Kucher, and M. Lindström, "From word clouds to Word Rain: Revisiting the classic word cloud to visualize climate change texts," *Inf. Vis.*, vol. 23, pp. 217–238, 2024, doi: 10.1117/14738716241236188.
- [24] N. A. Pandia, P. Ramadani, S. P. U. Ritonga, F. Aulia, and Mhd.Furqan, "Analisis Sentimen Masyarakat terhadap Penggunaan Teknologi AI dengan Metode Machine Learning," *J. Ilm. Sist. Inf. dan Ilmu Komput.*, p., 2025, doi: 10.55606/juisik.v5i2.1198.
- [25] O. Ediansa, Zulfikri Akbar, Saleh Yaakub, Agesha Marsyaf, and Iwan Eka Putra, "Pemanfaatan Teknologi Dalam Pemasaran Produk Tenan Inkubator Bisnis Teknologi LPPM Universitas Muhammadiyah Jambi," *J. Pengabd. UntukMu NegeRI*, vol. 8, no. 2, pp. 263–267, Aug. 2024, doi: 10.37859/jpumri.v8i2.7430.
- [26] P. Raval, "NLP-Based Sentiment Analysis of Social Media Data on Public Perception of Environmental Policies," *Int. J. Environ. Sci.*, p., 2025, doi: 10.64252/1ygfxv29.
- [27] L. Chen, L. Huang, Wanqi, Suwei, and Y. Li, "The study on the dissemination of waste sorting policies on social media and the public's feedback attitudes: a text analysis based on comment data of policies in 46 key cities in China," *Front. Environ. Sci.*, p., 2025, doi: 10.3389/fenvs.2025.1546136.
- [28] Y. Qi and Z. Shabrina, "Sentiment analysis using Twitter data: a comparative application of lexicon- and machine-learning-based approach," *Soc. Netw. Anal. Min.*, vol. 13, no. 1, pp. 1–14, Feb. 2023, doi: 10.1007/s13278-023-01030-x.
- [29] I. D. Baga, G. K. Pati, and L. L. Momo, "ANALYSIS OF PUBLIC SENTIMENT TOWARDS VILLAGE OFFICIALS IN DISTRIBUTION OF SOCIAL ASSISTANCE TO THE COMMUNITY," *JTH J. Technol. Heal.*, p., 2025, doi: 10.61677/jth.v2i2.90.
- [30] F. Omojowo, "Hybrid data mining for pandemic public opinion analysis: integrating sentiment, topic, and geolocation data," *J. Electr. Syst. Inf. Technol.*, vol. 12, p., 2025, doi: 10.1186/s43067-025-00281-2.
- [31] K. Khadafi, F. Irhamn, D. Fatah, and U. Madura, "Analisis Sentimen Opini Masyarakat Terhadap Kebijakan Kenaikan PPN (Pajak Pertambahan Nilai) pada Media Sosial Twitter dengan Metode Support Vector Machine," *Com. Commun. Inf. Technol. J.*, p., 2025, doi: 10.47467/comit.v3i2.8778.