

ANALISIS TOPIK KELUHAN PROGRAM MAKAN BERGIZI GRATIS PADA PLATFORM X MENGGUNAKAN *LATENT DIRICHLET ALLOCATION*

Afifah Risti Ardi^{1*}, Nina Sulistiyowati², Ahmad Khusaeri³

^{1,2,3}Universitas Singaperbangsa Karawang; Jl. HS. Ronggo Waluyo, Puseurjaya, Telukjambe Timur, Karawang, Jawa Barat 41361; (0267) 641177

Keywords:

Program Makan Bergizi Gratis; *Topic Modeling*; *Latent Dirichlet Allocation*; CRISP-DM; Platform X.

Correspondent Email:

afifahrst1@gmail.com

Abstrak. Program Makan Bergizi Gratis (MBG) yang diluncurkan pada 6 Januari 2025 merupakan program prioritas pemerintah untuk meningkatkan status gizi masyarakat. Namun, pelaksanaannya masih menghadapi permasalahan seperti kasus keracunan makanan, kendala distribusi, dan keterbatasan anggaran. Hal tersebut memicu tanggapan dan keluhan masyarakat di platform X yang bersifat tidak terstruktur dan berjumlah besar sehingga sulit dianalisis secara manual. Penelitian ini bertujuan untuk mengidentifikasi topik keluhan masyarakat terhadap pelaksanaan Program MBG menggunakan model LDA (Latent Dirichlet Allocation) dengan metodologi CRISP-DM. Data penelitian terdiri dari 6.748 tweet yang dikumpulkan melalui teknik web scraping pada periode 6 Januari hingga 31 Desember 2025. Setelah melalui preprocessing, pelabelan sentimen menggunakan Leksikon InSet, serta seleksi *tweet* berlabel negatif, diperoleh 2.075 *tweet* yang dianalisis menggunakan model LDA dengan pembobotan TF-IDF. Hasil penelitian menunjukkan jumlah topik optimal adalah 5 topik dengan coherence score sebesar 0,386. Berdasarkan model tersebut, diidentifikasi 5 tema utama, yaitu Akuntabilitas & Tata Kelola Anggaran, Dampak bagi Kesehatan Anak di Sekolah, Kualitas & Kelayakan Menu Makanan, Kendala Distribusi & Pemenuhan Gizi, serta Insiden Keracunan & Keamanan Pangan.



Copyright © 2026 The Author(s), Published by Jurnal Informatika dan Teknik Elektro Terapan. This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (CC BY-NC 4.0).

Abstract. The Free Nutritious Meal Program (MBG), launched on January 6, 2025, is a government priority program to improve the nutritional status of the community. However, its implementation still faces problems such as food poisoning cases, distribution constraints, and budget constraints. This triggers public responses and complaints on platform X which are unstructured and large in number, making it difficult to analyze manually. This study aims to identify topics of public complaints regarding the implementation of the MBG Program using the LDA (Latent Dirichlet Allocation) model with the CRISP-DM methodology. The research data consists of 6,748 tweets collected through web scraping techniques from January 6 to December 31, 2025. After preprocessing, sentiment labeling using the InSet Lexicon, and selection of negatively labeled comments, 2,075 comments were obtained which were analyzed using the LDA model with TF-IDF weighting. The results show the optimal number of topics is 5 topics with a coherence score of 0.386. Based on this model, 5 main themes were identified, namely Accountability & Budget Governance, Impact on Children's Health in Schools, Quality & Adequacy of Food Menus, Distribution Constraints & Nutrition Fulfillment, and Food Poisoning & Safety Incidents.

1. PENDAHULUAN

Masalah gizi masih menjadi salah satu tantangan utama di Indonesia, terutama pada kelompok rentan seperti anak-anak dan ibu hamil. Di berbagai daerah, stunting serta keterbatasan akses terhadap makanan bergizi masih menjadi permasalahan yang mendesak [1]. Untuk mengatasi hal tersebut, pemerintah meluncurkan Program Makan Bergizi Gratis (MBG) yang mulai diimplementasikan pada 6 Januari 2025. Program ini menargetkan sekitar 82,9 juta penerima manfaat, meliputi ibu hamil, ibu menyusui, balita, serta peserta didik dari berbagai jenjang pendidikan [2], [3].

Seiring pelaksanaannya, masyarakat mulai memberikan beragam respons terhadap Program MBG. Berdasarkan laporan Litbang Kompas, pemantauan terhadap 74.688 unggahan di lima platform media sosial selama periode 17-25 September 2025 menemukan bahwa 66% warganet memberikan sentimen negatif terhadap program ini [4]. Tingginya persentase sentimen negatif menunjukkan adanya ketidakpuasan masyarakat terhadap efektivitas dan pelaksanaan Program MBG. Hal ini sejalan dengan penelitian Nango et al. (2025) yang menyebutkan bahwa pelaksanaan Program MBG di beberapa daerah masih menghadapi kendala, seperti keterbatasan infrastruktur, kurangnya tenaga kerja terlatih, lemahnya pengawasan, dan isu transparansi anggaran [5].

Di era digital saat ini, platform X (sebelumnya Twitter) merupakan media sosial yang banyak digunakan untuk menyampaikan pendapat secara terbuka [6]. Berdasarkan laporan We Are Social dan Meltwater pada Juli 2025, Indonesia menempati posisi keempat dengan sekitar 23,76 juta pengguna aktif [7]. Namun, data unggahan di platform X bersifat tidak terstruktur, jumlahnya sangat besar, dan terus bertambah sehingga sulit dianalisis secara manual [8].

Untuk menganalisis data tersebut secara sistematis, penelitian ini menggunakan metodologi CRISP-DM (*Cross-Industry Standard Process for Data Mining*) yang terdiri dari enam tahap terstruktur, serta model LDA (*Latent Dirichlet Allocation*) sebagai metode *topic modeling*. LDA merupakan model probabilistik yang efektif untuk menemukan topik tersembunyi dalam kumpulan dokumen teks [9]. Berbagai penelitian sebelumnya

menunjukkan keefektifan LDA dalam menganalisis data media sosial. Sembiring (2025) mengidentifikasi delapan topik diskusi Pemilu 2024 dengan coherence 0,4467 [10]. Muhaimin et al. (2023) menemukan empat topik dominan stunting dengan coherence 0,6 [11]. Nulhakim & Shiddieq (2025) mengidentifikasi tujuh topik kebijakan PPN 12% dengan coherence tertinggi 0,88 [12]. Namun, hingga saat ini belum ditemukan penelitian yang menggunakan LDA untuk menganalisis topik keluhan masyarakat terhadap pelaksanaan Program MBG melalui platform X. Penelitian ini mengisi kesenjangan tersebut dengan menerapkan metodologi CRISP-DM dan model LDA untuk mengidentifikasi topik keluhan masyarakat, mengevaluasi kualitas model menggunakan *coherence score*, serta memanfaatkan hasil analisis sebagai masukan untuk perbaikan program.

2. TINJAUAN PUSTAKA

2.1. Program Makan Bergizi Gratis

Program Makan Bergizi Gratis (MBG) adalah program prioritas pemerintah yang diluncurkan pada 6 Januari 2025 untuk meningkatkan status gizi masyarakat, khususnya peserta didik dan kelompok rentan. Program ini menjadi salah satu upaya pemerintah dalam mendukung pembangunan sumber daya manusia menuju Indonesia Emas 2045 [13].

2.2. Platform X

X (sebelumnya Twitter) adalah platform media sosial yang digunakan penggunanya untuk menyampaikan opini, pengalaman, dan tanggapan terhadap isu publik secara cepat melalui teks singkat. Karakteristik data yang tidak terstruktur dan bersifat *real-time* menjadikan platform ini relevan sebagai sumber data analisis opini publik [14].

2.3. Text Mining

Text mining merupakan proses mengekstraksi informasi penting dari data teks tidak terstruktur melalui identifikasi pola dan hubungan dalam teks. Proses ini digunakan untuk mengubah teks mentah dari sumber seperti media sosial, situs web, ulasan, dan komentar menjadi informasi yang bermakna [15].

2.4. Analisis Sentimen

Analisis sentimen merupakan proses mendeteksi serta mengkategorikan opini maupun emosi yang terdapat dalam suatu teks. Proses ini bertujuan untuk menilai apakah teks mengandung sentimen positif, negatif, atau netral, sehingga dapat memberikan gambaran pandangan atau sikap seseorang terhadap suatu topik [13].

2.5. TF-IDF

Term Frequency-Inverse Document Frequency (TF-IDF) merupakan metode klasik yang digunakan dalam analisis teks untuk menilai tingkat kepentingan suatu kata pada dokumen. TF mengukur frekuensi kemunculan kata dalam dokumen, sedangkan IDF mengukur tingkat kelangkaan kata pada seluruh dokumen. Kombinasi keduanya menghasilkan bobot kata yang lebih representatif [16], [17]. Untuk mengurangi pengaruh panjang dokumen, bobot TF-IDF dinormalisasi menggunakan *cosine normalization* agar setiap dokumen memiliki skala vektor yang seragam. Secara matematis dituliskan sebagai berikut [18].

$$W_{norm}(d, t) = \frac{TF(d, t) \times \ln\left(\frac{D}{df(t)}\right)}{\sqrt{\sum_{i=1}^n \left(TF(d, i) \times \ln\left(\frac{D}{df(i)}\right)\right)^2}}$$

Keterangan:

$W_{norm}(d, t)$ = bobot kata t pada dokumen d setelah normalisasi

$TF(d, t)$ = frekuensi kemunculan kata t dalam dokumen d

D = jumlah seluruh dokumen dalam korpus

$df(t)$ = jumlah dokumen yang memuat kata t

$TF(d, i)$ = frekuensi kata ke- i dalam dokumen d

$df(i)$ = jumlah dokumen yang memuat kata ke- i

n = jumlah seluruh kata unik dalam dokumen

2.6. Latent Dirichlet Allocation (LDA)

Latent Dirichlet Allocation (LDA) adalah model probabilistik yang digunakan untuk menemukan tema atau topik tersembunyi dalam sekumpulan dokumen. Model ini mengasumsikan bahwa setiap dokumen terdiri dari beberapa topik, dan setiap topik memiliki distribusi kata dengan probabilitas tertentu. Distribusi *Dirichlet* digunakan untuk menentukan proporsi topik dalam setiap dokumen serta distribusi kata dalam setiap topik [19].

Dalam proses inferensi, LDA bertujuan menghitung distribusi posterior $p(z | w)$. Namun, karena bersifat intractable, estimasi dilakukan menggunakan metode Gibbs Sampling yang secara matematis dituliskan sebagai berikut [20], [21].

$$P(z_i = k | z_{-i}, w_i, d_i) \propto \frac{n_{d_i k} + \alpha}{\sum_{k=1}^K n_{d_i k} + K\alpha} \times \frac{n_{kw_i} + \beta}{\sum_{w=1}^W n_{kw} + W\beta}$$

Keterangan:

$(z_i = k)$ = penugasan topik k untuk kata ke- i

z_{-i} = penugasan topik untuk semua kata selain kata ke- i

w_i = kata yang sedang diamati

d_i = dokumen yang sedang diamati

$n_{d_i k}$ = jumlah kemunculan topik k pada dokumen d

$\sum_{k=1}^K n_{d_i k}$ = jumlah kemunculan semua topik pada dokumen d

n_{kw_i} = jumlah kemunculan kata w pada topik k

$\sum_{w=1}^W n_{kw}$ = jumlah kemunculan semua kata pada topik k

K = jumlah topik

W = jumlah kosakata

2.7. Coherence Score

Coherence score merupakan metrik yang digunakan untuk menilai kualitas topik yang dihasilkan oleh *topic modeling* seperti LDA. Metrik ini mengukur konsisten atau relevan kata-kata dalam setiap topik. Semakin tinggi nilai *coherence score*, semakin baik kualitas topik karena kata-kata di dalamnya memiliki keterkaitan semantik [22]. Secara matematis, *coherence score* untuk setiap topik dituliskan sebagai berikut [23].

$$C_v(T_k) = \frac{1}{|S|} \sum_{(w_i, w_j) \in S} \text{cosine_sim}(V(w_i), V(w_j))$$

Keterangan:

$C_v(T_k)$ = *coherence score* untuk topik ke- k

$|S|$ = jumlah total pasangan kata (w_i, w_j) yang diamati dalam satu topik

$V(w)$ = *context vector* kata

$\text{cosine_sim}(V(w_i), V(w_j))$ = kemiripan antara dua *context vector*

2.8. pyLDAvis

pyLDAvis merupakan pustaka Python untuk memvisualisasikan hasil *topic modeling* secara interaktif sehingga memudahkan interpretasi topik pada model LDA [24]. Visualisasi ini terdiri dari dua panel, yaitu peta jarak antar topik yang menunjukkan keterpisahan dan

kesamaan antar topik berdasarkan ukuran dan jarak lingkaran, serta panel istilah yang menampilkan kata paling dominan berdasarkan *saliency* dengan pengaturan relevansi menggunakan parameter λ [9].

2.9. Word Cloud

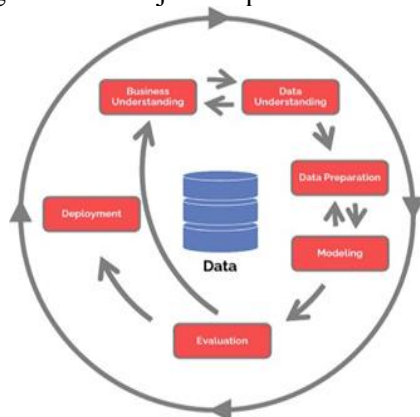
Word cloud adalah visualisasi teks yang menampilkan kata berdasarkan frekuensi kemunculannya, di mana kata yang sering muncul ditampilkan dengan ukuran lebih besar [25].

2.10. CRISP-DM

Cross-Industry Standard Process for Data Mining (CRISP-DM) adalah metodologi yang banyak digunakan pada proyek *data mining* dan *data science*. Metodologi ini memberikan langkah-langkah terstruktur, mulai dari identifikasi kebutuhan bisnis hingga penerapan hasil analisis, yang terdiri dari enam tahapan, yaitu (1) *Business Understanding*, (2) *Data Understanding*, (3) *Data Preparation*, (4) *Modeling*, (5) *Evaluation*, dan (6) *Deployment* [26].

3. METODE PENELITIAN

Penelitian ini menggunakan metodologi CRISP-DM (*Cross-Industry Standard Process for Data Mining*) melalui enam tahapan, sebagaimana ditunjukkan pada Gambar 1.



Gambar 1 Tahapan CRISP-DM

3.1. Business Understanding

Pada tahap *business understanding* akan dilakukan pemahaman terhadap permasalahan yang menjadi fokus penelitian serta tujuan penelitian yang ingin dicapai terkait keluhan terhadap pelaksanaan Program Makan Bergizi Gratis (MBG).

3.2. Data Understanding

Pada tahap *data understanding* akan dilakukan pemahaman terhadap kondisi awal *dataset* yang digunakan dalam penelitian. Tahap ini akan meliputi proses pengumpulan data dari platform X menggunakan teknik *web scraping* dengan kata kunci “makan bergizi gratis” pada periode 6 Januari hingga 31 Desember 2025. Selanjutnya akan dilakukan identifikasi terhadap jumlah data, atribut, dan tipe data untuk memahami karakteristik *dataset*.

3.3. Data Preparation

Pada tahap *data preparation* akan dilakukan pengolahan *dataset* agar siap digunakan untuk analisis. Tahap ini akan meliputi seleksi data, *preprocessing* untuk pelabelan sentimen, pelabelan dan seleksi sentimen, *preprocessing* untuk *modeling*, serta ekstraksi fitur. Seleksi data akan dilakukan dengan memilih atribut yang relevan terhadap tujuan penelitian serta menyesuaikan struktur *dataset* jika diperlukan. Tahap *preprocessing* untuk pelabelan sentimen akan mencakup pembersihan data awal (*missing value* dan duplikasi), *text cleaning*, *case folding*, serta normalisasi kata. Selanjutnya akan dilakukan pelabelan sentimen menggunakan *lexicon-based* dengan Leksikon InSet untuk mengklasifikasikan teks menjadi positif, netral, dan negatif, kemudian difokuskan pada data berlabel negatif sebagai representasi keluhan. Pada tahap *preprocessing* untuk pemodelan akan mencakup *tokenizing*, *stopword removal*, pembentukan frasa (bigram dan trigram), serta pembersihan data akhir sebelum digunakan dalam analisis. Ekstraksi fitur akan dilakukan menggunakan TF-IDF untuk menghasilkan bobot kata yang mencerminkan tingkat kepentingannya dalam korpus, yang selanjutnya digunakan sebagai input pemodelan.

3.4. Modeling

Pada tahap *modeling* akan dilakukan identifikasi topik keluhan masyarakat terhadap pelaksanaan Program Makan Bergizi Gratis (MBG) menggunakan model LDA (*Latent Dirichlet Allocation*). Tahap ini meliputi penentuan jumlah topik optimal menggunakan *coherence score* serta implementasi model LDA untuk menghasilkan kumpulan topik beserta kata kunci yang merepresentasikan keluhan masyarakat.

3.5. Evaluation

Pada tahap *evaluation* akan dilakukan evaluasi terhadap pemodelan topik menggunakan *coherence score* untuk menentukan jumlah topik yang paling koheren dan mudah diinterpretasikan.

3.6. Deployment

Pada tahap *deployment* akan dilakukan penyajian hasil pemodelan topik secara visual serta interpretasi hasil analisis sebagai dasar penyusunan rekomendasi perbaikan Program Makan Bergizi Gratis (MBG). Tahap ini akan meliputi visualisasi topik menggunakan *word cloud* dan *pyLDAvis*, interpretasi topik, serta penyusunan rekomendasi berdasarkan hasil analisis

4. HASIL DAN PEMBAHASAN

4.1. Business Understanding

Hasil tahap *business understanding* menunjukkan bahwa pelaksanaan Program Makan Bergizi Gratis (MBG) masih menghadapi berbagai permasalahan, seperti kasus keracunan makanan, belum meratanya operasional SPPG, keterlambatan pelayanan, serta keterbatasan anggaran yang memengaruhi kualitas pelaksanaan program. Selain itu, munculnya tagar seperti #StopMBG di media sosial menunjukkan adanya ketidakpuasan masyarakat terhadap program tersebut. Berdasarkan kondisi tersebut, penelitian ini berfokus pada analisis keluhan masyarakat terhadap pelaksanaan program MBG dengan menggunakan data dari platform X. Data tersebut bersifat tidak terstruktur dan berjumlah besar sehingga sulit dianalisis secara manual. Oleh karena itu, penelitian ini bertujuan untuk mengidentifikasi topik keluhan masyarakat menggunakan metodologi CRISP-DM dan model LDA, mengevaluasi kualitas topik menggunakan *coherence score*, serta memanfaatkan hasil analisis topik sebagai masukan untuk mendukung perbaikan pelaksanaan program.

4.2. Data Understanding

Hasil pengumpulan data dari platform X menggunakan teknik *web scraping* dengan kata kunci “makan bergizi gratis” pada periode 6 Januari hingga 31 Desember 2025 diperoleh sebanyak 6.748 tweet. Data tersebut disimpan

dalam format CSV sebagai dataset awal untuk proses analisis.

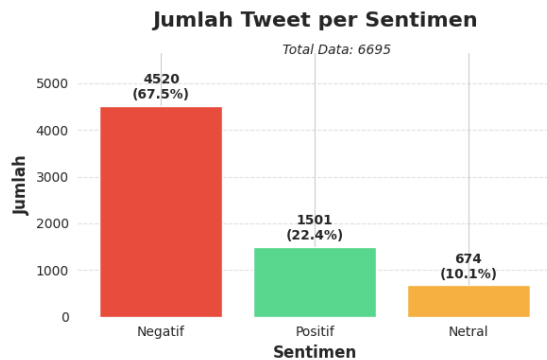
Hasil identifikasi menunjukkan dataset awal terdiri dari 6.748 baris data dengan 15 kolom atribut yang mencakup tipe data numerik (*int64* dan *float64*) dan tipe data teks (*object*), seperti *full_text* yang berisi teks *tweet*, *created_at* yang menunjukkan waktu publikasi, serta *favorite_count*, *retweet_count*, dan *reply_count* yang menggambarkan tingkat interaksi pengguna. Selain itu, terdapat atribut lain seperti *id_str*, *conversation_id_str*, *username*, *lang*, *location*, *image_url*, *tweet_url*, *in_reply_to_screen_name*, dan *quote_count* yang berfungsi sebagai metadata pelengkap dari setiap *tweet*.

4.3. Data Preparation

Hasil seleksi data menunjukkan dari 15 atribut yang tersedia pada dataset awal, hanya atribut *full_text* yang dipilih sebagai atribut yang relevan dengan tujuan penelitian karena analisis topik berfokus pada isi teks *tweet* sebagai dasar pembentukan topik.

Hasil tahap *preprocessing* untuk pelabelan sentimen dilakukan melalui pembersihan data awal, *text cleaning*, *case folding*, dan normalisasi kata. Pada proses pembersihan data awal tidak ditemukan data kosong, namun terdapat 53 data duplikat sehingga jumlah data berkurang menjadi 6.695 *tweet* setelah penghapusan duplikat. Selanjutnya, *text cleaning* dilakukan dengan menghapus elemen tidak relevan seperti URL, emoji, hashtag, simbol, dan angka, kemudian dilanjutkan dengan *case folding* untuk menyeragamkan huruf menjadi huruf kecil serta normalisasi kata menggunakan kamus normalisasi agar kata tidak baku diubah menjadi bentuk baku.

Hasil pelabelan sentimen terhadap 6.695 *tweet* menunjukkan 4.520 *tweet* (67,5%) termasuk kategori sentimen negatif, 1.501 *tweet* (22,4%) kategori sentimen positif, dan 674 *tweet* (10,1%) kategori sentimen netral, sebagaimana ditampilkan pada Gambar 2. Dari 4.250 *tweet* berlabel negatif yang menjadi fokus analisis, dipilih *tweet* dengan skor sentimen *InSet* kurang dari -2 karena skor yang mendekati nol rentan terhadap kesalahan klasifikasi dan kurang dapat diandalkan [27]. Berdasarkan proses seleksi tersebut, diperoleh sebanyak 2.237 *tweet*.



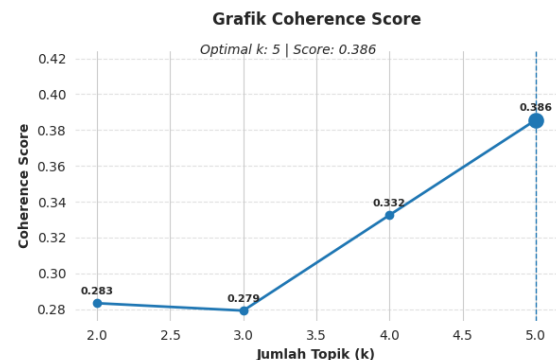
Gambar 2 Hasil Pelabelan Sentimen

Hasil tahap *preprocessing* untuk *modeling* dilakukan melalui *tokenizing*, *stopword removal*, identifikasi frasa, dan pembersihan data akhir. *Tokenizing* digunakan untuk memecah teks menjadi kumpulan token, kemudian *stopword removal* dilakukan untuk menghapus kata umum yang tidak informatif serta menyaring token dengan panjang kurang dari tiga karakter guna mengurangi noise. Selanjutnya, identifikasi frasa menggunakan model bigram dan trigram dari pustaka Gensim untuk menggabungkan pasangan kata yang sering muncul menjadi satu token. Pada tahap pembersihan data akhir tidak ditemukan data kosong, namun terdapat 162 data duplikat yang dihapus sehingga tersisa 2.075 *tweet* untuk proses ekstraksi fitur.

Hasil ekstraksi fitur berupa matriks TF-IDF berukuran 2.075 dokumen $\times$ 560 fitur. Matriks ini menunjukkan terdapat 560 kata unik yang digunakan sebagai representasi numerik dalam pemodelan.

4.4. Modeling

Hasil pengujian jumlah topik menunjukkan model LDA dengan 5 topik merupakan jumlah topik yang optimal dengan *coherence score* sebesar 0,386. Nilai tersebut lebih tinggi dibandingkan model dengan 2 hingga 4 topik, sehingga menghasilkan keterkaitan antar kata yang lebih koheren sebagaimana ditunjukkan pada Gambar 3. Pembatasan rentang jumlah 2 hingga 5 topik dilakukan untuk menjaga keterpahaman model dan mengurangi tumpang tindih kata antar topik karena terlalu banyak topik dapat menyulitkan interpretasi [28]



Gambar 3 Grafik *Coherence Score*

Hasil pemodelan berupa 5 topik utama yang masing-masing direpresentasikan oleh kumpulan kata kunci dengan bobot probabilitas tertinggi. Bobot tersebut menunjukkan probabilitas kemunculan kata pada suatu topik. Pemodelan dilakukan dengan melatih model LDA menggunakan korpus teks hasil *preprocessing* dengan parameter $\alpha=0,5$ dan $\eta=0,1$. Hasil pemodelan topik ditampilkan pada Tabel 1.

Tabel 1 Hasil Pemodelan Topik

Topik	Kata Kunci
1	0.067*"korupsi" + 0.040*"pemerintah" + 0.035*"rakyat" + 0.035*"negara" + 0.032*"pendidikan" + 0.030*"anggaran" + 0.023*"pajak" + 0.022*"layak" + 0.020*"program" + 0.020*"janji"
2	0.067*"anak" + 0.042*"sekolah" + 0.026*"buang" + 0.021*"kaya" + 0.020*"orang" + 0.020*"pakai" + 0.019*"salah" + 0.018*"mubazir" + 0.017*"sakit" + 0.017*"kesehatan"
3	0.042*"makanan" + 0.038*"sayur" + 0.032*"sehat" + 0.026*"sia" + 0.026*"menu" + 0.025*"menunya" + 0.025*"suka" + 0.018*"bayar" + 0.015*"siswa" + 0.015*"pejabat"
4	0.033*"ribu" + 0.032*"gizi" + 0.031*"mentah" + 0.029*"gagal" + 0.028*"anak" + 0.024*"vendor" + 0.022*"kali" + 0.021*"susu" + 0.018*"program" + 0.018*"telat"
5	0.039*"keracunan" + 0.029*"mbg" + 0.022*"makanan" + 0.020*"beracun" + 0.017*"siswa" + 0.016*"korban" + 0.016*"banget" + 0.014*"basi" + 0.014*"program" + 0.013*"mending"

4.5. Evaluation

Hasil evaluasi pemodelan topik menggunakan *coherence score* menunjukkan model LDA dengan 5 topik menghasilkan nilai tertinggi sebesar 0,386, dibandingkan model dengan 2 hingga 4 topik sebagaimana ditunjukkan pada Tabel 2. Nilai ini

menunjukkan bahwa model 5 topik menghasilkan topik yang lebih koheren dan mudah diinterpretasikan, sehingga dipilih sebagai jumlah topik final.

Tabel 2 Hasil *Coherence Score*

Jumlah Topik	Coherence Score
2 Topik	0.283
3 Topik	0.279
4 Topik	0.332
5 Topik	0.386

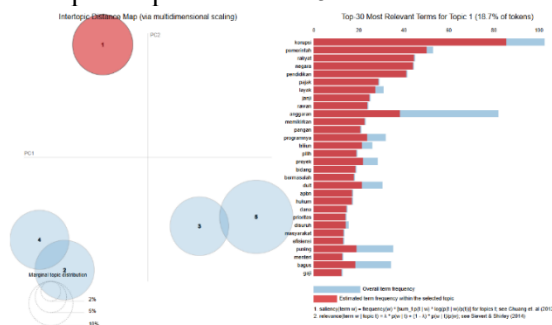
4.6. Deployment

Hasil visualisasi pemodelan topik menggunakan *word cloud* pada Gambar 4 menunjukkan perbedaan kata dominan pada setiap topik.



Gambar 4 Word Cloud Topik

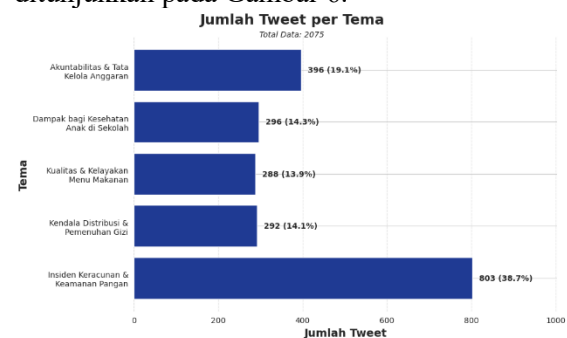
Hasil visualisasi pemodelan topik menggunakan pyLDAvis menunjukkan setiap topik memiliki posisi dan ukuran berbeda pada *Intertopic Distance Map* yang mencerminkan perbedaan kosakata antar tema. Selain itu, *Top-30 Most Relevant Terms* pada setiap topik memperjelas fokus pembahasan masing-masing tema. Contoh visualisasi pyLDAvis ditampilkan pada Gambar 5.



Gambar 5 pyLDAvis Topik 1

Hasil interpretasi topik menghasilkan lima tema utama keluhan masyarakat terhadap pelaksanaan Program MBG, yaitu: (1) akuntabilitas dan tata kelola anggaran, (2) dampak bagi kesehatan anak di sekolah, (3) kualitas dan kelayakan menu makanan, (4) kendala distribusi dan pemenuhan gizi, serta (5) insiden keracunan dan keamanan pangan. Tema insiden keracunan dan keamanan pangan menjadi topik dengan proporsi tertinggi sebesar

38,7% dari total *tweet*, sebagaimana ditunjukkan pada Gambar 6.



Gambar 6 Grafik Distribusi *Tweet* per Tema

Berdasarkan hasil analisis topik, dirumuskan beberapa rekomendasi perbaikan program, antara lain peningkatan transparansi anggaran, pemantauan kesehatan anak secara berkala, penyusunan standar menu berbasis gizi, perbaikan sistem distribusi makanan, serta penerapan standar keamanan pangan untuk mencegah kasus keracunan makanan.

5. KESIMPULAN

- Penelitian ini berhasil mengidentifikasi topik keluhan masyarakat terhadap pelaksanaan Program Makan Bergizi Gratis (MBG) secara sistematis menggunakan metodologi CRISP-DM dan model LDA. Dari 6.748 tweet yang dikumpulkan di platform X, diperoleh 2.075 *tweet* keluhan digunakan dalam pemodelan topik. Hasil pengujian menunjukkan bahwa model dengan 5 topik merupakan jumlah topik optimal dengan *coherence score* sebesar 0,386.
- Hasil interpretasi topik menghasilkan lima tema utama keluhan masyarakat, yaitu: (1) akuntabilitas dan tata kelola anggaran, (2) dampak bagi kesehatan anak di sekolah, (3) kualitas dan kelayakan menu makanan, (4) kendala distribusi dan pemenuhan gizi, serta (5) insiden keracunan dan keamanan pangan. Tema insiden keracunan dan keamanan pangan menjadi topik dengan proporsi tertinggi sebesar 38,7% sehingga menunjukkan bahwa isu keamanan pangan merupakan perhatian utama masyarakat terhadap pelaksanaan program MBG.
- Hasil penelitian menunjukkan bahwa analisis topik menggunakan model LDA (*Latent Dirichlet Allocation*) dapat mengidentifikasi permasalahan utama dalam pelaksanaan program MBG

berdasarkan opini masyarakat di media sosial. Temuan penelitian ini diharapkan dapat menjadi masukan dalam meningkatkan transparansi anggaran, pemantauan kesehatan anak secara berkala, penyusunan standar menu berbasis gizi, perbaikan sistem distribusi makanan, serta penerapan standar keamanan pangan pada pelaksanaan program.

- d. Penelitian selanjutnya disarankan untuk memperluas sumber data dari berbagai platform media sosial serta mengeksplorasi metode pemodelan topik lain seperti BERTopic, NMF, atau LSA sebagai pembanding model LDA dengan variasi jumlah topik yang lebih luas untuk menghasilkan analisis yang lebih stabil dan optimal.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada pihak-pihak terkait yang telah memberi dukungan terhadap penelitian ini.

DAFTAR PUSTAKA

- [1] M. A. Rojabi, *Mitra BGN: Syarat, Cara Daftar Online, dan Kunci Lolos Verifikasi Program Makan Bergizi Gratis*. 2026-03-01, 2025. [Online]. Available: <https://www.bgn.go.id/news/berita/bgn-198-sppg-sudah-kantongi-sertifikat-laik-higiene-sanitasi>
- [2] N. Azzahra, A. D. Dharmawan, A. F. Mardatilah, M. I. Habibi, Virliana Amanda Putri, and Suhardi, "Pelaksanaan Uji Coba Program Makan Bergizi Gratis di SMP Negeri 4 Tangerang," *J. Pengabd. Masy. dan Ris. Pendidik.*, vol. 3, no. 4, pp. 5036–5044, 2025, doi: 10.31004/jerkin.v3i4.1335.
- [3] M. M. Laia, F. N. Hasan, and A. Y. Kuntoro, "Analisis Sentimen Program Makan Gratis Pada Platform X Menggunakan Algoritma Naïve Bayes," *J. Ilm. Inform.*, vol. 13, no. 2, pp. 259–264, 2025.
- [4] Y. A. Krisdamarjati, "Sentimen Negatif Warganet Mencapai 66 Persen, MBG Mendesak Dievaluasi," *kompas.id*.
- [5] M. Nango, B. Purwoko, M. Hazim, and Karwanto, "Makan Bergizi Gratis: Strategi Kebijakan Publik Menuju Generasi Emas Indonesia 2045," *J. Pendidik. Agama Katolik*, vol. 25, no. 2, pp. 354–368, 2025.
- [6] R. N. Muhammad, L. Wulandari S, and B. Tanggahma, "Pengaruh Media Sosial Pada Persepsi Publik Terhadap Sistem Peradilan: Analisis Sentimen di Twitter," *UNES Law Rev.*, vol. 7, no. 1, pp. 507–516, 2024, [Online]. Available: <https://review-unes.com/index.php/law/article/view/2327>
- [7] Nouvan, "Negara dengan Pengguna X (Twitter) Terbanyak Juli 2025, Indonesia Urutan Keempat," *dataloka*. [Online]. Available: <https://dataloka.id/humaniora/4627/negara-dengan-pengguna-x-twitter-terbanyak-juli-2025-indonesia-urutan-keempat/>
- [8] A. Amrullah, "Analisis Sentimen Publik terkait Migrasi Tenaga Kerja Indonesia di Platform X menggunakan SVM-IndoBERT," *Unisda J. Math. Comput. Sci.*, vol. 11, no. 1, pp. 53–63, 2025.
- [9] A. F. I. Ridat and A. R. Isnain, "Pemodelan Topik dengan LDA untuk Memahami Persepsi Publik terhadap Apple Vision Pro melalui Analisis Berita Twitter," *JUPI (Jurnal Ilm. Penelit. dan Pembelajaran Inform.)*, vol. 10, no. 4, pp. 3233–3246, 2025.
- [10] A. R. Sembiring, "Sentiment Analysis dan Topic Modeling pada Tweet Bahasa Indonesia Mengenai Pemilihan Umum 2024," 2025.
- [11] A. Muhaimin *et al.*, "Social Media Analysis and Topic Modeling: Case Study of Stunting in Indonesia," *Telemat. J. Inform. dan Teknol. Inf.*, vol. 20, no. 3, pp. 406–415, 2023, doi: 10.31515/telematika.v20i3.10797.
- [12] R. F. Nulhakim and D. F. Shiddieq, "Analisis Sentimen Publik X Terhadap Kenaikkan Ppn 12 % pada Twitter Menggunakan Latent Dirichlet Allocation," *J. Pendidik. Indones.*, vol. 6, no. 8, pp. 4090–4103, 2025.
- [13] A. Sitanggang, Y. Umaidah, and R. I. Adam, "Analisis Sentimen Masyarakat terhadap Program Makan Siang Gratis pada Media Sosial X Menggunakan Algoritma Naïve Bayes," *JITET (Jurnal Inform. dan Tek. Elektro Ter.)*, vol. 12, no.

