

Algoritma K Means dan TF-IDF untuk Pengelompokan Opini Publik terhadap Program Makan Bergizi Gratis pada Komentar TikTok

Devi Sugianti¹, Ari Putra Wibowo², Anas Syaifudin³, Rizqi Wijonarko⁴

¹Teknik Informatika, fakultas Teknologi Informasi, Institut Widya Pratama

¹devi.sugianti027@iwima.ac.id, ²ariputra@iwima.ac.id *, ³anas.sya@iwima.ac.id , ⁴rizqiwijonarko@iwima.ac.id

Abstract

In Indonesia, 14% of children suffer from stunting due to malnutrition. To address this issue, the government launched the Free Nutritious Meal (MBG) Program, which has generated diverse public opinions on social media, particularly TikTok. This study aims to cluster public opinions regarding the MBG program using the K-Means Clustering algorithm combined with Term Frequency-Inverse Document Frequency (TF-IDF) without requiring manual labeling. A total of 57,362 comments were collected, of which 53,204 valid comments remained after preprocessing. Truncated Singular Value Decomposition (SVD) was applied for dimensionality reduction, while the optimal number of clusters ($K = 5$) was determined using the Elbow Method and Silhouette Score. The clustering results identified five main discussion themes: general program discussion (28.4%), child nutrition and school access (7.9%), spontaneous reactive responses (45.2%), criticism and rejection of the program (15.0%), and support for public figures (3.6%). The model achieved a Silhouette Score of 0.0385 and a Davies-Bouldin Index of 4.9507, reflecting the characteristics of short and informal social media text. The findings demonstrate that the unsupervised clustering approach effectively maps public opinion into meaningful thematic groups and provides valuable insights for the National Nutrition Agency to improve menu quality, budget transparency, and distribution standards of the MBG program.

Keywords: K-Means Clustering, TF-IDF, public opinion, MBG program, TikTok, unsupervised learning

Abstrak

Di Indonesia, 14% anak mengalami stunting akibat kekurangan gizi. Untuk mengatasi permasalahan tersebut, pemerintah meluncurkan Program Makan Bergizi Gratis (MBG) yang memunculkan beragam opini publik di media sosial, khususnya TikTok. Penelitian ini bertujuan mengelompokkan opini publik terkait program MBG menggunakan algoritma K-Means Clustering yang dikombinasikan dengan Term Frequency-Inverse Document Frequency (TF-IDF) tanpa memerlukan pelabelan manual. Sebanyak 57.362 komentar dikumpulkan, dengan 53.204 komentar valid setelah melalui tahap prapemrosesan. Truncated Singular Value Decomposition (SVD) diterapkan untuk reduksi dimensi, sedangkan jumlah kluster optimal ($K = 5$) ditentukan menggunakan Elbow Method dan Silhouette Score. Hasil clustering mengidentifikasi lima tema utama, yaitu diskusi umum program (28,4%), gizi anak dan akses sekolah (7,9%), respons reaktif spontan (45,2%), kritik dan penolakan terhadap program (15,0%), serta dukungan terhadap tokoh publik (3,6%). Model menghasilkan nilai Silhouette Score sebesar 0,0385 dan Davies-Bouldin Index sebesar 4,9507, yang mencerminkan karakteristik data teks media sosial yang pendek dan informal. Temuan penelitian menunjukkan bahwa pendekatan *unsupervised clustering* mampu memetakan opini publik ke dalam kelompok tematik yang bermakna serta memberikan masukan bagi Badan Gizi Nasional dalam meningkatkan kualitas menu, transparansi anggaran, dan standar distribusi program MBG.

Kata kunci: K-Means Clustering, TF-IDF, opini publik, program MBG, TikTok, unsupervised learning

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

1. Pendahuluan

Di Indonesia pada tahun 2024 mencatat angka stunting sebesar 14 %, dari akibat kekurangan gizi yang dialami oleh anak-anak dari usia dini di berbagai daerah [1]. Pemerintah menetapkan salah satu kebijakan yang bertujuan untuk meningkatkan kualitas gizi dan untuk mengurangi beban ekonomi dari rumah tangga miskin adalah dengan cara program makan bergizi gratis (MBG) [2]. MBG adalah salah satu program unggulan yang diinisiatif Presiden Prabowo Subianto, dalam meningkatkan sumber daya manusia dengan menyasar anak-anak serta ibu hamil [3]. Program MBG untuk meningkatkan kesiapan peningkatan belajar serta pemenuhan gizi siswa [4].

Dengan adanya program tersebut dapat mengubah kebiasaan makan sehat dan mengurangi angka

stunting[5]. MBG dapat mendorong dan membangun ekonomi lokal dengan melibatkan UMKM, akan tetapi program MBG menghadapi permasalahan seperti ketidaksesuaian menu dengan standar gizi, kurangnya variasi menu, infrastruktur yang terbatas, serta makanan yang tidak layak konsumsi [3], [6]. Program MBG memiliki tujuan mulia, akan tetapi pelaksanaannya tidak lepas dari kontroversi dan beragam respons masyarakat. Adanya opini masyarakat yang mendukung dan menolak keberlangsungan dari program MBG [7]. Pelaksanaan MBG memunculkan berbagai respon dan pandangan di ruang publik khususnya di media sosial [8].

Ruang digital, terutama media sosial kini menjadi telah menjadi arena utama bagi masyarakat untuk menuangkan prespektif, terhadap kebijakan publik

pemerintah. Salah satu media sosial yang terkenal adalah TikTok [9]. Di Indonesia pengguna tiktok sekitar 99,1 juta pengguna dengan di dominasi Gen Z atau kalangan remaja media sosial TikTok paling sering digunakan untuk berkomunikasi [10]-[11] TikTok telah mengalami pergeseran fungsi yang signifikan dan bertransformasi menjadi ruang diskusi publik dan politik untuk mengeskpresikan pandangan terhadap isu dan kebijakan dari pemerintah [12]. Konten-konten pada platform TikTok tentang MBG menghasilkan komentar yang tinggi, sehingga ada tantangan dan peluang untuk dapat diteliti [13].

Untuk melakukan analisis opini publik dengan berskala besar dari media sosial dapat menggunakan pendekatan *Natural Language Processing* yang mampu memproses dan mengekstrak pola dari data teks tidak terstruktur secara otomatis [14]. Pada penelitian [15]-[17]. NLP digunakan untuk menganalisis opini publik dengan menggunakan IndoBERT, *Gated Recurrent Unit* (GRU) dan *Naïve bayes* tentang isu MBG di platform Twitter/X dengan menggunakan pendekatan *supervised learning* dengan melabelkan secara manual dan 3 label (positif, negatif, netral). Pada penelitian tersebut belum dapat menggali tema-tema diskusi dan tidak memanfaatkan TikTok sebagai sumber data.

Pada penelitian sebelumnya yang dilakukan oleh [18] untuk mengukur kebijakan dari Bapak Presiden dengan menggunakan media sosial X, pada penelitian [19] sumber data yang digunakan dari Twitter (X) dengan jumlah 10.524 dengan metode yang diusulkan *lexicon based* serta *SVM* dengan pendekatan *supervised learning*, akurasi sebesar 89,26 %. Kesenjangan ini mendorong untuk melakukan penelitian ini akan menggali opini publik dengan menggunakan media sosial TikTok serta data yang digunakan sebanyak 57.326 data, pendekatan yang digunakan *unsupervised* yang lebih eksploratif yaitu K-Means.

K-Means menjadi metode pengelompokan yang dipilih karena mampu mempartisi data dengan skala besar, serta algoritma *clustering* yang paling sederhana dan paling umum dengan tipe data yang bersifat numerik [20]. Ditahap awal ditetapkan jumlah K untuk melakukan pengelompokan intertieve dengan melakukan partisi set data[21]. Pada penelitian yang dilakukan oleh [22] bahwa K- Menas dapat digunakan untuk mengelompokkan buku pada perpustakaan yang berbasis teks dengan melalui tahap *preprocessing text mining*. K-Means dapat mengelompokkan dokumen dan prediksi topik di media social.[23]-[24]. Dalam pengelompokkan ke topik yang bermakna dalam algoritma K-Means untuk merepresentasikan fitur nya menggunakan TF-IDF agar menghasilkan vektor bobot dari setiap kata yang penting [25]. Berdasarkan uraian diatas, maka tujuan penelitian ini menerapkan algoritma K-Means clustering dengan TF-IDF untuk melakukan pengelompokkan opini publik pada platform TikTok tentang komentar MBG, agar dapat

menginterpretasikan tema-tema utama yang sedang didiskusikan publik tentang kebijakan pemerintah.

2. Metode Penelitian

Penelitian ini menggunakan *text mining* dan *unsupervised learning*. Data yang digunakan berupa tekstual tidak terstruktur yang dikumpulkan dari media sosial TikTok, kemudian dianalisis untuk menemukan pola pengelompokkan opini public terhadap kebijakan pemerintah tentang program MBG, tanpa menggunakan label [26]. Dengan metode TF-IDF merupakan salah satu teknik untuk pengolahan teks, dengan melakukan evaluasi tingkat kepentingan suatu kata dalam dokumen.

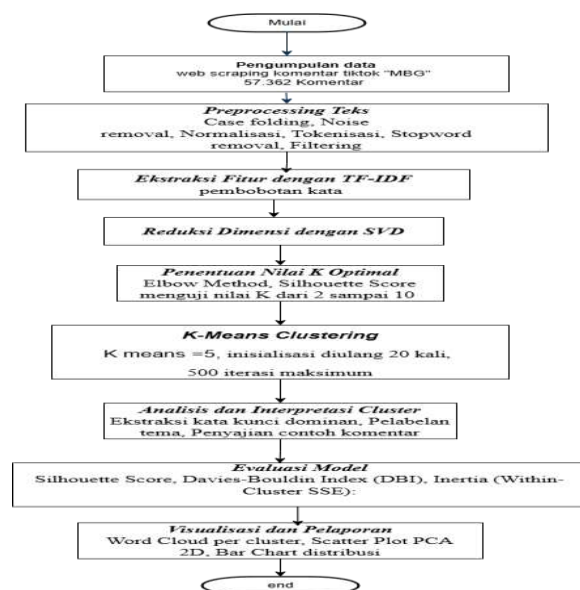
Metode TF-IDF menghitung dua faktor penting, yaitu: *Term Frequency* (TF) digunakan mengukur tingkat keseringan kemunculan kata dalam sebuah dokumen dan *Inverse Documen Frequency* (IDF) digunakan untuk mengukur tingkat seberapa penting suatu kata didalam sebuah dokumen. Jika IDF tinggi maka kata-kata yang muncul jarang dikoleksi dalam sebuah dokumen.

2.1. Pengumpulan Data

Data diambil dari data public *Kaggle* yang dikumpulkan dari komentar pengguna TikTok dengan *web scraping* yang berhasil dikumpulkan sejumlah 57.362 komentar, Adapun data yang valid 53.204 komentar setelah dilakukan proses *preprocessing*. Pengumpulan data dilakukan dari Januari sampai Oktober 2025.

2.2. Tahapan Penelitian

Tahapan penelitian untuk melakukan pengelompokkan opini public tentang MBG dengan menggunakan K-Means, dengan (TF-IDF) untuk mengolah teks pada pemodelan bahasa alami, adapun alur penelitian akan digambarkan pada Gambar 1.



Gambar 1. Alur Penelitian *Clustering K-Means*

2.3. Preprocessing Teks

Tahap *preprocessing* adalah tahap yang paling kritis atau sebagai fondasi dalam penelitian NLP, karena kualitas data yang dihasilkan akan menentukan kualitas *clustering*[27]. Pada tahap ini dilakukan proses *preprocessing* untuk membersihkan dan menyiapkan data agar siap dianalisis. Tahap pertama adalah *case folding*, yaitu mengonversi seluruh teks komentar menjadi huruf kecil untuk menyeragamkan representasi kata. Selanjutnya dilakukan *noise removal*, yaitu menghapus *mention*, URL/tautan, karakter ASCII, emoji, angka, dan tanda baca yang tidak memiliki nilai semantik. Tahap berikutnya adalah normalisasi, yaitu menyeragamkan kata-kata tidak baku dan singkatan dalam bahasa Indonesia informal menjadi bentuk bakunya. Setelah itu dilakukan tokenisasi, yaitu memecah teks menjadi token atau kata-kata individual sebagai unit analisis dasar. Tahap terakhir adalah *filtering*, yaitu menghapus komentar yang menjadi kosong setelah proses *preprocessing*, sehingga diperoleh 53.204 komentar valid yang digunakan pada tahap analisis selanjutnya.

2.4. Ekstraksi Fitur dengan TF-IDF

Term Frequency-Inverse Document Frequency (TF-IDF) merupakan metode pembobotan kata yang digunakan untuk mengukur tingkat kepentingan relatif suatu kata dalam sebuah dokumen terhadap keseluruhan koleksi dokumen [28]–[29]. Metode ini memberikan bobot yang lebih tinggi pada kata yang sering muncul dalam suatu dokumen, tetapi jarang muncul pada dokumen lainnya. Tahap pertama dalam perhitungan TF-IDF adalah menghitung nilai *Term Frequency* (TF), yaitu frekuensi kemunculan suatu istilah dalam sebuah dokumen.

$$TF(t, d) = \frac{f(t, d)}{\sum f(d)} \quad (1)$$

Keterangan

$f(t, d)$: merupakan jumlah kemunculan dari istilah t dalam dokumen d .

Keunikan kata dari semua teks yang akan ditunjukkan oleh frekuensi dokumen terbalik (IDF).

$$IDF(t) = \log\left(\frac{n}{df(t)}\right) \quad (2)$$

Keterangan:

N = total jumlah dokumen

$df(t)$ = jumlah dokumen yang mengandung term t .

Bobot TF_IDF

$$TF - IDF(t, d) = TF(t, d) \times IDF(t) \quad (3)$$

2.5. Reduksi Dimensi dengan SVD

Untuk mereduksi dimensi matriks dapat diterapkan *Truncated Singular Value Decomposition* (SVD), agar komputasi K-Means langsung pada matriks tersebut tidak efisien dan rentan terhadap *curse of dimensionality*.

2.6. Penentuan Nilai K optimal

Pada penelitian ini menggunakan dua metode untuk menentukan K optimal: *Elbow Method*, dalam menentukan jumlah *cluster* terbaik dengan melihat presentase dari hasil perbandingan antara jumlah *cluster* yang membentuk siku pada grafik [30], dan *Silhouette Score*, *Silhouette* di gunakan untuk melihat kuliatas serta kekuatan dari *cluster*[31].

2.7. K-Means Clustering

K-Means clustering merupakan salah satu algoritma *unsupervised learning* yang digunakan untuk melakukan pengelompokkan data, dengan meminimalkan jumlah jarak kuadrat antara setiap titik data dengan centroid clusternya [32]. Adapun langkah-langkah K-Means adalah: Menentukan jumlah *cluster* (k), Inisialisasi titik pusat (centroid) awal, Menghitung jarak setiap data ke *cluster* terdekat, Pembaruan posisi centroid dan Ulangi iterasi ke 3 hingga kondisi konvergen tercapai, atau ketika tidak ada satu pun data yang berpindah *cluster*, atau pergeseran centroid sudah sangat kecil di bawah ambang batas [33].

2.8. Analisis dan Interpretasi Cluster

Pada tahap ini dilakukan Analisis yang mendalam terhadap isi di tiap *cluster* untuk menghasilkan interpretasi tematik yang bermakna.

2.9. Evaluasi Model

Dalam penelitian ini evaluasi model dengan menggunakan *Silhouette Score*, dan *Inertia (Within-Cluster SSE)*. *Silhouette Score* adalah untuk mengevaluasi dan mengukur tingkat mirip suatu data dengan clusternya sendiri, nilai dari *silhouette Score* dari range -1 sampai dengan 1. Jika nilai mendekati nilai 1 berarti menunjukkan *cluster* yang baik [34]. *Davies-Bouldin Index (DBI)* merupakan sebuah nilai yang digunakan untuk acuan dalam menentukan kualitas dari *clustering*, jika nilai dari DBI minimal maka *cluster* dianggap sudah optimal[35]. Sedangkan *Inertia(Within-Cluster SSE)* merupakan jumlah kuadrat jarak seluruh titik data ke centroid *cluster* nya, jika nilainya lebih kecil maka menunjukkan *cluster* lebih kompak.

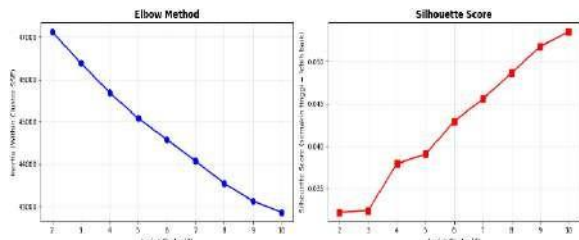
3. Hasil dan Pembahasan

3.1. Data Set

Data yang dikumpulkan sebanyak 57.326 komentar, setelah dilakukan proses: *Case folding*, *Noise removal*, *Normalisasi*, *Tokenisasi*, *Stopword removal* data yang valid menjadi 53.204 komentar, atau sekitar 92,7 %. Dari input data yang valid sebanyak 53.204 komentar, dengan *parameter max_fatures= 5.000 min_df=3* serta *max_df=0.85* menghasilkan matriks TF-IDF= 53.187 x 5.000. Pada fase reduksi dimensi diterapkan SVD *Truncated Singular Value Decomposition* yang mereduksi dimensi matriks menjadi 100 komponen utama.

3.2. Penentuan jumlah cluster

Untuk dapat mencapai cluster yang optimal maka melakukan pengujian terhadap K dari 2 sampai dengan 10 dengan menggunakan *Elbow* dan *Method Silhouette Score*. Hasil pengujian disajikan pada Gambar 2.



Gambar 2. Penentuan jumlah cluster dengan *Elbow* dan *Method Silhouette Score*

Penentuan jumlah cluster dengan metode *Elbow* dan *Silhouette* juga dapat dilihat dari Tabel 1.

Tabel 1. Pengujian Nilai K dengan *Elbow Method* dan *Silhouette Score*

Nilai K	Inertia SSE	Shilhouette Score
2	47,125	0.0322
3	46,384	0.0324
4	45,690	0.0379
5	45.078	0.0391
6	44,578	0,0429
7	44,070	0,0456
8	43,540	0,0486
9	43,124	0,0518
10	42,855	0,0535

Dari Tabel 1 dipilih nilai K=5 untuk jumlah cluster yang optimal. Sedangkan Tabel 2 merupakan distribusi komentar per cluster.

Tabel 2. Distribusi Komentar Percluster

Cluster	Jumlah komentar	Presentase
0	15.096 komentar	28.4 %
1	4.180 komentar	7,9 %
2	24.027 komentar	45,2%
3	7.960 komentar	15,0 %
4	1.941 komentar	3,6 %

3.3. Cluster dengan K-Means

Penerapan algoritma K-Means clustering dengan K=5, data komentar sebanyak 53.204 komentar dari TikTok untuk distribusi komentar per cluster dapat dilihat dari Tabel 3.

Tabel 3. Top Kata Tiap Cluster

Cluster	Jumlah	Presentase	Kata Kunci dominan
C0	15.096	28.4 %	program, pak, banyak, mbg, buat, orang, dapur, korupsi, salah, bagus, gizi, bukan, lebih, ahli, semua
C1	4.180	7.9 %	anak, mbg, makan, sekolah, makanan, kasih, anak anak, menu, buat, burger, sama, hari, mbg anak, suka, sekolah anak
C2	24.027	45.2%	bener, sticker, burger, makan, sama, buk,

C3	7.960	15.0%	sekolah, betul, tutup, gratis, apa, setuju, ayam, tan, god
C4	1.941	3.6%	mbg, stop, stop mbg, sekolah, belum, menu, banyak, tutup, program, dapur, menu mbg, makan, hari, buat, korupsi
			ibu, mbg, keren, setuju, suka, sama, ngomong, keren ibu, mewakili, bener, sama ibu, sehat, ibu ngomong, apa, anak

Berikut contoh komentar di tiap cluster dapat dilihat pada Gambar 3.

CONTOH KOMENTAR ASLI PER CLUSTER

--- Cluster 0 ---
 [1] Jagan repeat trus... coba DM yg ud suka pake jeans itu lgsg dicera fakesi trussas ekeksi. Jgn rilez trus lgsg kerja
 [2] sda ltra mada hulu caca untuk program tnd
 [3] bilang aja no nonton anime
 [4] oiaaaa yaa belat itu diambil sendiri @ di sini juga gitu kak investor nya org lala mnd kades nya mada begitn what
 [5] lra ilimadililik sekele mendingan reney juga gile di mabain rana apun belian ya lalulur ga mabersitkarna orang lula

--- Cluster 1 ---
 [1] anak saya sda saya larang mtuk makan H&G.
 [2] ga setuju sama si ibu, jngpood gah sehat buat anak
 [3] anak saya sda saya larang mtuk makan H&G.
 [4] memang menu H&G awal nya jga aladadanya mabaw pernah gah mda nalinge
 [5] kssir juga anak klo ga bebel makan lagi, berrap dori makan gratis dapi mabawnya sda yg badi, dan terpalita mabaw kan

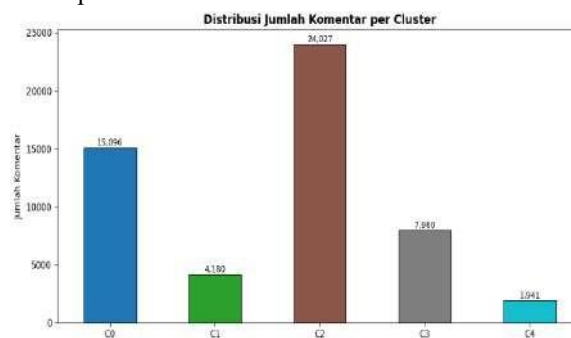
--- Cluster 2 ---
 [1] genti raket eseng paling beier
 [2] tina yg mli giri senaga saya tinda list anda mada ase gacaa, target @ mda nya mada ya... yg kaha tahu mabaw
 [3] cembay, mengobasi,, RALASI,,
 [4] Oh my goodness
 [5] kalo mada collab sama superind gen si

--- Cluster 3 ---
 [1] banyak juga plosok plosok di tnda yang belum pernah dapat mg
 [2] bantakan H&G
 [3] om mda yg di lombo mendingan rupanya dari pada di jma bant
 [4] H&G MADAAN BERAKAWI GRATIS
 [5] aku cwe aja loh mada sapa 4 cepeng sangking ananya, kha nyak mada' tni ga bersukur bngt jrrrrr, ah mada, padaha

--- Cluster 4 ---
 [1] mantap itu ini @
 [2] caw @kali: kaku-dhat vaa
 [3] bu, tenng aja, sy ga mda kaku, justry lagi herna kempa itu ngelasy sampai mda-muter bwa personal smpai. padahal
 [4] marta mng mngalah bli ayas
 dan ga mndong kya itu
 [5] bener sanggettt sanggettt sanggettt asli doiter gati itu int the best

Gambar 3. Contoh Komentar Asli percluster

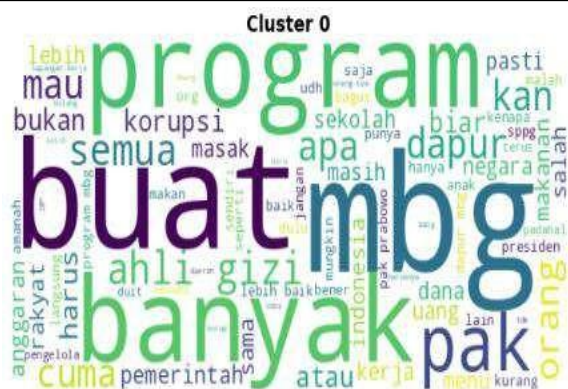
Berikut ini merupakan distribusi di tiap cluster dapat dilihat pada Gambar 4.



Gambar 4. Distribusi Cluster

3.4. Hasil dan Evaluasi Model K Means

Hasil analisis dengan menggunakan word cloud per cluster sebagai visualisasi kata-kata yang dibuat secara terpisah untuk disetiap cluster. Tujuannya untuk dapat mengetahui kata-kata yang paling dominan atau representatif pada masing-masing cluster.



Gambar 5. Wordcloud Cluster 0

Gambar 5 menunjukkan hasil *wordcloud* untuk cluster 0 dengan kata kunci adalah program, pak, banyak, mbg, buat, orang, dapur, korupsi, salah, bagus, gizi, bukan, lebih, ahli, semua.



Gambar 6. Wordcloud Cluster 1

Untuk cluster 1 dapat dilihat pada Gambar 6 dengan kata kunci anak, mbg, makan, sekolah, makanan, kasih, anak anak, menu, buat, burger, sama, hari, mbg anak, suka, sekolah anak



Gambar 7. Wordcloud Cluster 2

Untuk cluster 2 dapat dilihat pada Gambar 7 dengan kata kunci bener, sticker, burger, makan, sama, buk, sekolah, betul, tutup, gratis, apa, setuju, ayam, tan, god.



Gambar 8. Wordcloud Cluster 3

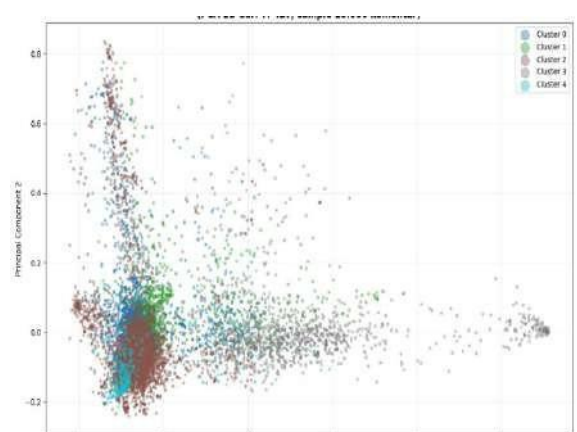
Pada Gambar 8 menunjukkan hasil *wordcloud* untuk cluster 3 dengan kata kunci adalah: mbg, stop, stop mbg, sekolah, belum, menu, banyak, tutup, program, dapur, menu mbg, makan, hari, buat, korupsi.



Gambar 9. Wordcloud Cluster 4

Hasil *wordcloud* untuk cluster 4 ditunjukkan pada Gambar 9 di atas, dengan kata kunci adalah: ibu, mbg, keren, setuju, suka, sama, ngomong, keren ibu, mewakili, bener, sama ibu, sehat, ibu ngomong, apa, anak.

Evaluasi model K-Means dengan menggunakan Silhouette dan Davies-Bouldin.



Gambar 10. Visualisasi Cluster

Silhouette Score digunakan untuk mengukur kemiripan objek dengan cluster nilai berikisar antara -1 sampai dengan +1[34]. Adapun Score mencapai =0,0385. Davies-Bouldin digunakan untuk mengukur

rasio antara dispersi intra cluster dan jarak antar cluster, untuk Davies-Bouldin sebesar = 4,9507.

4. Kesimpulan

Pada penelitian menggunakan data berupa komentar dari social media tiktok tentang kebijakan pemerintah terhadap program MBG, pengelompokkan opini publik ini menerapkan algoritma K means dengan berbasis TF_IDF. Data yang terkumpul sejumlah 57.362 komentar, setelah melakukan proses preprocessing data yang valid menjadi 53.204 komentar. Evaluasi kuantitatif menghasilkan Silhouette Score sebesar 0,0385 dan Davies-Bouldin Index sebesar 4,9507. Nilai tersebut tergolong rendah secara absolut, namun merupakan kondisi yang lazim pada data teks informal dan pendek seperti komentar media sosial, dan telah dikonfirmasi melalui evaluasi kualitatif yang menunjukkan koherensi tematik yang jelas pada setiap cluster. Penelitian ini membuktikan bahwa pendekatan unsupervised clustering menghasilkan peta tematik opini publik yang lebih informatif dan granular dibandingkan pendekatan klasifikasi sentimen konvensional yang hanya menghasilkan tiga kelas predefinisi.

Secara praktis, dominasi cluster kritik dan penolakan (15,0%) serta cluster isu kualitas menu mengindikasikan bahwa transparansi anggaran dan peningkatan standar kualitas distribusi makanan perlu mendapat perhatian prioritas dari Badan Gizi Nasional. Untuk penelitian selanjutnya, disarankan penggabungan pendekatan K-Means dengan pemodelan topik berbasis Latent Dirichlet Allocation (LDA) atau representasi fitur berbasis word embedding seperti IndoBERT, serta sumber data ke berbagai platform dan rentang waktu yang lebih panjang guna menangkap dinamika perubahan opini publik secara temporal.

Daftar Rujukan

- [1] A. Kiftiyah, F. Ayu Palestina, F. Ulul Abshar, K. Rofiah, and M. Bergizi Gratis, "Program Makan Bergizi Gratis (MBG) dalam Perspektif Keadilan Sosial dan Dinamika Sosial – Politik," *Pancasila: Jurnal Keindonesiaan*, vol. 5, no. 1, pp. 101–112, Apr. 2025, doi: 10.52738/PJK.V5I1.726.
- [2] J. I. Psikologi *et al.*, "Implementasi Kebijakan Program Makan Bergizi Gratis Dalam Meningkatkan Kesejahteraan Masyarakat Miskin," *Jurnal Ilmu Psikologi dan Kesehatan | E-ISSN : 3063-1467*, vol. 3, no. 1, pp. 778–781, May 2026, doi: 10.62379/JIPK.V3I1.1910.
- [3] A. Azzahra, S. Marmoah, I. / Lembaga Penulis, and P. S. Subang, "ANALISIS PERMASALAHAN MAKAN BERGIZI GRATIS PADA SEKOLAH DASAR DI INDONESIA," *Didaktik : Jurnal Ilmiah PGSD STKIP Subang*, vol. 11, no. 04, pp. 242–255, 2026, doi: 10.36989/DIDAKTIK.V11I04.8960.
- [4] A. A. Saputri and D. M. Utari, "EFEKTIVITAS PROGRAM MAKAN BERGIZI GRATIS (MBG) TERHADAP STATUS GIZI, KESIAPAN BELAJAR, DAN LUARAN PENDIDIKAN PESERTA DIDIK: TINJAUAN SISTEMATIS," *JKM (Jurnal Kesehatan Masyarakat) Cendekia Utama*, vol. 14, no. 1, pp. 35–48, May 2026, Accessed: Jul. 17, 2026. [Online]. Available: <https://jurnal.stikeskendekiautamakudus.ac.id/index.php/JKM/article/view/3045>
- [5] A. Sophian, "Integrasi MBG dengan Program Edukasi Gizi untuk Perubahan Perilaku Sehat", doi: 10.5281/ZENODO.17140209.
- [6] J. Manajemen *et al.*, "Implementasi Kebijakan Program Makan Bergizi Gratis sebagai Pendorong dan Pembangunan Ekonomi Lokal," *Jurnal Manajemen, Bisnis dan Kewirausahaan*, vol. 6, no. 1, pp. 95–112, Jan. 2026, doi: 10.55606/JUMBIKU.V6I1.6403.
- [7] K. Pemerintah *et al.*, "Komunikasi Pemerintah Dan Fragmentasi Opini Publik Di Media Sosial: Kasus Program Makan Bergizi Gratis," *Khazanah Multidisiplin*, vol. 7, no. 1, pp. 14–34, Mar. 2026, doi: 10.15575/KL.V7I1.54253.
- [8] K. H. Pratiawan and D. Yuniarto, "Analisis Sentimen Publik terhadap Program Makan Bergizi Gratis dengan Algoritma Naive Bayes," *RIGGS: Journal of Artificial Intelligence and Digital Business*, vol. 4, no. 4, pp. 5412–5419, Dec. 2025, doi: 10.31004/riggs.v4i4.3652.
- [9] S. V. Mahardhika, I. Nurjannah, I. I. Ma'una, and Z. Islamiyah, "Faktor-Faktor Penyebab Tingginya Minat Generasi Post-Millennial Di Indonesia Terhadap Penggunaan Aplikasi Tik-Tok," *SOSEARCH : Social Science Educational Research*, vol. 2, no. 1, pp. 40–53, Dec. 2021, doi: 10.26740/SOSEARCH.V2N1.P40-53.
- [10] A. Maretha, A. Khairin Pebrianti Nur, R. Dwi Harry Saputra, S. Tinggi Ilmu Ekonomi Balikpapan, and K. Timur, "Analisis Penggunaan Tiktok Terhadap Perubahan Perilaku Komunikasi Generasi Z," *Indo-MathEdu Intellectuals Journal*, vol. 6, no. 8, pp. 12053–12064, Dec. 2025, doi: 10.54373/IMEIJ.V6I8.4673.
- [11] N. Fadillah and M. Syukur, "Pemanfaatan TikTok dalam Meningkatkan Literasi Sosial pada Remaja di Desa Jauh Pandang Kecamatan Pitumpanua Kabupaten Wajo," *Pinisi Journal of Sociology Education Review*, vol. 5, no. 3, pp. 389–396, Nov. 2025, Accessed: Aug. 11, 2026. [Online]. Available: <http://ojs.unm.ac.id/jser/article/view/75159>
- [12] R. C. Rizanul, D. Eginis, S. Sofiana, A. Jasmine, and D. Juniarto, "Analisis Peran Media Sosial (TikTok) dalam Dinamika Partisipasi Politik pada Pemilu Presiden 2024," *Journal of Social Contemplativa*, vol. 3, no. 1, pp. 51–61, Feb. 2025, doi: 10.61183/JSC.V3I1.91.
- [13] K. Pemerintah *et al.*, "Komunikasi Pemerintah Dan Fragmentasi Opini Publik Di Media Sosial: Kasus Program Makan Bergizi Gratis," *Khazanah Multidisiplin*, vol. 7, no. 1, pp. 14–34, Mar. 2026, doi: 10.15575/KL.V7I1.54253.
- [14] M. Fazri and A. Voutama, "Analisis Sentimen Publik Terhadap Danantara Di Media Sosial X Menggunakan Nlp Dan Pembelajaran Mesin," *JOISIE (Journal Of Information Systems And Informatics Engineering)*, vol. 9, no. 1, pp. 197–206, Jul. 2025, doi: 10.35145/JOISIE.V9I1.4924.
- [15] F. Haerani, "Analisis Sentimen Program Makan Bergizi Gratis Berdasarkan Komentar Di Media Sosial X Menggunakan Gru," *Jurnal Informatika dan Teknik Elektro Terapan*, vol. 14, no. 1, pp. 2830–7062, Jan. 2026, doi: 10.23960/JITET.V14I1.8754.
- [16] M. Haikal and A. Erfina, "Analisis Sentimen Warganet terhadap Program Makan Bergizi Gratis (MBG) Menggunakan Model IndoBERT," *Jurnal Locus Penelitian dan Pengabdian*, vol. 5, no. 3, pp. 1626–1636, Mar. 2026, doi: 10.58344/LOCUS.V5I3.5253.
- [17] M. Bergizi Gratis dengan Algoritma Naive Bayes, K. Hapidz Pratiawan, and D. Yuniarto, "Analisis Sentimen Publik terhadap Program Makan Bergizi Gratis dengan Algoritma Naive Bayes," *RIGGS: Journal of Artificial Intelligence and Digital Business*, vol. 4, no. 4, pp. 5412–5419, Dec. 2025, doi: 10.31004/riggs.v4i4.3652.
- [18] I. Najiyah and M. Rizal, "Analisis Sentimen Media Sosial X Terhadap Kebijakan Presiden Republik Indonesia Prabowo Subianto," *JURNAL FASILKOM*, vol. 15, no. 3, pp. 443–455, Dec. 2025, doi: 10.37859/JF.V15I3.10385.
- [19] Z. Akbar, Z. Akbar, I. Riadi, and R. Umar, "Analisis Sentimen Program Makan Bergizi Gratis Menggunakan Lexicon-Based dan Support Vector Machine," *JURNAL FASILKOM*, vol. 16, no. 1, pp. 150–156, Apr. 2026, doi: 10.37859/JF.V16I1.10948.
- [20] S. Mutofin, T. Wicaksono, and A. Murtadho, "Perbandingan Kinerja Algoritma Kmeans dengan Kmeans Median pada

- Deteksi Kanker Payudara,” *Jurnal Informasi dan Teknologi*, pp. 88–91, Feb. 2023, doi: 10.37034/JIDT.V5I1.274.
- [21] M. Norshahlan, H. Jaya, and R. Kustini, “Penerapan Metode Clustering Dengan Algoritma K-means Pada Pengelompokan Data Calon Siswa Baru,” *Jurnal Sistem Informasi Triguna Dharma (JURSI TGD)*, vol. 2, no. 6, pp. 1042–1053, Nov. 2023, doi: 10.53513/JURSI.V2I6.9148.
- [22] A. A. Riyanto, D. Daryanto, and G. Abdurrahman, “Text Mining Untuk Clustering Buku Di Perpustakaan Menggunakan Metode K-Means,” *National Multidisciplinary Sciences*, vol. 1, no. 6, pp. 835–845, Nov. 2022, doi: 10.32528/NMS.V1I6.239.
- [23] M. Darwis, G. T. Pranoto, Y. E. Wicaksana, and Y. Yaddarabullah, “Implementation of TF-IDF Algorithm and K-mean Clustering Method to Predict Words or Topics on Twitter,” *JISA(Jurnal Informatika dan Sains)*, vol. 3, no. 2, pp. 49–55, Dec. 2020, doi: 10.31326/JISA.V3I2.831.
- [24] M. Darwis, G. T. Pranoto, Y. E. Wicaksana, and Y. Yaddarabullah, “Implementation of TF-IDF Algorithm and K-mean Clustering Method to Predict Words or Topics on Twitter,” *JISA(Jurnal Informatika dan Sains)*, vol. 3, no. 2, pp. 49–55, Dec. 2020, doi: 10.31326/JISA.V3I2.831.
- [25] D. Remawati, H. Wijayanto, Y. Retno, W. Utami, and B. D. Raharja, “Pengelompokan Film Trending di Youtube Menggunakan TF-IDF dan K-Means Clustering,” *Jurnal Sistem Informasi Triguna Dharma (JURSI TGD)*, vol. 4, no. 1, pp. 65–74, Jan. 2025, doi: 10.53513/JURSI.V4I1.10614.
- [26] A. M. Ikotun, A. E. Ezugwu, L. Abualigah, B. Abuhaija, and J. Heming, “K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data,” *Inf. Sci. (N. Y.)*, vol. 622, pp. 178–210, Apr. 2023, doi: 10.1016/J.INS.2022.11.139.
- [27] V. Agustina, A. Herliana, and E. P. Korespondensi, “Analisis Sentimen Publik atas Kebijakan Efisiensi Anggaran 2025 dengan Text Mining dan Natural Language Processing,” *Jurnal Media Informatika*, vol. 6, no. 3, pp. 2182–2194, Jul. 2025, doi: 10.55338/JUMIN.V6I3.6301.
- [28] W. N. I. Al-Obaydy, H. A. Hashim, Y. A. Najm, and A. A. Jalal, “Document classification using term frequency-inverse document frequency and K-means clustering,” *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 27, no. 3, pp. 1517–1524, 2022, doi: 10.11591/ijeecs.v27.i3.pp1517-1524.
- [29] M. Darwis, G. T. Pranoto, Y. E. Wicaksana, and Y. Yaddarabullah, “Implementation of TF-IDF Algorithm and K-mean Clustering Method to Predict Words or Topics on Twitter,” *JISA(Jurnal Informatika dan Sains)*, vol. 3, no. 2, pp. 49–55, Dec. 2020, doi: 10.31326/JISA.V3I2.831.
- [30] N. Putu, E. Merliana, and A. J. Santoso, “Prosiding Seminar Nasional Multi Disiplin Ilmu & Call For Papers Unisbank (SENDI U) Kajian Multi Disiplin Ilmu untuk Mewujudkan Poros Maritim dalam Pembangunan Ekonomi Berbasis Kesejahteraan Rakyat ANALISA PENENTUAN JUMLAH CLUSTER TERBAIK PADA METODE K-M...”.
- [31] P. K. Ritonga and M. S. Hasibuan, “Analisis Perbandingan Silhouette dengan Elbow pada Algoritma K-Means dan DBSCAN,” *METIK Jurnal*, vol. 9, no. 1, pp. 64–71, Jun. 2025, doi: 10.47002/METIK.V9I1.1027.
- [32] M. Darwis, G. T. Pranoto, Y. E. Wicaksana, and Y. Yaddarabullah, “Implementation of TF-IDF Algorithm and K-mean Clustering Method to Predict Words or Topics on Twitter,” *JISA(Jurnal Informatika dan Sains)*, vol. 3, no. 2, pp. 49–55, Dec. 2020, doi: 10.31326/JISA.V3I2.831.
- [33] M. A. Nasution *et al.*, “ALGORITMA K-MEANS DALAM PENGELOMPOKAN SURAT KELUAR DI KANTOR KEMENTERIAN AGAMA KOTA PEMATANG SIANTAR,” *Jurnal Manajemen Informatika Jayakarta*, vol. 4, no. 1, pp. 61–71, Feb. 2024, doi: 10.52362/JMIJAYAKARTA.V4I1.1304.
- [34] J. M. Guntur, R. Kurniawan, and U. Hayati, “Algoritma K-Means untuk Meningkatkan Silhouette Score pada Pengelompokan Data Stok Bahan Manufaktur di PT. XYZ Kabupaten Majalengka,” *MULTINETICS*, vol. 11, no. 1, pp. 11–21, May 2025, doi: 10.32722/MULTINETICS.V11I1.7259.
- [35] M. R. Syahkur, D. Hartama, and S. Solikhun, “Evaluasi Jumlah Cluster pada Algoritma K-Means++ Menggunakan Silhouette dan Elbow dengan Validasi Nilai DBI dalam Mengelompokkan Gizi Balita,” *JST (Jurnal Sains dan Teknologi)*, vol. 13, no. 3, pp. 487–496, Oct. 2024, doi: 10.23887/JSTUNDIKSHA.V13I3.86419.