

Implementation of BERTopic for Topic Modeling Analysis of the Free Nutritious Meal Program Based on YouTube Comments

Widya Wahyuni^{1*}, Tri Putri Lestari^{2*}, Milla Apriliana^{3*}, Riyang Gumelta^{4*}**

* Teknologi Informasi, Politeknik Negeri Padang

** Bisnis Digital, Universitas Indraprasta PGRI

widyawahyuni@pnp.ac.id¹, Tplestar189@gmail.com², millaapriliana@pnp.ac.id³, riyang@pnp.ac.id⁴

Article Info

Article history:

Received 2025-06-04

Revised 2025-07-21

Accepted 2025-08-09

Keyword:

BERTopic,

Topic Modeling,

YouTube Comments,

Free Nutritious Meal Program,

Public Perception.

ABSTRACT

The Free Nutritious Meal Program (Makan Bergizi Gratis), represents a significant national effort aimed at mitigating stunting rates in Indonesia, having commenced its operations in January 2025. As the program progressed, public sentiment towards it evolved, resulting in a diverse array of opinions that were extensively debated on various social media platforms, notably YouTube. This study was conducted with the objective of examining the perceptions of the public regarding Makan Bergizi Gratis through a topic modeling methodology employing the BERTopic approach, which analyzed 19,843 comments from YouTube. The analytical framework entailed several stages, including data preprocessing, sentence-based embedding representation, dimensionality reduction via UMAP, clustering through HDBSCAN, and topic interpretation grounded in c-TF-IDF. The findings indicate that public commentary is categorizable into ten primary themes, encompassing issues such as the involvement of political figures, concerns over budget transparency, the program's educational benefits, and the need for equitable access in underserved regions. Evaluation results show that BERTopic outperformed the traditional LDA model, with a coherence score of 0.46 compared to 0.39 and topic diversity of 76 percent compared to 71 percent. This analysis reveals that public perception of Makan Bergizi Gratis is multifaceted, shaped by social experience, political context, and economic expectations. These insights may serve as a valuable foundation for a more comprehensive understanding of public opinion, thereby supporting more targeted and responsive policy development.



This is an open access article under the [CC-BY-SA](#) license.

I. PENDAHULUAN

Program Makan Bergizi Gratis (MBG), yang pertama kali diluncurkan pada 6 Januari 2025, hingga kini masih terus berjalan [1]. Program ini termasuk salah satu gagasan yang dibawa oleh Presiden terpilih Prabowo Subianto dan Wakil Presiden terpilih Gibran Rakabuming saat kampanye Pemilihan Presiden dan Wakil Presiden tahun 2024 dengan tujuan agar angka stunting yang terjadi di Indonesia dapat berkurang [2].

Stunting merupakan gangguan pertumbuhan berdasarkan usia anak [3]. Secara global, berdasarkan laporan WHO, UNICEF dan Word Bank Group di tahun 2020 diperkirakan terdapat 144 juta anak dengan usia di bawah 5 tahun

terdampak stunting [4], dan Indonesia dengan prevalensi stunting balita sebesar 31,8%, Indonesia menempati peringkat kedua tertinggi di kawasan Asia Tenggara [5]. Meskipun pada tahun 2019 angka prevalensi stunting di Indonesia berhasil turun menjadi 21,6%, berdasarkan hasil Survei Status Gizi Indonesia (SSGI), angka tersebut masih berada di atas ambang batas yang dikelompokkan sebagai kategori rendah oleh WHO, yaitu 20% [6].

Banyak faktor yang memengaruhi terjadinya stunting pada anak, mulai dari keadaan sosial ekonomi, status gizi ibu ataupun asupan makanan [7]. Berdasarkan penelitian yang dilakukan oleh Esty dan Mifahul, kondisi stunting memiliki potensi menimbulkan kerugian ekonomi bagi Indonesia,

sehingga penanganan stunting sebaiknya dimulai sejak masa kehamilan, bahkan sejak masa remaja [8].

Dalam masa kampanye Pemilihan Presiden dan Wakil Presiden, program Makanan Bergizi Gratis (MBG) menjadi salah satu program unggulan pasangan Prabowo Subianto dan Gibran Rakabuming dalam bidang Kesehatan [9]. Program ini menyasar anak sekolah, termasuk santri dan juga ibu hamil [10], terutama wilayah yang termasuk dalam 3T (Terdepan, Terluar dan tertinggal) [11]. Selain untuk mengatasi angka terjadinya stunting, program MBG ini juga bertujuan mencegah masalah gizi lainnya serta mendukung pertumbuhan dan perkembangan generasi muda secara optimal [12]. Pemerintah telah menyiapkan anggaran sebesar Rp71 triliun sebagai tahap awal pelaksanaan program ini dan ditargetkan dapat menjangkau hingga 82,9 juta penerima manfaat saat dilaksanakan secara penuh [13].

Seiring dengan diberlakukannya program MBG di seluruh Indonesia, masyarakat mulai menyuarakan beragam pendapatnya, salah satunya melalui media sosial. Platform media sosial memberikan ruang bagi pengguna dari berbagai belahan dunia untuk menghasilkan serta membagikan informasi. Hal ini membentuk sebuah jejaring besar yang memudahkan pengumpulan, penyimpanan, pemrosesan, dan analisis data terkait isu atau topik tertentu secara cepat dan efisien [14]. Salah satunya adalah platform youtube.

YouTube diluncurkan pada tahun 2005 sebagai platform kencan online yang gagal, lalu beralih menjadi layanan berbagi video umum [15]. Kemudian YouTube diakuisisi Google pada tahun 2006 dengan nilai 1,65 miliar dolar AS [16]. Pada tahun 2021, YouTube menjadi salah satu platform media sosial paling berpengaruh di dunia, menempati posisi kedua dalam hal penggunaan sehari-hari dan akses informasi berita [17]. Platform ini memungkinkan pengguna mengunggah video, sementara pengguna lain dapat memberikan komentar, serta menekan tombol like atau dislike pada konten tersebut [18].

Dengan banyaknya respon masyarakat terhadap program MBG yang bermunculan di platform YouTube, komentar pengguna menjadi salah satu sumber data teks yang kaya akan opini publik. Komentar tersebut mencakup berbagai bentuk reaksi, mulai dari dukungan, pertanyaan, hingga kritik terhadap implementasi program. Banyaknya data dan bentuknya yang tidak terstruktur menjadikan analisis secara manual kurang efisien dan memerlukan waktu yang lama. Pemodelan topik merupakan salah satu metode dalam analisis data teks yang bertujuan untuk mengelompokkan dokumen ke dalam beberapa tema berdasarkan kesamaan konten.

Topic modeling adalah metode *unsupervised machine learning* yang digunakan untuk mengelompokkan dokumen berdasarkan kemiripan makna. Teknik ini banyak diterapkan dalam bidang *Natural Language Processing* (NLP) dan *Information Retrieval* [19]. Metode ini membantu menganalisis data teks dalam jumlah besar secara otomatis, serta menghasilkan kelompok topik yang biasanya

direpresentasikan sebagai kumpulan kata-kata yang sering muncul bersama dalam satu tema tertentu [20].

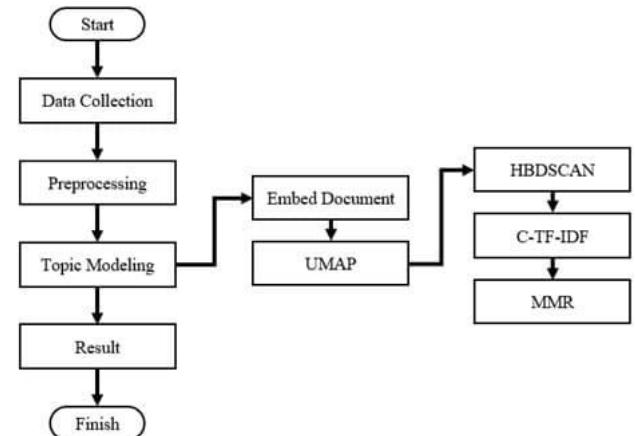
Dalam penelitian sebelumnya yang dilakukan oleh Dini dkk., *topic modeling* diterapkan untuk menganalisis pemberitaan Pemilu 2024 menggunakan metode BERTopic pada situs web Detik.com. Penelitian tersebut memanfaatkan data berupa 15.019 artikel berita yang diunduh dari laman detik.com. Proses analisis menggunakan kombinasi BERTopic dan algoritma klasterisasi K-Means untuk mengidentifikasi tema-tema utama dalam pemberitaan pemilu. Hasil penelitian menunjukkan beberapa topik dominan yang memberikan gambaran mengenai isu-isu yang berkembang selama periode Pemilu 2024 [21].

Dalam penelitiannya, Kristine dkk. berfokus pada eksplorasi opini publik tentang kendaraan listrik yang diperoleh dari komentar-komentar YouTube. Penelitian ini menggunakan metode BERTopic dengan total data sebanyak 38.896 komentar dari sejumlah video terkait. Hasilnya menunjukkan bahwa BERTopic mampu menghasilkan topik-topik yang bermakna dan beragam, sehingga membantu identifikasi tema dominan dalam respons Masyarakat [22].

Oleh karena itu, penelitian ini dilakukan untuk menganalisis komentar publik di YouTube terkait MBG dengan menggunakan pendekatan BERTopic. Metode ini dipilih karena kemampuannya dalam mengekstrak tema-tema utama dari data teks tidak terstruktur secara otomatis dan kontekstual. Hasil penelitian diharapkan dapat memberikan gambaran mengenai persepsi publik terhadap pelaksanaan MBG.

II. METODE

Gambar 1 merupakan metode penelitian yang digunakan dalam penelitian ini.



Gambar 1. Alur Penelitian

A. Data Collecting

Proses pengumpulan data dalam penelitian ini dilakukan secara otomatis melalui metode web scraping terhadap komentar publik pada sebuah video di platform YouTube. Video yang menjadi objek penelitian dipilih berdasarkan

relevansinya dengan kata kunci Makan Bergizi Gratis dan MBG , dengan rentang waktu unggahan dari November 2024 hingga April 2025. Data yang dikumpulkan mencakup informasi seperti isi komentar, nama pengguna, waktu penerbitan komentar, serta jumlah interaksi berupa *likes*.

Tujuan dari proses pengumpulan data ini adalah untuk memperoleh perspektif masyarakat terkait program MBG berdasarkan komentar-komentar yang muncul pada video-video tersebut. Dari proses ini, diperoleh dataset sebanyak 19.843 komentar yang berasal dari 71 video. Gambar 4 menampilkan visualisasi dari dataset hasil pengumpulan data tersebut.

	video_title	comment
0	Respons soal Makan Bergizi Gratis, Ekonom: Say...	Yang dibutuhkan masyarakat secara tegas: bukan...
1	Respons soal Makan Bergizi Gratis, Ekonom: Say...	Memang alam Indonesia kaya raya ... tanah air...
2	Respons soal Makan Bergizi Gratis, Ekonom: Say...	Ini mah program abal2 , media korupsi , dana 4...
3	Respons soal Makan Bergizi Gratis, Ekonom: Say...	Sebaik apapun programnya selama korupsi tidak ...
4	Respons soal Makan Bergizi Gratis, Ekonom: Say...	Pak Prabowo alokasikan saja dananya, pendidika...
...
18995	Dugaan keracunan karena MBG: 'Perlu evaluasi b...	MBG (makanan berasukan gratis) 🍽
18996	Dugaan keracunan karena MBG: 'Perlu evaluasi b...	Ini kan sesuai keinginan mayoritas rakyat. Mi...
18997	Dugaan keracunan karena MBG: 'Perlu evaluasi b...	Investigasi oii kepala daerah nya
18998	Dugaan keracunan karena MBG: 'Perlu evaluasi b...	🤣🤣 pokonya jogedin ajah ok gas ok gas nyampe ...
18999	Dugaan keracunan karena MBG: 'Perlu evaluasi b...	Stop mbg 🤦, ganti dgn msg 🥰

Gambar 2. Dataset Komentar Dari Video Youtube.

B. Preprocessing

Dataset yang diperoleh masih berupa data mentah. Untuk memastikan kualitas dan kesiapan data sebelum analisis lebih lanjut, dilakukan tahap pra-pemrosesan (*preprocessing*). Tahap ini bertujuan untuk membersihkan data dari karakter atau informasi yang tidak relevan, mengurangi dimensi data, serta membentuk data menjadi lebih terstruktur dan siap untuk dianalisis lebih lanjut[23]. Adapun tahapan yang dilakukan adalah [24]:

1) *Case Folding*: Proses mengubah semua huruf dalam teks menjadi bentuk yang sama, seperti seluruhnya huruf kecil atau seluruhnya huruf besar. Tujuannya adalah mengurangi perbedaan pada kata-kata yang sebenarnya sama, hanya karena adanya variasi huruf kapital dan huruf kecil [21]. Dalam penelitian ini, seluruh huruf diubah menjadi huruf kecil, seperti yang terlihat pada Gambar 3.

	comment	case_folding
0	Yang dibutuhkan masyarakat secara tegas: bukan...	yang dibutuhkan masyarakat secara tegas: bukan...
1	Memang alam Indonesia kaya raya ... tanah air...	memang alam Indonesia kaya raya ... tanah air...
2	Ini mah program abal2 , media korupsi , dana 4...	ini mah program abal2 , media korupsi , dana 4...
3	Sebaik apapun programnya selama korupsi tidak ...	sebaik apapun programnya selama korupsi tidak ...
4	Pak Prabowo alokasikan saja dananya, pendidika...	pak prabowo alokasikan saja dananya, pendidika...

Gambar 3.Hasil Proses Case Folding Pada Dataset.

2) *Cleaning Data*: Proses mengidentifikasi, memperbaiki, atau menghapus data yang tidak benar, tidak lengkap, atau mengandung gangguan dari kumpulan data [25]. Dalam penelitian ini, tahap pembersihan data dilakukan dengan menghilangkan berbagai elemen yang dapat mengganggu analisis, seperti simbol, emoji, angka, *mention*, URL, serta komentar spam. Selain itu, dilakukan pula penghapusan

komentar yang terlalu pendek dan tidak informatif, serta baris kosong yang muncul sebagai hasil dari proses sebelumnya, seperti yang terlihat pada Gambar 4.

	case_folding	cleaned
1	memang alam indonesia kaya raya ... tanah air...	memang alam indonesia kaya raya tanah air indo...
3	sebaik apapun programnya selama korupsi tidak ...	sebaik apapun programnya selama korupsi tidak...

Gambar 4. Hasil Cleaning Data

3) *Normalization*: Normalisasi dalam penelitian ini bertujuan untuk mengubah bentuk-bentuk kata yang tidak baku, seperti kata gaul, singkatan, dan kesalahan penulisan, menjadi kata-kata yang sesuai dengan kaidah bahasa Indonesia yang baku dan mudah dipahami secara luas.

	cleaned_6	normalized_text
0	memang alam indonesia kaya raya tanah air indo...	memang alam indonesia kaya raya tanah air indo...
1	sebaik apapun programnya selama korupsi tidak...	sebaik apapun programnya selama korupsi tidak...
2	pak prabowo alokasikan saja dananya pendidikan...	pak prabowo alokasikan saja dananya pendidikan...
3	jika dana desa dialihkan makanan bergizi untuk...	jika dana desa dialihkan makanan bergizi untuk...
4	ribu udah cukup itu juga kalau dana bantuan...	ribu sudah cukup itu juga kalau dana bantuan...

Gambar 5. Hasil Normalization

4) *Stopword Removal*: Proses ini bertujuan untuk menghilangkan kata-kata atau frasa yang umum digunakan tetapi tidak membawa makna atau informasi signifikan dalam analisis dokumen [26]. Contoh kata tersebut meliputi *yang, dan, di, itu, nya*, serta preposisi atau konjungsi lainnya.

	normalized_text	stopwords
0	memang alam indonesia kaya raya tanah air indo...	memang alam indonesia kaya raya tanah air indo...
1	sebaik apapun programnya selama korupsi tidak...	sebaik apapun programnya selama korupsi tidak...
2	pak prabowo alokasikan saja dananya pendidikan...	pak prabowo alokasikan saja dananya pendidikan...
3	jika dana desa dialihkan makanan bergizi untuk...	dana desa dialihkan makanan bergizi anak...
4	ribu sudah cukup itu juga kalau dana bantuan...	ribu cukup kalau dana bantuan...

Gambar 6. Hasil Stopword Removal

C. Topic Modeling

Pemodelan topik adalah metode analisis teks yang digunakan untuk menemukan tema atau topik utama dari sekumpulan dokumen secara otomatis berdasarkan pola penggunaan kata di dalamnya [21]. Dalam penelitian ini, model yang digunakan adalah BERTopic. BERTopic merupakan pustaka *Python* untuk pemodelan topik yang memanfaatkan representasi semantik teks. Metode ini menggunakan *Sentence Transformers* untuk menghasilkan vektor representasi makna teks, HDBSCAN untuk membentuk kluster topik yang relevan, serta *Class-based TF-IDF* guna menghasilkan topik yang mudah diinterpretasi dengan tetap mempertahankan kata-kata kunci [27].

BERTopic menghasilkan representasi topik melalui tiga tahap utama. Pertama, setiap dokumen diubah menjadi vektor embedding menggunakan *pre-trained language model*. Selanjutnya, dimensi *vector embedding* tersebut dikurangi untuk meningkatkan efisiensi proses *clustering*. Tahap terakhir adalah ekstraksi representasi topik dari kelompok dokumen yang terbentuk dengan menggunakan variasi khusus dari metode *c-TF-IDF* [28].

III. HASIL DAN PEMBAHASAN

Dalam penelitian ini dilakukan analisis topic modeling terhadap data yang berkaitan dengan kata kunci MBG dengan menggunakan model BERTopic. Dataset diperoleh dari komentar publik yang dikumpulkan menggunakan teknik *web crawling* pada *platform* YouTube, dengan rentang waktu unggahan dari November 2024 hingga April 2025.

Dari proses pengumpulan data, diperoleh total 19.843 komentar yang berasal dari 71 video. Video-video tersebut bersumber dari berbagai kanal YouTube, mulai dari kanal berita hingga kanal yang membahas topik seputar MBG. Seluruh data kemudian disimpan dalam format CSV guna memudahkan proses analisis lebih lanjut.

Dataset yang telah dikumpulkan selanjutnya melalui tahapan *preprocessing*, yang mencakup *case folding*, *cleaning data*, normalisasi teks, serta *stopword removal*. Setelah proses *preprocessing* selesai dilakukan, analisis topic modeling diterapkan pada data yang telah dibersihkan guna mengungkap topik-topik utama yang terkandung dalam komentar pengguna mengenai program MBG.

A. Pemodelan BERTopic

Setelah proses *preprocessing* dilakukan pada *dataset*, tahap selanjutnya adalah pemodelan topik menggunakan BERTopic untuk mengidentifikasi tema-tema utama dalam komentar publik. Hasil dari pemodelan ini menunjukkan bahwa komentar berhasil dikelompokkan ke dalam sejumlah kluster topik yang saling berbeda.

1) Document Embedding

Proses ini diimplementasikan menggunakan model *pre-trained Sentence-BERT (SBERT) paraphrase-multilingual-MiniLM-L12-v2*.

Preview hasil embedding dan teks:										...		\
	text	0	1	2	3	4	5	6	7	8	...	\
0	alam indonesia kaya raya tanah air indonesia s...	0.278713	0.052800									
1	sebaik apapun programnya selama korupsi tidak...	0.056800	-0.003749									
2	prabowo alokasiakan dananya pendidikan kesehatan...	0.071908	0.243858									
3	dana desa dialihkan makanan bergizi anak anak ...	-0.005241	-0.030157									
4	ribu dana bantuananya korupsi desa setempat	0.137579	-0.195950									
		-0.045174	-0.017129	0.229234	-0.207823	0.014281	0.147856	-0.003938	...			
0		-0.184867	0.015190	0.039447	-0.231192	-0.053495	-0.122719	-0.019169	...			
1		-0.048484	0.010136	0.151883	-0.130552	0.210136	0.282015	0.037532	...			
2		-0.198598	0.300734	0.275874	0.374209	-0.193015	-0.165977	-0.141287	...			
3		-0.192787	0.189454	-0.014372	0.014781	-0.047491	0.016504	-0.044302	...			
		374	375	376	377	378	379	380	...			
0		-0.048209	-0.422242	0.089368	0.309198	-0.028493	0.061115	0.566522				
1		0.101027	-0.191904	0.108048	0.221975	-0.087096	0.089337	0.343923				
2		-0.052046	-0.176678	-0.084055	0.0380335	-0.085551	0.016434	0.292803				
3		0.110216	-0.288479	0.090324	0.036730	0.152681	0.228875	0.050495				
4		0.154634	-0.409881	0.072565	0.177227	0.089151	-0.007941	0.196739				
		381	382	383								
0		0.003906	0.187141	-0.077386								
1		-0.072981	0.330721	-0.010017								
2		0.058259	-0.018149	0.134905								
3		-0.274037	0.154955	-0.028796								
4		-0.066350	0.017902	-0.033529								

Gambar 7. Hasil *document embedding* yang menghasilkan *vector embedding* berdimensi 384

Model ini dipilih karena memiliki ukuran yang ringan namun efektif, serta mendukung berbagai bahasa, termasuk Bahasa Indonesia. Melalui proses ini, setiap komentar yang

telah melalui tahapan *preprocessing* dikonversi menjadi representasi numerik atau vektor berdimensi tinggi. Model ini menghasilkan *embedding vector* dengan dimensi 384. Gambar 7 menampilkan hasil proses *document embedding* dalam bentuk representasi numerik untuk setiap komentar.

2) Dimensionality Reduction

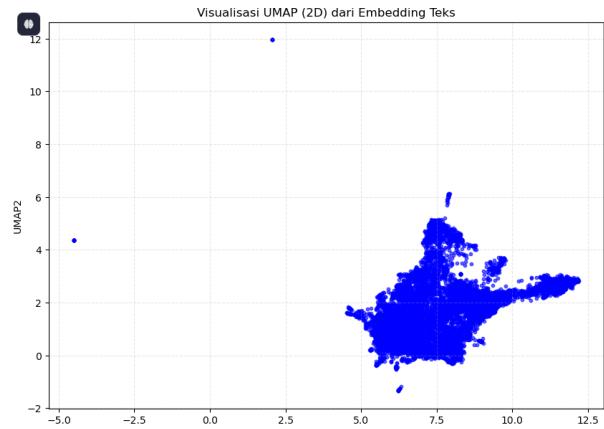
UMAP (*Uniform Manifold Approximation and Projection*) digunakan dalam penelitian ini untuk mereduksi dimensi *vector embedding* hasil *Sentence-BERT*. Tujuan utama dari proses ini adalah menyederhanakan representasi *embedding* yang berdimensi tinggi sehingga sulit untuk dianalisis secara langsung. Selain itu, proses ini juga bertujuan memungkinkan pelaksanaan klusterisasi yang lebih efisien.

Konfigurasi parameter yang digunakan adalah *n_neighbors=15*, *n_components=4*, *min_dist=0.0*, dan *metric='cosine'*. Parameter-parameter tersebut menentukan bagaimana UMAP melakukan reduksi dimensi terhadap data. Hasil dari proses dimensionality reduction menggunakan UMAP ditampilkan pada Gambar 8.

	UMAP1	UMAP2	UMAP3	UMAP4	text
0	9.427979	2.495186	2.604158	4.012278	alam indonesia kaya raya tanah air indonesia s...
1	11.077528	3.061589	5.562970	3.861470	sebaik apapun programnya selama korupsi tidak...
2	8.348460	2.532370	2.483518	4.087186	prabowo alokasiakan dananya pendidikan kesehatan...
3	8.007891	3.046287	4.710370	5.298572	dana desa dialihkan makanan bergizi anak anak ...
4	11.067844	2.583335	5.363358	4.336613	ribu dana bantuananya korupsi desa setempat

Gambar 8. Hasil dimentionality reduction menggunakan UMAP

Gambar 9 menunjukkan hasil visualisasi *embedding* komentar publik setelah reduksi dimensi dilakukan menggunakan algoritma UMAP ke dalam ruang dua dimensi. Setiap titik pada grafik merepresentasikan satu komentar yang dipetakan berdasarkan kedekatan makna atau semantiknya dengan komentar lainnya. Tampak bahwa dokumen-dokumen yang memiliki kesamaan makna cenderung membentuk gugus (*cluster*) atau area yang padat, menunjukkan adanya struktur topik yang dapat diidentifikasi secara visual.

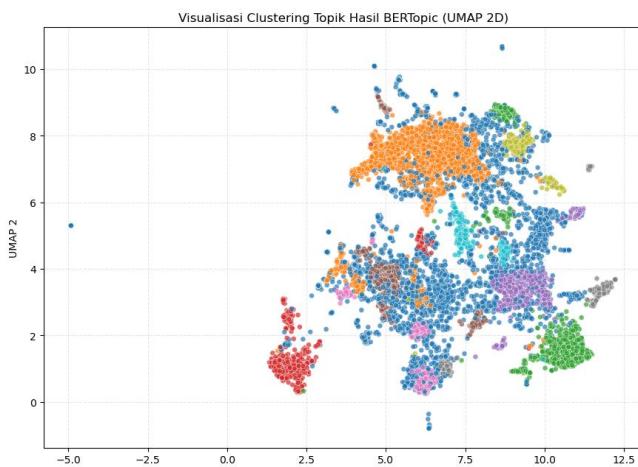


Gambar 9. Visualisasi hasil dimentionality reduction menggunakan UMAP

3) Clustering

Setelah reduksi dimensi dilakukan menggunakan UMAP, tahap selanjutnya adalah proses klusterisasi menggunakan algoritma HDBSCAN untuk mengelompokkan komentar berdasarkan kemiripan semantiknya. Dengan parameter $\text{min_cluster_size}=50$, proses ini membagi vektor-vektor tersebut ke dalam beberapa kelompok tematik yang konsisten sekaligus mengisolasi komentar-komentar yang tidak memiliki koherensi semantik.

Kluster yang terbentuk menjadi dasar dalam penyusunan topik utama yang muncul dari opini publik. Hasil dari proses ini divisualisasikan pada Gambar 10, di mana setiap warna merepresentasikan kelompok topik yang berbeda, sedangkan titik-titik menunjukkan persebaran komentar dalam ruang dua dimensi. Visualisasi ini mempermudah pemahaman terhadap struktur topik yang terbentuk serta membantu dalam proses eksplorasi dan interpretasi isi percakapan publik yang berkaitan dengan program MBG.



Gambar 10. Hasil dari proses clustering menggunakan HBDSCAN

4) Topic representation

Setelah proses klusterisasi dilakukan menggunakan HDBSCAN, metode $c\text{-TF-IDF}$ diimplementasikan untuk menyusun representasi topik. Proses ini bertujuan menentukan kata-kata yang paling mewakili setiap topik. Dengan representasi tersebut, setiap topik dapat diinterpretasikan secara lebih jelas dan mendalam, sehingga mempermudah analisis lanjutan terhadap tema-tema yang berkembang dalam komentar mengenai program MBG.

Gambar 11 menunjukkan sepuluh kata kunci teratas beserta skor $c\text{-TF-IDF}$ -nya untuk salah satu topik yang dihasilkan oleh pemodelan BERTopic. Kata "indonesia" memiliki skor tertinggi (0,177), menunjukkan bahwa kata ini paling representatif dalam topik tersebut.

```
[('indonesia', 0.17732731053250836),
 ('papua', 0.07587720026307415),
 ('anak', 0.044003477528142876),
 ('rakyat', 0.03500898723485097),
 ('negara', 0.0335866692683978),
 ('program', 0.024832834592915187),
 ('daerah', 0.024432243357452612),
 ('java', 0.023047942343894193),
 ('mbg', 0.02187205630051675),
 ('prabowo', 0.020475323660025073)]
```

Gambar 11. Hasil dari proses $c\text{-TF-IDF}$

Representasi topik menggunakan $c\text{-TF-IDF}$ menghasilkan sepuluh topik utama dengan kata-kata kunci berperingkat tertinggi, yang menjadi dasar dalam mengidentifikasi dan menganalisis persepsi publik terhadap program MBG.

Pemodelan awal menggunakan BERTopic menghasilkan 174 topik yang terbentuk secara otomatis dari data komentar YouTube. Jumlah ini terlalu kompleks untuk dianalisis secara interpretatif, sehingga dilakukan proses penyederhanaan. Untuk tujuan tersebut, fungsi `reduce_topics()` digunakan dari pustaka BERTopic, yang secara sistematis menggabungkan topik-topik yang memiliki tingkat kesamaan semantik tinggi berdasarkan representasi embedding-nya. sehingga didapatkanlah sepuluh topik yang digunakan untuk mengambil kesimpulan dari program MBG

```
Representation \
0 [anak, makan, mbg, program, sekolah, makanan, ...]
1 [prabowo, cahyadi, anak, mbg, bersyukur, rakyat...]
2 [korupsi, yayasan, uang, koruptor, program, mb...
3 [makan, makanan, nasi, anak, bergizi, ribu, sa...
4 [indonesia, papua, anak, rakyat, negara, progr...
5 [presiden, pemerintah, kampanye, negara, janji...
6 [program, mbg, programnya, berjalan, prabowo, ...]
7 [sekolah, siswa, anak, program, kantin, uang, ...]
8 [desa, daerah, kota, petani, kampung, terpenc...
9 [kerja, lapangan, bos, pekerjaan, phk, gaji, r...
```

Gambar 12. Hasil dari pemodelan BERTopic

B. Evaluasi Model

Efektivitas model BERTopic dalam menganalisis komentar publik dievaluasi menggunakan dua metrik utama yang secara konvensional digunakan dalam topik modeling, yaitu coherence score dan topic diversity. Coherence score mengukur tingkat keterkaitan semantik antar kata dalam suatu topik, sedangkan topic diversity menilai sejauh mana keragaman kata kunci yang muncul di antara topik-topik yang dihasilkan.

Hasil evaluasi menunjukkan bahwa model BERTopic memperoleh nilai rata-rata coherence score sebesar 0,46 dan topic diversity sebesar 76%, lebih tinggi dibandingkan model LDA yang hanya mencapai 0,39 dan 71%.

Meskipun perbedaannya tidak terlalu jauh, hasil ini mengindikasikan bahwa BERTopic cenderung menghasilkan topik yang lebih koheren secara semantik dan tidak redundan. Penilaian ini memperkuat keunggulan BERTopic dalam membentuk kluster topik yang lebih representatif dan bermakna, khususnya dalam konteks data komentar YouTube yang umumnya bersifat singkat dan informal.

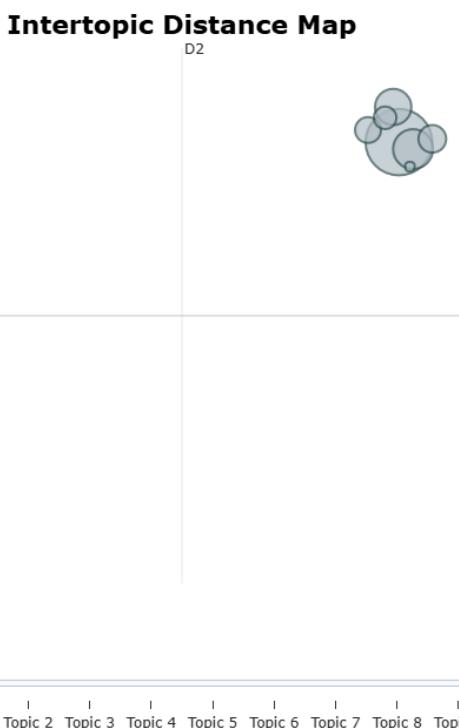
Secara metodologis, BERTopic menawarkan keunggulan dibandingkan model-model lama seperti *Latent Dirichlet Allocation* (LDA). LDA mengandalkan pendekatan *bag-of-words* dan metode probabilistik yang mengabaikan urutan kata serta makna kontekstual, sehingga kurang efektif untuk menganalisis teks pendek seperti komentar di YouTube. Sebaliknya, BERTopic menggunakan representasi berbasis kalimat (*sentence embeddings*) yang dibentuk oleh model transformer (*Sentence-BERT*), yang mampu menangkap makna secara menyeluruh dalam konteks kalimat.

Metode ini mengelompokkan komentar berdasarkan kesamaan makna, bukan sekadar kemunculan kata yang sama. Dengan kinerja yang unggul dan landasan teoritis yang kuat, BERTopic terbukti menjadi pilihan yang lebih tepat untuk menganalisis opini publik dalam konteks program Makan Bergizi Gratis.

C. Visualisasi Hasil

1) Intertopic Distance Map

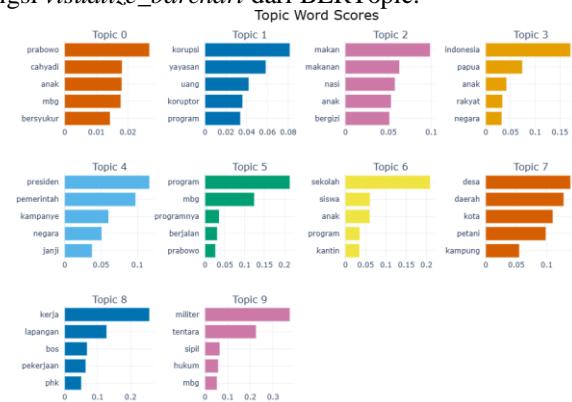
Intertopic Distance Map merupakan grafik yang menunjukkan seberapa dekat atau jauhnya antar topik. Pada Gambar 13, tampak beberapa titik berbentuk lingkaran. Setiap lingkaran dalam grafik merepresentasikan satu topik, dan jarak antar lingkaran mencerminkan kedekatan semantik antar topik tersebut. Berdasarkan visualisasi pada Gambar 13, terlihat bahwa Topik 2, 3, 5, 6, 7, dan 9 saling berdekatan, membentuk satu kelompok topik yang memiliki keterkaitan semantik. Sementara itu, Topik 1, 4, dan 8 membentuk kelompok tersendiri yang relatif berjauhan dari kelompok sebelumnya.



Gambar 13. Visualisasi grafik menggunakan *intertopic distance map*

2) Topic Word Scores

Topic Word Score merupakan salah satu visualisasi hasil pemodelan topik yang menunjukkan skor atau tingkat kepentingan setiap kata kunci dalam suatu topik. Visualisasi ini biasanya disajikan dalam bentuk bar chart menggunakan fungsi *visualize_barchart* dari BERTopic.



Gambar 14. Visualisasi hasil pemodelan topik dalam bentuk barchar

Berdasarkan hasil pemodelan topik menggunakan BERTopic dengan kalsterisasi HDBSCAN, ditemukan 10 topik utama yang merepresentasikan opini publik secara menyeluruh. Setiap topik ditentukan berdasarkan skor *c-TF-IDF* yang mengidentifikasi kata-kata kunci paling menonjol dalam dataset komentar. Gambar 15 merupakan topik yang di dapatkan dari hasil Pemodelan BERTopic.

No.	Topik	Representative Keyword
0	Tokoh politik & dukungan publik	prabowo, cahyadi, anak, mbg, bersyukur
1	Kekhawatiran korupsi dan pengawasan dana	korupsi, yayasan, uang, koruptor, program
2	Penerimaan makanan bergizi untuk anak	makanan, nasi, anak, bergizi
3	Keadilan distribusi wilayah	indonesia, papua, anak, rakyat, negara
4	Janji kampanye dan ekspektasi politik	presiden, pemerintah, kampanye, negara, janji
5	Pelaksanaan program MBG	program, mbg, programnya, berjalan, prabowo
6	Dampak langsung ke sekolah dan siswa	sekolah, siswa, anak, program, kantin
7	Akses di wilayah pedesaan	desa, daerah, kota, petani, kampung
8	Harapan terhadap ekonomi & lapangan kerja	kerja, lapangan, bos, pekerjaan, phk
9	Peran militer dalam pengawasan program	militer, tentara, sipil, hukum, mbg

Gambar 15. Interpretasi topik komentar Youttube

Topik-topik yang muncul menunjukkan keberagaman sudut pandang masyarakat terkait program MBG. Pada topik 2 dan 6, aspek makanan, sekolah, dan anak-anak menjadi fokus utama, yang menggambarkan keterkaitan langsung masyarakat dengan manfaat program di lingkungan pendidikan. Di sisi lain, topik 1 dan 8 mencerminkan adanya kekhawatiran masyarakat terhadap isu-isu pengelolaan dana, transparansi, dan ketimpangan ekonomi yang mungkin timbul dalam pelaksanaannya. Hal ini menunjukkan perlunya upaya untuk meningkatkan kepercayaan publik

terhadap lembaga atau pihak yang bertanggung jawab atas program MBG.

Selain itu, terdapat pula topik yang menyoroti peran Prabowo sebagai sosok penting dalam pelaksanaan program MBG, seperti pada topik 0 dan 5. Hal ini mengindikasikan bahwa sebagian masyarakat mengaitkan pencapaian maupun tantangan program dengan aspek kepemimpinan politik. Sementara itu, topik 9 yang didominasi oleh kata-kata seperti "militer" dan "tentara" menunjukkan adanya pembahasan mengenai peran pemerintah dalam pengawasan dan pelaksanaan program melalui pendekatan keamanan.

IV. KESIMPULAN

Temuan yang diperoleh dari analisis pemodelan topik menghasilkan sepuluh topik yang mencerminkan beragam tema yang mewakili perbedaan pendapat masyarakat. Setiap topik yang ditemukan dibentuk berdasarkan kumpulan kata kunci yang berhubungan dengan program Makan Bergizi Gratis (MBG). Topik 0 dan 4 menunjukkan hubungan antara inisiatif MBG dengan tokoh politik terkenal serta janji-janji kampanye, yang menunjukkan bahwa sebagian masyarakat memandang program ini bukan hanya sebagai upaya meningkatkan gizi, tetapi juga sebagai bagian dari agenda politik. Sementara itu, Topik 1 menggambarkan kekhawatiran masyarakat terkait kemungkinan penyalahgunaan anggaran, yang menjadi isu penting dalam pelaksanaan program berskala nasional. Sebaliknya, Topik 2, 5, dan 6 menyoroti dampak positif dari program MBG bagi anak-anak dan lingkungan sekolah. Topik 3 dan 7 menekankan pentingnya pemerataan akses, khususnya di wilayah Indonesia Timur dan daerah pedesaan. Terakhir, Topik 8 dan 9 mencerminkan perubahan harapan masyarakat terhadap dampak ekonomi dan cara pengawasan pelaksanaan program MBG.

Temuan ini memperjelas sejumlah hal penting yang mencerminkan cara masyarakat memandang program MBG. Masalah transparansi anggaran muncul sebagai perhatian utama, menunjukkan bahwa tanggung jawab kepada publik dianggap sama pentingnya dengan pelaksanaan teknis program. Selain itu, masih ada anggapan adanya ketimpangan antara wilayah kota dan desa dalam hal penyebaran program, yang menunjukkan harapan masyarakat terhadap akses yang lebih adil. Keterkaitan antara program MBG dan tokoh politik juga menunjukkan bahwa inisiatif ini tidak hanya dipahami sebagai kebijakan di bidang gizi, tetapi juga sebagai bagian dari konteks politik dan kepemimpinan. Di sisi lain, pembahasan mengenai keterlibatan militer dalam pengawasan program menunjukkan tingginya perhatian masyarakat terhadap cara pemerintah melaksanakan kebijakan publik.

Temuan ini memperjelas sejumlah hal penting yang mencerminkan cara masyarakat memandang program MBG. Masalah transparansi anggaran muncul sebagai perhatian utama, menunjukkan bahwa tanggung jawab kepada publik dianggap sama pentingnya dengan pelaksanaan teknis

program. Selain itu, masih ada anggapan adanya ketimpangan antara wilayah kota dan desa dalam hal penyebaran program, yang menunjukkan harapan masyarakat terhadap akses yang lebih adil. Keterkaitan antara program MBG dan tokoh politik juga menunjukkan bahwa inisiatif ini tidak hanya dipahami sebagai kebijakan di bidang gizi, tetapi juga sebagai bagian dari konteks politik dan kepemimpinan. Di sisi lain, pembahasan mengenai keterlibatan militer dalam pengawasan program menunjukkan tingginya perhatian masyarakat terhadap cara pemerintah melaksanakan kebijakan publik.

Secara keseluruhan, analisis topik menggunakan BERTopic memberikan gambaran yang komprehensif mengenai opini publik terhadap MBG, mulai dari apresiasi terhadap tujuan program hingga kekhawatiran terhadap pelaksanaannya. Metode ini terbukti efektif dalam mengelompokkan percakapan publik menjadi topik-topik yang bermakna sehingga dapat digunakan sebagai dasar evaluasi program, penyusunan kebijakan, dan strategi komunikasi publik yang lebih tepat sasaran.

DAFTAR PUSTAKA

- [1] A. Andin *et al.*, "Penerapan Nilai Pancasila Melalui Program Makan Bergizi Gratis," vol. 3, no. 1, pp. 370–383, 2025.
- [2] A. Kiftiyah, F. A. Palestina, F. U. Abshar, and K. Rofiah, "Program Makan Bergizi Gratis (MBG) dalam Perspektif Keadilan Sosial dan Dinamika Sosial-Politik," *J. Keindonesian*, vol. 5, no. 1, 2025, doi: <https://doi.org/10.52738/pjk.v5i1.726>.
- [3] S. H. Quamme and P. O. Iversen, "Prevalence of child stunting in Sub-Saharan Africa and its risk factors," *Clin. Nutr. Open Sci.*, vol. 42, no. 2022, pp. 49–61, 2022, doi: 10.1016/j.nutos.2022.01.009.
- [4] I. Govender, S. Rangiah, R. Kaswa, and D. Nzaumvila, "Erratum to: Malnutrition in children under the age of 5 years in a primary health care setting (S Afr Fam Pract. 2021;63(1), a5337. 10.4102/safp.v63i1.5337)," *South African Fam. Pract.*, vol. 63, no. 1, pp. 1–6, 2021, doi: 10.4102/SAFP.V63I1.5416.
- [5] H. Hatijar, "The Incidence of Stunting in Infants and Toddlers," *J. Ilm. Kesehat. Sandi Husada*, vol. 12, no. 1, pp. 224–229, 2023, doi: 10.35816/jiskh.v12i1.1019.
- [6] A. Andiani, T. Lestari, and T. Sumiati, "Gambaran Pengetahuan Remaja Tentang Stunting," *J. Biosaintek*, vol. 5, no. 2, pp. 17–20, 2023, doi: 10.5204/biosaintek.v5i2.1641.
- [7] F. Agusybana, A. Pratiwi, P. L. Kurnia, N. Nandini, J. Santoso, and A. Setyo, "Reducing Stunting Prevalence: Causes, Impacts, and Strategies," *BIO Web Conf.*, vol. 54, pp. 1–6, 2022, doi: 10.1051/bioconf/20225400009.
- [8] E. A. Suryana and M. Azis, "the Potential of Economic Loss Due To Stunting in Indonesia," *J. Ekon. Kesehat. Indones.*, vol. 8, no. 1, p. 52, 2023, doi: 10.7454/eki.v8i1.6796.
- [9] D. L. Lestari, Y. H. Gusmira, A. Mashdarul, and A. Y. Amelia, "Free Nutritious Meal Policy As A Solution To Overcoming The Stunting Problem In Indonesia," vol. 4, pp. 10021–10031, 2024.
- [10] S. S. Fasha and D. Tesniyadi, "Analisis Wacana Kritis Pada Artikel Tempo.co Yang Berjudul "Dana BOS Untuk Program Makan Siang Gratis"," vol. 4, no. 3, pp. 15077–15089, 2024, doi: <https://doi.org/10.31004/innovative.v4i3.12362>.
- [11] A. Purnomo, A. Putikadyanto, and A. Nur, "Program Makan Bergizi Gratis: Analisis Kritis Transformasi Pendidikan Indonesia Menuju Generasi Emas 2045," 2025.
- [12] K. Andreas *et al.*, "Tinjauan Kritis Tentang Program Makan Bergizi Gratis Terhadap Produktivitas Belajar Siswa," vol. 2,

- 2025.
- [13] 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," vol. 7, no. 2, pp. 1364–1373, 2025.
- [14] X. Huang *et al.*, "Social media mining under the COVID-19 context: Progress, challenges, and opportunities," *Int. J. Appl. Earth Obs. Geoinf.*, vol. 113, no. March, p. 102967, 2022, doi: 10.1016/j.jag.2022.102967.
- [15] L. Kauffman, E. M. Weisberg, J. Eng, and E. K. Fishman, "YouTube and Radiology: The Viability, Pitfalls, and Untapped Potential of the Premier Social Media Video Platform for Image-Based Education," vol. 29, no. 5, pp. 1–8, 2022, doi: <https://doi.org/10.1016/j.acra.2020.12.018>.
- [16] A. Khakim and M. Zubaidi, "The Necessity of Educational Content for Indonesian Society Through YouTube Channels," pp. 296–309, 2024, doi: <https://doi.org/10.38142/ijssc.v2i3.225>.
- [17] T. S. Fischer, C. Kolo, and C. Mothes, "Political Influencers on YouTube: Business Strategies and Content Characteristics," *Media Commun.*, vol. 10, no. 1, pp. 259–271, 2022, doi: 10.17645/mac.v10i1.4767.
- [18] W. Sui, A. Sui, and R. E. Rhodes, "What to watch: Practical considerations and strategies for using YouTube for research," *Digit. Heal.*, vol. 8, 2022, doi: 10.1177/20552076221123707.
- [19] A. Abuzayed and H. Al-Khalifa, "BERT for Arabic Topic Modeling: An Experimental Study on BERTopic Technique," *Procedia CIRP*, vol. 189, pp. 191–194, 2021, doi: 10.1016/j.procs.2021.05.096.
- [20] C. C. Silva, M. Galster, and F. Gilson, "Topic modeling in software engineering research," *Empir. Softw. Eng.*, vol. 26, no. 6, 2021, doi: 10.1007/s10664-021-10026-0.
- [21] D. Aryani, I. Lucia Kharisma, A. Sujada, and K. Kamdan, "Topic Modeling of the 2024 Election Using the BERTopic Method on Detik.com News Articles," *Inf. J. Ilm. Bid. Teknol. Inf. dan Komun.*, vol. 9, no. 2, pp. 171–180, 2024, doi: 10.25139/inform.v9i2.8429.
- [22] T. Informasi, "Jurnal Identifikasi Opini Publik Terhadap Kendaraan Listrik dari Data Komentar YouTube: Pemodelan Topik Menggunakan BERTopic," vol. 10, pp. 195–203, 2024.
- [23] A. A. Syam, G. H. M, A. Salim, D. F. Surianto, and M. F. B, "Analisis teknik preprocessing pada sentimen masyarakat terkait konflik israel-palestina menggunakan support vector machine," vol. 9, no. 3, pp. 1464–1472, 2024.
- [24] N. Charibaldi, A. Harfiani, and O. Samuel Simanjuntak, "Comparison of the Effect of Word Normalization on Naïve Bayes Classifier and K-Nearest Neighbor Methods for Sentiment Analysis," *Inf. J. Ilm. Bid. Teknol. Inf. dan Komun.*, vol. 9, no. 1, pp. 25–31, 2023, doi: 10.25139/inform.v9i1.7111.
- [25] K. Maharanan, S. Mondal, and B. Nemade, "A review: Data pre-processing and data augmentation techniques," *Glob. Transitions Proc.*, vol. 3, no. 1, pp. 91–99, 2022, doi: 10.1016/j.gltip.2022.04.020.
- [26] P. Wahyuni and M. A. Romli, "Comparison of Naïve Bayes Classifier and Decision Tree Algorithms for Sentiment Analysis on the House of Representatives' Right of Inquiry on Twitter," vol. 8, no. 2, pp. 523–530, 2024.
- [27] C. S. Ponay, "Topic Modeling on Customer Feedback from an Online Ticketing System using Latent Dirichlet Allocation and BERTopic," *Proc. - 2022 2nd Int. Conf. Inf. Comput. Res. iCORE 2022*, no. April, pp. 1–6, 2022, doi: 10.1109/iCORE58172.2022.00020.
- [28] M. Grootendorst, "BERTopic: Neural topic modeling with a class-based TF-IDF procedure," 2022, [Online]. Available: <http://arxiv.org/abs/2203.05794>