

Implementasi BERTopic untuk Menganalisis Isu Program Makan Bergizi Gratis pada Media Sosial X

Implementation of BERTopic for Analyzing Issues on the Free Nutritious Meal Program on Social Media X

Marcellio Aurel Christian*¹, Eka Dyar Wahyuni², Mohammad Al Hafidz³

^{1,2,3}Sistem Informasi, Universitas Pembangunan Nasional "Veteran" Jawa Timur

E-mail: ¹marcellioach@gmail.com, ²ekawahyuni.si@upnjatim.ac.id,

³hafidz.si@upnjatim.ac.id

Abstrak

Program Makan Bergizi Gratis menjadi salah satu kebijakan pemerintah yang banyak dibahas masyarakat di media sosial X. Banyaknya unggahan membuat identifikasi isu secara manual kurang efisien, sehingga diperlukan metode yang mampu mengelompokkan percakapan publik secara otomatis. Penelitian ini bertujuan mengimplementasikan BERTopic untuk memetakan isu Program Makan Bergizi Gratis pada media sosial X. Data dikumpulkan melalui proses scraping dengan kata kunci "Program Makan Bergizi Gratis" atau "MBG" pada periode 6 Januari 2025 hingga 6 Oktober 2025. Dari 37.368 data mentah, diperoleh 10.637 data setelah tahap filtering. Tahapan penelitian meliputi prapemrosesan teks, pemodelan topik menggunakan BERTopic, evaluasi, visualisasi, dan pelabelan topik. Hasil terbaik diperoleh pada min_cluster_size 40 dengan nilai coherence 0,740873, stability 0,753521, dan final score 0,747197. BERTopic menghasilkan 10 isu utama yang mencakup manfaat gizi anak, kebijakan dan anggaran, pelaksanaan di sekolah, keamanan makanan, pengawasan, operasional SPPG, serta stabilitas pangan. Hasil penelitian menunjukkan bahwa BERTopic mampu memetakan isu kebijakan publik berbasis teks pendek berbahasa Indonesia pada media sosial X secara representatif.

Kata kunci: BERTopic, topic modeling, Program Makan Bergizi Gratis, media sosial X, text mining

Abstract

The Free Nutritious Meal Program is one of the government policies widely discussed on social media X. The large number of posts makes manual issue identification inefficient; therefore, an automatic topic grouping method is needed. This study aims to implement BERTopic to map issues related to the Free Nutritious Meal Program on social media X. Data were collected through scraping using the keywords "Program Makan Bergizi Gratis" or "MBG" from January 6, 2025, to October 6, 2025. From 37,368 raw posts, 10,637 data remained after filtering. The research stages include text preprocessing, topic modeling using BERTopic, evaluation, visualization, and topic labeling. The best result was obtained with min_cluster_size 40, producing a coherence score of 0.740873, stability score of 0.753521, and final score of 0.747197. BERTopic generated 10 main issues, including child nutrition benefits, policy and budget, school implementation, food safety, supervision, SPPG operations, and food price stability. The results show that BERTopic can representatively map public policy issues from short Indonesian social media texts on X.

Keywords: BERTopic, topic modeling, Free Nutritious Meal Program, social media X, text mining

1. PENDAHULUAN

Program Makan Bergizi Gratis merupakan salah satu kebijakan prioritas nasional yang menjadi perhatian publik karena berkaitan langsung dengan peningkatan kualitas gizi pelajar dan penyiapan generasi unggul di masa depan [1]. Program ini dilaksanakan dengan melibatkan berbagai pihak, seperti sekolah, UMKM, serta instansi pemerintah dalam proses produksi dan pendistribusian makanan bergizi kepada masyarakat [2],[3]. Namun, seiring dengan implementasinya di berbagai daerah, program ini juga memunculkan berbagai respons masyarakat. Respons tersebut tidak hanya berkaitan dengan manfaat program, tetapi juga mencakup persoalan pelaksanaan di lapangan, seperti distribusi makanan, kualitas makanan, kesiapan dapur produksi, pengawasan, serta keamanan pangan. Beberapa pemberitaan mengungkapkan bahwa kasus keracunan makanan dalam pelaksanaan Program Makan Bergizi Gratis diduga disebabkan oleh kontaminasi bakteri, lemahnya standar operasional prosedur pada dapur produksi, serta keterbatasan pengawasan dalam skala produksi besar [4]. Secara nasional, tercatat sedikitnya 17 kasus keracunan Program Makan Bergizi Gratis yang terjadi di 10 provinsi, sehingga memicu perhatian publik terhadap efektivitas dan keamanan pelaksanaan program [5].

Perkembangan media sosial membuat respons masyarakat terhadap kebijakan pemerintah dapat diamati secara lebih cepat dan luas. Media sosial menjadi ruang bagi masyarakat untuk menyampaikan opini, membagikan informasi, serta mengikuti perkembangan isu yang sedang ramai dibicarakan. Berdasarkan laporan [6], jumlah pengguna aktif media sosial di Indonesia mencapai 143 juta pengguna atau sekitar 50,2% dari total populasi. Selain itu, laporan [7] menunjukkan bahwa pengguna media sosial X di Indonesia mencapai sekitar 25,16 juta pengguna. Jumlah tersebut menunjukkan tingginya keterlibatan masyarakat dalam aktivitas digital dan menjadikan media sosial X sebagai salah satu *platform* yang relevan untuk mengamati percakapan publik mengenai kebijakan pemerintah.

Media sosial X memiliki karakteristik berbasis teks singkat yang memungkinkan informasi dan opini menyebar dengan cepat. Fitur seperti balasan, *reply*, tagar, serta unggahan gambar membuat percakapan mengenai suatu isu dapat berkembang dalam waktu singkat [8],[9]. Dalam konteks Program Makan Bergizi Gratis, percakapan di X tidak hanya berisi informasi umum mengenai pelaksanaan program, tetapi juga mencerminkan berbagai isu yang menjadi perhatian masyarakat. Isu tersebut dapat berkaitan dengan manfaat program bagi anak, peran pemerintah, penggunaan anggaran, pelaksanaan di sekolah, operasional dapur, keterlibatan instansi, hingga kasus keracunan makanan. Banyaknya unggahan yang membahas Program Makan Bergizi Gratis membuat proses identifikasi isu secara manual menjadi kurang efisien. Selain itu, teks media sosial umumnya bersifat pendek, tidak terstruktur, dan memiliki variasi bahasa yang tinggi, sehingga diperlukan metode yang mampu mengelompokkan teks berdasarkan kesamaan makna.

Salah satu pendekatan yang dapat digunakan untuk mengidentifikasi isu dalam kumpulan teks adalah *topic modeling*. *Topic modeling* merupakan metode dalam *text mining* yang bertujuan menemukan tema atau topik utama dari sekumpulan

dokumen [10]. Metode ini dapat membantu merangkum percakapan dalam jumlah besar menjadi beberapa kelompok topik yang lebih mudah dipahami. Dalam konteks media sosial, *topic modeling* bermanfaat untuk mengetahui isu dominan yang muncul dalam percakapan masyarakat, sehingga hasil analisis tidak hanya menunjukkan banyaknya pembahasan, tetapi juga menggambarkan tema utama yang berkembang di ruang digital.

BERTopic merupakan salah satu metode *topic modeling* modern yang memanfaatkan representasi semantik berbasis *embedding* untuk mengelompokkan dokumen berdasarkan kedekatan makna [11]. BERTopic bekerja melalui beberapa tahapan, yaitu pembentukan representasi dokumen, reduksi dimensi menggunakan UMAP, pengelompokan dokumen menggunakan HDBSCAN, serta pembentukan representasi topik menggunakan *class-based TF-IDF* atau *c-TF-IDF* [11]. Melalui tahapan tersebut, BERTopic tidak hanya mengandalkan frekuensi kemunculan kata, tetapi juga mempertimbangkan kemiripan makna antardokumen. Kemampuan ini membuat BERTopic sesuai digunakan untuk menganalisis teks media sosial yang pendek, tidak terstruktur, dan memiliki konteks pembahasan yang beragam.

Beberapa penelitian sebelumnya menunjukkan bahwa BERTopic efektif digunakan untuk menganalisis topik pada data media sosial. Penelitian yang membandingkan LDA, NMF, Top2Vec, dan BERTopic menunjukkan bahwa BERTopic mampu menghasilkan topik yang lebih konsisten, representatif, dan mudah diinterpretasikan dibandingkan beberapa metode lainnya pada data Twitter [12], [13]. Penelitian lain mengenai kebijakan keanekaragaman hayati Indonesia pada media sosial X menunjukkan bahwa BERTopic dapat menghasilkan topik yang koheren dan informatif, dengan nilai koherensi tertinggi sebesar 0,7733 [14]. Di sisi lain, penelitian terdahulu mengenai Program Makan Bergizi Gratis di media sosial X sudah pernah dilakukan, namun fokusnya masih terbatas pada analisis sentimen menggunakan algoritma *machine learning* dan *deep learning* [15],[16]. Meskipun demikian, kajian khusus pemetaan isu Program Makan Bergizi Gratis (MBG) di media sosial X menggunakan BERTopic masih terbatas. Penelitian ini menawarkan kebaruan melalui tiga kontribusi spesifik. Pertama, menerapkan penyaringan ketat (*graph clustering*) untuk membuang ribuan data duplikat agar topik yang dihasilkan tidak bias. Kedua, melakukan pembersihan teks yang disesuaikan secara khusus dengan bahasa tidak baku dan singkatan. Ketiga, menghasilkan pemetaan isu yang dapat digunakan oleh pembuat kebijakan untuk mendeteksi masalah krusial di lapangan, seperti kasus keracunan makanan dan stabilitas harga pangan. Dengan pendekatan ini, BERTopic mampu memetakan percakapan publik mengenai Program MBG secara lebih bersih dan akurat.

Berdasarkan permasalahan tersebut, penelitian ini mengimplementasikan BERTopic untuk menganalisis isu Program Makan Bergizi Gratis pada media sosial X. Penelitian ini tidak bertujuan mengevaluasi pelaksanaan Program Makan Bergizi Gratis secara langsung, tetapi berfokus pada pemetaan isu yang muncul dalam percakapan masyarakat di media sosial X. Data yang digunakan berupa unggahan X dengan kata kunci “Program Makan Bergizi Gratis” atau “MBG” pada periode 6 Januari 2025 hingga 6 Oktober 2025. Dari proses pengumpulan data diperoleh 37.368 data mentah, kemudian setelah melalui proses *filtering* diperoleh 10.637

data yang digunakan untuk analisis lebih lanjut. Tujuan penelitian ini adalah mengimplementasikan BERTopic untuk mengelompokkan unggahan masyarakat ke dalam beberapa isu utama serta menginterpretasikan topik-topik dominan yang muncul terkait Program Makan Bergizi Gratis. Kontribusi penelitian ini adalah memberikan pemetaan isu publik mengenai Program Makan Bergizi Gratis berbasis data media sosial X, serta menunjukkan penerapan BERTopic dalam analisis kebijakan publik berbasis teks pendek berbahasa Indonesia.

2. METODOLOGI PENELITIAN

Penelitian ini menggunakan pendekatan *text mining* dengan metode *topic modeling* untuk mengidentifikasi isu-isu yang muncul dalam percakapan masyarakat mengenai Program Makan Bergizi Gratis pada media sosial X. Metode yang digunakan adalah BERTopic karena mampu mengelompokkan dokumen berdasarkan kemiripan makna dan menghasilkan representasi topik yang lebih kontekstual pada data teks pendek seperti unggahan media sosial [12], [11]. Tahapan penelitian meliputi pengumpulan data, *filtering* data, prapemrosesan teks, pemodelan topik menggunakan BERTopic, evaluasi topik, visualisasi, dan pelabelan topik.

2.1. Pengumpulan Data

Data yang digunakan dalam penelitian ini berupa unggahan dari media sosial X yang membahas Program Makan Bergizi Gratis. Pengumpulan data dilakukan menggunakan *Tweet Harvest* melalui *Google Colaboratory* dengan kata kunci “Program Makan Bergizi Gratis” atau “MBG”. Rentang waktu pengambilan data ditetapkan mulai 6 Januari 2025 hingga 6 Oktober 2025. Periode tersebut dipilih karena mencakup masa ketika Program Makan Bergizi Gratis mulai banyak dibahas di media sosial, baik terkait pelaksanaan program, manfaat, anggaran, distribusi, maupun isu keamanan makanan. Hasil proses pengumpulan data menghasilkan 37.368 data mentah. Data hasil *scraping* memuat beberapa atribut, seperti *conversation_id_str*, *created_at*, *favorite_count*, *full_text*, *id_str*, *image_url*, *reply_count*, *retweet_count*, *tweet_url*, *user_id_str*, dan *username*. Namun, penelitian ini hanya memanfaatkan atribut yang relevan dengan kebutuhan analisis, yaitu *created_at*, *full_text*, dan *image_url*. Atribut *created_at* digunakan untuk mengetahui waktu unggahan, *full_text* digunakan sebagai sumber utama analisis teks, sedangkan *image_url* digunakan untuk mengidentifikasi unggahan yang memuat gambar, khususnya gambar yang memuat teks. Ringkasan data yang digunakan dalam penelitian ini ditunjukkan pada Tabel 1.

Tabel 1. Ringkasan *Dataset* Penelitian

Keterangan	Deskripsi
Sumber data	Media sosial X
Kata kunci	“Program Makan Bergizi Gratis” atau “MBG”
Periode data	6 Januari 2025–6 Oktober 2025
Jumlah data mentah	37.368 unggahan
Jumlah data setelah <i>filtering</i>	10.637 unggahan
Atribut yang digunakan	<i>created_at</i> , <i>full_text</i> , <i>image_url</i>

2.2. Filtering Data

Tahap *filtering* data dilakukan untuk memastikan bahwa data yang digunakan sesuai dengan kebutuhan penelitian dan memiliki kualitas yang layak untuk dianalisis. Pada tahap awal, hanya kolom *created_at*, *full_text*, dan *image_url* yang dipertahankan, sedangkan kolom lain yang tidak digunakan dalam analisis dihapus. Selanjutnya, dilakukan pembersihan terhadap data yang tidak lengkap, seperti data tanpa teks atau data yang tidak sesuai dengan kebutuhan penelitian. Tahap berikutnya dilakukan dengan menghapus unggahan yang mengandung unsur *spam* atau promosi yang tidak relevan, seperti kata “promo”, “diskon”, “voucher”, “*flash sale*”, dan istilah sejenis. Proses ini dilakukan agar data yang dianalisis benar-benar berkaitan dengan pembahasan Program Makan Bergizi Gratis, bukan unggahan iklan atau konten promosi. Selain itu, dilakukan penghapusan data duplikat dan data yang memiliki kemiripan teks sangat tinggi. Pengukuran kemiripan dilakukan menggunakan pendekatan TF-IDF dan *cosine similarity*. Teks dengan nilai kemiripan lebih dari 0,85 dianggap terlalu mirip, kemudian dikelompokkan menggunakan *graph clustering* agar hanya satu teks yang dipertahankan dari kelompok teks yang serupa. Setelah seluruh proses *filtering* selesai, jumlah data yang digunakan dalam penelitian berkurang dari 37.368 data mentah menjadi 10.637 data.

2.3. Prapemrosesan Teks

Prapemrosesan teks dilakukan untuk menyiapkan data agar lebih bersih, seragam, dan siap digunakan dalam pemodelan topik. Data media sosial umumnya memiliki karakteristik tidak terstruktur, seperti mengandung tautan, tanda baca, angka, simbol, tagar, *mention*, singkatan, serta variasi penulisan tidak baku. Oleh karena itu, prapemrosesan teks diperlukan agar model dapat mengenali pola pembahasan secara lebih optimal. Tahapan prapemrosesan yang dilakukan meliputi *cleaning*, *case folding*, *normalization*, dan *stopword removal*. Pada tahap *cleaning*, teks dibersihkan dari URL, *mention*, tagar, simbol HTML, angka, tanda baca, dan karakter yang tidak relevan. Tahap *case folding* dilakukan dengan mengubah seluruh huruf menjadi huruf kecil agar kata yang sama tidak dianggap berbeda hanya karena perbedaan kapitalisasi. Selanjutnya, tahap *normalization* dilakukan untuk mengubah kata tidak baku atau singkatan menjadi bentuk baku, misalnya kata “gak” menjadi “tidak” dan “tp” menjadi “tapi”. Tahap terakhir adalah *stopword removal*, yaitu menghapus kata-kata umum yang kurang memberikan makna penting dalam analisis, seperti “yang”, “dan”, “di”, serta kata hubung lainnya. Hasil dari tahap prapemrosesan berupa teks bersih yang digunakan sebagai masukan dalam pemodelan topik BERTopic. Dengan tahapan ini, data yang digunakan menjadi lebih ringkas dan lebih fokus pada kata-kata yang merepresentasikan isi pembahasan.

2.4. Pemodelan Topik Menggunakan BERTopic

Pemodelan topik dilakukan menggunakan BERTopic untuk mengelompokkan unggahan media sosial X ke dalam beberapa isu utama berdasarkan kemiripan makna. BERTopic dipilih karena metode ini mampu memanfaatkan representasi semantik berbasis *embedding*, sehingga tidak hanya bergantung pada frekuensi kemunculan kata, tetapi juga mempertimbangkan kedekatan makna antardokumen [12]. Tahapan pemodelan BERTopic dalam penelitian ini terdiri atas empat proses utama, yaitu *document embedding*, reduksi dimensi, *document clustering*, dan

representasi topik. Pada tahap *document embedding*, setiap teks diubah menjadi representasi vektor menggunakan model paraphrase-multilingual-MiniLM-L12-v2. Model ini digunakan karena mendukung pemrosesan teks multibahasa, termasuk bahasa Indonesia, serta mampu merepresentasikan makna dokumen dalam bentuk numerik.

Setelah dokumen direpresentasikan dalam bentuk vektor, tahap berikutnya adalah reduksi dimensi menggunakan UMAP. Tahap ini bertujuan menyederhanakan representasi vektor berdimensi tinggi agar lebih mudah dikelompokkan, tetapi tetap mempertahankan hubungan kemiripan antar dokumen. Selanjutnya, proses *clustering* dilakukan menggunakan HDBSCAN untuk mengelompokkan dokumen yang memiliki kesamaan makna ke dalam topik tertentu. Dokumen yang tidak memiliki kedekatan makna dengan kelompok mana pun dikategorikan sebagai *outlier*. Setelah klaster terbentuk, representasi topik dilakukan menggunakan *class-based Term Frequency-Inverse Document Frequency* atau *c-TF-IDF*. Metode ini digunakan untuk menentukan kata-kata yang paling mewakili setiap topik. Kata dengan bobot tertinggi pada setiap topik kemudian digunakan sebagai dasar dalam proses interpretasi dan pelabelan topik. Konfigurasi utama BERTopic yang digunakan dalam penelitian ini ditunjukkan pada Tabel 2.

Tabel 2. Konfigurasi BERTopic

Komponen	Metode/Model
Model <i>embedding</i>	paraphrase-multilingual-MiniLM-L12-v2
Reduksi dimensi	UMAP
Algoritma <i>clustering</i>	HDBSCAN
Representasi topik	<i>c-TF-IDF</i>
Pengujian parameter	<i>min_cluster_size</i>
Pengujian jumlah topik	5 sampai 20 topik
Evaluasi	<i>Coherence, stability, dan final score</i>

2.5. Evaluasi Topik

Evaluasi topik dilakukan untuk menentukan konfigurasi BERTopic yang paling sesuai dengan karakteristik data. Evaluasi dilakukan menggunakan nilai *coherence*, *stability*, dan *final score*. Nilai *coherence* yang digunakan dalam penelitian ini adalah jenis Coherence Value (C_v) untuk mengukur keeratan hubungan makna antar kata utama dalam satu topik. Sementara itu, nilai *stability* dihitung menggunakan indeks kemiripan *Jaccard* (*Jaccard similarity*) untuk menilai konsistensi topik yang dihasilkan dari beberapa kali percobaan model. *Final score* dihitung dari rata-rata nilai *coherence* dan *stability*, sehingga konfigurasi terbaik dipilih berdasarkan keseimbangan antara kualitas makna topik dan kestabilan hasil pemodelan. Evaluasi dilakukan dalam dua tahap. Tahap pertama adalah pengujian beberapa nilai *min_cluster_size* pada HDBSCAN. Parameter ini digunakan untuk menentukan jumlah minimum dokumen yang dapat membentuk satu klaster. Nilai yang terlalu kecil dapat menghasilkan jumlah topik terlalu banyak dan sulit diinterpretasikan, sedangkan nilai yang terlalu besar dapat membuat topik menjadi terlalu umum. Oleh karena itu, beberapa nilai *min_cluster_size* diuji untuk memperoleh konfigurasi yang menghasilkan jumlah topik, jumlah *outlier*, *coherence*, dan *stability* yang paling

seimbang. Tahap kedua adalah pengujian jumlah topik akhir pada rentang 5 hingga 20 topik. Pengujian ini dilakukan agar topik yang dihasilkan tidak terlalu sedikit sehingga menjadi terlalu umum, tetapi juga tidak terlalu banyak sehingga sulit diinterpretasikan. Konfigurasi akhir dipilih berdasarkan nilai *final score* tertinggi.

2.6. Visualisasi dan Pelabelan Topik

Setelah model BERTopic terbaik diperoleh, hasil pemodelan divisualisasikan untuk mempermudah proses interpretasi. Visualisasi yang digunakan meliputi informasi jumlah dokumen pada setiap topik, *intertopic distance map*, dan *topic word scores*. Informasi jumlah dokumen digunakan untuk melihat dominasi masing-masing topik, *intertopic distance map* digunakan untuk melihat hubungan atau kedekatan antar topik, sedangkan *topic word scores* digunakan untuk mengetahui kata-kata dengan bobot tertinggi pada setiap topik. Pelabelan topik dilakukan secara manual berdasarkan kata-kata representatif yang dihasilkan oleh BERTopic dan contoh dokumen pada setiap kluster. Topik yang memiliki kata kunci dengan makna serupa diberi label sesuai dengan isu yang paling menggambarkan isi pembahasan. Sementara itu, topik dengan nomor -1 dikategorikan sebagai *outlier* karena berisi dokumen yang tidak memiliki kedekatan makna yang cukup kuat dengan kelompok topik tertentu. Oleh karena itu, *outlier* tidak digunakan sebagai topik utama dalam proses interpretasi.

2.7. Penyimpanan Model

Tahap akhir dari metodologi ini adalah penyimpanan model BERTopic. Model yang disimpan merupakan model terbaik yang telah melalui proses evaluasi dan digunakan dalam pembentukan topik akhir. Penyimpanan model dilakukan agar model dapat digunakan kembali pada data baru tanpa perlu melakukan proses pelatihan ulang. Dengan demikian, hasil pemodelan dapat dimanfaatkan kembali untuk analisis lanjutan maupun pengembangan sistem analisis berbasis web.

3. HASIL DAN PEMBAHASAN

Bagian ini menjelaskan hasil implementasi BERTopic dalam mengidentifikasi isu-isu yang muncul pada percakapan pengguna media sosial X mengenai Program Makan Bergizi Gratis. Hasil yang dibahas meliputi hasil pengumpulan dan *filtering* data, evaluasi konfigurasi BERTopic, hasil pembentukan topik, interpretasi isu, serta pembahasan terhadap makna isu yang ditemukan. Sesuai dengan format jurnal, hasil penelitian disajikan dalam bentuk tabel dan pembahasan agar lebih mudah dipahami pembaca.

3.1. Hasil Pengumpulan dan *Filtering* Data

Data penelitian diperoleh dari media sosial X menggunakan kata kunci “Program Makan Bergizi Gratis” atau “MBG” pada periode 6 Januari 2025 hingga 6 Oktober 2025. Proses pengumpulan data menghasilkan 37.368 data mentah. Data hasil *scraping* memuat beberapa atribut, seperti *created_at*, *full_text*, *image_url*, *username*, *tweet_url*, *reply_count*, *retweet_count*, dan atribut lainnya. Namun, penelitian ini hanya menggunakan atribut yang relevan dengan kebutuhan analisis, yaitu *created_at*, *full_text*, dan *image_url*. Tahap *filtering* dilakukan untuk memastikan bahwa data yang digunakan memiliki kualitas yang sesuai untuk pemodelan topik. Proses ini meliputi pemilihan atribut utama, penghapusan data

kosong, penghapusan konten *spam*, penghapusan data duplikat, serta penghapusan unggahan yang memiliki tingkat kemiripan teks sangat tinggi. Pengukuran kemiripan teks dilakukan menggunakan TF-IDF dan cosine similarity dengan ambang batas 0,85. Teks yang memiliki kemiripan di atas ambang batas tersebut dikelompokkan menggunakan graph clustering melalui algoritma connected components. Dalam proses ini, setiap dokumen dipetakan sebagai titik dan hubungan kemiripan antar-dokumen sebagai garis hubungannya. Untuk menyaring data, dari setiap kelompok teks yang serupa hanya diambil satu dokumen yang memiliki koneksi terbanyak karena dokumen tersebut terbukti paling representatif dalam mewakili keseluruhan isi percakapan. Teks dengan skor kemiripan lebih dari 0,85 dianggap terlalu mirip, kemudian dikelompokkan agar hanya satu teks yang dipertahankan. Setelah seluruh proses *filtering* selesai, jumlah data berkurang dari 37.368 data mentah menjadi 10.637 data.

Tabel 3. Hasil Pengumpulan dan *Filtering* Data

Tahap Pengolahan Data	Jumlah Data
Data hasil <i>scraping</i>	37.368
Data setelah <i>filtering</i>	10.637

Berdasarkan Tabel 3, proses *filtering* berperan penting dalam mengurangi data yang tidak relevan dan berulang. Pengurangan data tidak hanya menunjukkan adanya data kosong atau *spam*, tetapi juga memperlihatkan bahwa percakapan di media sosial X sering memuat unggahan yang sama atau sangat mirip, misalnya unggahan berita, *repost*, atau teks promosi yang berulang. Dengan menghapus data tersebut, *dataset* yang digunakan dalam pemodelan menjadi lebih bersih dan lebih mampu merepresentasikan variasi percakapan publik mengenai Program Makan Bergizi Gratis.

3.2. Hasil Evaluasi Konfigurasi BERTopic

Pemodelan topik dilakukan menggunakan BERTopic melalui beberapa tahapan, yaitu *document embedding*, reduksi dimensi, *clustering*, dan representasi topik. Pada tahap *document embedding*, setiap teks diubah menjadi representasi vektor menggunakan model paraphrase-multilingual-MiniLM-L12-v2. Model ini digunakan karena mendukung pemrosesan teks multibahasa, termasuk bahasa Indonesia, serta mampu merepresentasikan makna dokumen dalam bentuk numerik. Setelah itu, UMAP digunakan untuk mereduksi dimensi, HDBSCAN digunakan untuk membentuk kluster topik, dan *c-TF-IDF* digunakan untuk menghasilkan kata-kata representatif pada setiap topik. Evaluasi BERTopic dilakukan dalam dua tahap. Tahap pertama adalah pengujian beberapa nilai *min_cluster_size* untuk menentukan konfigurasi HDBSCAN yang paling sesuai dengan karakteristik data. Parameter ini berpengaruh terhadap jumlah minimum dokumen dalam sebuah kluster. Nilai yang terlalu kecil dapat menghasilkan jumlah topik yang terlalu banyak dan sulit diinterpretasikan, sedangkan nilai yang terlalu besar dapat membuat beberapa isu berbeda tergabung ke dalam topik yang terlalu umum. Berdasarkan hasil pengujian, konfigurasi terbaik diperoleh pada nilai *min_cluster_size* sebesar 40. Konfigurasi ini menghasilkan 34 topik awal dengan jumlah *outlier* sebanyak 2.025 data atau 21,15%. Nilai *coherence* yang diperoleh sebesar 0,705934, nilai *stability* sebesar

0,562666, dan *final score* sebesar 0,634300. Hasil tersebut menunjukkan bahwa nilai *min_cluster_size* 40 mampu memberikan keseimbangan antara jumlah topik, keterkaitan kata dalam topik, kestabilan hasil, dan jumlah *outlier*.

Tabel 4. Hasil Evaluasi Min Cluster Terbaik

Parameter	Hasil
<i>min_cluster_size</i> terbaik	40
Jumlah topik awal	34
Jumlah <i>outlier</i>	2.025
Persentase <i>outlier</i>	21,15%
<i>Coherence</i>	0,705934
<i>Stability</i>	0,562666
<i>Final score</i>	0,634300

Setelah nilai *min_cluster_size* terbaik diperoleh, evaluasi dilanjutkan dengan menguji jumlah topik akhir pada rentang 5 hingga 20 topik. Pengujian ini dilakukan untuk menghasilkan topik yang lebih ringkas dan mudah diinterpretasikan tanpa menghilangkan variasi isu utama dalam data. Jumlah topik yang terlalu sedikit berpotensi membuat beberapa isu berbeda tergabung menjadi satu, sedangkan jumlah topik yang terlalu banyak dapat menyulitkan proses interpretasi dan pelabelan. Hasil terbaik diperoleh pada *target_nr_topics* = 11, yang menghasilkan 10 topik aktual di luar *outlier*. Konfigurasi ini memperoleh nilai *coherence* sebesar 0,740873, *stability* sebesar 0,753521, dan *final score* sebesar 0,747197. Nilai tersebut menunjukkan bahwa hasil akhir memiliki keterkaitan kata yang baik dan kestabilan topik yang cukup kuat. Oleh karena itu, 10 topik aktual digunakan sebagai hasil akhir dalam proses interpretasi isu.

Tabel 5. Hasil Evaluasi Jumlah Topik Akhir

Parameter	Hasil
<i>target_nr_topics</i> terbaik	11
Jumlah topik aktual	10
Jumlah <i>outlier</i>	2.025
Persentase <i>outlier</i>	21,15%
<i>Coherence</i>	0,740873
<i>Stability</i>	0,753521
<i>Final score</i>	0,747197

Berdasarkan Tabel 5, penyesuaian jumlah topik berhasil meningkatkan kualitas hasil pemodelan. Nilai *coherence* meningkat dari 0,705934 menjadi 0,740873, sedangkan nilai *stability* meningkat dari 0,562666 menjadi 0,753521. Peningkatan ini menunjukkan bahwa pengurangan jumlah topik membuat kata-kata dalam setiap topik menjadi lebih saling berkaitan dan hasil topik menjadi lebih stabil. Dengan demikian, konfigurasi akhir BERTopic dinilai layak digunakan untuk memetakan isu Program Makan Bergizi Gratis pada media sosial X. Persentase *outlier* sebesar 21,15% menunjukkan bahwa sebagian unggahan tidak memiliki kedekatan makna yang cukup kuat dengan klaster utama. Kondisi ini wajar pada data media sosial karena unggahan di X cenderung pendek, tidak terstruktur, mengandung variasi bahasa, serta sering memuat konteks implisit. *Outlier* tidak

digunakan sebagai topik utama dalam interpretasi agar hasil pembahasan lebih fokus pada kelompok isu yang memiliki pola makna lebih jelas.

3.3. Hasil Visualisasi Topik

Hasil pemodelan BERTopic divisualisasikan untuk membantu memahami sebaran topik dan hubungan antar kelompok isu. Visualisasi yang digunakan meliputi informasi topik, *intertopic distance map*, dan *topic word scores*. Informasi topik digunakan untuk melihat jumlah dokumen pada setiap topik, *intertopic distance map* digunakan untuk menunjukkan kedekatan antar topik, sedangkan *topic word scores* digunakan untuk melihat kata-kata yang paling merepresentasikan setiap topik.

	Topic	Count	Name	Representation	Representative_Docs
0	-1	2025	-1_gizi_spgg_dapur_sekolah	[gizi, spgg, dapur, sekolah, makanan, satuan, ...]	[badan gizi nasional bgn mendistribusikan maka...
1	0	2642	0_anak_sehat_gizi_anak	[anak, sehat, gizi, anak, generasi, ekono...	[dukung program makan bergizi gratis anak anak...
2	1	2096	1_indonesia_anak_anak indonesia papua	[indonesia, anak, anak indonesia, papua, gener...	[program makan bergizi gratis bantu anak anak ...]
3	2	995	2_presiden_pemerintah_prabowo_presiden prabowo	[presiden, pemerintah, prabowo, presiden prabo...	[pemerintah menaikkan anggaran program makan b...
4	3	938	3_sekolah_babinsa_siswa_koramil	[sekolah, babinsa, siswa, koramil, bhabinkamti...	[babinsa koramil spong pendampingan makan bergi...
5	4	453	4_keracunan_siswa_makanan_mengalami	[keracunan, siswa, makanan, mengalami, diduga...	[belasan siswa smk pembangunan bandung barat c...
6	5	120	5_kepala_intelijen_rapat_kejaksanaan	[kepala, intelijen, rapat, kejaksaan, negeri, ...]	[kamis januari kejaksaan negeri aceh singkil m...
7	6	117	6_hamil_menyusui_hamil menyusui_balita	[hamil, menyusui, hamil menyusui, balita, anak...	[program makan bergizi gratis hamil menyusui b...
8	7	70	7_polri_maksimal_ratusan_spgg	[polri, maksimal, ratusan, spgg, pelayanan, pe...	[polri mendirikan ratusan dapur satuan pelayan...
9	8	60	8_dapur_dapur_spgg_spgg_dapur sehat	[dapur, dapur spgg, spgg, dapur sehat, tinjau...	[kapolres sukoharjo tinjau lokasi rencana dapu...
10	9	57	9_tito_karnavian_karnavian_tito_sphp	[tito karnavian, karnavian, tito, sphp, beras...	[makan bergizi gratis kuat tito karnavian jaga...

Gambar 1. Informasi Topik Hasil BERTopic

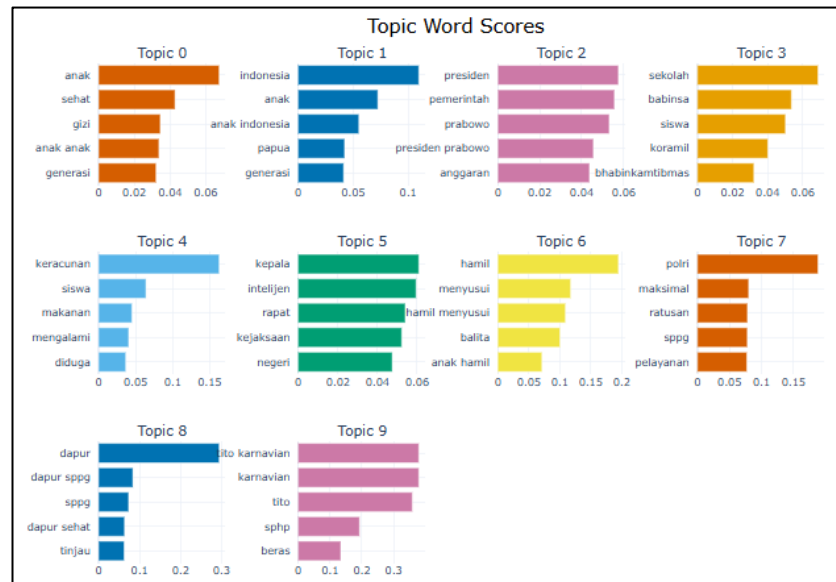
Gambar 1 menunjukkan informasi topik hasil pemodelan BERTopic. Berdasarkan hasil tersebut, Topik -1 dikategorikan sebagai *outlier* dengan jumlah 2.025 dokumen. Topik terbesar adalah Topik 0 dengan 2.642 dokumen, diikuti Topik 1 dengan 2.096 dokumen. Dominasi kedua topik tersebut menunjukkan bahwa pembahasan mengenai pemenuhan gizi, kesehatan anak, serta program gizi untuk anak Indonesia menjadi isu yang paling banyak muncul dalam percakapan publik mengenai Program Makan Bergizi Gratis.



Gambar 2. Intertopic distance map

Gambar 2 menunjukkan hubungan antar topik yang terbentuk. Setiap lingkaran merepresentasikan satu topik, sedangkan ukuran lingkaran menunjukkan jumlah

dokumen dalam topik tersebut. Semakin besar lingkaran, semakin banyak dokumen yang tergabung dalam topik tersebut. Jarak antar lingkaran menunjukkan tingkat kemiripan pembahasan. Topik yang berdekatan memiliki isi pembahasan yang relatif mirip, sedangkan topik yang berjauhan menunjukkan tema yang lebih berbeda.



Gambar 3. Topic word scores

Gambar 3 menunjukkan kata-kata dengan bobot tertinggi pada setiap topik. Kata-kata tersebut digunakan sebagai dasar dalam proses interpretasi dan pelabelan topik. Misalnya, Topik 0 didominasi oleh kata “anak”, “sehat”, “gizi”, “generasi”, dan “ekonomi”, sehingga topik ini diarahkan pada isu pemenuhan gizi dan kesehatan anak. Sementara itu, Topik 4 ditandai oleh kata “keracunan”, “siswa”, “makanan”, dan “korban”, sehingga topik ini diarahkan pada isu kasus keracunan makanan siswa.

3.4. Hasil Interpretasi Topik

Setelah konfigurasi terbaik diperoleh, dilakukan interpretasi terhadap topik yang terbentuk. Interpretasi dilakukan berdasarkan kata kunci utama dari *c-TF-IDF*, contoh dokumen dalam setiap klaster, serta konteks pembahasan Program Makan Bergizi Gratis. Topik -1 tidak dimasukkan ke dalam interpretasi utama karena merupakan *outlier*. Hasil interpretasi 10 topik utama ditampilkan pada Tabel 6.

Tabel 6. Interpretasi Topic BERTopic

Topik	Kata Kunci Utama	Label Isu
0	anak, sehat, gizi, anak-anak, generasi, ekonomi, dukung, tumbuh, lokal, bangsa	Pemenuhan Gizi dan Kesehatan Anak
1	indonesia, anak, anak indonesia, papua, generasi, sehat, emas, gizi, indonesia emas	Program Gizi untuk Anak Indonesia
2	presiden, pemerintah, prabowo, presiden prabowo, anggaran, subianto, rp, triliun	Kebijakan dan Anggaran Program

3	sekolah, babinsa, siswa, koramil, bhabinkamtibmas, kegiatan, monitoring, wilayah	Pelaksanaan Program di Sekolah
4	keracunan, siswa, makanan, mengalami, diduga, menyantap, korban, massal	Kasus Keracunan Makanan Siswa
5	kepala, intelijen, rapat, kejaksaan, negeri, staf, koordinasi, rapat koordinasi	Koordinasi dan Pengawasan Program
6	hamil, menyusui, balita, anak hamil, balita hamil, sekolah hamil	Penerima Manfaat Ibu Hamil dan Balita
7	polri, maksimal, ratusan, sppg, pelayanan, pengabdian, mendukung, memperluas	Dukungan Polri terhadap SPPG
8	dapur, dapur sppg, sppg, dapur sehat, tinjau, jateng, dandim, lahan	Operasional Dapur SPPG
9	tito karnavian, karnavian, tito, sphp, beras, pemda, harga, inflasi	Stabilitas Harga Beras dan Pangan

Berdasarkan Tabel 6, isu yang muncul dalam percakapan masyarakat cukup beragam. Topik yang terbentuk tidak hanya menggambarkan manfaat Program Makan Bergizi Gratis, tetapi juga menunjukkan perhatian terhadap kebijakan, anggaran, pelaksanaan di sekolah, pengawasan, keamanan makanan, operasional dapur, serta stabilitas pangan. Dengan demikian, hasil pemodelan BERTopic menunjukkan bahwa percakapan publik mengenai Program Makan Bergizi Gratis memiliki cakupan isu yang luas.

3.5. Pembahasan Hasil Pemodelan Topik

Hasil pemodelan menggunakan BERTopic menunjukkan bahwa percakapan masyarakat mengenai Program Makan Bergizi Gratis pada media sosial X terbagi ke dalam 10 isu utama. Isu yang terbentuk tidak hanya menggambarkan manfaat program, tetapi juga mencakup aspek kebijakan, pelaksanaan, pengawasan, keamanan makanan, operasional dapur, serta stabilitas pangan. Hal ini menunjukkan bahwa pembahasan publik terhadap Program Makan Bergizi Gratis memiliki cakupan yang luas dan tidak hanya berpusat pada tujuan program. Topik yang paling dominan berkaitan dengan pemenuhan gizi dan kesehatan anak serta program gizi untuk anak Indonesia. Dominasi topik ini menunjukkan bahwa narasi manfaat program masih menjadi perhatian utama dalam percakapan publik. Program Makan Bergizi Gratis banyak dikaitkan dengan peningkatan kualitas kesehatan anak, pemerataan manfaat, serta harapan terhadap generasi Indonesia yang lebih sehat. Munculnya isu seperti “Papua” dan “Indonesia emas” juga menunjukkan bahwa sebagian masyarakat menghubungkan program ini dengan pemerataan pembangunan dan tujuan jangka panjang peningkatan sumber daya manusia.

Selain manfaat program, hasil pemodelan juga menunjukkan adanya perhatian terhadap kebijakan dan anggaran. Topik mengenai pemerintah, Presiden Prabowo, serta anggaran menunjukkan bahwa masyarakat tidak hanya membahas tujuan sosial program, tetapi juga memperhatikan aspek tata kelola dan pembiayaan. Hal ini memperlihatkan bahwa program berskala nasional dinilai bukan hanya dari

manfaatnya, tetapi juga dari kesiapan pelaksanaan, transparansi, dan efektivitas penggunaan anggaran.

Isu pelaksanaan program di sekolah juga menjadi salah satu topik penting. Munculnya pembahasan mengenai sekolah, siswa, monitoring, dan keterlibatan aparat wilayah menunjukkan bahwa masyarakat memperhatikan bagaimana program dijalankan di lapangan. Topik ini memperlihatkan bahwa pelaksanaan Program Makan Bergizi Gratis dipahami sebagai kegiatan lintas sektor yang melibatkan sekolah, pemerintah daerah, dan berbagai instansi pendukung. Temuan penting lainnya adalah munculnya topik mengenai kasus keracunan makanan siswa. Isu ini menunjukkan bahwa keamanan pangan menjadi perhatian kritis dalam percakapan publik. Pembahasan mengenai keracunan, korban, dan makanan memperlihatkan adanya kekhawatiran masyarakat terhadap kualitas pengolahan dan distribusi makanan. Dengan demikian, BERTopic mampu menangkap isu sensitif yang berhubungan langsung dengan kepercayaan publik terhadap pelaksanaan program.

Selain itu, topik mengenai koordinasi, pengawasan, dukungan Polri terhadap SPPG, dan operasional dapur SPPG menunjukkan bahwa aspek teknis pelaksanaan juga menjadi bagian penting dalam diskusi publik. Masyarakat tidak hanya membahas hasil program, tetapi juga memperhatikan proses penyediaan makanan, kesiapan dapur, layanan, serta peran lembaga pendukung. Isu ini penting karena keberhasilan program makan berskala besar sangat bergantung pada kesiapan sistem produksi, distribusi, dan pengawasan. Topik mengenai stabilitas harga beras dan pangan menunjukkan bahwa Program Makan Bergizi Gratis juga dikaitkan dengan isu ekonomi pangan. Hal ini menunjukkan bahwa masyarakat memahami program ini tidak hanya sebagai kebijakan sosial dan kesehatan, tetapi juga sebagai program yang berhubungan dengan kebutuhan pasokan bahan pangan, harga, dan inflasi. Oleh karena itu, pelaksanaan program perlu dilihat sebagai kebijakan yang memiliki keterkaitan dengan sektor pangan dan ekonomi.

Hasil ini sejalan dengan penelitian [12] yang menunjukkan bahwa BERTopic sesuai digunakan untuk menganalisis teks pendek dan tidak terstruktur pada media sosial. Temuan ini juga mendukung penelitian [14] yang menunjukkan bahwa konfigurasi BERTopic pada teks berbahasa Indonesia perlu diuji agar menghasilkan topik yang koheren dan informatif. Dengan nilai *coherence* sebesar 0,740873 dan *stability* sebesar 0,753521, hasil penelitian ini menunjukkan bahwa BERTopic mampu menghasilkan pemetaan isu yang cukup representatif terhadap percakapan publik mengenai Program Makan Bergizi Gratis pada media sosial X. Secara keseluruhan, tujuan penelitian untuk mengimplementasikan BERTopic dalam mengidentifikasi isu Program Makan Bergizi Gratis pada media sosial X telah tercapai. Sepuluh topik yang dihasilkan menunjukkan bahwa percakapan publik mencakup isu manfaat gizi, kebijakan dan anggaran, pelaksanaan di sekolah, keamanan makanan, pengawasan, operasional SPPG, serta stabilitas pangan. Dengan demikian, BERTopic dapat digunakan sebagai pendekatan yang efektif untuk memetakan isu kebijakan publik berbasis teks pendek di media sosial.

3.6. Keterbatasan Hasil

Meskipun hasil pemodelan menunjukkan bahwa BERTopic mampu menghasilkan topik yang representatif, terdapat beberapa keterbatasan yang perlu diperhatikan. Pertama, hasil topik sangat dipengaruhi oleh kualitas data dan proses prapemrosesan. Data media sosial X memiliki karakteristik teks pendek, tidak terstruktur, serta sering mengandung bahasa tidak baku, sehingga masih memungkinkan adanya informasi yang tidak sepenuhnya tertangkap oleh model. Kedua, proses pelabelan topik masih dilakukan secara manual berdasarkan kata kunci dan konteks dokumen dalam setiap kluster. Hal ini membuat interpretasi topik tetap memerlukan pemahaman peneliti terhadap konteks Program Makan Bergizi Gratis. Ketiga, data penelitian hanya berasal dari media sosial X, sehingga hasil pemodelan belum dapat merepresentasikan seluruh opini atau percakapan publik di *platform* lain. Oleh karena itu, penelitian selanjutnya dapat memperluas sumber data, membandingkan beberapa model *embedding*, serta melibatkan lebih dari satu penilai dalam proses pelabelan topik agar hasil interpretasi menjadi lebih objektif.

KESIMPULAN

Berdasarkan hasil penelitian, tujuan untuk mengimplementasikan BERTopic dalam mengidentifikasi isu Program Makan Bergizi Gratis pada media sosial X telah tercapai. Dari 37.368 data mentah yang dikumpulkan melalui kata kunci “Program Makan Bergizi Gratis” atau “MBG” pada periode 6 Januari 2025 hingga 6 Oktober 2025, diperoleh 10.637 data setelah melalui tahap *filtering*. Data tersebut kemudian digunakan untuk pemodelan topik menggunakan BERTopic. Hasil pemodelan menunjukkan bahwa BERTopic mampu menghasilkan 10 topik utama setelah data yang tidak masuk ke kluster utama dikategorikan sebagai *outlier*. Konfigurasi terbaik diperoleh pada nilai *min_cluster_size* 40 dengan nilai *coherence* sebesar 0,740873, *stability* sebesar 0,753521, dan *final score* sebesar 0,747197. Nilai tersebut menunjukkan bahwa topik yang terbentuk memiliki keterkaitan kata dan kestabilan yang cukup baik untuk merepresentasikan isu dalam percakapan publik.

Isu yang ditemukan dapat dikelompokkan ke dalam beberapa bahasan utama, yaitu manfaat gizi dan kesehatan anak, sasaran penerima manfaat, kebijakan dan anggaran, pelaksanaan program di sekolah, keamanan makanan, koordinasi dan pengawasan, operasional SPPG, serta stabilitas pangan. Temuan ini menunjukkan bahwa percakapan masyarakat mengenai Program Makan Bergizi Gratis tidak hanya berfokus pada manfaat program, tetapi juga mencakup aspek tata kelola, teknis pelaksanaan, risiko keamanan pangan, dan dampak terhadap sektor pangan.

Dengan demikian, penelitian ini menunjukkan bahwa BERTopic dapat digunakan sebagai metode untuk memetakan isu kebijakan publik berbasis teks pendek berbahasa Indonesia pada media sosial X. Hasil penelitian ini terbatas pada percakapan di media sosial X, sehingga belum merepresentasikan seluruh respons publik di *platform* lain. Penelitian selanjutnya dapat memperluas sumber data, membandingkan model *embedding*, serta melibatkan lebih dari satu penilai dalam proses pelabelan topik agar interpretasi isu menjadi lebih objektif.

DAFTAR PUSTAKA

- [1] www.presidentri.go.id, "Presiden Prabowo: Program Makan Bergizi Gratis Adalah Investasi untuk Masa Depan Bangsa," 15 Oktober. Diakses: 18 Oktober 2025. [Daring]. Tersedia pada: <https://www.presidentri.go.id/siaran-pers/president-prabowo-program-makan-bergizi-gratis-adalah-investasi-untuk-masa-depan-bangsa/>
- [2] Albaburrahim, A. P. A. Putikadyanto, A. N. Efendi, M. A. Alatas, S. Romadhon, dan L. R. Wachidah, "Program Makan Bergizi Gratis: Analisis Kritis Transformasi Pendidikan Indonesia Menuju Generasi Emas 2045," *Entita J. Pendidik. Ilmu Pengetah. Sos. dan Ilmu-Ilmu Sos.*, hal. 767–780, 2025, doi: 10.19105/ejpis.v1i.19191.
- [3] I. W. Lendra, D. Husni, dan Y. Fitriani, "Kebijakan Makan Bergizi Gratis dan Relevansinya terhadap Nilai-nilai Good Governance: Analisis Kualitatif dalam Administrasi Publik," *Arus J. Sos. dan Hum.*, vol. 5, no. 1, 2025.
- [4] news.detik.com, "Ini Dugaan Penyebab Ratusan Siswa Keracunan MBG di Bandung Barat," 23 September. Diakses: 18 Oktober 2025. [Daring]. Tersedia pada: <https://news.detik.com/berita/d-8126149/ini-dugaan-penyebab-ratusan-siswa-keracunan-mbg-di-bandung-barat>
- [5] www.metrotvnews.com, "17 Kasus Keracunan MBG di 10 Provinsi: Kontaminasi, SOP Lemah, hingga Perlu Evaluasi Dapur," 15 Mei. Diakses: 18 Oktober 2025. [Daring]. Tersedia pada: <https://www.metrotvnews.com/read/NA0CEzqA-17-kasus-keracunan-mbg-di-10-provinsi-kontaminasi-sop-lemah-hingga-perlu-evaluasi-dapur>
- [6] We Are Social, "Digital 2025 Indonesia the Essential Guide To Digital Trends," *Weare.Social*, hal. 89–90, 2025, [Daring]. Tersedia pada: <https://wearesocial.com/id/blog/2025/02/digital-2025/>
- [7] S. Kemp, *DIGITAL 2025 Global Overview Report*. Meltwater, 2025.
- [8] D. M. Boyd dan N. B. Ellison, "Social Network Sites : Definition , History , and Scholarship," *J. Comput. Commun.*, vol. 13, hal. 210–230, 2008, doi: 10.1111/j.1083-6101.2007.00393.x.
- [9] R. Nikmah dan E. Rasyid, "Public Perception of Islamic Higher Education Institutions on Twitter in Indonesia : A Social Media Sentiment Analysis," *CHANNEL J. Komun.*, vol. 11, no. 2, hal. 129–136, 2023.
- [10] R. Churchill dan L. Singh, "The Evolution of Topic Modeling," *ACM Comput. Surv.*, vol. 54, no. 10, 2022, doi: 10.1145/3507900.
- [11] M. Grootendorst, "BERTopic: Neural topic modeling with a class-based TF-IDF procedure," 2022, [Daring]. Tersedia pada: <http://arxiv.org/abs/2203.05794>
- [12] R. Egger dan J. Yu, "A Topic Modeling Comparison Between LDA, NMF, Top2Vec, and BERTopic to Demystify Twitter Posts," *Front. Sociol.*, vol. 7, no. May, hal. 1–16, 2022, doi: 10.3389/fsoc.2022.886498.
- [13] T. I. Ifada, N. Ratnasari, Z. N. Rahmawati, F. Wahyudi, A. H. Widanigrum, dan R. I. Tarecha, "Analisis Pengaruh Kualitas Layanan Shopee Terhadap Kepuasan Konsumen Pada Mahasiswa," *JUSIFOR J. Sist. Inf. dan Inform.*, vol. 3, no. 1, hal. 38–44, Jun 2024, doi: 10.33379/jusifor.v3i1.4415.
- [14] N. N. Hidayanti dan S. Shaleha, "BERTopic Analysis of Indonesian Biodiversity Policy on Social Media," *ECTI Trans. Comput. Inf. Technol.*, vol. 18, no. 3, hal.

- 260–271, 2024, doi: 10.37936/ecti-cit.2024183.255058.
- [15] E. Triningsih, M. Afdal, I. Permana, dan N. Evrilyan Rozanda, “Analisis Sentimen Terhadap Program Makan Bergizi Gratis Menggunakan Algoritma Machine Learning Pada Sosial Media X,” *Build. Informatics, Technol. Sci.*, vol. 6, no. 4, hal. 2240–2250, 2025, doi: 10.47065/bits.v6i4.6534.
- [16] E. Setiadi dan D. A. Sulisty, “Analisis Sentimen Kebijakan Makan Bergizi Gratis Menggunakan IndoBERT dan Machine Learning,” *J. FASILKOM*, vol. 15, no. 3, hal. 507–513, 2025.