

Pemodelan Isu Dan Sikap Publik Terhadap Program Makan Bergizi Gratis Menggunakan Arsitektur Hibrida BERTopic Dan IndoBERT

¹Bayu Tri Nugroho, ²Arief Hermawan, ³Donny Avianto, ⁴Ari Risnanto

^{1,2,3,4}Universitas Teknologi Yogyakarta, Indonesia

bayutri2002@gmail.com; ariefdb@uty.ac.id; donny@uty.ac.id; arieriz@gmail.com

Article Info

Article history:

Received, 2026-05-23

Revised, 2026-06-20

Accepted, 2026-07-07

Kata Kunci:

BERTopic
IndoBERT,
Program MBG,
Stance Classification,
Topic Modeling

Keywords:

BERTopic
IndoBERT,
MBG Program,
Stance Classification,
Topic Modeling

ABSTRAK

Program Makan Bergizi Gratis (MBG) merupakan kebijakan sosial berskala nasional yang memicu diskursus luas di YouTube. Namun, analisis opini publik umumnya masih terbatas pada analisis sentimen sehingga belum mampu membedakan kritik operasional dan penolakan terhadap program. Penelitian ini mengusulkan pipeline analitik hibrida berbasis *machine learning* untuk menganalisis topik dan sikap publik dari 5.509 komentar YouTube yang telah diproses. Pemodelan topik dilakukan menggunakan BERTopic dengan *embedding paraphrase-multilingual-MiniLM-L12-v2*, UMAP, dan HDBSCAN, sedangkan klasifikasi sikap menggunakan model IndoBERT-base-p1 yang di-fine-tuning. Hasil menunjukkan BERTopic berhasil mengidentifikasi 19 topik dengan nilai koherensi (C_v) 0,4918. Model IndoBERT mencapai akurasi 70,00% dan *Macro F1-Score* 0,7002. Analisis menunjukkan sikap Oppositional mendominasi (50,0%), terutama pada isu keamanan pangan (91,1%) dan dugaan korupsi proyek (86,1%). Sebaliknya, dukungan tertinggi (22,2%) muncul pada pembahasan pemerataan manfaat program secara nasional. Temuan ini menunjukkan bahwa penguatan SOP higienitas, transparansi pengelolaan anggaran, dan komunikasi publik yang efektif menjadi faktor penting untuk meningkatkan kepercayaan masyarakat terhadap implementasi MBG.

ABSTRACT

The Free Nutritious Meal (MBG) program is a large-scale social policy that has generated extensive public discussion on YouTube. However, existing public opinion analyses primarily rely on sentiment analysis, which often fails to distinguish operational criticism from fundamental opposition to the policy. This study proposes a hybrid machine learning pipeline to analyze discussion topics and public stance using 5,509 preprocessed YouTube comments. Topic modeling was performed using BERTopic with the *paraphrase-multilingual-MiniLM-L12-v2* embedding model, UMAP, and HDBSCAN, while stance classification was conducted by fine-tuning the IndoBERT-base-p1 model. The results identified 19 coherent topics with a C_v coherence score of 0.4918. The fine-tuned IndoBERT achieved an accuracy of 70.00% and a Macro F1-score of 0.7002. Oppositional stances dominated the discussions (50.0%), particularly on food safety concerns (91.1%) and allegations of project corruption (86.1%). In contrast, supportive opinions (22.2%) primarily focused on the program's nationwide equity and social welfare objectives. These findings suggest that strengthening kitchen hygiene standard operating procedures (SOPs), enhancing budget transparency, and improving public communication are critical to increasing public trust and supporting the effective implementation of the MBG program.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-nc-nd/4.0/) license.



Penulis Korespondensi:

Bayu Tri Nugroho,
Program Studi Sistem Informasi,
Universitas Teknologi Yogyakarta,
Email: bayutri2002@gmail.com

1. PENDAHULUAN

Pemenuhan nutrisi dan gizi yang optimal sejak fase awal kehidupan merupakan fondasi esensial yang mendasari pembentukan kualitas sumber daya manusia (SDM) sebuah bangsa. Secara makroekonomi, permasalahan malnutrisi dan gangguan pertumbuhan kronis (*stunting*) bukan sekadar krisis kesehatan publik, melainkan sebuah beban struktural. Merespons urgensi sistemik tersebut, Pemerintah Republik Indonesia meluncurkan Program Makan Bergizi Gratis (MBG) sebagai agenda prioritas nasional dengan operasionalisasi yang terus diekspansi untuk menjangkau lebih dari 36,7 juta penerima manfaat.

Dengan implikasi fiskal yang masif, implementasi teknis program MBG secara alamiah memicu pengawasan dan polarisasi opini publik yang sangat tajam di ruang digital. Platform berbagi video seperti YouTube menempati posisi sentral dalam transmisi diskursus ini. Laporan pandangan mata, liputan jurnalistik warga, hingga pemberitaan kanal arus utama mengenai uji coba dapur Satuan Pelayanan Pemenuhan Gizi (SPPG) memantik ribuan komentar berantai. Volume opini tekstual yang tidak terstruktur ini merekam denyut nadi persepsi masyarakat terhadap program, menjadikannya aset *big data* yang amat berharga bagi perumusan evaluasi kebijakan publik secara *data-driven*.

Pemrosesan Bahasa Alami (*Natural Language Processing/NLP*) untuk mengekstraksi wawasan dari opini publik di media sosial telah menjadi area penelitian yang berkembang pesat. Berbagai studi telah mengaplikasikan algoritma *machine learning* untuk memetakan diskursus publik, mulai dari klasifikasi sentimen hingga pemodelan topik [1]. Beberapa penelitian terdahulu menjadi sumber literatur yang relevan dan esensial dalam membangun kerangka metodologis pada penelitian ini.

Kajian yang komprehensif mengenai pemodelan topik pada teks media sosial dan berita daring menunjukkan dominasi penggunaan algoritma *Latent Dirichlet Allocation* (LDA). Muslim et al. [2] membandingkan model *Probabilistic Latent Semantic Analysis* (PLSA) dan turunannya untuk menganalisis data pemilihan umum, sementara Septiara et al. [3] mengaplikasikan LDA untuk menganalisis transkrip pidato politik Presiden. Ekstraksi topik menggunakan LDA juga terbukti mampu memetakan persepsi publik terhadap isu strategis lainnya, seperti analisis *reshuffle* Menteri Keuangan pada komentar YouTube [4], pemodelan topik pariwisata Yogyakarta [5], serta analisis pemberitaan telemedicine di Indonesia [6]. Meskipun LDA terbukti stabil, model probabilistik konvensional ini kerap menunjukkan kelemahan saat menangani teks media sosial yang singkat, asimetris, dan sarat dengan bahasa informal, yang berdampak pada rendahnya nilai *coherence score*.

Untuk mengatasi keterbatasan tersebut, pendekatan pemodelan topik mulai bergeser ke arah arsitektur berbasis *transformer*. Penelitian yang dilakukan oleh Riyansyah et al. [7] membuktikan bahwa algoritma BERTopic secara signifikan mengungguli LDA dalam menganalisis berita konflik internasional, menghasilkan topik yang jauh lebih kohesif secara semantik. Keunggulan representasi kontekstual dari algoritma ini juga dikonfirmasi oleh Putri et al. [8] dalam analisis ulasan aplikasi perbankan digital, di mana BERTopic mencatat skor koherensi yang lebih tinggi dibandingkan metode probabilistik. Integrasi BERTopic dengan pra-pemrosesan model bahasa berbasis *transformer* memungkinkannya mengelompokkan dokumen berdasar kedekatan semantik, bukan sekadar frekuensi kemunculan kata seperti pada pemodelan klasik [9].

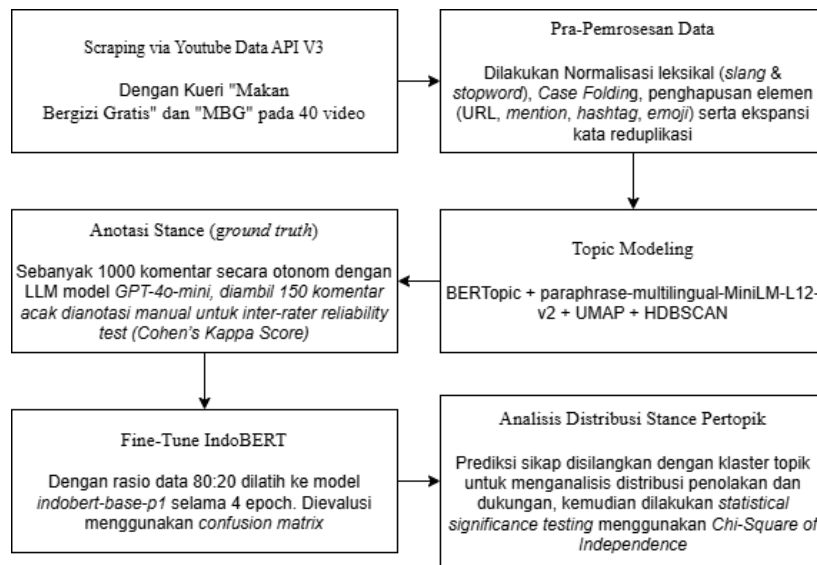
Selain pemodelan topik, memahami opini masyarakat memerlukan pendekatan klasifikasi teks. Sebagian besar penelitian terdahulu mengawinkan pemodelan topik dengan Analisis Sentimen (positif, negatif, netral) untuk mengukur respons emosional publik. Kombinasi metode ini telah diterapkan secara luas pada berbagai kasus, seperti evaluasi ulasan aplikasi urun dana (*crowdfunding*) menggunakan LDA dan *Support Vector Machine* [10], analisis kepuasan pengguna aplikasi perbankan [11], hingga pemetaan respons netizen terhadap konflik geopolitik menggunakan ekstraksi leksikon [12]. Penggunaan model pra-catih seperti IndoBERT terbukti sangat andal untuk teks berbahasa Indonesia. Irmawan et al. [13] yang melakukan *fine-tuning* IndoBERT untuk klasifikasi ulasan aplikasi agen perjalanan daring berhasil mencapai akurasi sebesar 91%, melampaui algoritma tradisional.

Dalam konteks kebijakan program Makan Bergizi Gratis (MBG), beberapa studi terkini telah berupaya memetakan diskursus yang terjadi di ruang digital. Suhaeni et al. [14] mengaplikasikan LDA pada komentar YouTube untuk menemukan kekhawatiran masyarakat terkait harga dan distribusi pangan. Di platform X (Twitter), Sul-toni et al. [15] menggunakan metode Lexicon dan SVM dipadukan dengan LDA untuk optimalisasi distribusi program MBG berdasarkan analisis sentimen. Serupa dengan itu, Yanna dan Yustanti [16] mengawinkan IndoBERT untuk analisis sentimen dan LDA pada data Twitter terkait MBG, menghasilkan temuan sentimen negatif yang didominasi oleh kekhawatiran terhadap prioritas APBN. Kajian yang paling mendekati penelitian ini dilakukan oleh Wahyuni et al. [17], yang mengimplementasikan BERTopic pada komentar YouTube terkait MBG. Hasil evaluasinya membuktikan BERTopic memiliki performa ekstraksi topik yang lebih baik dibanding LDA dalam memahami kompleksitas opini publik.

Kajian dari berbagai literatur di atas menunjukkan bahwa masalah evaluasi wacana kebijakan publik di media sosial belum terjawab secara memuaskan. Pertama, penelitian-penelitian mengenai kebijakan MBG [15], [16] masih terpaku pada pendekatan Analisis Sentimen emosional. Padahal, untuk evaluasi kebijakan politis, analisis *Stance Classification* (mendukung/*supportive*, menentang/*oppositional*, atau netral/*neutral*) jauh lebih relevan karena mampu membedakan warga yang mungkin menggunakan sentimen linguistik negatif (misalnya marah pada teknis pelaksanaan) namun secara esensial mendukung keberadaan program tersebut. Kedua, meskipun penelitian Wahyuni et al. [17] telah sukses mengaplikasikan BERTopic pada program MBG di YouTube, penelitian tersebut hanya berfokus murni pada ekstraksi topik (*unsupervised*) tanpa integrasi model klasifikasi terarah (*supervised*) terhadap opini pada setiap komentarnya.

2. METODE PENELITIAN

Penelitian ini dilaksanakan melalui serangkaian tahapan komputasional kronologis dimana terdapat dua tahap utama yaitu, analisis pemodelan topik menggunakan BERTopic dan juga klasifikasi sikap (*stance*) menggunakan IndoBERT. Pada Gambar 1 dijelaskan tahapan lengkap penelitian ini:



Gambar 1 Alur Penelitian Pemodelan Isu Dan Sikap Publik Terhadap Program Makan Bergizi Gratis

Data dikumpulkan dari platform YouTube melalui teknik *scraping* menggunakan aplikasi terprogram (YouTube Data API v3) dengan menggunakan bahasa pemrograman Python. Kueri pencarian difokuskan secara spesifik pada entitas "Makan Bergizi Gratis" dan "MBG", dengan menerapkan filter pada video-video berpopularitas tinggi dari kanal dengan rencana verifikasi dan dipublikasikan pada rentang bulan Januari 2025 sampai bulan April 2026. Proses ekstraksi ini juga melibatkan penyaringan komentar duplikat dan juga pembatasan 1 username hanya diambil 1 komentar, dengan hasil akhir dapat mengakuisisi sebanyak 5.510 komentar mentah. Selanjutnya, dilakukan pra-pemrosesan pada data teks tersebut untuk mereduksi derau linguistik. Tahapan ini meliputi pengecilan aksara (*case folding*), purifikasi teks dari tautan URL, sebutan akun (*mention*), *hashtag*, dan emoji, serta ekspansi untuk kata reduplikasi. Normalisasi leksikal juga diterapkan menggunakan kamus kustom untuk memperbaiki 470 struktur *slang* atau bahasa tidak baku, yang kemudian dilanjutkan dengan penghapusan 247 tipe *stopword* yang tidak memiliki nilai diskriminatif untuk pemodelan.

Proses pemodelan wacana publik menggunakan algoritma BERTopic bergantung pada konfigurasi matematis dari sub-komponen penyusunnya, yaitu *Uniform Manifold Approximation and Projection* (UMAP) untuk reduksi dimensi dan *Hierarchical Density-Based Spatial Clustering of Applications with Noise* (HDBSCAN) untuk klusterisasi tekstual. Agar eksperimen ini dapat direplikasi secara presisi, konfigurasi *hyperparameter* pada kedua algoritma tersebut didokumentasikan secara eksplisit. Pada tahap reduksi dimensi, algoritma UMAP dikonfigurasi dengan parameter *n_neighbors* sebesar 30, yang menentukan cara algoritma menyeimbangkan fokus antara struktur lokal dan global dari spasial data. Selain itu, parameter *min_dist* diatur pada nilai 0.1 guna mengontrol tingkat kepadatan representasi vektor dalam ruang dimensi rendah. Proses perhitungan kemiripan antarvektor dokumen di dalam UMAP didasarkan pada metrik jarak *cosine distance*. Selanjutnya, matriks dimensi rendah yang dihasilkan oleh UMAP diproses oleh algoritma HDBSCAN untuk pembentukan kluster isu. Eksplorasi kelompok dokumen diatur secara ketat dengan menetapkan parameter *min_cluster_size* sebesar 50, yang bertindak sebagai ambang batas minimum jumlah komentar YouTube untuk dapat diakui sebagai sebuah kluster topik yang valid. Adapun penentuan fungsi jarak metrik (*metric distance*)

yang digunakan untuk mengukur kedekatan kepadatan spasial antardokumen pada tahap klusterisasi ini menggunakan metrik *euclidean*.

Selain parameter pada tahap reduksi dimensi dan klusterisasi, kontrol hierarki penyusutan topik juga diimplementasikan untuk meningkatkan kualitas interpretasi model. Parameter *nr_topics* pada BERTopic diatur pada angka 20. Konfigurasi ini menginstruksikan algoritma untuk secara iteratif menggabungkan kluster-kluster yang memiliki kemiripan semantik paling identik yang dievaluasi berdasarkan pengukuran *cosine similarity* pada representasi matriks c-TF-IDF, sehingga hasil topik dapat menyusut menjadi batas target tersebut. Pembatasan target ke 20 entitas ini diputuskan guna meminimalkan fragmentasi dan tumpang tindih wacana (*topic overlapping*) pada data bervolume menengah. Melalui konfigurasi penyusutan ini, algoritma pada akhirnya mendistribusikan korpus ke dalam 19 topik diskursus sentral dan 1 kluster anomali derau.

Sebagai model pembanding (*baseline*), konfigurasi *hyperparameter* pada algoritma *Latent Dirichlet Allocation* (LDA) juga ditetapkan secara presisi guna menjamin validitas replikasi. Proses pelatihan LDA dikonfigurasi dengan melakukan 10 kali putaran pelatihan penuh terhadap keseluruhan korpus (*passes*=10), dengan batas konvergensi maksimal 100 iterasi per dokumen (*iterations*=100). Untuk mengontrol kelangkaan (*sparsity*) distribusi ruang topik, parameter probabilitas prior untuk sebaran topik dalam dokumen (α) dan sebaran kata dalam topik (η) diatur pada mode komputasi otomatis (*auto-compute*). Konfigurasi dinamis ini memungkinkan algoritma mempelajari struktur kepadatan asimetris secara empiris langsung dari data dengan tidak menggunakan nilai simetris statis.

Guna mengukur kualitas pemodelan secara objektif dan memfasilitasi komparasi yang berimbang (*fair comparison*) dengan model *baseline Latent Dirichlet Allocation* (LDA), penelitian ini mengimplementasikan metrik koherensi semantik berbasis *embedding*. Proses kalkulasi diawali dengan mengekstraksi 15 kata kunci teratas berdasarkan peringkat relevansi dari masing-masing topik. Seluruh kata kunci tersebut kemudian diproyeksikan ke dalam ruang vektor menggunakan model pra-latih (*paraphrase-multilingual-MiniLM-L12-v2*). Nilai koherensi per topik dievaluasi melalui kalkulasi rata-rata *cosine similarity* antar seluruh pasangan kata kunci di dalam matriks tersebut, dengan mengecualikan kesamaan mandiri (*self-similarity*). Nilai akhir dinormalisasi pada rentang 0 hingga 1, di mana nilai 1 mengindikasikan identitas semantik penuh dan 0 menunjukkan ketiadaan relasi. Penggunaan metode evaluasi ini bertujuan untuk memperoleh estimasi kedekatan semantik yang lebih stabil, serta membatasi pengaruh derau (*noise*) komputasional dari kata-kata berperingkat bawah.

Setelah model topik terbentuk, penelitian dilanjutkan dengan analisis prediktif berbasis *Stance Classification* (Klasifikasi Sikap). Guna mengatasi masalah ketidakseimbangan kelas (*class imbalance*) yang umum terjadi pada data opini, pelabelan data *ground truth* sebanyak 1.000 komentar dilakukan secara otonom menggunakan *Large Language Model* (GPT-4o-mini). Skema anotasi menggunakan *quota-based sampling* yang secara ketat membatasi penerimaan target kelas menjadi distribusi yang merata: 333 data *Supportive*, 333 data *Neutral*, dan 334 data *Oppositional*. Guna memastikan validitas dan reliabilitas pelabelan otonom tersebut, penelitian ini mengimplementasikan uji kesepakatan antar-penilai (*inter-rater reliability test*). Sebuah subsampel acak diambil sebanyak 150 dokumen dari total *ground truth* untuk dianotasi ulang secara manual oleh peneliti manusia melalui prosedur *blind annotation* (anotasi buta tanpa melihat prediksi mesin). Keselarasan komputasional antara label yang dihasilkan oleh arsitektur GPT dan kognisi manusia kemudian dievaluasi menggunakan metrik *Cohen's Kappa Score* untuk mengukur reliabilitas anotasi artifisial sekaligus mengeliminasi probabilitas kebetulan (*chance agreement*).

Dataset yang telah diproses kemudian dibagi dengan rasio 80% untuk pelatihan dan 20% untuk validasi. Data pelatihan disuplai ke dalam arsitektur jaringan *transformer* indobert-base-p1 untuk proses *fine-tuning* selama 4 iterasi (*epochs*). Kinerja klasifikasi divalidasi menggunakan matriks evaluasi *accuracy*, *precision*, *recall*, dan metrik harmonik *Macro F1-Score*. Model sentimen yang telah terlatih kemudian digunakan untuk memprediksi sikap atas seluruh dataset komentar. Tahap komputasi terakhir adalah fusi data, di mana probabilitas prediksi sikap disilangkan dengan kluster topik untuk menganalisis distribusi penolakan dan dukungan secara spesifik pada masing-masing wacana program. Guna memvalidasi korelasi silang tersebut, pengujian signifikansi statistik (*statistical significance testing*) diimplementasikan menggunakan uji *Chi-Square of Independence*. Pengujian ini bertujuan untuk mengetahui secara matematis apakah variasi distribusi dukungan maupun penolakan pada setiap kluster isu terjadi akibat probabilitas kebetulan atau memang merepresentasikan kecenderungan sikap publik yang berbeda secara signifikan.

3. HASIL DAN ANALISIS

Scraping awal mengakuisisi 5.510 komentar. Setelah intervensi serangkaian proses pra-pemrosesan terdapat 1 baris teks nihil terhapus sehingga menyisakan 5.509 baris data absolut. Pembersihan *noise* sintaktis ini menghasilkan reduksi efisiensi panjang karakter rata-rata dari 27,9 kata per dokumen menjadi 21,3 kata (penghematan komputasional sebesar 23,7%). Pada tabel 1 dipaparkan informasi mengenai video yang

dilakukan scraping serta jumlah komentar yang diambil dan di lewati (filter duplicate username serta filter kata kunci “makan bergizi gratis” dan “mbg”).

Tabel 1. Daftar Informasi Video Youtube

No	Link YouTube	Judul Video	Komentar Diambil	Dilewati (Total)
1	https://www.youtube.com/watch?v=a2WXt0aW76g	Kroni Prabowo dalam Proyek Makan Bergizi Gratis (MBG) Bocor Alus Politik	796	2834
2	https://www.youtube.com/watch?v=v12Ec7xHuo4	Keracunan Massal Proyek Makan Bergizi Gratis Bocor Alus Politik	279	550
3	https://www.youtube.com/watch?v=5MGRcdpr67g	Makan BERGIZI Apa BER4CUN Gratis!?- Aduan Masyarakat	473	951
4	https://www.youtube.com/watch?v=KIC1Jcl9HIQ	Ahli Gizi dr Tan Shot Yen Kritik Keras MBG di DPR: Yang Dibagi Adalah Burger, Oh My God..	500	1343
5	https://www.youtube.com/watch?v=0BRFIBVWH3c	TAK ADA MAKAN SIANG GRATIS Kelindan Kepentingan di Balik Program MBG	324	1081
6	https://www.youtube.com/watch?v=xefYpQRqxUU	Makan Bergizi Gratis yang Problematis Jelasin Dong!	212	327
7	https://www.youtube.com/watch?v=aYEY-L0EDaf	MBG di Purworejo Disajikan Sepenuh Hati, Menu Makanan Bervariasi Lengkap Kandungan Gizi Liputan 6	213	607
8	https://www.youtube.com/watch?v=9GUcjgVubpU	Mendalami Kejagalan Program Makan Bergizi Gratis (MBG)	89	831
...
...
36	https://www.youtube.com/watch?v=pDzm7bVcJrA	Pemerintah Tutup Sementara 2.162 SPPG demi Perbaikan Layanan Program Makan Bergizi Gratis	35	54
37	https://www.youtube.com/watch?v=hEiaoa8Yb0E	Viral! Dapur Makan Bergizi Gratis di Cipatat Kena Semprot, Cuci Nampun Ala Kadarnya	14	64
38	https://www.youtube.com/watch?v=rXBBiwa6Dew	Ratusan Siswa Keracunan Makanan Bergizi Gratis, Siapa Tanggung Jawab?	76	190
39	https://www.youtube.com/watch?v=d2sVZonCJMg	Sepekan, Kasus Keracunan Makan Bergizi Gratis Terus Bermunculan	27	107
40	https://www.youtube.com/watch?v=gJdq64jqNFA	Trauma Akibat Keracunan Massal Makan Bergizi Gratis Explained	13	62

Tabel 2. Contoh Data Setelah Pra-Pemrosesan

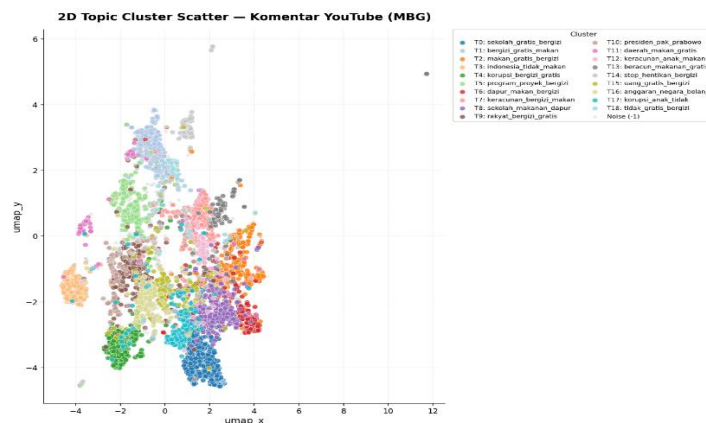
username	created_at	text	clean_text
@LusianaKurniawati-v1c6z	2026-04-16T14:24:53Z	MBg ...koperasi merah putih...lambung suara grindra...pemilu ke depan ..	makan bergizi gratis koperasi merah putih lambung suara grindra pemilu
@hartutituti-u8t	2026-04-16T14:12:32Z	Bukan ada lagi..Aroma korupsi di MBG, memangbitu ladang basah	bukan aroma korupsi makan bergizi gratis memangbitu ladang basah
@syafiihs2	2026-04-15T04:34:01Z	Mhn maaf sy rakyat kecil melihat program yg kt nya mbg ini tdk lebih hny u persiapan th 2029.. dan itu fik mnurut sy pribadi...	mohon maaf rakyat kecil melihat program makan bergizi gratis tidak u persiapan th 2029 fik mnurut pribadi
@PakdeMantoo	2026-03-28T01:00:12Z	MBG adalah ladang korupsi model baru Prabowo yang lama ngga di apa apain para koruptor baru dari MBG	makan bergizi gratis ladang korupsi model baru prabowo tidak apain koruptor baru makan bergizi gratis

Model paraphrase-multilingual-MiniLM-L12-v2 mengekstraksi data ke dalam matriks berdimensi spasial (5509, 384). Fungsi *reduce_outliers* berbasis *embeddings* menunjukkan kinerja yang sangat tangguh dengan merestorasi 1.645 dokumen, menekan volume *noise* dari 1.987 menjadi hanya 342 komentar. Model ini mencetak skor *Overall Coherence* (C_v) di angka 0.4918. Meskipun nilai koherensi C_v sebesar 0.4918 berada pada kategori moderat, capaian ini secara empiris dinilai valid dan representatif untuk pemodelan korpus media sosial.

Karakteristik *User-Generated Content* (UGC) pada kolom komentar YouTube memiliki tingkat derau (*noise*) yang tinggi, asimetris, sangat pendek, serta sarat akan penggunaan bahasa informal dan slang lokal. Dalam studi pemodelan topik teks pendek, nilai C_v pada rentang 0,4 hingga 0,5 terbukti secara ilmiah tetap mampu menghasilkan representasi klaster isu yang kohesif secara semantik dan dapat diinterpretasikan dengan baik untuk analisis sosiolinguistik lanjutan. Pada tabel 2 merangkum hasil pemodelan topik beserta kata kunci (*keyword*) yang teratas yang muncul pada topik tersebut. Kemudian Gambar 2 memberikan visualisasi proyeksi UMAP yang menunjukkan persebaran dan kedekatan semantik antar 19 klaster topik secara spasial.

Tabel 3. Hasil Topic Modeling

topic_id	doc_count	top_words
-1	342	noise
0	583	sekolah, gratis, bergizi, anak, makan, tidak, pendidikan, guru, siswa, menengah, dasar, orang, program, buat, dana
1	465	bergizi, gratis, makan, pak, prabowo, bagus, semoga, biar, banget, tidak, bubarkan, tutup, kasih, orang, baru
2	472	makan, gratis, bergizi, anak, beras, tidak, makanan, ayam, nasi, basi, menu, susu, orang, kasih, ibu
3	339	indonesia, tidak, makan, gratis, bergizi, anak, negara, rakyat, program, nasional, tentara, sekolah, prabowo, seluruh, orang
4	440	korupsi, bergizi, gratis, makan, ladang, koruptor, proyek, program, lahan, nepotisme, kolusi, tidak, pejabat, rakyat, negara
...
...
15	209	uang, gratis, bergizi, makan, tidak, orang, tua, kasih, rakyat, uangnya, buang, biar, pakai, buat, dana, anggaran, negara, belanja, pendapatan, rakyat, gratis, bergizi, makan, pajak, tidak, program, proyek, ekonomi, pemerintah, buat
16	229	korupsi, anak, tidak, gratis, makan, orang, bergizi, uang, sekolah, dana, program, tua, pejabat, buat, rakyat
17	157	rakyat
18	99	tidak, gratis, bergizi, makan, pak, jangan, koar, usah, doyan, an, he, bukan, ww, merika, qodari



Gambar 2 Hasil Scatter 2D Setiap Topik

Untuk menguji validitas arsitektur pemodelan wacana, kinerja BERTopic dikomparasikan secara empiris terhadap model probabilistik LDA pada dataset yang sama. Agar perbandingan berjalan setara, model dasar LDA dikonfigurasi untuk mengekstraksi jumlah topik yang identik, yakni $K=19$. Ringkasan hasil komparasi komputasional pada kedua algoritma tersebut dipaparkan pada Tabel 4.

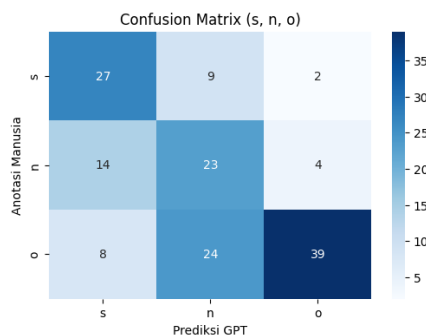
Tabel 4. Perbandingan Kinerja Pemodelan Topik (BERTopic vs LDA)

Metrik Evaluasi	BERTopic	Latent Dirichlet Allocation (LDA)
Total Dokumen	5.509	5.509
Jumlah Topik	19	19
Klaster Derau (Noise)	342 dokumen (6,2%)	Tidak Tersedia (N/A)
Mean Semantic Coherence	0,4918	0,4786

Merujuk pada Tabel 4, representasi wacana yang dibangun oleh arsitektur berbasis *Transformer* mencatat performa yang lebih unggul dibandingkan pemodelan probabilistik klasik. BERTopic menghasilkan skor rata-rata koherensi semantik sebesar 0,4918, melampaui capaian model LDA yang berada di angka 0,4786. Peningkatan kohesi matriks semantik ini membuktikan bahwa mekanisme BERTopic memiliki kemampuan yang lebih baik dalam menangkap konteks dan mengelompokkan leksikon media sosial Indonesia yang asimetris ke dalam kluster yang bermakna. Keunggulan spasial BERTopic juga terlihat pada kemampuannya mengisolasi dokumen yang tidak relevan. Melalui algoritma HDBSCAN, *pipeline* ini berhasil menyaring 342 dokumen (6,2% dari total korpus) ke dalam kluster anomali derau. Mekanisme penyaringan anomali (*noise filtering*) ini berkontribusi dalam mempertahankan tingkat koherensi pada 19 topik sentral. Pendekatan spasial ini menawarkan keunggulan komparatif dibandingkan model LDA, di mana model probabilistik konvensional tersebut secara bawaan akan mengalokasikan setiap dokumen ke dalam distribusi topik tanpa memiliki fungsi eksklusif untuk teks dengan relevansi semantik yang rendah.

Penggunaan teknik *quota-based sampling* pada proses anotasi GPT diimplementasikan untuk mendistribusikan proporsi kelas sentimen secara merata. Langkah prapemrosesan ini dilakukan untuk membatasi kecenderungan model (*model bias*) yang sering muncul akibat ketimpangan distribusi data. Kualitas dataset *ground truth* yang

dianotasi oleh GPT dievaluasi melalui uji *Cohen's Kappa*. Pengujian silang terhadap 150 sampel validasi manual mencatat tingkat akurasi sebesar 59,33% dengan skor metrik kappa di angka 0,3959. Merujuk pada batas konvensi akademis, nilai tersebut merepresentasikan tingkat kesepakatan moderat (*moderate agreement*). Analisis terhadap matriks kebingungan (*confusion matrix*) pada Gambar 3 mengindikasikan bahwa disparitas interpretasi antara model mesin dan anotator manusia secara dominan berpusat pada demarkasi kelas *Neutral* dan *Oppositional*. Ketidaksiharan ini merupakan fenomena sosiolinguistik yang lazim dalam ranah analisis sentimen pada Konten Buatan Pengguna (*User-Generated Content / UGC*). Secara spesifik, diskursus di platform YouTube terkait regulasi publik sering kali memuat kritik implisit, pertanyaan investigatif, maupun majas sarkasme yang menghasilkan derajat ambiguitas yang tinggi. Model komputasi generatif memiliki kecenderungan untuk mengklasifikasikan tuturan tanpa leksikon negatif eksplisit ke dalam kelas netral, sedangkan kognisi manusia memiliki kapabilitas untuk mengidentifikasi muatan pragmatis yang bersifat oposisional di balik struktur tuturan tersebut. Meskipun dihadapkan pada kompleksitas interpretasi linguistik ini, dataset artifisial yang dihasilkan dinilai telah memenuhi ambang batas reliabilitas yang memadai (*acceptable reliability*) untuk diimplementasikan sebagai basis korpus dalam pelatihan arsitektur IndoBERT.



Gambar 3 Heatmap Presentase Klasifikasi Sikap Setiap Topik

Tahapan pelatihan (*fine-tuning*) pada arsitektur IndoBERT selanjutnya dilakukan secara iteratif dengan mempertimbangkan stabilitas generalisasi. Rincian metrik pelatihan selama 4 *epoch* dapat diamati pada Tabel 5, yang kemudian dapat diperoleh capaian kinerja prediksi akhir terhadap validation set seperti yang disajikan pada Tabel 6 berikut:

Tabel 5. Metrik Riwayat Pelatihan Model IndoBERT

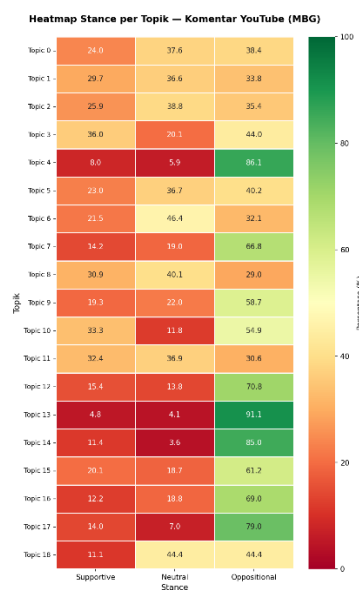
Epoch	Training Loss	Validation Loss	Validation Accuracy
1	1.0347	0.9057	59.5%
2	0.6549	0.7587	70.0%
3	0.3477	0.8313	66.0%
4	0.1779	0.8685	68.0%

Tabel 6. Laporan Klasifikasi Validation Set Fine-Tuning IndoBERT

Kelas Stance	Precision	Recall	F1-Score	Support
Supportive (Mendukung)	0.6044	0.8209	0.6962	67
Neutral (Netral)	0.7857	0.6567	0.7154	67
Oppositional (Menentang)	0.7736	0.6212	0.6891	66
Metrik Global	Akurasi: 70%	Macro F1:	0.7002	200

Pelatihan model IndoBERT-base-p1 yang difasilitasi oleh dataset penyeimbang otonom (1.000 sampel merata) diselesaikan dalam 4 *epoch*. Model *fine-tuning* IndoBERT-base-p1 dievaluasi kinerjanya pada setiap *epoch* menggunakan kombinasi metrik akurasi dan fungsi kerugian (*loss function*). Berdasarkan laporan klasifikasi pada Tabel 4, model mencapai tingkat akurasi moderat sebesar 70,00% dengan Macro F1-Score 0,7002. Untuk menguji validitas proses pelatihan dan mendeteksi potensi *overfitting*, pergerakan nilai fungsi kerugian selama 4 *epoch* dipetakan pada Gambar 3. Hasil plot data menunjukkan bahwa *training loss* mengalami penurunan linier yang konsisten, dari nilai 1,0347 pada *epoch* 1 menjadi 0,1779 pada *epoch* 4. Namun, tren berbeda ditunjukkan oleh *validation loss*, di mana titik konvergensi optimal tercapai pada *epoch* 2 dengan nilai kerugian terendah sebesar 0,7587. Setelah melewati *epoch* 2, nilai *validation loss* terdeteksi merangkak naik secara bertahap menjadi 0,8313 pada *epoch* 3 dan menyentuh angka 0,8685 pada *epoch* 4. Divergensi spasial di mana *training loss* terus menurun sementara *validation loss* mulai meningkat mengindikasikan adanya gejala *overfitting* ringan (*mild overfitting*) akibat keterbatasan ukuran sampel pembanding. Guna mengantisipasi penurunan kemampuan generalisasi tersebut, penelitian ini menerapkan pendekatan *early stopping* secara implisit. Bobot parameter (*weights*) model terbaik yang diekstrak dari konfigurasi *epoch* 2 ditetapkan sebagai model final untuk melakukan inferensi prediktif terhadap keseluruhan korpus data. Pada *Classification Report* menunjukkan IndoBERT memiliki *Precision* tertinggi dalam mengidentifikasi posisi Netral (78,57%),

sementara *Recall* tertinggi dicetak pada pengenalan posisi *Supportive* (82,09%). Saat inferensi model diterapkan ke 5.509 dataset utuh, peta makro opini publik menunjukkan bahwa posisi *Oppositional* mendominasi dengan 50,0% (2.754 komentar). Sikap objektif *Neutral* sebesar 27,8% (1.533 komentar), sementara dukungan *Supportive* sebesar 22,2% (1.222 komentar). Rata-rata confidence prediksi mesin tercatat pada probabilitas yang meyakinkan di angka 0,7519. Matriks kebingungan (*confusion matrix*) mengungkap fenomena menarik, yaitu kesalahan prediksi tertinggi terjadi pada tumpang tindih antara posisi *Neutral* dan *Oppositional*. Fenomena tumpang tindih klasifikasi ini dapat diidentifikasi secara konkret pada beberapa sampel komentar di dalam dataset. Sebagai contoh, komentar dari akun @syafiihs2 yang berbunyi: *'Mhn maaf sy rakyat kecil melihat program yg kt nya mbg ini tdk lebih hny u persiapan th 2029.. dan itu fik mnurut sy pribadi...'*. Secara anotasi manual (*ground truth*), komentar tersebut dikategorikan sebagai posisi *Oppositional* karena mengandung kecurigaan politis (program dianggap hanya sebagai alat kampanye pemilu 2029). Namun, model IndoBERT memprediksinya sebagai posisi *Neutral*. Hal ini terjadi karena struktur kalimat tersebut tidak mengekspresikan leksikon emosional yang agresif (seperti kata makian atau kemarahan), melainkan menggunakan konstruksi kalimat bergaya opini investigatif-spekulatif yang secara semantik bias sebagai pernyataan netral oleh representasi vektor arsitektur *transformer*.



Gambar 4 Heatmap Presentase Klasifikasi Sikap Setiap Topik

Untuk memastikan bahwa perbedaan distribusi persentase sikap publik antar-topik pada visualisasi peta panas (*heatmap*) memiliki validitas ilmiah, uji statistik *Chi-Square of Independence* dilakukan terhadap matriks frekuensi persilangan antara 19 topik sentral dan tiga kelas sentimen. Hasil pengujian komputasional mencatat nilai *Chi-Square* sebesar 897,7762 dengan derajat kebebasan (*degrees of freedom*) sebesar 38. Pengujian ini menghasilkan probabilitas (*p-value*) di angka $< 0,001$. Mengingat nilai *p-value* jauh berada di bawah ambang batas signifikansi standar ($\alpha = 0,05$), hipotesis nol (H_0) ditolak. Secara statistik, hal ini mengonfirmasi bahwa terdapat dependensi yang sangat signifikan secara independen antara kategori wacana (topik) dan kecenderungan polarisasi opini (*stance*). Dengan kata lain, tingginya tingkat oposisi pada kluster tata kelola atau tingginya dukungan pada kluster pemerataan merupakan pola representatif yang valid secara empiris, bukan anomali probabilitas.

Berdasarkan silangan data di atas, fenomena anomali respons publik dapat dijabarkan secara komprehensif ke dalam tiga klasifikasi sosiologis utama yang merangkum keseluruhan 19 kluster isu yang diuji:

Episentrum Penolakan (Skeptisisme Tata Kelola & Fobia Medis): Berdasarkan distribusi data, sentimen oposisi tertinggi dipicu oleh dua variabel utama: kekhawatiran terhadap keselamatan konsumsi dan rendahnya tingkat kepercayaan terhadap tata kelola anggaran. Topik 13 (Ketakutan terhadap Makanan Beracun) mencatat proporsi penolakan tertinggi sebesar 91,1%. Tingkat resistensi ini berkorelasi langsung dengan tingginya sentimen pada Topik 14 (Desakan Penghentian Program) sebesar 85,0%, yang didukung oleh isu turunan pada Topik 12 (Insiden Keracunan Siswa, 70,8%) dan Topik 7 (Korban Kasus Keracunan, 66,8%). Pada variabel tata kelola birokrasi, pesimisme warganet terefleksi pada tingginya sentimen oposisi di Topik 4 (Potensi Korupsi Pejabat) sebesar 86,1% dan Topik 17 (Korupsi Dana Pendidikan) sebesar 79,0%. Keraguan terhadap manajemen fiskal (APBN) juga terekam signifikan pada Topik 16 (Anggaran Negara, 69,0%) dan Topik 15 (Pemborosan Anggaran, 61,2%). Selain itu, sentimen negatif bermuatan politis teridentifikasi pada Topik 9 (Tudingan Program Kampanye, 58,7%), Topik 10 (Kritik terhadap Presiden, 54,9%), dan Topik 5 (Tuntutan

Evaluasi Proyek, 40,2%). Temuan ini mengindikasikan bahwa resistensi warganet pada video Youtube yang diteliti tidak ditujukan pada substansi pemenuhan gizi, melainkan terpusat pada risiko kegagalan implementasi dan defisit akuntabilitas fiskal.

Afirmasi terhadap Ekosistem dan Pemerataan: Konsentrasi sentimen positif (dukungan) ditemukan pada topik yang berkaitan dengan keadilan spasial (pemerataan) dan kesiapan infrastruktur. Topik 3 (Pemerataan Nasional) mencatat sentimen positif tertinggi di kategori ini dengan 36,0%, yang relevan dengan temuan pada Topik 11 (Perluasan ke Daerah dan Desa) sebesar 32,4%. Hal ini menunjukkan bahwa dimensi perluasan akses gizi ke wilayah non-urban mendapatkan penerimaan yang relatif baik dari warganet. Pada aspek operasional, dukungan tercatat pada Topik 8 (Infrastruktur Sekolah dan Dapur) sebesar 30,9%, serta Topik 0 (Fasilitas Pendidikan Dasar) sebesar 24,0%.

Kontroversi Investigatif dan Objektivitas Teknis: Penelitian ini juga mengidentifikasi kluster diskursus yang berfokus pada aspek teknis dengan distribusi sentimen yang lebih berimbang. Topik 1 (Evaluasi Keberlanjutan Program) mencatat Indeks Kontroversi tertinggi (0,6344), dengan proporsi sentimen yang terbagi secara proporsional: Netral (36,6%), Oposisi (33,8%), dan Dukungan (29,7%). Objektivitas percakapan warganet juga mendominasi aspek manajemen fungsional. Hal ini terlihat pada Topik 6 (Manajemen Dapur dan Yayasan) dengan 46,4% sentimen netral, serta Topik 2 (Substitusi Lauk dan Susu) dengan 38,8% sentimen netral. Data ini menunjukkan bahwa percakapan mengenai operasional teknis dan regulasi penyedia logistik cenderung analitis dan berfokus pada pengawasan jalannya program. Sementara itu, pada Topik 18 (Komentar Tidak Relevan/Sarkas), distribusi sentimen terbagi sama rata antara Netral dan Oposisi (masing-masing 44,4%).

4. KESIMPULAN

Penelitian ini mengintegrasikan BERTopic dan IndoBERT untuk menganalisis topik dan sikap publik terhadap Program Makan Bergizi Gratis (MBG) berdasarkan 5.509 komentar *YouTube*. BERTopic berhasil mengidentifikasi 19 topik dengan nilai koherensi (C_v) sebesar 0,4918, sedangkan model IndoBERT yang di-fine-tuning mencapai akurasi 70,00% dan *Macro F1-score* 0,7002. Hasil analisis menunjukkan bahwa sikap Oppositional mendominasi (50,0%), terutama dipicu oleh isu keamanan pangan dan dugaan korupsi, sementara dukungan tertinggi muncul pada harapan pemerataan manfaat program. Temuan ini mengindikasikan bahwa resistensi publik lebih dipengaruhi oleh persoalan implementasi dan tata kelola daripada tujuan program itu sendiri. Oleh karena itu, pemerintah perlu memperkuat SOP higienitas, transparansi pengelolaan anggaran, serta komunikasi publik untuk meningkatkan kepercayaan masyarakat. Penelitian ini masih memiliki keterbatasan berupa potensi bias sumber data *YouTube* dan pelabelan otomatis berbasis GPT. Penelitian selanjutnya disarankan memanfaatkan data lintas *platform*, anotasi *human-in-the-loop*, dan analisis longitudinal untuk memperoleh gambaran opini publik yang lebih representatif dan dinamis.

REFERENSI

- [1] R. Egger and J. Yu, "A Topic Modeling Comparison Between LDA, NMF, Top2Vec, and BERTopic to Demystify Twitter Posts," *Frontiers in Sociology*, vol. 7, May 2022, doi: 10.3389/fsoc.2022.886498.
- [2] M. P. Muslim, N. T. Hadi, and M. Adrezo, "Graph Regularized Probabilistic Latent Semantic Analysis for Topic Analysis Using Social Media Data," *Sinkron : Jurnal dan Penelitian Teknik Informatika*, vol. 9, no. 1, pp. 329–337, Jan. 2025, doi: 10.33395/sinkron.v9i1.14348.
- [3] D. I. Septiara, Deni Arifianto, and Wiwik Suharso, "Topic Analysis in Political Speech Video Transcripts Using the Latent Dirichlet Allocation (LDA) Method," *JUSTINDO (Jurnal Sistem dan Teknologi Informasi Indonesia)*, vol. 11, no. 1, pp. 1–10, Feb. 2026, doi: 10.32528/justindo.v11i1.4044.
- [4] N. F. A. Cendra, N. K. Wakhid, and H. Al Rosyid, "Analisis Persepsi Publik Terhadap Reshuffle Menteri Keuangan 2025 Berdasarkan Topic Modeling LDA Pada Komentar Youtube," *JATI (Jurnal Mahasiswa Teknik Informatika)*, vol. 10, no. 1, pp. 646–654, Feb. 2026.
- [5] Uray Nur Khadijah and Nuri Cahyono, "Analisis Topic Modelling Pariwisata Yogyakarta Menggunakan Latent Dirichlet Allocation (LDA)," *The Indonesian Journal of Computer Science*, vol. 13, no. 4, pp. 6075–6086, Jul. 2024, doi: 10.33022/ijcs.v13i4.3816.
- [6] S. B. Panuntun, D. Krismawati, S. Pramana, and E. T. Astuti, "Analisis Teks Pemberitaan Telemedicine di Indonesia: Pendekatan Sentimen, NER, Topic Modeling, dan Social Network dalam Memahami Isu dan Persepsi," *Indonesian of Health Information Management Journal (INOHIM)*, vol. 11, no. 1, pp. 56–67, Jun. 2023, doi: 10.47007/inohim.v11i1.500.
- [7] R. G. Riyansyah and S. Anggai, "Jurnal Impresi Indonesia (JII) Application of Latent Dirichlet Allocation (LDA) and BERTopic Algorithms for Headline and Topic Analysis of Palestine-Israel Conflict News in Indonesian Online Media," *Jurnal Impresi Indonesia (JII)*, vol. 5, no. 1, pp. 97–44, Jan. 2026.

- [8] S. A. Putri, K. D. Tania, and Naretha Kawadha Pasemah Gumay, "Knowledge Discovery Through Sentiment Analysis and Topic Modeling of BCA Mobile and MyBCA," *Journal of Mathematics, Computations and Statistics*, vol. 8, no. 2, pp. 669–682, Oct. 2025, doi: 10.35580/jmathcos.v8i2.9782.
- [9] L. Ma *et al.*, "AI-powered topic modeling: comparing LDA and BERTopic in analyzing opioid-related cardiovascular risks in women," *Exp. Biol. Med.*, vol. 250, 2025, doi: 10.3389/ebm.2025.10389.
- [10] I. N. Switrayana, D. Ashadi, H. Hairani, and A. Aminuddin, "Sentiment Analysis and Topic Modeling of Kitabisa Applications using Support Vector Machine (SVM) and Smote-Tomek Links Methods," *International Journal of Engineering and Computer Science Applications (IJECSA)*, vol. 2, no. 2, pp. 81–91, Sep. 2023, doi: 10.30812/ijecsa.v2i2.3406.
- [11] C. A. Sirait, D. I. Inan, R. Juita, and V. A. L. Sirait, "An Analysis Of User Satisfaction Using Sentiment Analysis And Topic Modeling: A Case Of Livin Mandiri Mobile Application," *JUPI (Jurnal Ilmiah Penelitian dan Pembelajaran Informatika)*, vol. 10, no. 4, pp. 2930–2940, Nov. 2025, doi: 10.29100/jupi.v10i4.6644.
- [12] J. P. Sosial, R. Kurniawan, I. Santoso, and A. K. Asrila, "Bagaimana netizen di Indonesia merespon konflik Israel-Palestina: Sebuah pendekatan topic modelling," *Universitas Negeri Padang Jl. Prof. Dr. Hamka, Air Tawar*, vol. 23, no. 02, pp. 65–77, Sep. 2025, doi: 10.7454/jps.2025.16.
- [13] O. A. Irmawan, I. Budi, A. B. Santoso, and P. K. Putra, "Improving Sentiment Analysis and Topic Extraction in Indonesian Travel App Reviews Through BERT Fine-Tuning," *Jurnal Nasional Pendidikan Teknik Informatika (JANAPATI)*, vol. 13, no. 2, pp. 359–370, Jul. 2024, doi: 10.23887/janapati.v13i2.77028.
- [14] Cici Suhaeni, L. N. A. Mualifah, and H. Wijayanto, "LDA Topic Modeling Analysis of Public Discourse on Indonesia's Free Nutritious Meals Program (MBG)," *IJID (International Journal on Informatics for Development)*, vol. 14, no. 1, pp. 587–600, Jun. 2025, doi: 10.14421/ijid.2025.5211.
- [15] A. Sultoni, D. Andhika Putra, H. Noor Wahidah, M. Mahathir Arief, and P. R. Jingga Putria, "Public Sentiment Analysis and Distribution Optimization MBG," *Jurnal Matematika Thales (JMT)*, vol. 07, no. 01, pp. 35–53, Sep. 2025, [Online]. Available: https://drive.google.com/drive/folders/1OAb9G2avR0fv_BL82uLixPeeXOVWG6qV
- [16] S. Mahmudah, P. Yanna, and W. Yustanti, "Sentiment Analysis and Topic Modeling Using BERT And LDA Methods (Case Study of Free Meal Program on Twitter)," *Journal of Emerging Information Systems and Business Intelligence*, vol. 7, no. 1, pp. 144–158, Mar. 2026.
- [17] W. Wahyuni, T. P. Lestari, M. Apriliana, and R. Gumelta, "Implementation of BERTopic for Topic Modeling Analysis of the Free Nutritious Meal Program Based on YouTube Comments," *Journal of Applied Informatics and Computing (JAIC)*, vol. 9, no. 4, pp. 2548–6861, Aug. 2025, [Online]. Available: <http://jurnal.polibatam.ac.id/index.php/JAIC>