

***TOPIC MODELLING* ASPIRASI MASYARAKAT MENGUNAKAN METODE BERTOPIC**

(Studi Kasus : Aspirasi Masyarakat di Dewan Perwakilan Daerah
Republik Indonesia)

TUGAS AKHIR

Diajukan Sebagai Salah Satu Syarat Untuk Memperoleh Gelar Sarjana Program
Studi Statistika



Disusun Oleh:
Nail Anfasa Fadla
22611050

**PROGRAM STUDI STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS ISLAM INDONESIA
YOGYAKARTA
2026**

HALAMAN PERSETUJUAN PEMBIMBING
TUGAS AKHIR

Judul : *Topic Modelling* Aspirasi Masyarakat
Menggunakan Metode BERTopic (Studi Kasus :
Aspirasi Masyarakat di Dewan Perwakilan
Daerah Republik Indonesia)

Nama Mahasiswa : Nail Anfasa Fadla

NIM : 22611050

**TUGAS AKHIR INI TELAH DIPERIKSA DAN DISETUJUI UNTUK
DIUJIKAN**

Mengetahui
Ketua Prodi Statistika

Yogyakarta, 30 April 2026
Dosen Pembimbing


(Dr. Atina Ahdika, S.Si., M.Si.)


(Dr. Edy Widodo, S.Si., M.Si.)

HALAMAN PENGESAHAN

TUGAS AKHIR

Topic Modelling Aspirasi Masyarakat Menggunakan Metode BERTopic

(Studi Kasus : Aspirasi Masyarakat di Dewan Perwakilan Daerah
Republik Indonesia)

Nama Mahasiswa : Nail Anfasa Fadla

NIM : 22611050

TUGAS AKHIR INI TELAH DIUJIKAN

PADA TANGGAL : 18 MEI 2026

Nama Penguji

Tanda Tangan

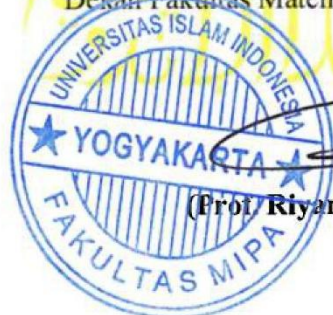
1. Ayundyah Kesumawati, S.Si., M.Si.

2. Dr. techn. Rohmatul Fajriyah, S.Si., M.Si.

3. Dr. Edy Widodo, S.Si., M.Si.

Mengetahui,

Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam



(Prof. Riyanto, S.Pd., M.Si., Ph.D.)



KATA PENGANTAR



Assalamu'alaikum Wr.Wb

Puji syukur kehadiran Allah Swt. atas segala rahmat, taufik, dan karunia-Nya, sehingga penulis dapat menyelesaikan Tugas Akhir yang berjudul “*Topic Modelling* Aspirasi Masyarakat Menggunakan Metode BERTopic (Studi Kasus : Aspirasi Masyarakat di Dewan Perwakilan Daerah Republik Indonesia)”. Tugas Akhir ini disusun sebagai salah satu syarat untuk memperoleh gelar sarjana pada program studi yang sedang penulis tempuh. Melalui penelitian ini, penulis berupaya memberikan gambaran mengenai pola dan kecenderungan topik aspirasi masyarakat dengan memanfaatkan pendekatan *topic modelling* dengan BERTopic.

Pada kesempatan ini, penulis menyampaikan ucapan terima kasih yang sebesar-besarnya kepada seluruh pihak yang telah berkontribusi dalam penyelesaian Tugas Akhir ini. Secara khusus, penulis menyampaikan terima kasih kepada dosen pembimbing yang telah meluangkan waktu, tenaga, dan pikiran untuk memberikan bimbingan, arahan, serta masukan selama proses penyusunan penelitian ini. Penulis juga menyampaikan terima kasih kepada seluruh Staf Kantor Dewan Perwakilan Daerah Republik Indonesia Daerah Istimewa Yogyakarta yang telah memberikan dukungan serta kesempatan kepada penulis dalam memperoleh data dan informasi yang dibutuhkan dalam penelitian ini. Selain itu, ucapan terima kasih penulis sampaikan kepada keluarga, sahabat, dan rekan-rekan yang senantiasa memberikan doa, semangat, dan dukungan selama proses penyusunan Tugas Akhir.

Penulis menyadari bahwa Tugas Akhir ini masih memiliki berbagai keterbatasan dan jauh dari kata sempurna. Oleh karena itu, penulis mengharapkan kritik dan saran yang membangun demi penyempurnaan penelitian ini di masa mendatang. Penulis berharap Tugas Akhir ini dapat memberikan manfaat, baik bagi pengembangan ilmu pengetahuan maupun bagi pihak-pihak yang berkepentingan.

Wassalamu'alaikum Wr.Wb

Yogyakarta, 30 April 2026

Penulis

DAFTAR ISI

HALAMAN JUDUL	i
HALAMAN PERSETUJUAN PEMBIMBING TUGAS AKHIR.....	ii
HALAMAN PENGESAHAN TUGAS AKHIR	iii
KATA PENGANTAR.....	iv
DAFTAR ISI	v
DAFTAR TABEL	vii
DAFTAR GAMBAR.....	viii
DAFTAR LAMPIRAN	ix
PERNYATAAN	x
INTISARI.....	xi
ABSTRACT	xii
BAB I PENDAHULUAN	1
1.1. Latar Belakang Masalah.....	1
1.2. Rumusan Masalah	3
1.3. Batasan Masalah.....	3
1.4. Tujuan Penelitian	3
1.5. Manfaat Penelitian	4
BAB II TINJAUAN PUSTAKA	5
2.1. Penelitian Terdahulu	7
BAB III LANDASAN TEORI	11
3.1. <i>Text Mining</i> dan <i>Natural Language Processing</i> (NLP)	11
3.2. Pra-pemrosesan Teks	11
3.2.1. <i>Case Folding</i>	12
3.2.2. Pembersihan Teks	12
3.2.3. <i>Stopword Removal</i>	12
3.2.4. <i>Stemming</i>	12
3.3. <i>Topic Modelling</i> : BERTopic	13
3.3.1. <i>Document Embedding</i>	13
3.3.2. Reduksi Dimensi (UMAP).....	14
3.3.3. <i>Clustering</i> (HDBSCAN)	15
3.3.4. <i>Topic Representation</i> (c-TF-IDF)	16
3.3.5. Visualisasi dan Evaluasi Struktur Topik	18
3.4. Metrik Evaluasi Klasterisasi	18
3.4.1. <i>Silhouette Score</i>	19
3.4.2. <i>Davies-Bouldin Index</i>	19
BAB IV METODOLOGI PENELITIAN.....	20
4.1. Populasi dan Sampel Penelitian	20
4.2. Tempat dan Waktu Penelitian	20
4.3. Definisi Operasional Variabel Penelitian.....	20
4.4. Alat dan cara organisasi data.....	20
4.5. Metode Penelitian.....	21
BAB V PEMBAHASAN	23
5.1. Deskripsi Data Penelitian	23
5.2. Pra-pemrosesan Teks Aspirasi	26

5.3. Representasi Semantik Dokumen dengan Model Embedding Multilingual	28
5.4. Reduksi Dimensi dengan UMAP	28
5.5. Pembentukan Klaster Topik dengan HDBSCAN	29
5.6. Identifikasi Representasi Topik melalui c-TF-IDF	30
5.7. Peningkatan Kualitas Representasi Topik dengan KeyBERT dan MMR.	31
5.8. Hasil Pemodelan Topik dan Interpretasi Tematik.....	31
5.9. Evaluasi Model.....	39
5.10. Interpretasi dalam Mendukung Fungsi Ketugasan DPD RI	40
BAB VI PENUTUP	42
6.1. Kesimpulan	42
6.2. Saran.....	43
DAFTAR PUSTAKA.....	44
RINGKASAN TA	48
DAFTAR PUSTAKA.....	57
LAMPIRAN	61

DAFTAR TABEL

Tabel 2.1 Tabel Penelitian Sebelumnya	8
Tabel 4.1 Tabel Definisi Operasional Variabel Penelitian.....	20
Tabel 5.1 Statistik Panjang Dokumen	23
Tabel 5.2 Contoh Dokumen Aspirasi (5 Pertama)	24
Tabel 5.3 Contoh Dokumen Mentah (3 Sampel Acak)	24
Tabel 5.4 Contoh Hasil Pra-pemrosesan Teks Aspirasi	26
Tabel 5.5 Konfigurasi UMAP	28
Tabel 5.6 Konfigurasi HBDSCAN.....	29
Tabel 5.7 Konfigurasi c-TF-IDF	30
Tabel 5.8 Hasil Pemodelan Topik	31
Tabel 5.9 Daftar Topik Teratas	32
Tabel 5.10 Ringkasan 3 Kata pada Setiap Topik Berdasarkan Nilai c-TF-IDF...	37
Tabel 5.11 Evaluasi Model BERTopic	39

DAFTAR GAMBAR

Gambar 4.1 Tahapan Penelitian	22
Gambar 5.1 Histogram Distribusi Panjang Dokumen.....	25
Gambar 5.2 <i>Barchart</i> Distribusi Kata Kunci.....	33
Gambar 5.3 Visualisasi <i>Intertopic Distance Map</i>	36
Gambar 5.4 <i>Heatmap</i> Kemiripan Antartopik	37

DAFTAR LAMPIRAN

Lampiran 1	61
Lampiran 2	62
Lampiran 3	76

PERNYATAAN

Dengan ini saya menyatakan bahwa dalam Tugas Akhir ini tidak terdapat karya karya yang sebelumnya pernah diajukan untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan orang lain, kecuali yang diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Yogyakarta, 30 April 2026

Penulis

INTISARI

TOPIC MODELLING ASPIRASI MASYARAKAT MENGGUNAKAN METODE BERTOPIC

(Studi Kasus : Aspirasi Masyarakat di Dewan Perwakilan Daerah
Republik Indonesia)

Nail Anfasa Fadla

Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Islam Indonesia

Peningkatan volume aspirasi masyarakat yang disampaikan kepada Dewan Perwakilan Daerah Republik Indonesia (DPD RI) menuntut adanya pendekatan analitis yang lebih sistematis dan berbasis data. Penelitian ini mengaplikasikan metode BERTopic untuk mengidentifikasi topik-topik utama dari data aspirasi masyarakat Daerah Istimewa Yogyakarta yang dihimpun selama Tahun Sidang 2024/2025. Dataset terdiri dari 486 dokumen aspirasi yang seluruhnya valid tanpa nilai kosong. Proses analisis meliputi pra-pemrosesan teks berbahasa Indonesia, pembentukan representasi semantik menggunakan model *sentence transformer* LazarusNLP/all-indo-e5-small-v4, reduksi dimensi dengan UMAP, klusterisasi berbasis kepadatan menggunakan HDBSCAN, serta ekstraksi kata kunci representatif melalui pendekatan c-TF-IDF yang diperkaya dengan KeyBERTInspired dan Maximal Marginal Relevance (MMR). Hasil pemodelan menghasilkan 15 topik tematik dengan 81 dokumen (16,7%) teridentifikasi sebagai *outlier*. Evaluasi kuantitatif menunjukkan Silhouette Score sebesar 0,4754 dan Davies-Bouldin Index sebesar 0,7238, yang mengindikasikan struktur kluster dengan kohesi internal yang cukup baik dan separasi antartopik yang memadai. Topik dengan dokumen terbanyak adalah Pendidikan, Perempuan dan Anak (98 dokumen), sementara skor c-TF-IDF tertinggi ditemukan pada topik berkarakter spesifik seperti Aktivitas Lansia (0,9628), Pertanian, Irigasi, Embung, dan Hasil Tani (0,8485), dan Alokasi Anggaran dan Pembangunan Infrastruktur (0,8173). Pemetaan lintas komite DPD RI menunjukkan bahwa Komite III memiliki cakupan aspirasi terluas dengan tujuh topik relevan, sedangkan Komite II menerima sinyal aspirasi terkuat pada sektor pertanian. Penelitian ini membuktikan bahwa BERTopic mampu mengolah data teks aspirasi tidak terstruktur menjadi representasi tematik yang informatif, sehingga berpotensi menjadi instrumen analitik berbasis data dalam mendukung fungsi representasi, legislasi, dan pengawasan DPD RI secara lebih responsif dan akuntabel.

Kata Kunci : BERTopic, *topic modeling*, aspirasi masyarakat, DPD RI, *Natural Language Processing*, c-TF-IDF, HDBSCAN.

ABSTRACT

TOPIC MODELING OF PUBLIC ASPIRATIONS USING BERTOPIC (Case Study : Public Aspirations at the Regional Representative Council of the Republic of Indonesia)

Nail Anfasa Fadla
Department of Statistics, Faculty of Matematics and Natural Sciences
Universitas Islam Indonesia

The increasing volume of public aspirations submitted to the Regional Representative Council of the Republic of Indonesia (DPD RI) demands a more systematic and data-driven analytical approach. This study applies the BERTopic method to identify key topics from public aspiration data collected in the Special Region of Yogyakarta during the 2024/2025 Legislative Session. The dataset comprises 486 aspiration documents, all of which were valid with no missing values. The analysis pipeline included Indonesian text preprocessing, semantic representation using the LazarusNLP/all-indo-e5-small-v4 sentence transformer model, dimensionality reduction via UMAP, density-based clustering using HDBSCAN, and representative keyword extraction through c-TF-IDF enriched with KeyBERTInspired and Maximal Marginal Relevance (MMR). The modeling produced 15 thematic topics, with 81 documents (16.7%) identified as outliers. Quantitative evaluation yielded a Silhouette Score of 0.4754 and a Davies-Bouldin Index of 0.7238, indicating a cluster structure with adequate internal cohesion and sufficient inter-topic separation. The topic with the highest document count was Education, Women, and Children (98 documents), while the highest c-TF-IDF scores were observed in highly specific topics such as Elderly Activities (0.9628), Agriculture, Irrigation, Reservoirs, and Agricultural Products (0.8485), and Budget Allocation and Infrastructure Development (0.8173). Cross-committee mapping across DPD RI revealed that Committee III holds the broadest aspiration coverage with seven relevant topics, while Committee II received the strongest aspiration signal in the agricultural sector. This study demonstrates that BERTopic is capable of transforming unstructured aspiration text data into informative thematic representations, positioning it as a promising data-driven analytical instrument to support DPD RI's functions of representation, legislation, and oversight in a more responsive and accountable manner..

Keywords: BERTopic, topic modeling, public aspirations, DPD RI, Natural Language Processing, c-TF-IDF, HDBSCAN.

BAB I

PENDAHULUAN

1.1. Latar Belakang Masalah

Dalam beberapa tahun terakhir, jumlah aspirasi masyarakat yang disampaikan melalui anggota Dewan Perwakilan Daerah Republik Indonesia (DPD RI) menunjukkan peningkatan yang signifikan. Beragam isu mulai dari ekonomi, pendidikan, kesehatan, hingga tata kelola pemerintahan kini banyak disampaikan oleh masyarakat kepada para anggota DPD RI sebagai representasi daerah. Dinamika ini mencerminkan meningkatnya kesadaran publik terhadap peran DPD RI sebagai saluran aspirasi politik dan kebijakan. Di sisi lain, peningkatan volume dan kompleksitas aspirasi tersebut menuntut adanya mekanisme pengelolaan informasi yang lebih sistematis, efisien, dan berbasis data agar setiap masukan masyarakat dapat direspons secara tepat dan berdaya guna (Rusdiyantara, 2025).

Dalam konteks lembaga perwakilan seperti Dewan Perwakilan Daerah Republik Indonesia (DPD RI), perubahan ini menghadirkan peluang sekaligus tantangan. Sebagai lembaga negara yang memiliki fungsi representasi daerah, DPD RI dituntut untuk menyerap dan menyalurkan aspirasi publik secara cepat, tepat, dan berbasis data. Akan tetapi, volume data aspirasi yang sangat besar, bersifat tidak terstruktur, dan beragam dari segi isu maupun gaya bahasa menjadikan proses analisis manual tidak lagi memadai. Tanpa alat bantu analitis yang efektif, potensi informasi penting dalam aspirasi publik dapat terabaikan, sehingga melemahkan peran DPD RI sebagai jembatan antara masyarakat daerah dan kebijakan nasional (Rusdiyantara, 2025).

Salah satu pendekatan yang relevan untuk menjawab tantangan ini adalah topic modeling, yaitu teknik analisis teks yang mampu mengidentifikasi tema-tema utama dari kumpulan data teks dalam jumlah besar. Model klasik seperti *Latent Dirichlet Allocation* (LDA) telah banyak digunakan untuk keperluan ini. Namun, LDA dan model serupa memiliki keterbatasan dalam menangkap makna kontekstual karena hanya bergantung pada frekuensi kata tanpa mempertimbangkan kedalaman relasi semantik. Perkembangan teknologi *transformer-based models* kemudian menghadirkan alternatif yang lebih unggul,

yaitu BERTopic, yang menggabungkan kekuatan representasi teks berbasis BERT, reduksi dimensi, dan *clustering* untuk menghasilkan topik yang lebih koheren dan interpretatif (Wahyuni et al., 2025).

Keunggulan BERTopic dibanding metode tradisional telah dibuktikan oleh berbagai penelitian. Wahyuni et al. (2025) menemukan bahwa BERTopic menghasilkan koherensi topik lebih baik dalam menganalisis komentar publik tentang program makan bergizi gratis. Simanjuntak et al. (2024) menunjukkan efektivitas BERTopic dalam memetakan opini publik terkait kendaraan listrik di Indonesia. Sementara itu, Myftaraj et al. (2024) menganalisis perkembangan riset *circular economy* dan *big data* menggunakan BERTopic yang dikombinasikan dengan embedding SciBERT, sehingga mampu mengidentifikasi struktur topik utama, keterkaitan antartema, serta dinamika perkembangan riset secara tematik dalam literatur ilmiah. Liu & Wan (2024) menerapkan BERTopic untuk memetakan tren penelitian pertanian presisi (*precision agriculture*) dengan menekankan dimensi temporal dan spasial, yang memungkinkan identifikasi evolusi topik riset dari waktu ke waktu serta perbedaan fokus penelitian antarwilayah. Di sisi lain, Nanyonga et al. (2025) menggunakan BERTopic untuk menganalisis laporan keselamatan penerbangan, dan menemukan bahwa pendekatan berbasis embedding transformer dan *density-based clustering* mampu menghasilkan topik yang lebih kontekstual dan mudah diinterpretasikan dibandingkan metode probabilistik tradisional.

Di Indonesia, penggunaan BERTopic dalam konteks lembaga negara masih relatif terbatas, sehingga penelitian ini menawarkan kebaruan sekaligus relevansi praktis. Dengan mengaplikasikan BERTopic pada data aspirasi masyarakat yang disampaikan kepada DPD RI, penelitian ini dapat membantu mengidentifikasi isu-isu utama yang berkembang di daerah secara lebih sistematis. Lebih jauh, hasil penelitian ini dapat memberikan kontribusi nyata bagi DPD RI dalam merumuskan rekomendasi kebijakan yang responsif, meningkatkan efektivitas fungsi representasi, serta memperkuat legitimasi lembaga di mata publik.

Bagi peneliti, ketertarikan mengangkat tema ini dilatarbelakangi oleh kebutuhan untuk menjawab kesenjangan antara ketersediaan data aspirasi masyarakat yang sangat besar dan keterbatasan metode tradisional dalam

mengolahnya. Penelitian ini juga diharapkan memperluas literatur ilmiah mengenai penerapan BERTopic dalam studi kebijakan publik di Indonesia, mengingat sebagian besar riset sebelumnya lebih banyak berfokus pada bidang pemasaran, layanan digital, atau isu kebijakan sektoral tertentu. Dengan demikian, penelitian ini bukan hanya berkontribusi pada pengembangan metode analisis teks, tetapi juga berperan dalam memperkuat kualitas demokrasi deliberatif, di mana suara masyarakat tidak hanya tersampaikan, tetapi juga benar-benar dipahami dan direspons dalam proses perumusan kebijakan.

1.2. Rumusan Masalah

1. Bagaimana distribusi dokumen dan skor representatif (*c-TF-IDF score*) pada setiap topik?
2. Apa saja topik utama yang terbentuk dari hasil klusterisasi aspirasi masyarakat menggunakan metode BERTopic, serta interpretasinya dalam mengidentifikasi isu prioritas masyarakat untuk mendukung fungsi representasi Dewan Perwakilan Daerah Republik Indonesia?

1.3. Batasan Masalah

1. Data yang digunakan terbatas pada aspirasi masyarakat yang diterima oleh anggota DPD RI, baik dari kanal resmi, kegiatan serap aspirasi maupun publikasi terbuka yang relevan.
2. Metode pemodelan topik yang digunakan adalah BERTopic, tanpa membandingkan secara teknis dengan metode *topic modeling* lainnya seperti LDA atau NMF
3. Analisis difokuskan pada identifikasi dan interpretasi tema/isu utama dari aspirasi masyarakat, bukan pada penilaian kebijakan atau rekomendasi implementatif secara langsung.
4. Ruang lingkup penelitian dibatasi pada aspirasi masyarakat dalam masa Tahun Sidang 2024/2025.

1.4. Tujuan Penelitian

1. Menganalisis distribusi dokumen dan skor representatif (*c-TF-IDF score*) pada setiap topik yang dihasilkan dari pemodelan aspirasi masyarakat menggunakan metode BERTopic.

2. Mengidentifikasi topik-topik utama yang terbentuk dari hasil klasterisasi aspirasi masyarakat serta menginterpretasikan temuan tersebut dalam rangka mengungkap isu prioritas masyarakat guna mendukung penguatan fungsi representasi Dewan Perwakilan Daerah Republik Indonesia.

1.5. Manfaat Penelitian

Penelitian ini diharapkan memberikan kontribusi teoretis dalam pengembangan kajian text mining dan topic modelling, khususnya melalui penerapan metode BERTopic pada data aspirasi masyarakat yang bersifat tidak terstruktur. Penelitian ini memperkaya literatur mengenai penggunaan pendekatan unsupervised learning dalam mengidentifikasi pola tematik dan struktur isu dari data teks, serta memperluas penerapannya dalam konteks representasi politik dan kelembagaan.

Secara praktis, penelitian ini menghasilkan pemetaan topik aspirasi masyarakat berdasarkan distribusi dokumen dan skor representatif (c-TF-IDF) pada setiap klaster yang terbentuk. Hasil analisis tersebut dapat dimanfaatkan oleh Dewan Perwakilan Daerah Republik Indonesia sebagai dasar identifikasi isu prioritas masyarakat secara lebih sistematis dan berbasis data. Selain itu, temuan penelitian ini dapat menjadi bahan pendukung dalam penyusunan rekomendasi kebijakan yang lebih responsif terhadap kebutuhan masyarakat serta mendukung peningkatan efektivitas pengelolaan dan pengolahan aspirasi dalam pelaksanaan fungsi representasi.

BAB II

TINJAUAN PUSTAKA

Topic modeling merupakan salah satu teknik dalam *Natural Language Processing (NLP)* yang digunakan untuk mengidentifikasi pola dan tema laten dari sekumpulan dokumen teks. Konsep ini berangkat dari kebutuhan untuk memahami data teks dalam jumlah besar yang tidak terstruktur, sehingga dapat dikelompokkan ke dalam topik-topik tertentu secara otomatis. Model klasik yang banyak digunakan antara lain Latent Dirichlet Allocation (LDA) dan Non-negative Matrix Factorization (NMF). LDA memodelkan dokumen sebagai distribusi topik dan topik sebagai distribusi kata, sementara NMF mendekomposisi matriks dokumen-kata menjadi representasi topik. Meskipun banyak digunakan, kedua metode ini memiliki keterbatasan. LDA misalnya, hanya mengandalkan frekuensi kata tanpa memperhatikan konteks semantik, sehingga topik yang dihasilkan sering kali sulit diinterpretasi (Grisales et al., 2023).

Untuk mengatasi keterbatasan tersebut, berkembang pendekatan baru yang memanfaatkan representasi berbasis *word embeddings* dan *transformer models*. Salah satu inovasi terkini adalah BERTopic, yang menggabungkan kekuatan BERT embeddings, teknik reduksi dimensi (UMAP), algoritma clustering (*HDBSCAN*), serta representasi topik berbasis c-TF-IDF. Pendekatan ini memungkinkan pengelompokan dokumen yang lebih koheren dan mampu menangkap relasi semantik antar kata maupun kalimat (Khodeir & Elghannam, 2025).

Sejak diperkenalkan oleh Grootendorst pada tahun 2022, BERTopic berkembang pesat dan telah digunakan di berbagai bidang. Dalam penelitian Wahyuni et al. (2025), metode ini digunakan untuk menganalisis persepsi publik terhadap program makan bergizi gratis di Indonesia melalui komentar YouTube. Hasil analisis menemukan sepuluh tema utama yang berkaitan dengan isu transparansi anggaran, keadilan distribusi, hingga dimensi politik program tersebut. BERTopic juga terbukti lebih unggul dibanding LDA, ditunjukkan dengan skor koherensi topik 0,46 dibandingkan 0,39 yang diperoleh LDA (Wahyuni et al., 2025).

Penelitian serupa dilakukan oleh Simanjuntak & Koyimatu (2024) yang mengidentifikasi opini publik terhadap kendaraan listrik di Indonesia menggunakan data komentar YouTube. Hasilnya menunjukkan bahwa BERTopic dengan *embedding* IndoBERTweet mampu menghasilkan topik yang memiliki koherensi tinggi (0,6–1,0) dan keberagaman (0,95), mencerminkan kemampuannya menangkap isu infrastruktur, ekonomi, dan lingkungan dalam transisi energi nasional (Simanjuntak et al., 2024).

Di bidang pendidikan, Khodeir & Elghannam (2024) membandingkan BERTopic dengan LDA, LSI, dan NMF dalam menganalisis forum MOOC. Hasil studi ini menemukan bahwa BERTopic lebih unggul ketika digunakan untuk mengidentifikasi banyak topik sekaligus, menghasilkan topik yang lebih semantik dan konsisten. Hal ini menegaskan keunggulan BERTopic dalam konteks diskusi daring yang memiliki data teks kompleks.

Sementara itu, Ma et al. (2025) menguji BERTopic untuk menganalisis risiko kardiovaskular terkait opioid pada perempuan menggunakan lebih dari 1.800 artikel ilmiah dari PubMed. Hasil penelitian menunjukkan bahwa BERTopic menghasilkan kluster topik yang lebih rapat, terpisah dengan jelas, dan memiliki koherensi yang lebih tinggi dibandingkan LDA. Integrasi BERTopic dengan AI bahkan memungkinkan interpretasi topik yang lebih cepat dan akurat.

Dalam ranah teknologi, Su et al. (2025) menggunakan BERTopic untuk mendeteksi tren penelitian Internet of Things (IoT) dan e-commerce berdasarkan 169 artikel dari Web of Science. Studi ini berhasil mengidentifikasi empat tema utama, yaitu transformasi model bisnis e-commerce, peran blockchain dalam keamanan data, integrasi logistik cerdas, serta perlindungan privasi. Penelitian ini menegaskan kemampuan BERTopic dalam menelusuri tren riset jangka panjang dan pola evolusi pengetahuan.

Selain itu, Rejeb et al. (2025) mengaplikasikan BERTopic untuk menganalisis literatur tentang blockchain dan smart cities. Hasil penelitian menunjukkan bahwa BERTopic unggul dalam menangkap hubungan semantik mendalam antar konsep, sehingga lebih baik dalam mengidentifikasi arah perkembangan riset kota pintar dibandingkan teknik *co-word analysis* tradisional.

Tidak hanya di ranah sosial atau teknologi, BERTopic juga mulai digunakan dalam konteks teknis seperti rekayasa perangkat lunak Gottumukkala et al. (2025a) menggabungkan BERTopic dengan *multi-output classifier* untuk memprediksi cacat perangkat lunak dan menemukan akar penyebab masalah. Hasilnya membuktikan bahwa BERTopic dapat digunakan tidak hanya untuk analisis opini publik, tetapi juga untuk mendukung proses teknis dalam *software engineering*.

Dari berbagai studi tersebut, terlihat jelas bahwa BERTopic memiliki fleksibilitas tinggi untuk diaplikasikan di berbagai domain, mulai dari pendidikan, kesehatan, teknologi, hingga isu-isu sosial-politik. Dalam konteks Indonesia, penelitian Wahyuni et al. (2025) dan Simanjuntak et al. (2024) memperlihatkan bahwa BERTopic efektif untuk menganalisis opini publik. Hal ini sangat relevan jika diaplikasikan pada analisis aspirasi masyarakat di DPD RI, di mana data aspirasi biasanya berupa teks tidak terstruktur dalam jumlah besar.

Melalui penerapan BERTopic, isu-isu utama yang berkembang di masyarakat dapat dipetakan secara sistematis dan berbasis data. Penelitian Aryani et al. (2024) yang menganalisis berita Pemilu 2024 dengan BERTopic juga menunjukkan bahwa metode ini mampu mengurai isu politik kompleks menjadi tema-tema yang lebih sederhana, mendukung perumusan kebijakan berbasis aspirasi masyarakat. Dengan demikian, penelitian ini akan memperkuat upaya pemanfaatan teknologi NLP modern dalam memperdalam pemahaman terhadap dinamika wacana publik serta mendukung pengambilan keputusan di lembaga legislatif.

2.1. Penelitian Terdahulu

Beberapa penelitian terdahulu yang relevan dengan topik penelitian ini disajikan pada **Tabel 2.1** berikut sebagai bahan rujukan dan pembandingan.

Tabel 2.1 Tabel Penelitian Sebelumnya

Tahun	Nama	Judul	Metode	Hasil Penelitian
2025	Nabila Khodeir, Fatma Elghannam	<i>Efficient Topic Identification for Urgent MOOC Forum Posts Using BERTopic and Traditional Topic Modeling Techniques</i>	Pemodelan topik menggunakan pendekatan komparatif antara LDA, LSI, NMF, dan BERTopic.	Meneliti penyebab urgensi pada postingan forum MOOC menggunakan LDA, LSI, NMF, dan BERTopic. Hasil menunjukkan NMF unggul untuk topik sedikit, sedangkan BERTopic menghasilkan topik lebih koheren dan lebih banyak ketika jumlah topik besar diperlukan. Model secara efektif mengungkap berbagai penyebab postingan mendesak.
2025	Devi Priya Gottumukkala, Prasad Reddy P.V.G.D, S. Krishna Rao	<i>Topic Modeling-Based Prediction of Software Defects and Root Cause Using BERTopic and Multioutput Classifier</i>	Integrasi BERTopic dengan <i>multi-output classifier</i> dan <i>voting ensemble</i> .	Mengembangkan model BERT-MOC dengan menggabungkan BERTopic dan multi-output classifier untuk memprediksi defect / non-defect sekaligus root cause defect. Model dengan voting ensemble mencapai akurasi hingga 97% untuk prediksi akar masalah dan 94% untuk prediksi keberadaan defect.
2025	Abderahman Rejeb, Karim Rejeb, Heba F. Zaher, Steve Simske	<i>Blockchain and Smart Cities: Co-Word Analysis and BERTopic Modeling</i>	Analisis bibliometrik menggunakan <i>co-word analysis</i> dan BERTopic.	Menggunakan co-word analysis dan BERTopic pada 576 artikel (2016–2025) untuk memetakan perkembangan riset blockchain di smart cities. Hasilnya menunjukkan blockchain berperan sebagai infrastruktur dasar yang aman dan interoperabel (terintegrasi dengan IoT, edge computing, smart contract) untuk layanan kota seperti transportasi, kesehatan, energi, dan waste management, serta adanya pergeseran fokus dari kajian umum ke aplikasi sektoral yang lebih spesifik.
2025	Yu-Sheng Su, Jun-qing Wang, Shou-Hsi Tu, Kuo-Ti Liao, Chien-Liang Lin	<i>Detecting Latent Topics and Trends in IoT and E-Commerce using BERTopic Modeling</i>	Pemodelan topik dan analisis tren temporal menggunakan BERTopic.	Menganalisis 169 artikel (2010–2024) Web of Science tentang IoT dan e-commerce menggunakan BERTopic untuk mengungkap tema utama dan tren evolusinya. Ditemukan empat tema inti: transformasi model bisnis e-commerce oleh IoT, peran blockchain dalam keamanan data dan trust, sinergi logistik cerdas dan e-commerce, serta isu privasi dan pengelolaan data pribadi. Studi juga menunjukkan pergeseran

Tahun	Nama	Judul	Metode	Hasil Penelitian
				fokus dari optimasi rantai pasok ke model bisnis cerdas berbasis data dan perlindungan privasi serta memberikan rekomendasi kebijakan dan strategi bagi pelaku industri.
2025	Widya Wahyuni, Tri Putri Lestari, Milla Apriliana, Riyang Gumelta	<i>Implementation of BERTopic for Topic Modeling Analysis of the Free Nutritious Meal Program Based on YouTube Comments</i>	Analisis komentar menggunakan BERTopic, dengan evaluasi koherensi dan <i>topic diversity</i> serta perbandingan dengan LDA.	Menggunakan BERTopic untuk menganalisis 19.843 komentar YouTube terkait Program Makan Bergizi Gratis di Indonesia. Dihasilkan 10 tema utama (mis. keterlibatan tokoh politik, transparansi anggaran, manfaat edukatif, dan pemerataan akses di daerah 3T). Evaluasi menunjukkan BERTopic mengungguli LDA (koherensi 0,46 vs 0,39; topic diversity 76% vs 71%). Temuan mengungkap bahwa persepsi publik bersifat multifaset, dipengaruhi konteks sosial, politik, dan ekonomi, dan dapat menjadi dasar perumusan kebijakan yang lebih responsif.
2025	Yiqiang Feng, Ziao Chen, Yuxin Zhang, Wenjuan Huang, Xuanming Zhang, Siyu He	<i>BERTopic_Teen: A Multi-Module Optimization Approach for Short Text Topic Modeling in Adolescent Health</i>	Pengembangan model BERTopic yang ditingkatkan dengan tiga modul: Popularity Deviation Regularizer (PDR), Dynamic Document Embedding Optimizer (DDEO), dan Probabilistic Reassignment Matrix (PRM), diuji pada 64.441 tweet kesehatan remaja.	BERTopic_Teen mengungguli LDA, NMF, Top2Vec, dan BERTopic orisinal di semua metrik. Mencapai peningkatan koherensi topik 16,1% (NPMI = 0,2184), keragaman topik lebih tinggi (TD = 0,9935), dan perplexity lebih rendah (1,7214). Berhasil mengklasifikasikan 61.039 dari 64.441 tweet menjadi 7 domain tematik utama.
2024	Kristine Angelina Simanjuntak, Muhamad Koyimatu, Yolla Putri Ervanisari, Tasmi	<i>Identifikasi Opini Publik Terhadap Kendaraan Listrik dari Data Komentar YouTube: Pemodelan Topik Menggunakan BERTopic</i>	Pemodelan topik berbasis BERTopic dengan <i>embedding</i> IndoBERTweet.	Mengumpulkan komentar YouTube tentang kendaraan listrik di Indonesia dan memodelkan topik opini publik menggunakan BERTopic dengan <i>embedding</i> IndoBERTweet. Topik-topik yang dihasilkan memiliki nilai koherensi sekitar 0,6–1 dan topic diversity 0,95838, yang menunjukkan kesamaan semantik yang kuat sekaligus keberagaman kata tinggi; hasil ini memberi gambaran isu utama dan persepsi pro-kontra masyarakat terhadap transisi kendaraan listrik.

Tahun	Nama	Judul	Metode	Hasil Penelitian
2023	Li Ma, Ru Chen, Weigong Ge1, Paul Rogers, Beverly Lyn-Cook, Huixiao Hong, Weida Tong, Ningning Wu, Wen Zou	<i>AI-Powered Topic Modeling: Comparing LDA and BERTopic in Analyzing Opioid-Related Cardiovascular Data</i>	Studi komparatif antara LDA dan BERTopic.	Membandingkan LDA dan BERTopic dalam menganalisis data kardiovaskular terkait opioid. BERTopic menghasilkan topik lebih stabil dan lebih mudah diinterpretasikan, sementara LDA memberikan struktur yang lebih generik. Studi menegaskan potensi model berbasis transformer untuk data medis.
2023	Andrés M. Grisales A., Sebastián Robledo, Martha Zuluaga	<i>Topic Modeling: Perspectives From a Literature Review</i>	<i>Literature review</i> dengan pendekatan bibliometrik dan scientometric terhadap perkembangan metode <i>topic modeling</i> .	Melakukan analisis bibliometrik dan scientometric terhadap perkembangan topic modeling, mengidentifikasi tren, model populer (LDA, pLSA, NMF, BERTopic), serta area aplikasi seperti media sosial, manajemen, dan politik. Menyimpulkan bahwa topic modeling sangat berkembang dan menjadi alat penting untuk analisis teks skala besar.
2015	Ricardo J. G. B. Campello, Davoud Moulavi, Arthur Zimek, Jörg Sander	<i>Hierarchical Density Estimates for Data Clustering, Visualization, and Outlier Detection</i>	Pengembangan metode clustering berbasis densitas HDBSCAN*.	Memperkenalkan HDBSCAN* sebagai metode clustering berbasis densitas yang menghasilkan hierarki lengkap untuk klaster, memungkinkan ekstraksi klaster “flat”, visualisasi struktur klaster, dan pendeteksian outlier global-lokal melalui GLOSH. Metode ini meningkatkan DBSCAN dan mengatasi keterbatasannya.
2015	Michael Röder, Andreas Both, Alexander Hinneburg	<i>Exploring the Space of Topic Coherence Measures</i>	Pengembangan dan evaluasi metrik topic coherence berbasis kombinasi komponen statistik dan korelasi dengan penilaian manusia.	Mengusulkan kerangka umum untuk membangun berbagai ukuran topic coherence berbasis kata, kemudian melakukan pencarian sistematis di ruang kombinasi komponen coherence dan menguji korelasinya dengan penilaian manusia. Ditemukan kombinasi baru ukuran coherence yang lebih baik daripada ukuran yang sudah ada dalam mendekati penilaian manusia, serta dibahas potensi aplikasinya pada text mining, information retrieval, dan web.

BAB III

LANDASAN TEORI

3.1. *Text Mining* dan *Natural Language Processing* (NLP)

Text Mining merupakan proses ekstraksi pola-pola bermakna dan pengetahuan tersembunyi dari sumber data teks yang tidak terstruktur, dengan tujuan mengubah data mentah tersebut menjadi informasi yang dapat dianalisis lebih lanjut (Aleqabie et al., 2024). Teknik-teknik yang digunakan dalam text mining mencakup tokenisasi, yang memecah teks menjadi unit-unit kecil seperti kata atau kalimat, serta *stemming* dan *lemmatization* yang mengurangi kata-kata ke bentuk dasarnya. Selain itu, teknik seperti klasterisasi dan klasifikasi digunakan untuk mengelompokkan dan mengategorikan teks sesuai dengan pola atau karakteristik tertentu. *Text Mining* sering kali diterapkan pada kumpulan data teks yang besar untuk menemukan tren atau hubungan yang tidak mudah dikenali secara manual, sehingga memberikan wawasan yang berguna untuk pengambilan keputusan atau penelitian lebih lanjut (Redddy et al., 2023).

Natural Language Processing (NLP), di sisi lain, adalah cabang dari kecerdasan buatan yang berfokus pada interaksi antara komputer dan bahasa manusia. NLP bertujuan untuk memungkinkan komputer untuk memahami, menganalisis, dan menghasilkan bahasa alami dalam bentuk teks atau suara. Dalam NLP, teknik yang digunakan meliputi analisis sintaksis untuk memahami struktur kalimat, serta analisis semantik untuk menafsirkan makna kata dan kalimat dalam konteks tertentu. Selain itu, NLP juga mencakup tugas-tugas seperti *part-of-speech tagging* untuk mengidentifikasi kategori kata, *named entity recognition* untuk mengenali entitas seperti nama orang atau tempat, dan analisis sentimen untuk memahami opini atau perasaan yang terkandung dalam teks. Dengan memanfaatkan NLP, komputer dapat menangani teks secara lebih cerdas, menginterpretasikan makna di balik kata-kata, dan menghasilkan respons yang relevan dalam interaksi manusia-komputer (Zare et al., 2025).

3.2. Pra-pemrosesan Teks

Sebelum data teks dapat dianalisis secara komputasional, diperlukan serangkaian tahapan pra-pemrosesan (preprocessing) guna mengubah teks mentah

yang bersifat tidak terstruktur menjadi representasi yang lebih bersih, konsisten, dan siap diolah oleh model. Tahapan ini merupakan salah satu komponen krusial dalam alur kerja Natural Language Processing (NLP), sebab kualitas teks masukan secara langsung memengaruhi akurasi dan kebermanaknaan hasil analisis yang diperoleh (HaCohen-Kerner et al., 2020).

3.2.1. Case Folding

Case folding adalah proses konversi seluruh karakter teks ke dalam bentuk huruf kecil (*lowercase*). Langkah ini bertujuan untuk menyeragamkan representasi kata sehingga variasi penulisan seperti Pendidikan dan pendidikan diperlakukan sebagai entitas yang identik oleh sistem (Manning et al., 2009).

3.2.2. Pembersihan Teks

Tahap pembersihan teks mencakup serangkaian operasi untuk menghilangkan elemen yang tidak membawa informasi semantik relevan. Penghapusan URL dan tautan daring dilakukan menggunakan ekspresi reguler (*regular expression*), mengingat tautan tersebut tidak berkontribusi pada representasi topik. Karakter numerik dieliminasi karena angka-angka dalam teks umumnya bersifat kontekstual dan tidak informatif secara tematik. Tanda baca dihapus secara menyeluruh untuk menghindari gangguan pada proses tokenisasi, diikuti dengan normalisasi spasi ganda menjadi spasi tunggal guna memastikan keseragaman struktur token (Manning et al., 2009)..

3.2.3. Stopword Removal

Stopword removal adalah proses penghapusan kata-kata yang sangat umum dan tidak membawa makna tematik yang diskriminatif. Proses ini umumnya dilakukan dalam dua lapisan: menggunakan daftar stopwords baku dari pustaka bahasa (seperti Sastrawi untuk Bahasa Indonesia), dan menggunakan daftar custom stopwords yang dikembangkan secara domain-spesifik sesuai karakteristik korpus yang dianalisis (Manning et al., 2009).

3.2.4. Stemming

Stemming adalah proses mereduksi setiap kata ke bentuk dasarnya (*root word*). Variasi morfologis seperti 'pembangunan', 'membangun', dan 'dibangun' direduksi menjadi representasi konsep yang sama, yaitu 'bangun'. Untuk Bahasa Indonesia, algoritma Sastrawi mengadaptasi pendekatan Nazief-Adriani yang telah

terbukti efektif dalam menangani kekayaan morfologi bahasa aglutinatif (Yudhana et al., 2019). Stemming secara signifikan memperkaya kepadatan semantik pada setiap dokumen dan mengurangi dimensi ruang fitur yang perlu dimodelkan.

3.3. *Topic Modelling: BERTopic*

BERTopic menghasilkan topik-topik yang dapat diinterpretasikan dan bermakna secara semantik dengan mengintegrasikan teknik *embedding* berbasis transformer dengan pendekatan *clustering* dan strategi *Term Frequency-Inverse Document Frequency* (TF-IDF) yang telah disesuaikan. Pendekatan ini memungkinkan identifikasi topik-topik yang tidak hanya relevan secara statistik, tetapi juga mencerminkan struktur semantik yang mendalam dalam kumpulan data teks. Dengan memanfaatkan *embedding transformer*, BERTopic mampu menangkap representasi kontekstual kata-kata dalam dokumen, sementara teknik clustering dan TF-IDF digunakan untuk mengelompokkan dokumen berdasarkan kemiripan makna, menghasilkan topik-topik yang lebih bermakna dan dapat ditafsirkan dalam konteks analisis teks yang lebih luas. (Gottumukkala et al., 2025).

3.3.1. *Document Embedding*

Langkah pertama dalam BERTopic adalah mengonversi data teks menjadi embedding BERT. Embedding BERT mengkodekan informasi kontekstual dan semantik dalam kalimat, yang sangat penting dalam aplikasi pemrosesan bahasa alami (*Natural Language Processing*, NLP). Untuk keperluan ini, digunakan model "Sentence-BERT", yang sangat efektif dalam mengubah kalimat atau paragraf menjadi representasi vektor berdimensi tinggi yang padat. Model pra-latih "all-MiniLM-L6-v2" digunakan untuk bahasa Inggris, sementara untuk bahasa multibahasa digunakan model "paraphrase-multilingual-MiniLM-L12-v2". (Reimers & Gurevych, 2019).

Representasi vektor yang dihasilkan bersifat dense dan terdistribusi dalam ruang geometris di mana kedekatan antarvektor merepresentasikan kemiripan semantik antardokumen. Karakteristik distribusi embedding ini menjadi prasyarat utama bagi algoritma clustering berbasis jarak seperti UMAP dan HDBSCAN, sebab performa klasterisasi sangat dipengaruhi oleh seberapa baik model embedding mampu memisahkan dokumen-dokumen dari topik yang berbeda ke

dalam wilayah-wilayah vektor yang terdistribusi secara terkluster dan tidak tumpang tindih.

3.3.2. Reduksi Dimensi (UMAP)

Dalam pipeline BERTopic, proses reduksi dimensi memegang peranan penting karena representasi vektor yang dihasilkan oleh model sentence-transformer memiliki dimensionalitas yang relatif tinggi, umumnya berada pada rentang 384 hingga 768 dimensi, sehingga belum optimal untuk langsung digunakan pada algoritma pengelompokan berbasis kepadatan seperti HDBSCAN. Oleh karena itu, digunakan metode *Uniform Manifold Approximation and Projection (UMAP)* untuk mereduksi dimensi data (McInnes et al., 2017).

UMAP mempertahankan struktur global data, memastikan bahwa titik data yang berdekatan dalam ruang asal juga berdekatan dalam ruang embedding setelah reduksi. Parameter `n_neighbors` mencerminkan keseimbangan antara struktur lokal dan global dalam ruang embedding: nilai yang terlalu kecil menghasilkan kluster yang terlalu granular dan tidak stabil, sementara nilai yang terlalu besar cenderung meratakan perbedaan antarkelompok topik yang secara semantik berbeda. Parameter `min_dist` yang mendekati nilai 0,0 mendorong titik-titik data dalam ruang laten untuk berkumpul lebih rapat (*tightly packed*), menghasilkan kluster yang lebih padat dan terpisah dengan jelas, konfigurasi yang direkomendasikan dalam implementasi BERTopic untuk meningkatkan kualitas pengelompokan (Grootendorst, 2022).

Pemilihan metric cosine dalam konfigurasi UMAP didasarkan pada sifat inheren representasi vektor berbasis transformer, di mana kemiripan semantik antardokumen lebih akurat direpresentasikan melalui sudut antarvektor dibandingkan jarak Euclidean absolut (Feng et al., 2025). Cosine similarity secara efektif mengabaikan perbedaan magnitudo vektor dan berfokus sepenuhnya pada arah semantiknya, sehingga relevan untuk data teks yang bervariasi panjangnya. Meski demikian, cosine similarity pada embedding tertentu dapat menghasilkan nilai yang arbitrer akibat regularisasi implisit model, sehingga penerapannya tetap memerlukan validasi empiris yang cermat (Steck et al., 2024).

Dasar matematika dari UMAP melibatkan aspek-aspek topologi, teori grafik, dan optimisasi. Berikut adalah rangkuman komponen-komponen utama dari UMAP.

Matriks Korelasi Sparse, yang dapat digunakan untuk membangun grafik KNN sebagai berikut: Hitung jarak pasangan antar titik data dalam dataset X. Misalkan $d(a, b)$ menyatakan jarak antara dua titik data a dan b dalam ruang n -dimensi yang sesuai (Gottumukkala et al., 2025).

$$d(a, b) = \sqrt{\sum_{i=1}^n (b_i - a_i)^2} \quad (3.1)$$

Pada persamaan (3.1), a_i, b_i adalah vektor Euclidean dalam ruang n -dimensi. Pilih $N(a)$ sebagai himpunan k tetangga terdekat dari data point a . Untuk itu, bangun *Simplicial Fuzzy Set*: Hitung derajat keanggotaan (berat) antara pasangan titik data menggunakan fungsi kernel. Misalkan P_{ij} mewakili derajat keanggotaan (berat) antara dua titik data x_i dan x_j , yang dapat dihitung menggunakan fungsi kernel:

$$P_{ij} = \exp\left(-\frac{d(x_i, x_j) - \rho_i}{\sigma_i}\right) \quad (3.2)$$

Pada persamaan (3.2), untuk setiap x_i , tentukan ρ_i dan σ_i , dimana ρ_i adalah jarak minimum dari x_i dan tentukan σ_i sebagai nilai jarak untuk yang memenuhi persamaan:

$$\sum_{j=1}^k \exp\left(\frac{-d(x_i, x_j) - \rho_i}{\sigma_i}\right) = \log_2(k) \quad (3.3)$$

(Gottumukkala et al., 2025)

3.3.3. Clustering (HDBSCAN)

DBSCAN adalah algoritma klusterisasi non-parametrik yang mengelompokkan titik data berdasarkan kerapatannya. Algoritma ini digunakan untuk mengelompokkan titik data ke dalam kluster berdasarkan kepadatan spasial menggunakan dua parameter yang didefinisikan pengguna: ϵ (epsilon) sebagai jarak terbesar antar pasangan dalam kluster yang sama dan *MinPts* sebagai jumlah minimum titik data yang diperlukan untuk mendefinisikan suatu area padat.

Berbeda dengan metode pengelompokan konvensional seperti K-Means yang mensyaratkan jumlah kluster ditentukan secara apriori, HDBSCAN (Hierarchical

DBSCAN) beroperasi berdasarkan estimasi kepadatan lokal dalam ruang embedding, sehingga mampu mengidentifikasi kelompok-kelompok topik secara otomatis tanpa asumsi terhadap bentuk geometris klaster maupun jumlah topik yang diharapkan. Karakteristik ini menjadikan HDBSCAN sangat sesuai untuk data yang bersifat tematik heterogen dan tidak terdistribusi secara merata (McInnes et al., 2017).

Titik data P dianggap sebagai titik utama jika terdapat paling tidak MinPts titik data dalam jarak eps dari P. Titik data yang bukan titik utama namun berada dalam jarak eps dari titik utama disebut titik batas. Dua titik data P dan Q dikatakan langsung terhubung-densitas jika keduanya adalah titik utama dan berada dalam jarak eps satu sama lain. Titik data yang tidak dapat dijangkau secara densitas oleh titik lainnya diklasifikasikan sebagai noise atau outlier (Stewart & Al-Khassaweneh, 2022).

Beberapa parameter kunci HDBSCAN yang mempengaruhi kualitas klasterisasi adalah sebagai berikut. Parameter `min_cluster_size` secara langsung mengendalikan granularitas topik yang dihasilkan: nilai yang lebih kecil menghasilkan lebih banyak klaster dengan cakupan tematik sempit, sementara nilai yang lebih besar mendorong terbentuknya klaster dengan tema yang lebih umum dan inklusif. Parameter `min_samples` berfungsi sebagai mekanisme kendali sensitivitas terhadap noise: semakin besar nilainya, semakin ketat persyaratan keanggotaan klaster, sehingga lebih banyak dokumen yang berpotensi dilabelkan sebagai noise (label -1). Metode seleksi klaster Excess of Mass (EOM) cenderung menghasilkan jumlah klaster lebih banyak dengan ukuran yang bervariasi dibandingkan metode alternatif 'leaf'. Parameter `prediction_data=True` memungkinkan kalkulasi probabilitas keanggotaan dokumen terhadap klaster yang terbentuk, yang dimanfaatkan oleh BERTopic dalam proses penugasan topik secara probabilistik.

3.3.4. *Topic Representation (c-TF-IDF)*

Setiap klaster dianggap sebagai sebuah topik, di mana dokumen-dokumen yang diklasterkan digabungkan menjadi sebuah dokumen tunggal (pseudo-document). Berbeda dengan TF-IDF konvensional yang beroperasi pada level dokumen individual, c-TF-IDF memperlakukan seluruh dokumen dalam satu

klaster sebagai pseudo-document tunggal, sehingga bobot term yang dihasilkan mencerminkan kekhasan leksikal suatu topik secara kolektif relatif terhadap topik-topik lainnya. Pendekatan ini lebih selaras dengan tujuan pemodelan topik, yakni mengidentifikasi kosakata yang paling diskriminatif dan representatif bagi masing-masing kelompok tematik dalam korpus.

Untuk membuat topik menjadi lebih mudah dipahami, BERTopic mengekstrak kata-kata yang paling sering atau mewakili dalam setiap klaster dokumen menggunakan skor c-TF-IDF $W_{x,c}$ (Gottumukkala et al., 2025b):

$$W_{x,c} = tf_{x,c} \times \log \left(1 + \frac{A}{f_x} \right) \quad (3.4)$$

di mana:

- $tf_{x,c}$ = frekuensi kata x dalam kelas c
- f_x = frekuensi kata x di seluruh kelas
- A = rata-rata jumlah kata per kelas

Parameter-parameter penting dalam konfigurasi CountVectorizer untuk c-TF-IDF meliputi: `ngram_range` yang mengakomodasi ekstraksi fitur pada level unigram maupun bigram secara simultan untuk bahasa yang kerap mengekspresikan konsep substantif melalui frasa dua kata; `min_df` yang menyaring term yang hanya muncul pada satu dokumen dan umumnya merupakan kesalahan pengetikan atau singkatan tidak informatif; serta `max_df` yang berfungsi sebagai mekanisme stopwords adaptif untuk mengeksklusi term yang distribusionalnya terlalu umum sehingga tidak berkontribusi pada diferensiasi antartopik .

Dua mekanisme pembobotan tambahan yang dapat meningkatkan kualitas representasi adalah `reduce_frequent_words`, yang mereduksi dominasi kata-kata dengan frekuensi global tinggi meskipun tidak termasuk stopwords standar; serta `bm25_weighting`, yang mengsubstitusi skema TF-IDF standar dengan fungsi Best Match 25 (BM25). BM25 mengatasi dua kelemahan fundamental TF-IDF konvensional: ia menerapkan saturasi frekuensi term yang mencegah dokumen panjang mendapatkan bobot berlebihan hanya karena mengulang sebuah kata berkali-kali, dan ia menormalisasi panjang dokumen secara lebih halus sehingga term dalam dokumen pendek mendapatkan bobot yang proporsional (Robertson & Zaragoza, 2009).

3.3.5. Visualisasi dan Evaluasi Struktur Topik

BERTopic menyediakan beberapa mekanisme visualisasi untuk memahami dan mengevaluasi struktur topik yang dihasilkan.

Barchart distribusi kata kunci menampilkan setiap topik yang direpresentasikan oleh sejumlah kata kunci beserta skor c-TF-IDF-nya dalam bentuk diagram batang horizontal. Visualisasi ini memungkinkan pembacaan langsung terhadap struktur leksikal setiap topik dan menjadi basis utama dalam pemberian nama tematik (Grootendorst, 2022).

Intertopic Distance Map memproyeksikan setiap topik sebagai titik dalam ruang vektor dua dimensi, di mana jarak antartitik mencerminkan tingkat kemiripan semantik antartopik. Topik-topik yang berdekatan dalam peta ini berbagi kesamaan tematik yang substansial, sementara topik yang berjauhan merepresentasikan domain isu yang secara semantik berbeda. Ukuran setiap titik umumnya proporsional terhadap jumlah dokumen anggota kluster (Grootendorst, 2022).

Heatmap Kemiripan Antartopik menyajikan matriks simetris berukuran $n \times n$ (dengan n adalah jumlah topik valid), di mana intensitas warna setiap sel merepresentasikan derajat kemiripan semantik antara pasangan topik yang bersesuaian. Diagonal utama matriks secara definitif bernilai 1,0 (kemiripan sempurna dengan diri sendiri), sementara nilai di luar diagonal yang tinggi mengindikasikan potensi tumpang tindih tematik antara dua topik. Heatmap ini berfungsi sebagai instrumen diagnostik yang membantu mengidentifikasi apakah struktur topik yang dihasilkan memiliki diferensiasi yang memadai, atau apakah terdapat pasangan topik yang secara semantik terlalu dekat sehingga dapat dipertimbangkan untuk digabungkan melalui tahap topic merging pada analisis lanjutan (Grootendorst, 2022).

3.4. Metrik Evaluasi Klasterisasi

Evaluasi kualitas model klasterisasi topik dilakukan melalui metrik-metrik statistik yang mengukur kualitas geometris struktur kluster yang terbentuk. Dua metrik yang umum digunakan secara komplementer adalah *Silhouette Score* dan *Davies-Bouldin Index*.

3.4.1. *Silhouette Score*

Silhouette Score mengukur seberapa baik setiap dokumen terposisi dalam klasternya sendiri relatif terhadap klaster tetangga yang paling dekat, dengan rentang nilai dari -1 hingga $+1$. Nilai mendekati $+1$ mengindikasikan bahwa dokumen-dokumen dalam suatu klaster sangat kohesif secara internal sekaligus terpisah dengan baik dari klaster-klaster lainnya, kondisi ideal yang mencerminkan struktur topik yang tajam dan terdefinisi dengan jelas. Nilai di sekitar 0 menandakan bahwa dokumen berada pada batas antara dua klaster, sementara nilai negatif mengindikasikan kemungkinan salah penugasan klaster (Rousseeuw, 1987).

3.4.2. *Davies-Bouldin Index*

Davies-Bouldin Index memberikan perspektif evaluasi yang komplementer dengan mengukur rasio antara dispersi intra-klaster dan separasi inter-klaster untuk setiap pasangan klaster, kemudian mengambil rata-ratanya. Berbeda dengan *Silhouette Score* yang bernilai lebih tinggi untuk klaster yang lebih baik, *Davies-Bouldin Index* bersifat terbalik: nilai yang lebih rendah mengindikasikan klaster yang lebih kompak secara internal sekaligus lebih terpisah satu sama lain. Kedua metrik ini bersama-sama menyediakan gambaran kuantitatif yang lebih komprehensif terhadap kualitas struktur pengelompokan, karena masing-masing menangkap aspek geometris yang berbeda dari kualitas klaster (Davies & Bouldin, 1979).

BAB IV

METODOLOGI PENELITIAN

4.1. Populasi dan Sampel Penelitian

Populasi dalam penelitian ini adalah seluruh aspirasi masyarakat yang diterima oleh anggota Dewan Perwakilan Daerah Republik Indonesia (DPD RI) melalui berbagai saluran, baik digital maupun non-digital, selama periode kerja anggota DPD RI. Populasi ini mencakup aspirasi yang diterima secara langsung oleh anggota DPD RI, maupun disalurkan melalui forum reses, rapat dengar pendapat, dan media sosial resmi anggota DPD RI.

Sampel dalam penelitian ini adalah 486 aspirasi masyarakat yang masuk dan ditampung oleh anggota DPD RI Dapil Daerah Istimewa Yogyakarta (DIY) selama Tahun Sidang 2024/2025.

4.2. Tempat dan Waktu Penelitian

Penelitian ini dilakukan di Kantor Perwakilan Daerah Istimewa Yogyakarta dengan rentang waktu November-Desember 2025.

4.3. Definisi Operasional Variabel Penelitian

Tabel 4.1 Tabel Definisi Operasional Variabel Penelitian

Variabel	Notasi	Skala	Definisi Operasional
Teks Aspirasi Masyarakat	T	Data string	Kalimat atau paragraf yang disampaikan oleh masyarakat DIY kepada anggota DPD RI DIY.

4.4. Alat dan cara organisir data

Penelitian ini menggunakan pendekatan komputasional berbasis bahasa pemrograman Python versi 3.9 ke atas, dengan dukungan pustaka (*library*) utama seperti BERTopic, UMAP-learn, HDBSCAN, HuggingFace Transformers, serta Scikit-learn dan Pandas untuk pengolahan data.

Pengumpulan data dilakukan melalui dokumentasi digital aspirasi masyarakat yang telah ditampung oleh anggota DPD RI Daerah Istimewa Yogyakarta (DIY) selama Tahun Sidang 2024/2025. Data ini diperoleh melalui hasil himpunan dari catatan hasil reses, forum dengar pendapat, serta sumber lain yang sah secara administratif. Data dikompilasi dalam bentuk teks digital yang siap diproses secara komputerisasi.

Setelah data terkumpul, dilakukan proses pra-pemrosesan teks, yang meliputi penghapusan karakter khusus, normalisasi huruf kapital, penghapusan stopwords dalam Bahasa Indonesia, dan konversi kalimat menjadi format bersih 8. Selanjutnya, setiap entri aspirasi masyarakat dikonversi menjadi representasi vektor menggunakan model *Sentence-BERT* (SBERT). Hasil embedding kemudian direduksi dimensinya menggunakan algoritma UMAP (*Uniform Manifold Approximation and Projection*) agar dapat dianalisis secara efisien pada tahap klusterisasi.

Dokumen yang telah direpresentasikan dalam ruang berdimensi rendah selanjutnya dikelompokkan ke dalam topik-topik menggunakan algoritma HDBSCAN (*Hierarchical Density-Based Spatial Clustering of Applications with Noise*). Setelah terbentuk kluster topik, dilakukan ekstraksi kata kunci menggunakan pendekatan *CountVectorizer* yang diikuti dengan penerapan *class-based* TF-IDF (c-TF-IDF), sehingga diperoleh kata-kata representatif yang menjelaskan masing-masing topik. Data hasil akhir akan divisualisasikan dalam bentuk grafik, word cloud, dan matriks koherensi untuk mendukung interpretasi secara kualitatif dan kuantitatif.

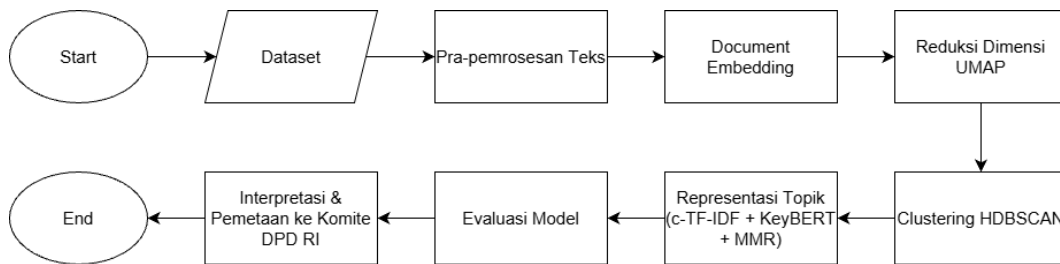
4.5. Metode Penelitian

Penelitian ini tergolong sebagai penelitian kuantitatif eksploratif yang menggunakan pendekatan *unsupervised learning* atau pembelajaran tanpa label untuk mengungkap pola tersembunyi (*latent patterns*) dari aspirasi masyarakat dalam bentuk teks tidak terstruktur. Fokus utama penelitian adalah penerapan metode statistik modern berbasis *Natural Language Processing (NLP)*, yaitu BERTopic, dengan tujuan mengeksplorasi dan memetakan tema utama dari aspirasi masyarakat di lingkup DPD RI DIY.

Metode pemilihan sampel yang digunakan adalah *non-probability sampling* dengan teknik *purposive sampling*, mengingat data yang digunakan bersifat spesifik, yaitu aspirasi masyarakat yang masuk kepada anggota DPD RI DIY selama Tahun Sidang 2024/2025. Teknik ini dipilih karena data hanya tersedia dalam rentang waktu tertentu dan dari wilayah representatif tunggal. Pemilihan sampel dilakukan dengan kriteria bahwa data harus berupa kalimat dalam Bahasa

Indonesia yang mengandung substansi aspiratif (opini, permintaan, usulan, atau kritik) yang relevan secara tematik dengan peran dan fungsi DPD RI.

Unit analisis dalam penelitian ini adalah setiap entri aspirasi masyarakat yang berupa satuan dokumen teks tunggal, baik itu komentar, paragraf aspirasi, atau narasi masyarakat yang masuk dalam arsip aspirasi DPD RI DIY. Metode ini diharapkan mampu memetakan topik-topik dominan dalam aspirasi masyarakat secara objektif berdasarkan teknik statistik yang telah dijelaskan sebelumnya.



Gambar 4.1 Tahapan Penelitian

Dalam pemodelan topik, metode BERTopic terdiri atas tiga tahap utama, yaitu pembentukan *document embedding*, pengelompokan dokumen ke dalam kluster berdasarkan kemiripan semantik, serta penyusunan representasi topik untuk setiap kluster. Tahapan awal dilakukan dengan menghasilkan *document embedding* guna merepresentasikan dokumen dalam ruang vektor. Selanjutnya, reduksi dimensi vektor dilakukan menggunakan Uniform Manifold Approximation and Projection (UMAP), kemudian proses klusterisasi diterapkan dengan metode Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN). Setelah proses pemodelan topik selesai, tahap berikutnya adalah evaluasi model yang dilakukan menggunakan metrik *topic coherence* dan *topic diversity*. (Mertayasa & Darmawan, 2022).

BAB V

PEMBAHASAN

5.1. Deskripsi Data Penelitian

Data yang digunakan merupakan Data Aspirasi Masyarakat (Asmas) DPD RI, khususnya yang berasal dari wilayah Daerah Istimewa Yogyakarta. Dataset dimuat dari berkas bertipe CSV dan disimpan dalam struktur dataframe, dengan kolom utama yang menjadi objek analisis adalah kolom aspirasi, yaitu teks aspirasi yang disampaikan oleh masyarakat secara langsung.

Langkah pertama yang dilakukan adalah memeriksa integritas data secara umum, meliputi jumlah kolom, jumlah total dokumen, serta keberadaan nilai kosong (*missing values*) pada kolom teks. Dokumen yang tidak mengandung teks aspirasi (nilai kosong) diidentifikasi dan dipisahkan, sehingga diperoleh jumlah dokumen valid yang dapat diproses.

Jumlah dokumen yang berhasil dihimpun sebanyak 486 entri, yang seluruhnya dinyatakan valid tanpa adanya nilai kosong (*missing values*). Hal ini menunjukkan bahwa kualitas data dari sisi kelengkapan tergolong sangat baik, sehingga tidak memerlukan tahapan penanganan data hilang (*missing value handling*) pada proses prapemrosesan.

Selanjutnya, dilakukan analisis statistik deskriptif terhadap panjang dokumen, baik dalam satuan jumlah kata maupun jumlah karakter sebagaimana ditampilkan pada **Tabel 5.1**.

Tabel 5.1 Statistik Panjang Dokumen

Statistik	Kata	Karakter
Rata-rata	36 kata	277,7 karakter
Minimum	6 kata	48 karakter
Maksimum	110 kata	779 karakter
Median	32 kata	250 karakter

Rata-rata panjang dokumen adalah 36 kata atau setara dengan 277,7 karakter. Dokumen terpendek terdiri dari 6 kata (48 karakter), sedangkan dokumen terpanjang mencapai 110 kata (779 karakter), mengindikasikan adanya perbedaan tingkat kedalaman dan detail dalam penyampaian aspirasi masyarakat. Nilai median

sebesar 32 kata (250 karakter) yang relatif mendekati rata-rata menunjukkan bahwa distribusi panjang dokumen cenderung tidak terlalu menceng ke arah tertentu (relatif simetris), sehingga tidak didominasi oleh dokumen yang sangat panjang maupun sangat pendek.

Sebagai bagian dari pemahaman awal terhadap konten dataset, ditampilkan pula lima sampel dokumen pertama beserta cuplikan teks dan jumlah katanya, serta tiga dokumen acak dalam bentuk teks mentah secara lengkap untuk memberikan gambaran kualitatif mengenai ragam gaya penulisan, panjang kalimat, dan topik yang diangkat dalam aspirasi masyarakat.

Tabel 5.2 Contoh Dokumen Aspirasi (5 Pertama)

No	Cuplikan Aspirasi	Jumlah Kata
1	Masyarakat meminta pendirian fasilitas pendidikan baru di wilayah yang lebih terjangkau. Maka, dalam hal ini, pemerintah...	34
2	Saat ini lokasi masih menyewa sehingga membutuhkan bantuan apabila berkenan barangkali pembangunan TK dan PAUD mandiri. Sa...	94
3	Dalam rangka memperkuat nilai-nilai keistimewaan dan kebudayaan, maka Pemerintah Daerah diharapkan untuk...	90
4	Disampaikan oleh Bapak Triyono dari Difabel bahwa pada Tahun 2014, selama ini banyak pelatihan yang diberikan kepada ka...	74
5	Dinas Koperasi dan UKM DIY menyediakan layanan pendampingan dan konsultasi bagi UMKM yang dikelola oleh perempuan dan pe...	60

Pada lima dokumen pertama, terlihat bahwa panjang teks bervariasi antara 34 hingga 94 kata, yang menunjukkan adanya perbedaan tingkat elaborasi dalam penyampaian aspirasi. Secara umum, isi dokumen mencerminkan permasalahan di berbagai sektor.

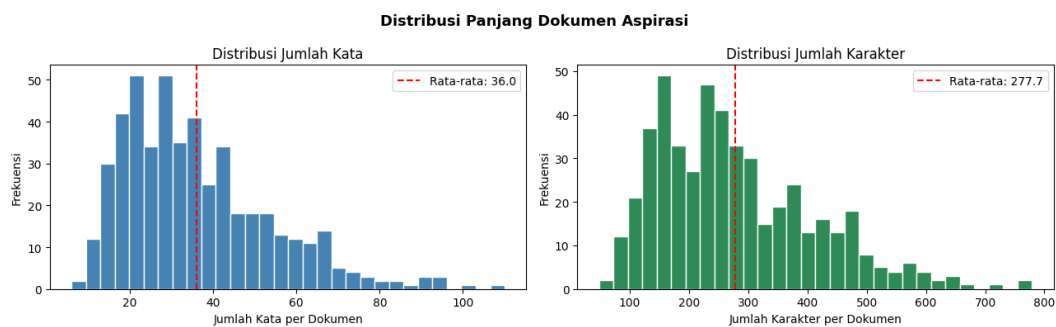
Tabel 5.3 Contoh Dokumen Mentah (3 Sampel Acak)

No Dokumen	Isi Dokomen
421	Infrastruktur pendidikan karakter di DIY masih terbatas, terutama di daerah pedesaan. Selain itu, alokasi anggaran untuk pengadaan sarana...
263	Menurut data Kementerian Pemuda dan Olahraga (Kemenpora), Indonesia membutuhkan lebih banyak pemain muda berbakat untuk bersaing di level...

No Dokumen	Isi Dokomen
432	Program Makan Bergizi Gratis perlu dukungan semua pihak, termasuk Organisasi Perempuan seperti Wanita Syarikat Islam, agar program pemerin...

Sementara itu, pada tiga sampel dokumen acak, terlihat bahwa aspirasi masyarakat tidak hanya berfokus pada isu lokal, tetapi juga mencakup perspektif yang lebih luas. Struktur kalimat yang digunakan cenderung deskriptif dan argumentatif, dengan penekanan pada penyampaian masalah sekaligus harapan atau usulan solusi.

Distribusi panjang dokumen divisualisasikan dalam dua histogram yang disajikan secara berdampingan, satu berdasarkan jumlah kata dan satu berdasarkan jumlah karakter. Pada masing-masing histogram, ditambahkan garis penanda nilai rata-rata sebagai referensi sebaran data.



Gambar 5.1 Histogram Distribusi Panjang Dokumen

Distribusi panjang dokumen aspirasi menunjukkan variasi yang cukup lebar baik dari sisi jumlah kata maupun jumlah karakter. Pada histogram jumlah kata, terlihat bahwa sebagian besar dokumen terkonsentrasi pada rentang sekitar 20 hingga 50 kata, dengan rata-rata sebesar 36 kata. Pola distribusi ini cenderung miring ke kanan (*positively skewed*), yang mengindikasikan adanya sejumlah dokumen dengan panjang yang relatif lebih tinggi dibandingkan mayoritas data. Hal ini terlihat dari keberadaan ekor distribusi yang memanjang hingga lebih dari 100 kata, meskipun frekuensinya semakin menurun seiring bertambahnya jumlah kata.

Sementara itu, distribusi jumlah karakter memperlihatkan pola yang sejalan dengan distribusi jumlah kata, dengan rata-rata sebesar 277,7 karakter. Mayoritas dokumen berada pada rentang sekitar 150 hingga 350 karakter. Distribusi ini juga menunjukkan kemencengan ke kanan, ditandai dengan adanya dokumen yang

memiliki jumlah karakter hingga mendekati 800, meskipun dengan frekuensi yang rendah.

5.2. Pra-pemrosesan Teks Aspirasi

Tahapan pra-pemrosesan diterapkan secara berurutan terhadap 486 dokumen aspirasi. Lapisan penghapusan stopword dilakukan dalam dua tahap: pertama menggunakan daftar baku dari pustaka Sastrawi, kemudian dilengkapi dengan daftar custom stopwords yang dikembangkan secara domain-spesifik untuk konteks teks aspirasi masyarakat. Daftar ini mencakup kata sapaan dan honorifik (seperti bapak, ibu, mohon), kata fungsi dan konjungsi (yang, dan, sehingga), kata keterangan umum (sudah, sangat, hanya), serta istilah meta-aspirasi yang muncul hampir di seluruh dokumen sehingga tidak bersifat diskriminatif secara topik (seperti masyarakat, pemerintah, aspirasi). Nama institusi dan wilayah geografis yang terlalu umum dalam konteks dataset ini, seperti yogyakarta, dpd, dan diy, turut dimasukkan ke dalam daftar eliminasi untuk mengurangi noise yang dapat mendistorsi struktur topik.

Stemming menggunakan algoritma Sastrawi terbukti efektif dalam menggabungkan variasi morfologis kata-kata Bahasa Indonesia ke dalam satu representasi dasar, sehingga memperkaya kepadatan semantik setiap dokumen dan mengurangi dimensi ruang fitur yang perlu dimodelkan.

Sebagai tahap akhir penyaringan, dokumen dengan kurang dari tiga token setelah seluruh proses pembersihan dieliminasi dari korpus. Ambang batas ini ditetapkan karena dokumen dengan token sangat sedikit tidak membawa konteks yang cukup untuk dikategorikan secara bermakna dalam pemodelan topik.

Tabel 5.4 Contoh Hasil Pra-pemrosesan Teks Aspirasi

No	Dokumen Asli	Hasil Pra-pemrosesan
0	Masyarakat meminta pendirian fasilitas pendidikan baru di wilayah yang lebih terjangkau. Maka, dalam hal ini, pemerintah diharapkan bisa lebih memperhatikan fasilitas pendidikan terutama di daerah terpencil seperti di Desa Girisubo, Panggang, Kabupaten Gunung Kidul.	diri fasilitas didik wilayah jangkau perhati fasilitas didik pencil desa girisubo panggang kabupaten gunung kidul
1	Saat ini lokasi masih menyewa sehingga membutuhkan bantuan	lokasi sewa butuh bantu kenan barangkali bangun tk paud damai

No	Dokumen Asli	Hasil Pra-pemrosesan
	apabila berkenan barangkali pembangunan TK dan PAUD damai. Saat ini siswa-siswi sudah 40 dan itu terus bertambah. TK ini awalnya didirikan oleh Bapak Hery dari Kapolda sejak tahun 2020. Ijin sudah ada, akreditasi saat ini B. Para wali hanya ditarik iuran sebesar 15.000. Sedangkan untuk gaji guru juga diberikan dari Pak Hery. Kami berharap TK ini bisa berkembang terus dengan memiliki tempat sendiri. Saat ini ada 5 guru yang mengajar membawahi 40 anak. Animo masyarakat cukup tinggi untuk menyekolahkan anaknya, kami sampai menolak karena keterbatasan fasilitas.	siswasiswi tambah tk diri kapolda ijin akreditasi wali tarik iur gaji guru tk kembang milik guru ajar bawah anak animo tinggi sekolah anak tolak batas fasilitas
2	dalam rangka memperkuat nilai-nilai keistimewaan dan kebudayaan, maka Pemerintah Daerah diharapkan untuk: 1) Setiap Dinas ada Saksi khusus yang menangani keistimewaan. 2) membentuk tim, yang dalam hal ini bisa melibatkan karang taruna. Kita sudah melakukan beberapa kerjasama, salah satunya dengan UNY. Dalam hal ini kami membuat Taman Pendidikan Aksara Jawa. Setiap anak wajib memahami bahasa jawa. Mungkin dari dinas nanti bisa membuat acara untuk hal ini, bisa dengan tema “GKR Hemas menyapa Pelajar Yogyakarta”. Kami berharap ada dukungan dari Gusti Kanjeng Ratu Hemas terkait ide-ide ini, yakni Pendidikan Keistimewaan Jogjakarta.	rangka kuat nilainilai istimewa budaya dinas saksi tangan istimewa bentuk tim libat karang taruna kerjasama uny taman didik aksara jawa anak wajib paham bahasa jawa dinas acara tema gkr has sapa ajar yogyakarta dukung gusti kanjeng ratu has ideide didik istimewa

Hasil pra-pemrosesan pada **Tabel 5.4** memperlihatkan transformasi yang signifikan, teks asli yang mengandung struktur bahasa alami kompleks berhasil direduksi menjadi representasi yang lebih ringkas dan terstandarisasi. Informasi yang bersifat kontekstual namun kurang relevan terhadap analisis topik, seperti kata penghubung atau penjelasan tambahan telah diminimalkan, sehingga mengurangi

kompleksitas data sekaligus meningkatkan konsistensi representasi teks antardokumen.

5.3. Representasi Semantik Dokumen dengan Model Embedding Multilingual

Model LazarusNLP/all-indo-e5-small-v4 dipilih sebagai *backbone* embedding berdasarkan dua pertimbangan utama. Pertama, model ini merupakan hasil fine-tuning dari arsitektur E5 pada korpus berbahasa Indonesia dalam skala besar, sehingga memiliki pemahaman leksikal dan morfologis yang jauh lebih baik dibandingkan model multilingual generik terhadap karakteristik bahasa aglutinatif seperti Bahasa Indonesia. Kedua, varian small-v4 menawarkan keseimbangan optimal antara efisiensi komputasi dan kualitas representasi semantik, menjadikannya praktis untuk dijalankan dalam lingkungan komputasi dengan sumber daya terbatas.

Proses encoding dilaksanakan dengan akselerasi GPU menggunakan parameter `device='cuda'` dan konfigurasi `batch_size=64` untuk memaksimalkan throughput pemrosesan dengan tetap mempertimbangkan kapasitas memori GPU yang tersedia.

Berdasarkan output embedding, diperoleh matriks berukuran (486, 384), yang menunjukkan bahwa seluruh 486 dokumen aspirasi berhasil direpresentasikan dalam vektor berdimensi 384. Dimensi 384 merupakan karakteristik model LazarusNLP/all-indo-e5-small-v4 yang dirancang untuk menghasilkan representasi yang ringkas namun tetap mampu menangkap makna kontekstual. Matriks embedding ini kemudian menjadi masukan utama bagi tahap reduksi dimensi dengan UMAP.

5.4. Reduksi Dimensi dengan UMAP

Konfigurasi UMAP yang diterapkan dalam penelitian ini dirancang dengan mempertimbangkan karakteristik khusus data aspirasi masyarakat, sebagaimana dirangkum pada **Tabel 5.5**.

Tabel 5.5 Konfigurasi UMAP

Parameter	Nilai
n_neighbors	15
n_components	5
metric	cosine

Nilai `n_neighbors=15` dipilih secara deliberatif dengan mempertimbangkan variasi tematik data aspirasi yang beragam namun dengan densitas distribusi yang tidak merata. Nilai ini dinilai mampu menangkap kemiripan lokal antardokumen secara representatif tanpa kehilangan struktur global antar kelompok topik.

Dimensi target ditetapkan sebesar 5 dimensi (bukan langsung ke 2 dimensi) berdasarkan rekomendasi implementasi BERTopic. Proyeksi prematur ke dimensi yang sangat rendah dapat menyebabkan hilangnya informasi struktural yang diperlukan HDBSCAN dalam membentuk klaster yang koheren.

Pemilihan `metric='cosine'` relevan mengingat panjang teks yang dianalisis bervariasi secara substansial dari satu kalimat singkat hingga beberapa paragraf, sehingga jarak Euclidean rentan terdistorsi oleh perbedaan skala panjang dokumen. Dampak konfigurasi ini bersifat kumulatif: cosine similarity juga berperan pada tahap pasca-klusterisasi untuk mendistribusikan ulang dokumen outlier ke kluster terdekat, yang terbukti meningkatkan koherensi topik (NPMI) hingga 16,1% dibandingkan konfigurasi standar, serta digunakan dalam pembangunan matriks kemiripan antartopik yang menjadi dasar visualisasi dan evaluasi model.

5.5. Pembentukan Klaster Topik dengan HDBSCAN

Karakteristik HDBSCAN yang mampu menentukan jumlah klaster secara otomatis tanpa asumsi terhadap bentuk geometris klaster sangat sesuai untuk data aspirasi masyarakat yang bersifat tematik heterogen dan tidak terdistribusi secara merata. Konfigurasi parameter yang diterapkan dirangkum pada **Tabel 5.6**.

Tabel 5.6 Konfigurasi HDBSCAN

Parameter	Nilai
<code>min_cluster_size</code>	10
<code>min_samples</code>	5
<code>cluster_selection_method</code>	eom

Nilai `min_cluster_size=10` ditetapkan untuk menghasilkan granularitas topik yang seimbang: cukup besar untuk menghindari klaster yang terlalu sempit dan tidak stabil, namun cukup kecil untuk tetap menangkap isu-isu yang spesifik namun cukup sering muncul dalam data aspirasi.

Nilai `min_samples=5` menetapkan tingkat konservatisme yang moderat dalam mengklasifikasikan dokumen sebagai outlier. Penggunaan `metric='euclidean'` pada

tahap ini konsisten dengan kondisi ruang vektor pasca-reduksi UMAP, di mana jarak Euclidean terbukti efektif setelah transformasi manifold dilakukan.

Metode *EOM (Excess of Mass)* dipilih karena menghasilkan jumlah kluster yang lebih banyak dengan ukuran yang bervariasi, sejalan dengan kebutuhan analisis data aspirasi yang mencakup beragam isu publik dengan tingkat kemunculan yang tidak seragam. Kombinasi konfigurasi ini bertujuan menghasilkan struktur kluster yang koheren dengan distribusi ukuran yang mencerminkan variasi frekuensi tematik dalam korpus aspirasi.

5.6. Identifikasi Representasi Topik melalui c-TF-IDF

Konfigurasi ekstraksi fitur teks menggunakan *CountVectorizer* disesuaikan terhadap karakteristik linguistik Bahasa Indonesia, sebagaimana dirangkum pada **Tabel 5.7**.

Tabel 5.7 Konfigurasi c-TF-IDF

Parameter	Nilai
ngram_range	(1, 2)
reduce_frequent_words	True
bm25_weighting	True

Penggunaan `ngram_range=(1, 2)` memiliki justifikasi linguistik yang kuat: Bahasa Indonesia kerap mengekspresikan konsep-konsep substantif melalui frasa dua kata, seperti 'jalan rusak', 'air bersih', 'pelayanan kesehatan', atau 'bantuan sosial'. Apabila didekomposisi menjadi unigram terpisah, frasa-frasa tersebut akan kehilangan makna kontekstualnya. Pengujian terhadap data ini menunjukkan bahwa inklusi bigram secara nyata memperkaya representasi topik, terutama untuk isu infrastruktur, layanan publik, dan kebijakan sosial yang lazimnya diformulasikan dalam bentuk frasa nominal majemuk.

Penetapan `min_df=2` efektif menyaring singkatan idiosinkratik dan nama proper yang tidak informatif secara tematik, sementara `max_df=0.85` secara adaptif mengeksklusi term yang distribusionalnya terlalu umum. Kombinasi kedua parameter ini mempersempit ruang fitur pada term-term dengan nilai informatif optimal.

Aktivasi `reduce_frequent_words` dan `bm25_weighting` secara bersama-sama memberikan dampak signifikan terhadap kualitas representasi. Dalam konteks data aspirasi masyarakat yang memiliki variasi panjang teks signifikan, keunggulan

BM25 dalam menangani heterogenitas panjang dokumen menjadi pertimbangan kritis yang menjustifikasi pemilihannya dibandingkan skema pembobotan standar.

5.7. Peningkatan Kualitas Representasi Topik dengan KeyBERT dan MMR

Meskipun c-TF-IDF telah menghasilkan representasi topik yang secara statistik diskriminatif, diperlukan dua lapisan representasi lanjutan yang bekerja secara komplementer, yaitu *KeyBERTInspired* dan *Maximal Marginal Relevance* (MMR).

Parameter `top_n_words=15` menetapkan jumlah kandidat kata kunci yang dievaluasi pada tahap *KeyBERTInspired*, sementara `nr_repr_docs=5` memastikan bahwa vektor referensi topik dikonstruksi dari lima dokumen paling representatif dalam setiap kluster — bukan sekadar rata-rata dari seluruh dokumen yang mungkin mengandung noise.

Parameter `diversity=0.3` pada MMR mengkalibrasi keseimbangan antara relevansi (70%) dan diversitas (30%), konfigurasi yang secara empiris menghasilkan daftar kata kunci yang interpretatif tanpa mengorbankan keterwakilan tematik.

Kedua model representasi dikonfigurasi sebagai *pipeline* bertingkat melalui struktur dictionary, yang memberikan fleksibilitas analitis, sehingga peneliti dapat membandingkan secara langsung representasi kata kunci dari c-TF-IDF, *KeyBERTInspired*, dan MMR untuk setiap topik, sehingga memungkinkan evaluasi kualitatif terhadap konsistensi dan koherensi semantik antartahap representasi.

5.8. Hasil Pemodelan Topik dan Interpretasi Tematik

Seluruh komponen pipeline BERTopic diintegrasikan dengan konfigurasi `top_n_words=10`, `language='multilingual'`, dan `calculate_probabilities=True`. Proses pelatihan dijalankan melalui metode `fit_transform()` yang menjalankan seluruh tahapan secara sekuensial, memanfaatkan representasi vektor embedding yang telah dikonstruksi sebelumnya guna menghindari komputasi ulang yang tidak perlu.

Tabel 5.8 Hasil Pemodelan Topik

Keterangan	Nilai
Total dokumen	486
Jumlah topik	15

Keterangan	Nilai
Dokumen <i>outlier</i>	81 (16.7%)

Dari total 486 dokumen aspirasi, berhasil diidentifikasi 15 topik utama di luar label outlier. Jumlah topik ini tidak ditetapkan secara manual, melainkan ditentukan secara otomatis oleh HDBSCAN berdasarkan struktur kepadatan data pada ruang laten hasil proyeksi UMAP, hal ini diperoleh dari cara perhitungan $\text{len}(\text{topic_info}) - 1$, di mana pengurangan satu dilakukan untuk mengeksklusi label -1. Dengan demikian, angka 15 topik merupakan hasil inferensi empiris model terhadap pola semantik yang terkandung dalam korpus, bukan parameter yang ditentukan peneliti di awal. Sebanyak 81 dokumen (16,7%) teridentifikasi sebagai outlier, proporsi moderat yang mengindikasikan sebagian besar dokumen telah berhasil terklaster dengan baik.

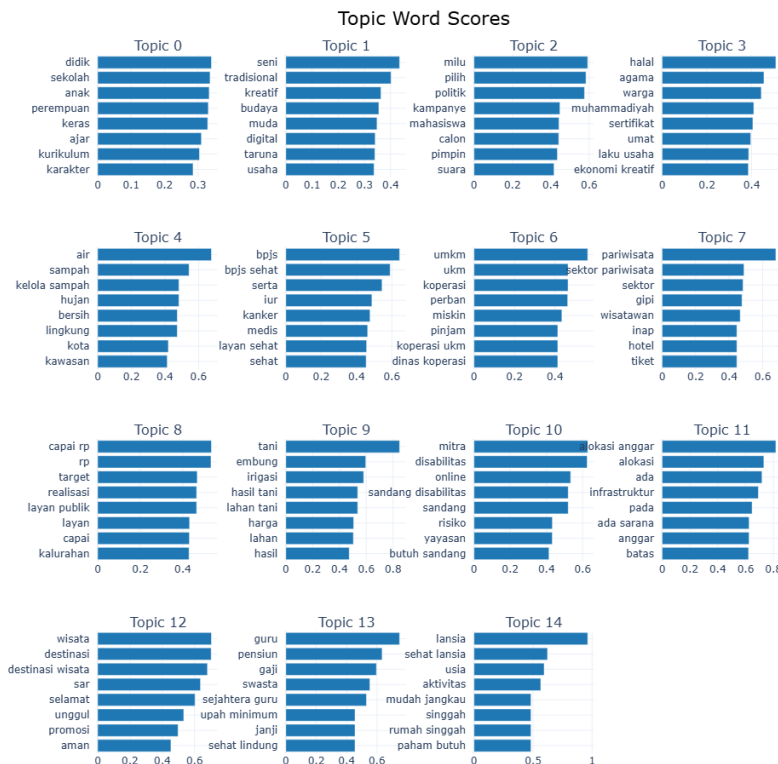
Tabel 5.9 Daftar Topik Teratas

Topik	Jumlah Dokumen	Label
0	98	0_didik_sekolah_anak_perempuan
1	55	1_seni_tradisional_kreatif_budaya
2	34	2_milu_pilih_politik_kampanye
3	26	3_halal_agama_warga_muhammadiyah
4	23	4_air_sampah_kelola_sampah_hujan
5	22	5_bpjs_bpjs_sehat_serta_iur
6	20	6_umkm_ukm_koperasi_perban
7	19	7_pariwisata_sektor_pariwisata_sektor_gipi
8	19	8_capai_pr_rp_target_realisasi
9	17	9_tani_embung_irigasi_hasil_tani
10	17	10_mitra_disabilitas_online_sandang_disabilitas
11	16	11_alokasi_anggar_alokasi_ada_infrastruktur
12	14	12_wisata_destinasi_destinasi_wisata_sar
13	13	13_guru_pensiun_gaji_swasta
14	12	14_lansia_sehat_lansia_usia_aktivitas

Distribusi jumlah dokumen per topik menunjukkan ketimpangan ukuran klaster yang mencerminkan variasi frekuensi isu dalam aspirasi masyarakat. Topik 0 (Pendidikan) mendominasi dengan 98 dokumen, diikuti Topik 1 (Seni dan Budaya) dengan 55 dokumen, sementara topik lainnya berkisar antara 12 hingga 34

dokumen. Pola ini mengindikasikan adanya isu dominan yang lebih sering disuarakan masyarakat, sementara isu lain bersifat lebih spesifik dan tersebar.

Barchart distribusi kata kunci pada **Gambar 5.2** memungkinkan pembacaan langsung terhadap struktur leksikal setiap topik dan menjadi basis utama dalam pemberian nama tematik. Proses penamaan dilakukan secara kualitatif dengan mempertimbangkan konsistensi semantik antara kata kunci c-TF-IDF, KeyBERT, dan MMR, serta dikonfirmasi melalui inspeksi dokumen sampel representatif dari masing-masing kluster.



Gambar 5.2 Barchart Distribusi Kata Kunci

Berikut adalah interpretasi dan penamaan dari masing-masing topik yang dihasilkan.

Topik 0: Pendidikan, Perempuan dan Anak, ditandai oleh kata kunci dominan seperti didik, sekolah, anak, dan perempuan. Topik ini merepresentasikan aspirasi masyarakat terkait pemerataan akses pendidikan, penguatan kurikulum berbasis karakter, serta perhatian khusus terhadap pendidikan dan perempuan. Dokumen-dokumen dalam kluster ini mengindikasikan adanya kebutuhan akan dukungan infrastruktur dan kebijakan afirmatif di sektor pendidikan.

Topik 1: Seni, Budaya, dan Usaha Kreatif, dicirikan oleh kata kunci seni, tradisional, kreatif, budaya, dan muda. Topik ini mencerminkan keprihatinan terhadap keberlanjutan seni dan kebudayaan lokal, sekaligus dorongan untuk mengintegrasikan nilai-nilai budaya ke dalam ekosistem kreatif generasi muda melalui platform digital dan program kepemudaan.

Topik 2: Pemilu, Politik dan Kebangsaan, didominasi oleh kata kunci milu (pemilu), pilih, politik, kampanye, dan mahasiswa. Topik ini merangkum aspirasi yang berkaitan dengan peningkatan partisipasi publik dalam proses demokrasi elektoral, termasuk harapan terhadap sinergi antara organisasi kemasyarakatan dan penyelenggara pemilu seperti KPU.

Topik 3: Keagamaan, mengangkat isu-isu yang berkaitan dengan halal, agama, warga, muhammadiyah, dan sertifikat. Topik ini merefleksikan tuntutan masyarakat terhadap pengembangan ekosistem usaha yang berlandaskan prinsip syariah, termasuk percepatan sertifikasi halal bagi pelaku usaha mikro serta penguatan toleransi antarumat beragama.

Topik 4: Pengelolaan Air, Sampah, dan Lingkungan, ditandai oleh kata kunci air, sampah, kelola sampah, hujan, dan bersih. Topik ini mengindikasikan kekhawatiran yang signifikan terhadap degradasi kualitas lingkungan, mencakup pengelolaan sampah, ketahanan terhadap banjir, serta aksesibilitas air bersih di kawasan perkotaan dan perdesaan.

Topik 5: BPJS dan Layanan Kesehatan, diidentifikasi melalui kata kunci bpjs, bpjs sehat, iur, kanker, dan medis. Topik ini mencerminkan aspirasi masyarakat terkait reformasi sistem jaminan kesehatan nasional, khususnya perluasan cakupan layanan bagi pasien dengan penyakit kronis seperti kanker, serta peningkatan kualitas fasilitas kesehatan primer.

Topik 6: UMKM, Koperasi, dan Pinjaman, dicirikan oleh kata kunci umkm, umk, koperasi, perban, dan miskin. Topik ini menggarisbawahi pentingnya pendampingan usaha, akses permodalan, dan penguatan kelembagaan koperasi sebagai instrumen pengentasan kemiskinan dan peningkatan daya saing ekonomi lokal.

Topik 7: Pariwisata dan Perhotelan, ditandai oleh kata kunci pariwisata, sektor pariwisata, gipi, wisatawan, hotel, dan tiket. Topik ini merepresentasikan

aspirasi pelaku industri pariwisata terhadap penyelesaian hambatan regulasi, peningkatan promosi destinasi, serta dukungan pemerintah dalam menarik investasi di sektor perhotelan dan jasa wisata.

Topik 8: Target dan Realisasi Anggaran Layanan Publik, diidentifikasi melalui kata kunci capai rp, target, realisasi, dan layan publik. Topik ini menyoroti persoalan tumpang tindih birokrasi, transparansi realisasi anggaran, serta desakan masyarakat terhadap pembentukan unit layanan rehabilitasi berbasis kelurahan yang lebih efisien dan akuntabel.

Topik 9: Pertanian, Irigasi, Embung, dan Hasil Tani, ditandai oleh kata kunci tani, embung, irigasi, hasil tani, dan lahan tani. Topik ini merangkum harapan komunitas petani terhadap pembenahan jaringan irigasi, pembangunan embung, serta peningkatan produktivitas lahan sebagai bentuk nyata keberpihakan pemerintah terhadap sektor agraris.

Topik 10: Inklusi Sosial dan Pemberdayaan Penyandang Disabilitas, dicirikan oleh kata kunci disabilitas, sandang disabilitas, mitra, yayasan, dan online. Topik ini merefleksikan kebutuhan akan regulasi yang lebih jelas, dukungan anggaran berkelanjutan, dan pengembangan program vokasional berbasis digital bagi penyandang disabilitas untuk meningkatkan kemandirian ekonomi mereka.

Topik 11: Alokasi Anggaran dan Pembangunan Infrastruktur, diidentifikasi melalui kata kunci alokasi anggar, infrastruktur, ada sarana, dan batas. Topik ini mencerminkan tuntutan masyarakat terhadap keberpihakan anggaran pemerintah dalam penyediaan sarana prasarana dasar, termasuk fasilitas pendidikan, teknologi informasi, dan infrastruktur wilayah yang masih tertinggal.

Topik 12: Destinasi Wisata dan Keselamatan, ditandai oleh kata kunci wisata, destinasi wisata, sar, selamat, dan promosi. Topik ini mengangkat pentingnya integrasi aspek keselamatan, termasuk keterlibatan tim SAR dalam pengelolaan destinasi wisata, serta pemberdayaan pelaku UMKM lokal sebagai bagian integral dari ekosistem pariwisata berbasis masyarakat.

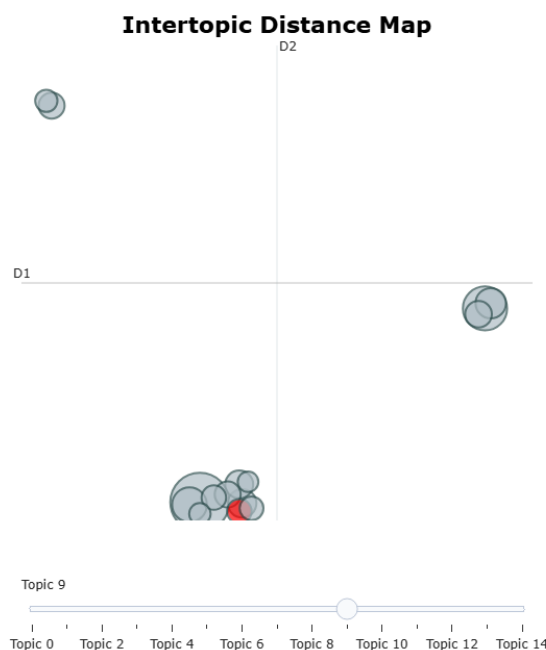
Topik 13: Gaji, Pensiun, dan Kesejahteraan Guru, dicirikan oleh kata kunci guru, pensiun, gaji, swasta, sejahtera guru, dan upah minimum. Topik ini menyuarakan aspirasi kalangan pendidik, khususnya guru swasta dan tenaga

honorer, yang menuntut peningkatan remunerasi, kejelasan status kepegawaian, serta jaminan perlindungan sosial yang setara dengan aparatur negara.

Topik 14: Aktivitas Lansia, diidentifikasi melalui kata kunci lansia, sehat lansia, usia, aktivitas, mudah jangkau, dan rumah singgah. Topik ini merepresentasikan kebutuhan spesifik kelompok lanjut usia akan fasilitas kesehatan yang aksesibel, program aktivitas sosial, serta regulasi daerah yang secara eksplisit berpihak pada pemenuhan hak-hak dasar warga lansia.

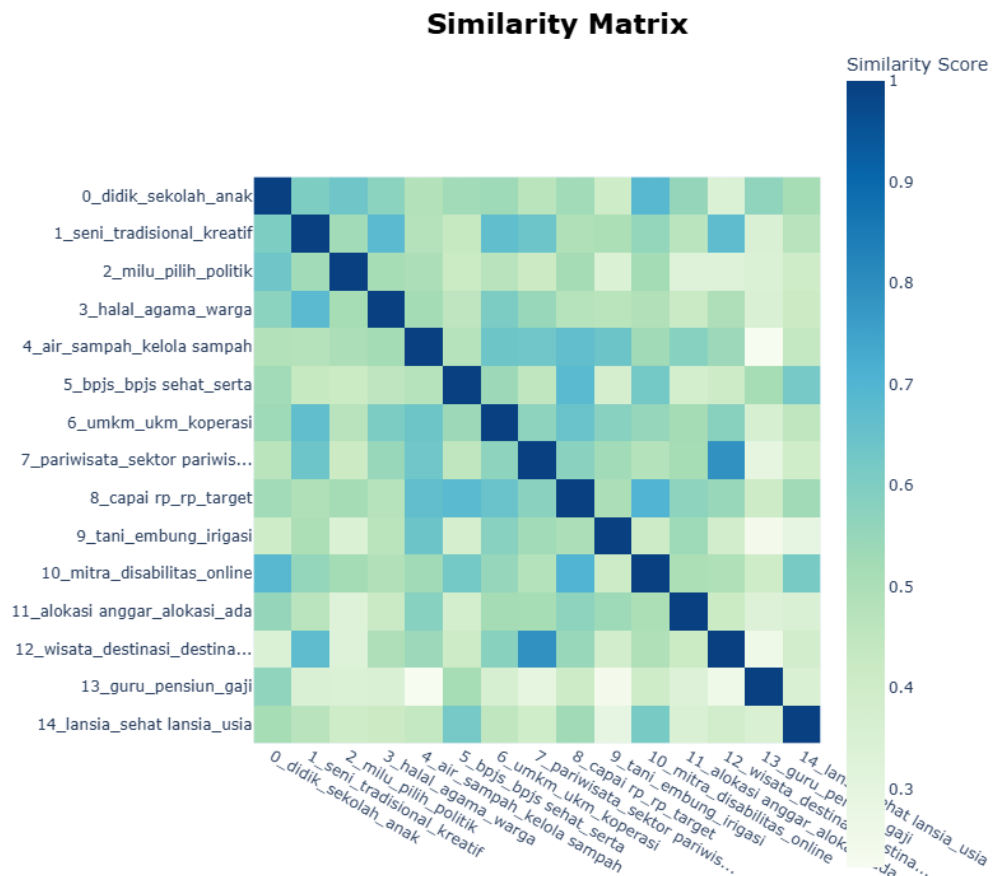
Secara keseluruhan, distribusi topik yang dihasilkan menunjukkan bahwa aspirasi masyarakat bersifat multidimensional, mencakup domain pendidikan, kesehatan, ekonomi, lingkungan, budaya, tata kelola pemerintahan, serta inklusi sosial. Keberagaman ini menegaskan bahwa proses penjaringan aspirasi publik tidak dapat direduksi ke dalam satu atau dua isu tunggal, melainkan menuntut pendekatan kebijakan yang komprehensif dan lintas sektor.

Visualisasi *intertopic distance map* pada **Gambar 5.3** menunjukkan bahwa sebagian besar topik terkonsentrasi dalam satu gugus utama, yang menandakan adanya kedekatan semantik antartopik yang dibahas. Beberapa topik tampak terpisah dari gugus tersebut, mengindikasikan karakteristik tematik yang lebih spesifik. Variasi ukuran lingkaran mencerminkan jumlah dokumen per topik, sehingga isu-isu dominan dalam data dapat teridentifikasi secara visual.



Gambar 5.3 Visualisasi *Intertopic Distance Map*

Selanjutnya, *heatmap* kemiripan antartopik pada **Gambar 5.4** menunjukkan bahwa struktur topik hasil pemodelan secara umum memiliki diferensiasi semantik yang cukup baik, yang terlihat dari dominasi nilai kemiripan rendah hingga menengah pada sebagian besar sel di luar diagonal utama. Beberapa pasangan topik memperlihatkan tingkat kemiripan yang lebih tinggi pada tema-tema yang secara substantif saling berkaitan, seperti seni tradisional dengan pariwisata, serta pengelolaan lingkungan dengan pertanian dan irigasi, irisan semantik yang wajar akibat kedekatan konteks pembahasan. Sebaliknya, topik seperti guru/pensiun, lansia, dan alokasi anggaran cenderung lebih terpisah, menunjukkan karakter tema yang lebih spesifik dan berdiri sendiri.



Gambar 5.4 *Heatmap* Kemiripan Antartopik

Tabel 5.10 Ringkasan 3 Kata pada Setiap Topik Berdasarkan Nilai c-TF-IDF

Topik	Jumlah Dokumen	Top-3 Kata (Skor c-TF-IDF)
0	98	didik (0.3404) sekolah (0.3363) anak (0.3337)
1	55	seni (0.4337) tradisional (0.4012) kreatif (0.3626)
2	34	milu (0.5923) pilih (0.5834) politik (0.5755)

Topik	Jumlah Dokumen	Top-3 Kata (Skor c-TF-IDF)
3	26	halal (0.5086) agama (0.4553) warga (0.4434)
4	23	air (0.6765) sampah (0.5435) kelola sampah (0.4832)
5	22	bpjs (0.6436) bpjs sehat (0.5901) serta (0.5444)
6	20	umkm (0.5618) ukm (0.4647) koperasi (0.4647)
7	19	pariwisata (0.6787) sektor pariwisata (0.4885) sektor (0.4829)
8	19	capai rp (0.5344) rp (0.5323) target (0.4678)
9	17	tani (0.8485) embung (0.5961) irigasi (0.5806)
10	17	mitra (0.6279) disabilitas (0.6249) online (0.5336)
11	16	alokasi anggaran (0.8173) alokasi (0.7310) ada (0.7178)
12	14	wisata (0.7044) destinasi (0.7027) destinasi wisata (0.6798)
13	13	guru (0.7488) pensiun (0.6338) gaji (0.5965)
14	12	lansia (0.9628) sehat lansia (0.6225) usia (0.5939)

Skor c-TF-IDF mengungkapkan pola yang konsisten: topik dengan kluster berukuran kecil cenderung menghasilkan skor lebih tinggi karena kata kuncinya bersifat lebih eksklusif. Hal ini terlihat pada Topik 14 yang mencatatkan skor tertinggi pada kata lansia (0,9628), diikuti Topik 9 dengan tani (0,8485) dan Topik 11 dengan alokasi anggaran (0,8173), ketiganya mengindikasikan kluster dengan koherensi tematik sangat tinggi. Sebaliknya, Topik 0 sebagai kluster terbesar menunjukkan skor yang lebih rendah (didik: 0,3404) sebagai konsekuensi dilusi leksikal akibat keberagaman dokumen yang lebih besar. Pola ini secara keseluruhan mengkonfirmasi bahwa model yang dihasilkan memiliki diferensiasi tematik yang valid sebagai landasan analitik pemetaan aspirasi masyarakat.

Pada tahap akhir, hasil penugasan topik diintegrasikan kembali ke dalam dataframe utama melalui kolom topik dan probabilitas. Inspeksi sampel dokumen per topik dengan menampilkan dua dokumen representatif beserta lima kata kunci utama berfungsi sebagai mekanisme validasi kualitatif untuk mengkonfirmasi bahwa dokumen yang ditugaskan ke dalam suatu topik memang secara substantif membahas tema yang sesuai dengan labelnya. Seluruh hasil pemodelan selanjutnya

disimpan dalam berkas CSV sebagai basis data terstruktur yang siap digunakan untuk analisis distribusi tematik dan pemetaan isu prioritas.

5.9. Evaluasi Model

Evaluasi dilakukan melalui pendekatan ganda yang mengintegrasikan penilaian kuantitatif berbasis metrik statistik dan penilaian kualitatif berbasis interpretabilitas tematik. Kedua metrik kuantitatif, Silhouette Score dan Davies-Bouldin Index dikomputasi dalam ruang vektor hasil reduksi UMAP, karena HDBSCAN sendiri beroperasi dan membentuk struktur klaster dalam ruang tersebut, sehingga evaluasi dalam ruang yang sama secara metodologis lebih konsisten dan representatif. Dokumen dengan label outlier (-1) dieksklusi sebelum kalkulasi metrik dilakukan.

Tabel 5.11 Evaluasi Model BERTopic

Indikator	Nilai
Silhouette Score (UMAP/Euclidean)	0.4754
Davies-Bouldin Index (UMAP)	0.7238
Jumlah topik	15
<i>Outlier</i>	81 (16.7%)

Silhouette Score sebesar 0,4754 berada pada kisaran menengah-positif, mengindikasikan bahwa dokumen dalam masing-masing topik memiliki kohesi internal yang cukup baik serta pemisahan antarklaster yang memadai. Adanya sebagian kedekatan antartopik adalah hal yang wajar mengingat karakter data aspirasi masyarakat yang cenderung kompleks dan saling beririsan secara tematik.

Davies-Bouldin Index sebesar 0,7238 yang tergolong rendah memperkuat indikasi bahwa klaster yang terbentuk cukup kompak secara internal sekaligus memiliki separasi yang baik. Kedua metrik ini secara bersama-sama memberikan gambaran kuantitatif yang komprehensif: model menghasilkan struktur topik yang tidak hanya koheren secara internal, tetapi juga cukup terdefinisi secara eksternal.

Model menghasilkan 15 topik valid, yang menunjukkan granularitas yang cukup seimbang dalam merepresentasikan keragaman isu aspirasi tanpa menghasilkan fragmentasi topik yang berlebihan. Proporsi outlier sebesar 16,7% berada dalam batas moderat, mencerminkan bahwa parameter clustering yang digunakan cenderung seimbang, tidak terlalu ketat hingga menghasilkan outlier

berlebihan, namun juga tidak terlalu longgar sehingga memaksakan seluruh dokumen masuk ke klaster tertentu.

Di samping evaluasi kuantitatif, proporsi dokumen outlier turut menjadi indikator diagnostik penting. Pada penelitian ini, proporsi 16,7% mengindikasikan keseimbangan yang baik antara ketepatan dan cakupan klasterisasi.

5.10. Interpretasi dalam Mendukung Fungsi Ketugasan DPD RI

Hasil pemodelan topik yang telah diidentifikasi memiliki relevansi langsung terhadap fungsi ketugasan Dewan Perwakilan Daerah Republik Indonesia (DPD RI). Secara kelembagaan, DPD RI membagi bidang kerjanya ke dalam empat komite dengan tugas pokok yang berbeda. Dengan memetakan 15 topik aspirasi masyarakat ke dalam kerangka komite tersebut, hasil pemodelan topik dapat berfungsi sebagai instrumen berbasis data dalam mengarahkan prioritas legislasi, pengawasan, dan representasi daerah yang lebih responsif dan terukur.

Komite I memiliki keterkaitan dengan Topik 2 (Pemilu, Politik dan Kebangsaan) dan Topik 10 (Inklusi Sosial dan Pemberdayaan Penyandang Disabilitas). Topik 2 relevan dalam konteks penguatan demokrasi elektoral di daerah, termasuk sinergi antara organisasi kemasyarakatan dan penyelenggara pemilu sebagai bagian dari mekanisme hubungan pusat-daerah. Sementara itu, Topik 10 berkorespondensi dengan fungsi Komite I dalam mendorong harmonisasi regulasi antara pemerintah pusat dan daerah guna memberikan perlindungan hukum yang lebih eksplisit bagi kelompok penyandang disabilitas.

Komite II memperoleh peta aspirasi yang paling langsung dari Topik 9 (Pertanian, Irigasi, Embung, dan Hasil Tani), yang mencatatkan skor c-TF-IDF tertinggi pada kata tani (0,8485). Topik ini secara langsung menunjuk pada urgensi perbaikan infrastruktur pertanian, ketahanan pangan lokal, dan peningkatan produktivitas lahan sebagai inti ketugasan komite ini. Topik 4 (Pengelolaan Air, Sampah, dan Lingkungan) turut relevan dalam konteks pengelolaan sumber daya alam dan ketahanan ekologis, sementara Topik 6 (UMKM, Koperasi, dan Pinjaman) mencerminkan aspirasi terhadap penguatan sumber daya ekonomi kerakyatan yang perlu didorong melalui kebijakan dan advokasi Komite II di tingkat nasional.

Komite III memiliki keterkaitan dengan spektrum topik yang paling luas di antara keempat komite. Dalam domain pendidikan, Topik 0 (Pendidikan, Perempuan dan Anak) dan Topik 13 (Gaji, Pensiun, dan Kesejahteraan Guru) mencakup isu pemerataan akses, penguatan kurikulum, hingga perlindungan remunerasi bagi guru swasta dan tenaga honorer. Topik 3 (Keagamaan) berkorespondensi dengan fungsi komite di bidang keagamaan, khususnya regulasi sertifikasi halal dan penguatan ekosistem usaha berbasis nilai syariah. Dalam dimensi sosial dan budaya, Topik 1 (Seni, Budaya, dan Usaha Kreatif), Topik 7 (Pariwisata dan Perhotelan), dan Topik 12 (Destinasi Wisata dan Keselamatan) mencerminkan aspirasi terhadap pengembangan pariwisata yang berakar pada kekayaan budaya lokal. Adapun Topik 5 (BPJS dan Layanan Kesehatan) dan Topik 14 (Aktivitas Lansia) melengkapi dimensi sosial komite ini, khususnya dalam mendorong kebijakan afirmatif di bidang kesehatan bagi kelompok rentan dan lanjut usia.

Komite IV memperoleh relevansi langsung dari Topik 8 (Target dan Realisasi Anggaran Layanan Publik) dan Topik 11 (Alokasi Anggaran dan Pembangunan Infrastruktur). Kedua topik ini secara eksplisit menyuarakan tuntutan transparansi realisasi anggaran daerah, ketimpangan distribusi infrastruktur antarwilayah, serta desakan terhadap keberpihakan alokasi fiskal bagi wilayah yang masih tertinggal, isu-isu yang secara langsung bersinggungan dengan fungsi pengawasan APBN dan kebijakan keuangan negara yang diemban Komite IV.

Dengan mengacu pada distribusi topik beserta bobot $c\text{-TF-IDF}$ -nya, setiap komite dapat memprioritaskan isu yang paling mendesak sesuai domainnya, sekaligus mengidentifikasi isu-isu lintas komite yang memerlukan koordinasi kelembagaan yang lebih sinergis. Pendekatan berbasis pemodelan topik ini dengan demikian berpotensi menjadi instrumen strategis dalam memperkuat kualitas representasi aspirasi daerah di tingkat nasional.

BAB VI

PENUTUP

6.1. Kesimpulan

Penelitian ini mengaplikasikan metode BERTopic untuk memodelkan aspirasi masyarakat yang disampaikan kepada Dewan Perwakilan Daerah Republik Indonesia (DPD RI), khususnya yang bersumber dari wilayah Daerah Istimewa Yogyakarta. Berdasarkan hasil pemodelan, korpus aspirasi berhasil diklasterisasi ke dalam 15 topik tematik yang secara kolektif merepresentasikan spektrum isu prioritas publik secara terstruktur dan terukur.

Menjawab rumusan masalah pertama, distribusi dokumen antarkluster menunjukkan ketidakmerataan yang signifikan, di mana Topik 0 (Pendidikan, Perempuan dan Anak) mendominasi dengan 98 dokumen, diikuti Topik 1 (Seni, Budaya, dan Usaha Kreatif) dengan 55 dokumen, dan Topik 2 (Pemilu, Politik dan Kebangsaan) dengan 34 dokumen. Distribusi ini mengindikasikan bahwa isu pendidikan merupakan domain aspirasi yang paling masif dan mendesak di mata publik. Di sisi lain, skor c-TF-IDF tertinggi justru ditemukan pada topik-topik dengan jumlah dokumen yang lebih sedikit namun berkarakter sangat spesifik, seperti Topik 14 (lansia, skor 0,9628), Topik 9 (tani, skor 0,8485), dan Topik 11 (alokasi anggaran, skor 0,8173). Fenomena ini menunjukkan bahwa tingginya skor c-TF-IDF mencerminkan kekohesian semantik dan kekhasan leksikal suatu kluster, bukan semata-mata volume dokumen yang dikandungnya. Dengan demikian, kedua dimensi tersebut (frekuensi dokumen dan skor representatif) perlu dibaca secara komplementer dalam menilai urgensi dan kedalaman suatu isu.

Menjawab rumusan masalah kedua, 15 topik yang terbentuk mampu merepresentasikan isu-isu prioritas masyarakat secara komprehensif, mulai dari domain pendidikan, kebudayaan, demokrasi elektoral, keagamaan, lingkungan hidup, jaminan kesehatan, pemberdayaan ekonomi kerakyatan, pariwisata, tata kelola anggaran, pertanian, inklusi sosial, infrastruktur dasar, hingga kesejahteraan kelompok rentan. Pemetaan lintas komite DPD RI terhadap 15 topik tersebut menunjukkan bahwa Komite III memiliki cakupan aspirasi paling luas, meliputi tujuh topik sekaligus, sementara Komite II mendapatkan sinyal aspirasi paling kuat

pada Topik 9 yang mencatatkan skor c-TF-IDF tertinggi secara keseluruhan. Komite I dan Komite IV masing-masing memperoleh dua topik dengan relevansi langsung terhadap fungsi ketugasannya, yakni tata kelola demokrasi dan inklusi sosial untuk Komite I, serta transparansi anggaran dan pemerataan infrastruktur untuk Komite IV. Secara keseluruhan, hasil penelitian ini membuktikan bahwa metode BERTopic mampu mengolah data teks aspirasi yang bersifat tidak terstruktur menjadi representasi tematik yang informatif, sehingga berpotensi menjadi instrumen analitik berbasis data dalam mendukung fungsi representasi, legislasi, dan pengawasan DPD RI secara lebih responsif dan akuntabel.

6.2. Saran

Berdasarkan temuan dan keterbatasan penelitian ini, beberapa saran dirumuskan sebagai berikut.

Pertama, cakupan penelitian dapat diperluas dengan mengintegrasikan data aspirasi dari seluruh wilayah dapil DPD RI, sehingga pemetaan isu prioritas yang dihasilkan bersifat nasional dan dapat menjadi basis kebijakan representasi yang lebih menyeluruh.

Kedua, untuk meningkatkan validitas interpretatif hasil klasterisasi, penelitian berikutnya perlu melibatkan pakar domain, seperti ahli kebijakan publik, pendidikan, dan kesehatan dalam proses validasi kualitatif label topik, sehingga temuan pemodelan tidak hanya sah secara komputasional tetapi juga relevan secara substantif bagi pemangku kepentingan.

DAFTAR PUSTAKA

- Aleqabie, H. J., Sfoq, M. S., Albeer, R. A., & Abd, E. H. (2024). A Review Of TextMining Techniques: Trends, and Applications In Various Domains. *Iraqi Journal for Computer Science and Mathematics*, 5(1). <https://doi.org/10.52866/ijcsm.2024.05.01.009>
- Davies, D. L., & Bouldin, D. W. (1979). A Cluster Separation Measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-1(2), 224–227. <https://doi.org/10.1109/TPAMI.1979.4766909>
- Feng, Y., Chen, Z., Zhang, Y., Huang, W., Zhang, X., & He, S. (2025). BERTopic_Teen: a Multi-module Optimization Approach for Short Text Topic Modeling in Adolescent Health. *Frontiers in Public Health*, 13. <https://doi.org/10.3389/fpubh.2025.1608241>
- Gottumukkala, D. P., Prasad, P. R., & Rao, S. K. (2025a). Topic modeling-based prediction of software defects and root cause using BERTopic, and multioutput classifier. *Scientific Reports*, 15(1). <https://doi.org/10.1038/s41598-025-11458-0>
- Gottumukkala, D. P., Prasad, P. R., & Rao, S. K. (2025b). Topic modeling-based prediction of software defects and root cause using BERTopic, and multioutput classifier. *Scientific Reports*, 15(1). <https://doi.org/10.1038/s41598-025-11458-0>
- Grisales, A. M., Robledo, S., & Zuluaga, M. (2023). Topic Modeling: Perspectives From a Literature Review. In *IEEE Access* (Vol. 11, pp. 4066–4078). Institute of Electrical and Electronics Engineers Inc. <https://doi.org/10.1109/ACCESS.2022.3232939>
- Grootendorst, M. (2022). *BERTopic: Neural topic modeling with a class-based TF-IDF procedure*. <http://arxiv.org/abs/2203.05794>
- HaCohen-Kerner, Y., Miller, D., & Yigal, Y. (2020). The influence of preprocessing on text classification using a bag-of-words representation. *PLoS ONE*, 15(5). <https://doi.org/10.1371/journal.pone.0232525>
- Khodeir, N., & Elghannam, F. (2025). Efficient topic identification for urgent MOOC Forum posts using BERTopic and traditional topic modeling

- techniques. *Education and Information Technologies*, 30(5), 5501–5527.
<https://doi.org/10.1007/s10639-024-13003-4>
- Liu, Y., & Wan, F. (2024). Unveiling temporal and spatial research trends in precision agriculture: A BERTopic text mining approach. *Heliyon*, 10(17).
<https://doi.org/10.1016/j.heliyon.2024.e36808>
- Ma, L., Chen, R., Ge, W., Rogers, P., Lyn-Cook, B., Hong, H., Tong, W., Wu, N., & Zou, W. (2025). AI-powered topic modeling: comparing LDA and BERTopic in analyzing opioid-related cardiovascular risks in women. *Experimental Biology and Medicine*, 250.
<https://doi.org/10.3389/ebm.2025.10389>
- Manning, C. D. ., Raghavan, Prabhakar., & Schütze, Hinrich. (2009). *Introduction to information retrieval*. Cambridge University Press.
- McInnes, L., Healy, J., & Astels, S. (2017). hdbscan: Hierarchical density based clustering. *The Journal of Open Source Software*, 2(11), 205.
<https://doi.org/10.21105/joss.00205>
- Mertayasa, I. K. T., & Darmawan, I. D. M. B. A. (2022). Pemodelan Topik Pada Ulasan Hotel Menggunakan Metode BERTopic Dengan Prosedur c-TF-IDF. *Jurnal Nasional Teknologi Informasi Dan Aplikasinya*, 1.
- Myftaraj, E., Fata, I., & Plasari, E. (2024). A Topic Modeling Analysis of Circular Economy and Big Data Research Using BERTopic and SciBERT. *International Journal of Innovative Technology and Interdisciplinary Sciences*. <https://doi.org/10.1515/IJITIS.2025.8.4.1195-1243>
- Nanyonga, A., Joiner, K., Turhan, U., & Wild, G. (2025). *Semantic Topic Modeling of Aviation Safety Reports: A Comparative Analysis Using BERTopic and PLSA*. <https://doi.org/10.20944/preprints202505.1509.v1>
- Redddy, B. U. J., Rami Reddy, C. R., & Kavya, B. (2023). Techniques, Applications and Issues of Text Mining. *Research Journal of Engineering and Technology*, 1–6. <https://doi.org/10.52711/2321-581X.2023.00001>
- Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. *ArXiv*.
<https://doi.org/https://doi.org/10.48550/arXiv.1908.10084>

- Rejeb, A., Rejeb, K., Zaher, H. F., & Simske, S. (2025). Blockchain and Smart Cities: Co-Word Analysis and BERTopic Modeling. In *Smart Cities* (Vol. 8, Number 4). Multidisciplinary Digital Publishing Institute (MDPI). <https://doi.org/10.3390/smartcities8040111>
- Robertson, S., & Zaragoza, H. (2009). The Probabilistic Relevance Framework: BM25 and Beyond. *Foundations and Trends® in Information Retrieval*, 4(1–2), 1–174. <https://doi.org/10.1561/15000000019>
- Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20, 53–65. [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7)
- Rusdiyantara, A. (2025). *Wawancara Penerimaan Aspirasi Masyarakat di DPD RI DIY*.
- Simanjuntak, K. A., Koyimatu, M., Ervanisari, Y. P., & Tasmi. (2024). Identifikasi Opini Publik Terhadap Kendaraan Listrik dari Data Komentar YouTube: Pemodelan Topik Menggunakan BERTopic. *TEMATIK*, 11(2), 195–203. <https://doi.org/10.38204/tematik.v11i2.2096>
- Steck, H., Ekanadham, C., & Kallus, N. (2024). Is Cosine-Similarity of Embeddings Really About Similarity? *Companion Proceedings of the ACM Web Conference 2024*, 887–890. <https://doi.org/10.1145/3589335.3651526>
- Stewart, G., & Al-Khassaweneh, M. (2022). An Implementation of the HDBSCAN* Clustering Algorithm. *Applied Sciences (Switzerland)*, 12(5). <https://doi.org/10.3390/app12052405>
- Su, Y. S., Wang, J. qing, Tu, S. H., Liao, K. T., & Lin, C. L. (2025). Detecting latent topics and trends in IoT and e-commerce using BERTopic modeling. *Internet of Things (The Netherlands)*, 32. <https://doi.org/10.1016/j.iot.2025.101604>
- Wahyuni, W., Lestari, T. P., Apriliana, M., & Gumelta, R. (2025). Implementation of BERTopic for Topic Modeling Analysis of the Free Nutritious Meal Program Based on YouTube Comments. In *Journal of Applied Informatics and Computing (JAIC)* (Vol. 9, Number 4). <http://jurnal.polibatam.ac.id/index.php/JAIC>
- Yudhana, A., Fadlil, A., & Rosidin, M. (2019). Indonesian Words Error Detection System using Nazief Adriani Stemmer Algorithm. In *IJACSA) International*

Journal of Advanced Computer Science and Applications (Vol. 10, Number 12). www.ijacsa.thesai.org

Zare, D., Fernandez-Sanz, L., Pospelova, V., & López-Baldominos, I. (2025). NLP and Text Mining for Enriching IT Professional Skills Frameworks. *Applied Sciences (Switzerland)*, 15(17). <https://doi.org/10.3390/app15179634>

RINGKASAN TA

Topic Modelling Aspirasi Masyarakat Menggunakan Metode BERTopic (Studi Kasus: Aspirasi Masyarakat di Dewan Perwakilan Daerah Republik Indonesia)

Nail Anfasa Fadla, Program Studi Statistika, Universitas Islam Indonesia

Abstrak

Peningkatan volume aspirasi masyarakat yang disampaikan kepada Dewan Perwakilan Daerah Republik Indonesia (DPD RI) menuntut adanya pendekatan analitis yang lebih sistematis dan berbasis data. Penelitian ini mengaplikasikan metode BERTopic untuk mengidentifikasi topik-topik utama dari data aspirasi masyarakat Daerah Istimewa Yogyakarta yang dihimpun selama Tahun Sidang 2024/2025. Dataset terdiri dari 486 dokumen aspirasi yang seluruhnya valid tanpa nilai kosong. Proses analisis meliputi pra-pemrosesan teks berbahasa Indonesia, pembentukan representasi semantik menggunakan model *sentence transformer* LazarusNLP/all-indo-e5-small-v4, reduksi dimensi dengan UMAP, klusterisasi berbasis kepadatan menggunakan HDBSCAN, serta ekstraksi kata kunci representatif melalui pendekatan c-TF-IDF yang diperkaya dengan KeyBERTInspired dan Maximal Marginal Relevance (MMR). Hasil pemodelan menghasilkan 15 topik tematik dengan 81 dokumen (16,7%) teridentifikasi sebagai *outlier*. Evaluasi kuantitatif menunjukkan Silhouette Score sebesar 0,4754 dan Davies-Bouldin Index sebesar 0,7238, yang mengindikasikan struktur kluster dengan kohesi internal yang cukup baik dan separasi antartopik yang memadai. Topik dengan dokumen terbanyak adalah Pendidikan, Perempuan dan Anak (98 dokumen), sementara skor c-TF-IDF tertinggi ditemukan pada topik berkarakter spesifik seperti Aktivitas Lansia (0,9628), Pertanian, Irigasi, Embung, dan Hasil Tani (0,8485), dan Alokasi Anggaran dan Pembangunan Infrastruktur (0,8173). Pemetaan lintas komite DPD RI menunjukkan bahwa Komite III memiliki cakupan aspirasi terluas dengan tujuh topik relevan, sedangkan Komite II menerima sinyal aspirasi terkuat pada sektor pertanian. Penelitian ini membuktikan bahwa BERTopic mampu mengolah data teks aspirasi tidak terstruktur menjadi

representasi tematik yang informatif, sehingga berpotensi menjadi instrumen analitik berbasis data dalam mendukung fungsi representasi, legislasi, dan pengawasan DPD RI secara lebih responsif dan akuntabel.

Kata kunci: BERTopic, topic modeling, aspirasi masyarakat, DPD RI, Natural Language Processing, c-TF-IDF, HDBSCAN

1. Pendahuluan

Dalam beberapa tahun terakhir, volume aspirasi masyarakat yang disampaikan melalui anggota Dewan Perwakilan Daerah Republik Indonesia (DPD RI) menunjukkan peningkatan yang signifikan. Beragam isu mulai dari ekonomi, pendidikan, kesehatan, hingga tata kelola pemerintahan kini banyak disampaikan oleh masyarakat kepada para anggota DPD RI sebagai representasi daerah. Dinamika ini mencerminkan meningkatnya kesadaran publik terhadap peran DPD RI sebagai saluran aspirasi politik dan kebijakan.

Sebagai lembaga negara yang memiliki fungsi representasi daerah, DPD RI dituntut untuk menyerap dan menyalurkan aspirasi publik secara cepat, tepat, dan berbasis data. Akan tetapi, volume data aspirasi yang sangat besar, bersifat tidak terstruktur, dan beragam dari segi isu maupun gaya bahasa menjadikan proses analisis manual tidak lagi memadai. Tanpa alat bantu analitis yang efektif, potensi informasi penting dalam aspirasi publik dapat terabaikan, sehingga melemahkan peran DPD RI sebagai jembatan antara masyarakat daerah dan kebijakan nasional.

Salah satu pendekatan yang relevan untuk menjawab tantangan ini adalah topic modeling, yaitu teknik analisis teks yang mampu mengidentifikasi tema-tema utama dari kumpulan data teks dalam jumlah besar. Model klasik seperti Latent Dirichlet Allocation (LDA) telah banyak digunakan untuk keperluan ini, namun memiliki keterbatasan dalam menangkap makna kontekstual karena hanya bergantung pada frekuensi kata. Perkembangan teknologi transformer-based models kemudian menghadirkan alternatif yang lebih unggul, yaitu BERTopic, yang menggabungkan kekuatan representasi teks berbasis BERT, reduksi dimensi, dan clustering untuk menghasilkan topik yang lebih koheren dan interpretatif.

Penelitian ini mengaplikasikan BERTopic pada data aspirasi masyarakat yang disampaikan kepada anggota DPD RI Daerah Istimewa Yogyakarta selama

Tahun Sidang 2024/2025. Tujuan penelitian adalah: (1) menganalisis distribusi dokumen dan skor representatif (c-TF-IDF score) pada setiap topik yang dihasilkan, dan (2) mengidentifikasi topik-topik utama serta menginterpretasikan temuan dalam rangka mengungkap isu prioritas masyarakat guna mendukung penguatan fungsi representasi DPD RI.

2. Tinjauan Pustaka

Topic modeling merupakan teknik dalam Natural Language Processing (NLP) yang digunakan untuk mengidentifikasi pola dan tema laten dari sekumpulan dokumen teks. Model klasik seperti Latent Dirichlet Allocation (LDA) dan Non-negative Matrix Factorization (NMF) telah banyak digunakan, namun keduanya memiliki keterbatasan dalam menangkap relasi semantik antarkata karena hanya mengandalkan frekuensi kemunculan kata tanpa mempertimbangkan konteks (Grisales et al., 2023).

BERTopic, yang diperkenalkan oleh Grootendorst pada tahun 2022, mengombinasikan BERT embeddings, reduksi dimensi UMAP, algoritma clustering HDBSCAN, serta representasi topik berbasis c-TF-IDF. Pendekatan ini memungkinkan pengelompokan dokumen yang lebih koheren dan mampu menangkap relasi semantik antarkalimat (Khodeir & Elghannam, 2025). Berbagai penelitian telah membuktikan keunggulan BERTopic: Wahyuni et al. (2025) menemukan BERTopic menghasilkan koherensi topik 0,46 dibandingkan LDA sebesar 0,39 dalam menganalisis komentar YouTube; Simanjuntak et al. (2024) menunjukkan efektivitasnya dalam memetakan opini publik terkait kendaraan listrik dengan koherensi 0,6–1,0 dan topic diversity 0,95; serta berbagai studi di bidang kesehatan, teknologi, dan rekayasa perangkat lunak (Ma et al., 2025; Su et al., 2025; Gottumukkala et al., 2025).

Di Indonesia, penggunaan BERTopic dalam konteks lembaga negara masih relatif terbatas. Aryani et al. (2024) menganalisis berita Pemilu 2024 menggunakan BERTopic dan menunjukkan bahwa metode ini mampu mengurai isu politik kompleks menjadi tema-tema yang lebih sederhana, mendukung perumusan kebijakan berbasis aspirasi masyarakat. Penelitian ini hadir untuk mengisi kesenjangan tersebut dengan mengaplikasikan BERTopic pada analisis aspirasi masyarakat DPD RI.

3. Metode Penelitian

3.1 Data dan Sampel

Penelitian ini menggunakan data aspirasi masyarakat (Asmas) yang dihimpun oleh anggota DPD RI Daerah Istimewa Yogyakarta selama Tahun Sidang 2024/2025. Data diperoleh dari hasil himpunan catatan reses, forum dengar pendapat, dan sumber administratif resmi lainnya, kemudian dikompilasi dalam format teks digital. Total dataset terdiri dari 486 dokumen aspirasi yang seluruhnya valid tanpa nilai kosong, dengan rata-rata panjang dokumen 36 kata (277,7 karakter), nilai minimum 6 kata, dan nilai maksimum 110 kata.

3.2 Alur Analisis BERTopic

Penelitian ini menggunakan pendekatan kuantitatif eksploratif berbasis unsupervised learning dengan metode BERTopic. Alur analisis terdiri atas tahapan berikut:

Pertama, pra-pemrosesan teks meliputi: case folding (konversi ke huruf kecil), penghapusan URL dan karakter numerik, penghapusan tanda baca, normalisasi spasi, penghapusan stopwords menggunakan Sastrawi dan custom stopwords domain-spesifik, serta stemming menggunakan algoritma Nazief-Adriani. Setelah pra-pemrosesan, seluruh 486 dokumen tetap dipertahankan karena tidak ada dokumen yang menghasilkan teks di bawah tiga token.

Kedua, representasi semantik dokumen dilakukan menggunakan model sentence transformer LazarusNLP/all-indo-e5-small-v4 yang telah di-fine-tune pada korpus berbahasa Indonesia. Setiap dokumen diproyeksikan ke dalam ruang vektor berdimensi 384, menghasilkan matriks embedding berukuran (486×384) .

Ketiga, reduksi dimensi dilakukan menggunakan UMAP dengan parameter: `n_neighbors=15`, `n_components=5`, `min_dist=0.0`, `metric='cosine'`, dan `random_state=42`. Pemilihan lima dimensi target (bukan dua dimensi) dimaksudkan untuk menjaga informasi struktural yang diperlukan oleh HDBSCAN dalam membentuk klaster yang koheren.

Keempat, klasterisasi topik menggunakan HDBSCAN dengan parameter: `min_cluster_size=10`, `min_samples=5`, `metric='euclidean'`, `cluster_selection_method='eom'`, dan `prediction_data=True`.

Kelima, representasi topik diekstraksi menggunakan c-TF-IDF melalui CountVectorizer dengan `ngram_range=(1,2)`, `min_df=2`, `max_df=0.85`, serta pembobotan BM25 dan `reduce_frequent_words=True`. Representasi diperkaya lebih lanjut menggunakan KeyBERTInspired (`top_n_words=15`, `nr_repr_docs=5`) dan Maximal Marginal Relevance/MMR (`diversity=0.3`).

Keenam, evaluasi model dilakukan secara kuantitatif menggunakan Silhouette Score dan Davies-Bouldin Index dalam ruang vektor UMAP, dengan mengeksklusi dokumen outlier (`label -1`) dari perhitungan.

4. Hasil dan Pembahasan

4.1 Hasil Pemodelan Topik

Dari total 486 dokumen aspirasi, BERTopic berhasil mengidentifikasi 15 topik tematik utama. Sebanyak 81 dokumen (16,7%) terklasifikasi sebagai outlier, proporsi yang berada pada tingkat moderat dan mengindikasikan bahwa sebagian besar dokumen telah berhasil terkaster dengan baik. Evaluasi kuantitatif menunjukkan Silhouette Score sebesar 0,4754 dan Davies-Bouldin Index sebesar 0,7238, yang mengindikasikan struktur klaster dengan kohesi internal yang cukup baik serta separasi antartopik yang memadai.

4.2 Distribusi Dokumen dan Skor c-TF-IDF

Distribusi dokumen antar topik menunjukkan ketimpangan yang signifikan. Topik 0 (Pendidikan, Perempuan dan Anak) mendominasi dengan 98 dokumen, diikuti Topik 1 (Seni, Budaya, dan Usaha Kreatif) dengan 55 dokumen, dan Topik 2 (Pemilu, Politik dan Kebangsaan) dengan 34 dokumen. Distribusi ini mengindikasikan bahwa isu pendidikan merupakan domain aspirasi yang paling masif dan mendesak.

Skor c-TF-IDF tertinggi justru ditemukan pada topik-topik dengan jumlah dokumen lebih sedikit namun berkarakter sangat spesifik: Topik 14 mencatatkan skor tertinggi pada kata 'lansia' (0,9628), diikuti Topik 9 dengan kata 'tani' (0,8485) dan Topik 11 dengan kata 'alokasi anggaran' (0,8173). Fenomena ini menunjukkan bahwa tingginya skor c-TF-IDF mencerminkan kekohesian semantik dan kekhasan leksikal suatu kluster, bukan semata-mata volume dokumennya.

4.3 Identifikasi 15 Topik Utama

Kelima belas topik yang terbentuk merepresentasikan spektrum isu prioritas publik secara menyeluruh, sebagaimana diuraikan sebagai berikut.

Topik 0: Pendidikan, Perempuan dan Anak, ditandai oleh kata kunci dominan seperti didik, sekolah, anak, dan perempuan. Topik ini merepresentasikan aspirasi masyarakat terkait pemerataan akses pendidikan, penguatan kurikulum berbasis karakter, serta perhatian khusus terhadap pendidikan dan perempuan. Dokumen-dokumen dalam kluster ini mengindikasikan adanya kebutuhan akan dukungan infrastruktur dan kebijakan afirmatif di sektor pendidikan.

Topik 1: Seni, Budaya, dan Usaha Kreatif, dicirikan oleh kata kunci seni, tradisional, kreatif, budaya, dan muda. Topik ini mencerminkan keprihatinan terhadap keberlanjutan seni dan kebudayaan lokal, sekaligus dorongan untuk mengintegrasikan nilai-nilai budaya ke dalam ekosistem kreatif generasi muda melalui platform digital dan program kepemudaan.

Topik 2: Pemilu, Politik dan Kebangsaan, didominasi oleh kata kunci milu (pemilu), pilih, politik, kampanye, dan mahasiswa. Topik ini merangkum aspirasi yang berkaitan dengan peningkatan partisipasi publik dalam proses demokrasi elektoral, termasuk harapan terhadap sinergi antara organisasi kemasyarakatan dan penyelenggara pemilu seperti KPU.

Topik 3: Keagamaan, mengangkat isu-isu yang berkaitan dengan halal, agama, warga, muhammadiyah, dan sertifikat. Topik ini merefleksikan tuntutan masyarakat terhadap pengembangan ekosistem usaha yang berlandaskan prinsip syariah, termasuk percepatan sertifikasi halal bagi pelaku usaha mikro serta penguatan toleransi antarumat beragama.

Topik 4: Pengelolaan Air, Sampah, dan Lingkungan, ditandai oleh kata kunci air, sampah, kelola sampah, hujan, dan bersih. Topik ini mengindikasikan kekhawatiran yang signifikan terhadap degradasi kualitas lingkungan, mencakup pengelolaan sampah, ketahanan terhadap banjir, serta aksesibilitas air bersih di kawasan perkotaan dan perdesaan.

Topik 5: BPJS dan Layanan Kesehatan, diidentifikasi melalui kata kunci bpjs, bpjs sehat, iur, kanker, dan medis. Topik ini mencerminkan aspirasi masyarakat terkait reformasi sistem jaminan kesehatan nasional, khususnya perluasan cakupan

layanan bagi pasien dengan penyakit kronis seperti kanker, serta peningkatan kualitas fasilitas kesehatan primer.

Topik 6: UMKM, Koperasi, dan Pinjaman, dicirikan oleh kata kunci umkm, ukm, koperasi, perban, dan miskin. Topik ini menggarisbawahi pentingnya pendampingan usaha, akses permodalan, dan penguatan kelembagaan koperasi sebagai instrumen pengentasan kemiskinan dan peningkatan daya saing ekonomi lokal.

Topik 7: Pariwisata dan Perhotelan, ditandai oleh kata kunci pariwisata, sektor pariwisata, gipi, wisatawan, hotel, dan tiket. Topik ini merepresentasikan aspirasi pelaku industri pariwisata terhadap penyelesaian hambatan regulasi, peningkatan promosi destinasi, serta dukungan pemerintah dalam menarik investasi di sektor perhotelan dan jasa wisata.

Topik 8: Target dan Realisasi Anggaran Layanan Publik, diidentifikasi melalui kata kunci capai rp, target, realisasi, dan layan publik. Topik ini menyoroti persoalan tumpang tindih birokrasi, transparansi realisasi anggaran, serta desakan masyarakat terhadap pembentukan unit layanan rehabilitasi berbasis kelurahan yang lebih efisien dan akuntabel.

Topik 9: Pertanian, Irigasi, Embung, dan Hasil Tani, ditandai oleh kata kunci tani, embung, irigasi, hasil tani, dan lahan tani. Topik ini merangkum harapan komunitas petani terhadap pembenahan jaringan irigasi, pembangunan embung, serta peningkatan produktivitas lahan sebagai bentuk nyata keberpihakan pemerintah terhadap sektor agraris.

Topik 10: Inklusi Sosial dan Pemberdayaan Penyandang Disabilitas, dicirikan oleh kata kunci disabilitas, sandang disabilitas, mitra, yayasan, dan online. Topik ini merefleksikan kebutuhan akan regulasi yang lebih jelas, dukungan anggaran berkelanjutan, dan pengembangan program vokasional berbasis digital bagi penyandang disabilitas untuk meningkatkan kemandirian ekonomi mereka.

Topik 11: Alokasi Anggaran dan Pembangunan Infrastruktur, diidentifikasi melalui kata kunci alokasi anggar, infrastruktur, ada sarana, dan batas. Topik ini mencerminkan tuntutan masyarakat terhadap keberpihakan anggaran pemerintah dalam penyediaan sarana prasarana dasar, termasuk fasilitas pendidikan, teknologi informasi, dan infrastruktur wilayah yang masih tertinggal.

Topik 12: Destinasi Wisata dan Keselamatan, ditandai oleh kata kunci wisata, destinasi wisata, sar, selamat, dan promosi. Topik ini mengangkat pentingnya integrasi aspek keselamatan, termasuk keterlibatan tim SAR dalam pengelolaan destinasi wisata, serta pemberdayaan pelaku UMKM lokal sebagai bagian integral dari ekosistem pariwisata berbasis masyarakat.

Topik 13: Gaji, Pensiun, dan Kesejahteraan Guru, dicirikan oleh kata kunci guru, pensiun, gaji, swasta, sejahtera guru, dan upah minimum. Topik ini menyuarakan aspirasi kalangan pendidik, khususnya guru swasta dan tenaga honorer, yang menuntut peningkatan remunerasi, kejelasan status kepegawaian, serta jaminan perlindungan sosial yang setara dengan aparatur negara.

Topik 14: Aktivitas Lansia, diidentifikasi melalui kata kunci lansia, sehat lansia, usia, aktivitas, mudah jangkau, dan rumah singgah. Topik ini merepresentasikan kebutuhan spesifik kelompok lanjut usia akan fasilitas kesehatan yang aksesibel, program aktivitas sosial, serta regulasi daerah yang secara eksplisit berpihak pada pemenuhan hak-hak dasar warga lansia.

4.4 Pemetaan Topik terhadap Fungsi Ketugasan DPD RI

Hasil pemetaan 15 topik ke dalam empat komite DPD RI menunjukkan distribusi yang proporsional. Komite I (otonomi daerah dan hubungan pusat-daerah) memperoleh relevansi dari Topik 2 (Pemilu, Politik dan Kebangsaan) dan Topik 10 (Inklusi Sosial dan Pemberdayaan Penyandang Disabilitas). Komite II (sumber daya alam dan ekonomi) mendapat sinyal aspirasi terkuat pada Topik 9 (Pertanian, Irigasi, Embung, dan Hasil Tani) dengan skor c-TF-IDF tertinggi (0,8485), dilengkapi Topik 4 (Pengelolaan Air, Sampah, dan Lingkungan) dan Topik 6 (UMKM, Koperasi, dan Pinjaman).

Komite III (pendidikan, agama, kesehatan, dan sosial budaya) memiliki cakupan aspirasi terluas dengan tujuh topik relevan: Topik 0 (Pendidikan, Perempuan dan Anak), Topik 13 (Gaji, Pensiun, dan Kesejahteraan Guru), Topik 3 (Keagamaan), Topik 1 (Seni, Budaya, dan Usaha Kreatif), Topik 7 dan 12 (Pariwisata dan Perhotelan; Destinasi Wisata dan Keselamatan), serta Topik 5 dan 14 (BPJS dan Layanan Kesehatan; Aktivitas Lansia). Komite IV (anggaran dan pengawasan keuangan negara) memperoleh relevansi langsung dari Topik 8 (Target dan Realisasi Anggaran Layanan Publik) dan Topik 11 (Alokasi Anggaran dan

Pembangunan Infrastruktur), yang keduanya mengindikasikan tuntutan kuat terhadap transparansi fiskal dan pemerataan pembangunan antarwilayah.

Visualisasi intertopic distance map menunjukkan bahwa sebagian besar topik terkonsentrasi dalam satu gugus utama yang menandakan kedekatan semantik antartopik, sementara beberapa topik seperti topik guru/pensiun, lansia, dan alokasi anggaran tampak lebih terpisah, mencerminkan karakter tematik yang lebih spesifik. Heatmap kemiripan antartopik mengkonfirmasi bahwa model memiliki diferensiasi semantik yang baik, dengan irisan semantik wajar hanya terjadi pada pasangan topik yang secara substantif berkaitan, seperti seni tradisional dengan pariwisata, serta pengelolaan lingkungan dengan pertanian.

5. Kesimpulan

Penelitian ini berhasil menerapkan metode BERTopic untuk memodelkan 486 dokumen aspirasi masyarakat DIY yang disampaikan kepada DPD RI selama Tahun Sidang 2024/2025, menghasilkan 15 topik tematik yang merepresentasikan spektrum isu prioritas publik secara terstruktur dan terukur.

Menjawab rumusan masalah pertama, distribusi dokumen menunjukkan dominasi Topik 0 (Pendidikan, Perempuan dan Anak, 98 dokumen), diikuti Topik 1 (Seni, Budaya, dan Usaha Kreatif, 55 dokumen) dan Topik 2 (Pemilu, Politik dan Kebangsaan, 34 dokumen). Sementara itu, skor c-TF-IDF tertinggi ditemukan pada topik-topik berkarakter spesifik: Topik 14 (lansia, 0,9628), Topik 9 (tani, 0,8485), dan Topik 11 (alokasi anggaran, 0,8173), menunjukkan bahwa kedua dimensi tersebut perlu dibaca secara komplementer.

Menjawab rumusan masalah kedua, 15 topik yang terbentuk mampu merepresentasikan isu-isu prioritas secara komprehensif. Pemetaan lintas komite menunjukkan Komite III memiliki cakupan terluas (7 topik), Komite II mendapat sinyal aspirasi terkuat pada sektor pertanian, serta Komite I dan IV masing-masing memperoleh dua topik dengan relevansi langsung. Hasil evaluasi dengan Silhouette Score 0,4754 dan Davies-Bouldin Index 0,7238 mengkonfirmasi bahwa BERTopic mampu mengolah data teks aspirasi tidak terstruktur menjadi representasi tematik yang valid dan dapat dipertanggungjawabkan sebagai landasan analitik bagi DPD RI.

DAFTAR PUSTAKA

- Aleqabie, H. J., Sfoq, M. S., Albeer, R. A., & Abd, E. H. (2024). A Review Of TextMining Techniques: Trends, and Applications In Various Domains. *Iraqi Journal for Computer Science and Mathematics*, 5(1). <https://doi.org/10.52866/ijcsm.2024.05.01.009>
- Davies, D. L., & Bouldin, D. W. (1979). A Cluster Separation Measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-1(2), 224–227. <https://doi.org/10.1109/TPAMI.1979.4766909>
- Feng, Y., Chen, Z., Zhang, Y., Huang, W., Zhang, X., & He, S. (2025). BERTopic_Teen: a Multi-module Optimization Approach for Short Text Topic Modeling in Adolescent Health. *Frontiers in Public Health*, 13. <https://doi.org/10.3389/fpubh.2025.1608241>
- Gottumukkala, D. P., Prasad, P. R., & Rao, S. K. (2025a). Topic modeling-based prediction of software defects and root cause using BERTopic, and multioutput classifier. *Scientific Reports*, 15(1). <https://doi.org/10.1038/s41598-025-11458-0>
- Gottumukkala, D. P., Prasad, P. R., & Rao, S. K. (2025b). Topic modeling-based prediction of software defects and root cause using BERTopic, and multioutput classifier. *Scientific Reports*, 15(1). <https://doi.org/10.1038/s41598-025-11458-0>
- Grisales, A. M., Robledo, S., & Zuluaga, M. (2023). Topic Modeling: Perspectives From a Literature Review. In *IEEE Access* (Vol. 11, pp. 4066–4078). Institute of Electrical and Electronics Engineers Inc. <https://doi.org/10.1109/ACCESS.2022.3232939>
- Grootendorst, M. (2022). *BERTopic: Neural topic modeling with a class-based TF-IDF procedure*. <http://arxiv.org/abs/2203.05794>
- HaCohen-Kerner, Y., Miller, D., & Yigal, Y. (2020). The influence of preprocessing on text classification using a bag-of-words representation. *PLoS ONE*, 15(5). <https://doi.org/10.1371/journal.pone.0232525>
- Khodeir, N., & Elghannam, F. (2025). Efficient topic identification for urgent MOOC Forum posts using BERTopic and traditional topic modeling

- techniques. *Education and Information Technologies*, 30(5), 5501–5527.
<https://doi.org/10.1007/s10639-024-13003-4>
- Liu, Y., & Wan, F. (2024). Unveiling temporal and spatial research trends in precision agriculture: A BERTopic text mining approach. *Heliyon*, 10(17).
<https://doi.org/10.1016/j.heliyon.2024.e36808>
- Ma, L., Chen, R., Ge, W., Rogers, P., Lyn-Cook, B., Hong, H., Tong, W., Wu, N., & Zou, W. (2025). AI-powered topic modeling: comparing LDA and BERTopic in analyzing opioid-related cardiovascular risks in women. *Experimental Biology and Medicine*, 250.
<https://doi.org/10.3389/ebm.2025.10389>
- Manning, C. D. ., Raghavan, Prabhakar., & Schütze, Hinrich. (2009). *Introduction to information retrieval*. Cambridge University Press.
- McInnes, L., Healy, J., & Astels, S. (2017). hdbscan: Hierarchical density based clustering. *The Journal of Open Source Software*, 2(11), 205.
<https://doi.org/10.21105/joss.00205>
- Mertayasa, I. K. T., & Darmawan, I. D. M. B. A. (2022). Pemodelan Topik Pada Ulasan Hotel Menggunakan Metode BERTopic Dengan Prosedur c-TF-IDF. *Jurnal Nasional Teknologi Informasi Dan Aplikasinya*, 1.
- Myftaraj, E., Fata, I., & Plasari, E. (2024). A Topic Modeling Analysis of Circular Economy and Big Data Research Using BERTopic and SciBERT. *International Journal of Innovative Technology and Interdisciplinary Sciences*. <https://doi.org/10.1515/IJITIS.2025.8.4.1195-1243>
- Nanyonga, A., Joiner, K., Turhan, U., & Wild, G. (2025). *Semantic Topic Modeling of Aviation Safety Reports: A Comparative Analysis Using BERTopic and PLSA*. <https://doi.org/10.20944/preprints202505.1509.v1>
- Redddy, B. U. J., Rami Reddy, C. R., & Kavya, B. (2023). Techniques, Applications and Issues of Text Mining. *Research Journal of Engineering and Technology*, 1–6. <https://doi.org/10.52711/2321-581X.2023.00001>
- Reimers, N., & Gurevych, I. (2019). Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks. *ArXiv*.
<https://doi.org/https://doi.org/10.48550/arXiv.1908.10084>

- Rejeb, A., Rejeb, K., Zaher, H. F., & Simske, S. (2025). Blockchain and Smart Cities: Co-Word Analysis and BERTopic Modeling. In *Smart Cities* (Vol. 8, Number 4). Multidisciplinary Digital Publishing Institute (MDPI). <https://doi.org/10.3390/smartcities8040111>
- Robertson, S., & Zaragoza, H. (2009). The Probabilistic Relevance Framework: BM25 and Beyond. *Foundations and Trends® in Information Retrieval*, 4(1–2), 1–174. <https://doi.org/10.1561/15000000019>
- Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20, 53–65. [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7)
- Rusdiyantara, A. (2025). *Wawancara Penerimaan Aspirasi Masyarakat di DPD RI DIY*.
- Simanjuntak, K. A., Koyimatu, M., Ervanisari, Y. P., & Tasmi. (2024). Identifikasi Opini Publik Terhadap Kendaraan Listrik dari Data Komentar YouTube: Pemodelan Topik Menggunakan BERTopic. *TEMATIK*, 11(2), 195–203. <https://doi.org/10.38204/tematik.v11i2.2096>
- Steck, H., Ekanadham, C., & Kallus, N. (2024). Is Cosine-Similarity of Embeddings Really About Similarity? *Companion Proceedings of the ACM Web Conference 2024*, 887–890. <https://doi.org/10.1145/3589335.3651526>
- Stewart, G., & Al-Khassaweneh, M. (2022). An Implementation of the HDBSCAN* Clustering Algorithm. *Applied Sciences (Switzerland)*, 12(5). <https://doi.org/10.3390/app12052405>
- Su, Y. S., Wang, J. qing, Tu, S. H., Liao, K. T., & Lin, C. L. (2025). Detecting latent topics and trends in IoT and e-commerce using BERTopic modeling. *Internet of Things (The Netherlands)*, 32. <https://doi.org/10.1016/j.iot.2025.101604>
- Wahyuni, W., Lestari, T. P., Apriliana, M., & Gumelta, R. (2025). Implementation of BERTopic for Topic Modeling Analysis of the Free Nutritious Meal Program Based on YouTube Comments. In *Journal of Applied Informatics and Computing (JAIC)* (Vol. 9, Number 4). <http://jurnal.polibatam.ac.id/index.php/JAIC>
- Yudhana, A., Fadlil, A., & Rosidin, M. (2019). Indonesian Words Error Detection System using Nazief Adriani Stemmer Algorithm. In *IJACSA) International*

Journal of Advanced Computer Science and Applications (Vol. 10, Number 12). www.ijacsa.thesai.org

Zare, D., Fernandez-Sanz, L., Pospelova, V., & López-Baldominos, I. (2025). NLP and Text Mining for Enriching IT Professional Skills Frameworks. *Applied Sciences (Switzerland)*, 15(17). <https://doi.org/10.3390/app15179634>

LAMPIRAN

Lampiran 1 Surat Permohonan Izin Penelitian & Pengambilan Data



FAKULTAS
MATEMATIKA &
ILMU PENGETAHUAN ALAM

Gedung Prof. Dr. H. Zanzawi Soejarto, M.Sc.
Kampus Terpadu Universitas Islam Indonesia
Jl. Kaliurang km 14,5 Yogyakarta 55584
T. (0274) 898444 ext. 3040, 3041
F. (0274) 896439
E. fmipa@uii.ac.id
W. fmipa.uui.ac.id

Nomor : 349/Dek/70-TA/Bag.TA/III/2026
Lamp : Satu Lembar
Hal : Permohonan Izin Penelitian & Pengambilan Data

Kepada Yth.
Bapak/Ibu Pimpinan
KEPALA KANTOR DPD RI DIY
JL. KUSUMANEGARA NO. 133, MUJA MUJU, UMBULHARJO, KOTA YOGYAKARTA,
DAERAH ISTIMEWA YOGYAKARTA

Assalamu'alaikum warahmatullahi wabarakaatuh

Bersama ini kami Pimpinan Fakultas Matematika dan Ilmu Pengetahuan Alam (MIPA) Universitas Islam Indonesia Yogyakarta bermaksud menyampaikan permohonan izin bagi mahasiswa kami di bawah ini :

Nama Mahasiswa	: NAIL ANFASA FADLA
NIM	: 22611050
Program Studi	: STATISTIKA
Maksud/Keperluan	: PERMOHONAN PERIZINAN PENELITIAN & PENGAMBILAN DATA
Judul Penelitian	: TOPIC MODELLING ASPIRASI MASYARAKAT MENGGUNAKAN METODE BERTOPIC (STUDI KASUS: ASPIRASI MASYARAKAT DI DEWAN PERWAKILAN DAERAH REPUBLIK INDONESIA)
Tempat Penelitian	: KANTOR DPD RI DIY
Pembimbing 1	: Dr. Edy Widodo, S.Si., M.Si.
Pembimbing 2	:

Selanjutnya mengenai pelaksanaan penelitian dan segala konsekuensi yang dipersyaratkan kami serahkan pada kebijaksanaan Bapak/Ibu Pimpinan.

Demikian permohonan ini kami sampaikan, atas perhatian dan kerjasamanya diucapkan terima kasih.

Wassalamu'alaikum warahmatullahi wabarakaatuh



Yogyakarta, 13 Maret 2026

Dekan,

Prof. Riyanto, S.Pd., M.Si., Ph.D.

Tembusan disampaikan kepada:

1. Kaprodi Statistika
2. Arsip

Lampiran 2 *Syntax* BERTopic

```
# =====
# INSTALASI DEPENDENCIES
# =====

!pip install bertopic
!pip install PySastrawi
!pip install umap-learn
!pip install hdbscan
!pip install sentence-transformers
!pip install keybert
!pip install plotly
!pip install -U scikit-learn
# =====
# 0. DESKRIPSI DATA PENELITIAN
# =====

import pandas as pd

# Load data
df_raw = pd.read_csv('/content/Data Asmas.csv')

# Nama kolom teks (sesuaikan jika berbeda)
KOLOM_TEKS = 'aspirasi'

# --- Gambaran Umum Dataset ---
print("=" * 55)
print("          DESKRIPSI DATA PENELITIAN")
print("=" * 55)

print(f"\n Sumber Data      : Data Aspirasi Masyarakat (Asmas) DPD
RI")
print(f" Wilayah           : Daerah Istimewa Yogyakarta")
print(f" Jumlah Kolom      : {df_raw.shape[1]}")
print(f" Jumlah Dokumen    : {len(df_raw)}")
print(f" Nilai Kosong      : {df_raw[KOLOM_TEKS].isna().sum()}
dokumen")
print(f" Dokumen Valid     : {df_raw[KOLOM_TEKS].notna().sum()}
dokumen")

# --- Statistik Panjang Teks ---
```

```

df_raw['panjang_kata'] = df_raw[KOLOM_TEKS].dropna().apply(lambda
x: len(str(x).split()))
df_raw['panjang_char'] = df_raw[KOLOM_TEKS].dropna().apply(lambda
x: len(str(x)))

print(f"\n{'-'*55}")
print("STATISTIK PANJANG DOKUMEN")
print(f"{'-'*55}")
print(f"  Rata-rata panjang : {df_raw['panjang_kata'].mean():.1f}
kata | {df_raw['panjang_char'].mean():.1f} karakter")
print(f"  Terpendek      : {df_raw['panjang_kata'].min()}
kata | {df_raw['panjang_char'].min()} karakter")
print(f"  Terpanjang     : {df_raw['panjang_kata'].max()}
kata | {df_raw['panjang_char'].max()} karakter")
print(f"  Median          :
{df_raw['panjang_kata'].median():.0f}
kata | {df_raw['panjang_char'].median():.0f} karakter")

# --- Tabel Contoh Data ---
print(f"\n{'-'*55}")
print("TABEL CONTOH DOKUMEN ASPIRASI (5 Pertama)")
print(f"{'-'*55}")

df_contoh = df_raw[[KOLOM_TEKS]].head(5).copy()
df_contoh.index = df_contoh.index + 1
df_contoh.index.name = 'No'
df_contoh['aspirasi_singkat'] = df_contoh[KOLOM_TEKS].apply(
    lambda x: str(x)[:120] + '...' if len(str(x)) > 120 else
str(x)
)
df_contoh['jumlah_kata'] = df_contoh[KOLOM_TEKS].apply(lambda x:
len(str(x).split()))

tabel_tampil = df_contoh[['aspirasi_singkat',
'jumlah_kata']].copy()
tabel_tampil.columns = ['Cuplikan Aspirasi', 'Jml Kata']

print(tabel_tampil.to_string())

```

```

# --- Contoh Dokumen Mentah Lengkap ---
print(f"\n{'-'*55}")
print("CONTOH DOKUMEN MENTAH (3 Sampel Acak)")
print(f"{'-'*55}")

sampel = df_raw[KOLOM_TEKS].dropna().sample(3, random_state=42)
for i, (idx, teks) in enumerate(sampel.items(), 1):
    print(f"\n [{i}] Dokumen #{idx + 1}")
    print(f" {teks}")

# --- Distribusi Panjang Dokumen (Histogram Teks) ---
import matplotlib.pyplot as plt

fig, axes = plt.subplots(1, 2, figsize=(13, 4))
fig.suptitle('Distribusi Panjang Dokumen Aspirasi', fontsize=13,
fontweight='bold')

axes[0].hist(df_raw['panjang_kata'].dropna(), bins=30,
color='steelblue', edgecolor='white')
axes[0].set_title('Distribusi Jumlah Kata')
axes[0].set_xlabel('Jumlah Kata per Dokumen')
axes[0].set_ylabel('Frekuensi')
axes[0].axvline(df_raw['panjang_kata'].mean(), color='red',
linestyle='--',
label=f"Rata-rata:
{df_raw['panjang_kata'].mean():.1f}")
axes[0].legend()

axes[1].hist(df_raw['panjang_char'].dropna(), bins=30,
color='seagreen', edgecolor='white')
axes[1].set_title('Distribusi Jumlah Karakter')
axes[1].set_xlabel('Jumlah Karakter per Dokumen')
axes[1].set_ylabel('Frekuensi')
axes[1].axvline(df_raw['panjang_char'].mean(), color='red',
linestyle='--',
label=f"Rata-rata:
{df_raw['panjang_char'].mean():.1f}")
axes[1].legend()

```

```

plt.tight_layout()
plt.savefig('/content/deskripsi_data.png', dpi=150)
plt.show()
print("\nGrafik disimpan ke: /content/deskripsi_data.png")
# =====
# 1. PRA-PEMROSESAN TEKS ASPIRASI
# =====

import pandas as pd
import re
import string

from Sastrawi.Stemmer.StemmerFactory import StemmerFactory
from Sastrawi.StopWordRemover.StopWordRemoverFactory import StopWordRemoverFactory

# --- Load Data ---
df = pd.read_csv('/content/Data Asmas.csv')
print(f"Jumlah data: {len(df)}")
print(df.head())

# Sesuaikan nama kolom dengan kolom teks aspirasi Anda
KOLOM_TEKS = 'aspirasi' # <-- ganti sesuai nama kolom di CSV Anda
df = df.dropna(subset=[KOLOM_TEKS])
df = df.reset_index(drop=True)

# --- Inisialisasi Tools Bahasa Indonesia ---
factory_stemmer = StemmerFactory()
factory_stopword = StopWordRemoverFactory()

stemmer = factory_stemmer.create_stemmer()
stopword_remover = factory_stopword.create_stop_word_remover()

# Tambahan stopwords domain aspirasi (sesuaikan jika perlu)
custom_stopwords = {
    # Sapaan & subjek
    'mohon', 'tolong', 'kepada', 'yth',
    'bapak', 'ibu', 'pak', 'bu',
    'saya', 'kami', 'kita',

    # Kata fungsi dasar

```

```

'yang', 'dan', 'di', 'ke', 'dari', 'untuk', 'dengan', 'pada',
'dalam', 'sebagai', 'oleh', 'agar', 'supaya',

# Konjungsi & transisi
'bahwa', 'serta', 'namun', 'hingga', 'sehingga',
'apabila', 'maupun', 'jika', 'maka', 'baik',

# Kata demonstratif & referensial
'ini', 'itu', 'tersebut', 'hal',

# Kata keterangan umum
'sudah', 'akan', 'juga', 'sangat', 'sekali',
'hanya', 'pun', 'pula', 'saat', 'setiap',
'antara', 'beberapa', 'salah', 'satunya',
'banyak', 'cukup', 'ada',

# Kata bantu (bermakna tipis)
'melalui', 'terkait', 'berbagai', 'seperti',

# Meta-aspirasi (muncul di hampir semua dokumen)
'masyarakat', 'pemerintah',
'berharap', 'harapan', 'diharapkan',
'aspirasi', 'menyampaikan', 'meminta',

# Institusional & geografis terlalu umum
'dpd', 'mpr', 'anggota', 'ri', 'daerah',
'yogyakarta', 'jogjakarta', 'diy', 'sleman', 'gunungkidul',
'kulon progo', 'bantul'

# Nama orang (noise)
'gus', 'hilmy', 'hery', 'triyono',

# Kuantitatif non-topik
'tahun', 'satu', 'dua',
}

def preprocess(teks):
    # Lowercase
    teks = teks.lower()

```

```

# Hapus URL
teks = re.sub(r'http\S+|www\S+', '', teks)
# Hapus angka
teks = re.sub(r'\d+', '', teks)
# Hapus tanda baca
teks = teks.translate(str.maketrans('', '',
string.punctuation))
# Hapus spasi ganda
teks = re.sub(r'\s+', ' ', teks).strip()
# Hapus stopword bawaan Sastrawi
teks = stopwords_remove.remove(teks)
# Hapus custom stopwords
teks = ' '.join([w for w in teks.split() if w not in
custom_stopwords])
# Stemming
teks = stemmer.stem(teks)
return teks

# Terapkan preprocessing
df['teks_bersih'] = df[KOLOM_TEKS].apply(preprocess)

# Filter teks yang terlalu pendek (kurang dari 3 kata)
df = df[df['teks_bersih'].apply(lambda x: len(x.split()) >= 3)]
df = df.reset_index(drop=True)

docss = df['teks_bersih'].tolist()

print(f"\nJumlah dokumen setelah preprocessing: {len(docss)}")
print("\nContoh hasil preprocessing:")
for i in range(3):
    print(f" Asli    : {df[KOLOM_TEKS][i]}")
    print(f" Bersih  : {docss[i]}")
    print()
# =====
# 2. REPRESENTASI SEMANTIK - MODEL EMBEDDING
# =====
from sentence_transformers import SentenceTransformer

MODEL_EMBEDDING = 'LazarusNLP/all-indo-e5-small-v4'

```

```

print(f"Memuat model embedding: {MODEL_EMBEDDING}")
embedding_model = SentenceTransformer(MODEL_EMBEDDING)

# Generate embeddings
print("Membuat embedding dokumen... (proses ini memanfaatkan GPU)")
embeddings = embedding_model.encode(
    docss,
    show_progress_bar=True,
    batch_size=64,          # turunkan ke 32 jika RAM habis
    device='cuda'          # otomatis pakai GPU Colab
)

print(f"\nBentuk embedding: {embeddings.shape}")
# Output: (jumlah_dokumen, dimensi_embedding)
# =====
# 4. PEMBENTUKAN KLASSTER - HDBSCAN
# =====
from hdbscan import HDBSCAN

hdbscan_model = HDBSCAN(
    min_cluster_size=10,      # minimum dokumen per topik;
    # sesuaikan dengan ukuran dataset
    min_samples=5,           # kontrol sensitivitas noise
    metric='euclidean',      # jarak setelah UMAP
    cluster_selection_method='eom', # 'eom' = lebih banyak
    # klaster kecil; 'leaf' = lebih sedikit
    prediction_data=True     # diperlukan untuk kalkulasi
    # probabilitas
)

print("Konfigurasi HDBSCAN:")
print(f"  min_cluster_size      : {hdbscan_model.min_cluster_size}")
print(f"  min_samples              : {hdbscan_model.min_samples}")
print(f"  cluster_selection_method : {hdbscan_model.cluster_selection_method}")
# =====

```

```

# 5. REPRESENTASI TOPIK - c-TF-IDF
# =====

from sklearn.feature_extraction.text import CountVectorizer
from bertopic.vectorizers import ClassTfidfTransformer

# CountVectorizer dengan konfigurasi Bahasa Indonesia
vectorizer_model = CountVectorizer(
    ngram_range=(1, 2),      # unigram + bigram untuk konteks lebih
    min_df=2,                # abaikan kata yang muncul < 2 dokumen
    max_df=0.85              # abaikan kata yang muncul di > 85%
    dokumen
)

# c-TF-IDF dengan BM25 weighting untuk representasi lebih baik
ctfidf_model = ClassTfidfTransformer(
    reduce_frequent_words=True,  # kurangi dominasi kata sangat
    umum
    bm25_weighting=True          # bobot BM25 lebih robust dari
    TF-IDF standar
)

print("Konfigurasi c-TF-IDF:")
print(f"  ngram_range           : {vectorizer_model.ngram_range}")
print(f"  reduce_frequent_words : True")
print(f"  bm25_weighting         : True")
# =====

# 6. REPRESENTASI LANJUTAN - KeyBERT + MMR
# =====

from bertopic.representation import KeyBERTInspired,
MaximalMarginalRelevance

# KeyBERT: memilih kata kunci berdasarkan kemiripan semantik
dengan topik
keybert_model = KeyBERTInspired(
    top_n_words=15,  # kandidat kata kunci yang dievaluasi
    nr_repr_docs=5   # jumlah dokumen representatif per topik
)

```

```

# MMR: menyeimbangkan relevansi dan keragaman kata kunci
mmr_model = MaximalMarginalRelevance(
    diversity=0.3      # 0=semua relevan (redundan), 1=semua
beragam; 0.3 adalah keseimbangan baik
)

# Gabungkan sebagai pipeline representasi
representation_model = {
    "KeyBERT": keybert_model,
    "MMR"      : mmr_model
}

print("Pipeline representasi:")
print("  [1] KeyBERTInspired → menangkap kata kunci semantik")
print("  [2] MaximalMarginalRelevance → mengurangi redundansi kata
kunci")

# =====
# 7. PEMODELAN TOPIK - FIT & INTERPRETASI
# =====

from bertopic import BERTopic

# Inisialisasi model BERTopic dengan semua komponen
topic_model = BERTopic(
    embedding_model      = embedding_model,
    umap_model           = umap_model,
    hdbscan_model        = hdbscan_model,
    vectorizer_model     = vectorizer_model,
    ctfidf_model         = ctfidf_model,
    representation_model = representation_model,
    top_n_words          = 10,
    language             = 'multilingual',
    calculate_probabilities = True,
    verbose              = True
)

# Fit model
print("Menjalankan BERTopic... (gunakan embedding yang sudah
dibuat)")
topics, probs = topic_model.fit_transform(docss, embeddings)

```

```

# ---- RINGKASAN HASIL ----
topic_info = topic_model.get_topic_info()
print(f"\n{'='*50}")
print(f"HASIL PEMODELAN TOPIK")
print(f"{'='*50}")
print(f"Total dokumen      : {len(docss)}")
print(f"Jumlah topik        : {len(topic_info) - 1}") # -1 karena
topik -1 adalah outlier
print(f"Dokumen outlier   : {sum(t == -1 for t in topics)} ({sum(t
== -1 for t in topics)/len(topics)*100:.1f}%)")

# Tampilkan info topik
print("\nDaftar Topik Teratas:")
print(topic_info[topic_info['Topic'] != -
1][['Topic', 'Count', 'Name']].head(20).to_string(index=False))

# ---- VISUALISASI ----
# Distribusi topik
fig_bar = topic_model.visualize_barchart(top_n_topics=15,
n_words=8)
fig_bar.show()

# Peta topik (2D)
fig_topics = topic_model.visualize_topics()
fig_topics.show()

# Heatmap kemiripan antar topik
fig_heatmap = topic_model.visualize_heatmap()
fig_heatmap.show()

# ---- SIMPAN HASIL KE DATAFRAME ----
df['topik']          = topics
df['probabilitas'] = [max(p) if p is not None else 0 for p in
probs]

# Tampilkan sampel per topik
print("\nContoh dokumen per topik:")
for topic_id in sorted(df['topik'].unique()):

```

```

        if topic_id == -1:
            continue

        sample = df[df['topik'] ==
topic_id][KOLOM_TEKS].head(2).tolist()
        words = [w for w, _ in topic_model.get_topic(topic_id)][[:5]
        print(f"\n Topik {topic_id} | Kata kunci: {'',
'.join(words)}")
        for s in sample:
            print(f"      → {s[:100]}...")

# Simpan ke CSV
df.to_csv('/content/hasil_bertopic.csv', index=False)
print("\nHasil disimpan ke: /content/hasil_bertopic.csv")
# =====
# 7b. TABEL c-TF-IDF NUMERIK PER TOPIK
# =====

import pandas as pd

print("=" * 65)
print("TABEL c-TF-IDF REPRESENTATIF PER TOPIK")
print("=" * 65)

# Kumpulkan semua topik valid (kecuali outlier -1)
topic_ids = sorted([t for t in topic_model.get_topics().keys() if
t != -1])

ctfidf_records = []

for topic_id in topic_ids:
    words_scores = topic_model.get_topic(topic_id) # list of
(word, score)
    topic_name = topic_model.get_topic_info().loc[
topic_model.get_topic_info()['Topic'] ==
topic_id, 'Name'
].values[0]
    count = topic_model.get_topic_info().loc[
topic_model.get_topic_info()['Topic'] ==
topic_id, 'Count'
].values[0]

```

```

        for rank, (word, score) in enumerate(words_scores[:10],
start=1):
            ctfidf_records.append({
                'Topik'      : topic_id,
                'Nama Topik' : topic_name,
                'Dokumen'    : count,
                'Peringkat'  : rank,
                'Kata Kunci' : word,
                'Skor c-TF-IDF': round(score, 4)
            })

df_ctfidf = pd.DataFrame(ctfidf_records)

# Tampilkan per topik
for tid in topic_ids:
    subset = df_ctfidf[df_ctfidf['Topik'] == tid][
        ['Peringkat', 'Kata Kunci', 'Skor c-TF-IDF']
    ].reset_index(drop=True)

    nama    = df_ctfidf[df_ctfidf['Topik'] == tid]['Nama
Topik'].values[0]
    jml     = df_ctfidf[df_ctfidf['Topik'] ==
tid]['Dokumen'].values[0]

    print(f"\n└─ Topik {tid} | {nama}")
    print(f"└─ Jumlah Dokumen: {jml}")
    print(subset.to_string(index=False))
    print("\n" + "-" * 50)

# ---- Ekspor ke CSV ----
df_ctfidf.to_csv('/content/ctfidf_per_topik.csv', index=False)
print("\n✓ Tabel c-TF-IDF disimpan ke:
/content/ctfidf_per_topik.csv")

# ---- Versi ringkas: top-3 kata per topik dalam satu tabel ----
print("\n" + "=" * 65)
print("RINGKASAN TOP-3 KATA PER TOPIK (c-TF-IDF)")
print("=" * 65)

```

```

summary_rows = []
for tid in topic_ids:
    top3 = df_ctfidf[df_ctfidf['Topik'] == tid].head(3)
    kata_skor = " | ".join(
        f"{r['Kata Kunci']} ({r['Skor c-TF-IDF']:.4f})"
        for _, r in top3.iterrows()
    )
    jml = df_ctfidf[df_ctfidf['Topik'] ==
tid]['Dokumen'].values[0]
    summary_rows.append({
        'Topik' : tid,
        'Dokumen' : jml,
        'Top-3 Kata (Skor c-TF-IDF)': kata_skor
    })

df_summary = pd.DataFrame(summary_rows)
print(df_summary.to_string(index=False))
# =====
# 8. EVALUASI MODEL
# =====

import numpy as np
from sklearn.metrics import silhouette_score, davies_bouldin_score
from sklearn.metrics.pairwise import cosine_similarity

# Ambil reduced embeddings dari model yang sudah fit
umap_embeddings = topic_model.umap_model.transform(embeddings)

topic_info = topic_model.get_topic_info()
n_topics = len(topic_info[topic_info['Topic'] != -1])
n_outlier = sum(t == -1 for t in topics)

# Filter outlier
mask = np.array(topics) != -1
umap_filtered = umap_embeddings[mask]
emb_original_filt = embeddings[mask] # tetap simpan
untuk referensi
label_filtered = np.array(topics)[mask]

```

```

# Silhouette di ruang UMAP (cosine)
silhouette_umap = silhouette_score(umap_filtered, label_filtered,
metric='euclidean')

davies_umap = davies_bouldin_score(umap_filtered, label_filtered)

print("=" * 55)
print("EVALUASI MODEL BERTOPIC – RUANG UMAP")
print("=" * 55)
print(f"  Silhouette (UMAP/euclidean)   : {silhouette_umap:.4f}")
print(f"  Davies-Bouldin (UMAP)            : {davies_umap:.4f}")
print(f"  Jumlah Topik                     : {n_topics}")
print(f"  Outlier                          : {n_outlier}
({n_outlier/len(topics)*100:.1f}%)")

```

Lampiran 3 Hasil BERTopic Untuk Tiap Aspirasi

Aspirasi	Topik	Probabilitas
Masyarakat meminta pendirian fasilitas pendidikan baru di wilayah yang lebih ...	9	0.495858
Saat ini lokasi masih menyewa sehingga membutuhkan bantuan apabila berkenan ...	0	0.160426
dalam rangka memperkuat nilai-nilai keistimewaan dan kebudayaan, maka Pemerintah Daerah ...	1	0.632414
Disampaikan oleh Bapak Triyono dari DIFabike bahwa pada Tahun 2014, ...	14	1
Dinas Koperasi dan UKM DIY menyediakan layanan pendampingan dan konsultasi ...	1	0.180809
Mendorong kebijakan yang memastikan akses pendidikan yang setara bagi perempuan, ...	0	0.436744
Mendorong lebih banyak perempuan untuk terlibat dalam politik dan pengambilan ...	0	1
FKPPI berharap dalam proses Pilkada tahun 2024, masyarakat Yogyakarta bisa ...	2	0.745671
FKPPI berharap adanya peningkatan sinergi antar komunitas dengan penyelenggara Pemilu ...	2	1
Perlunya edukasi kepada masyarakat yang lebih intensif terkait pentingnya Mewujudkan ...	2	0.056062
Bagaimana Pilkada 2024 bisa melahirkan pemimpin di Kabupaten Kota yang ...	2	0.563582
Sebagai daerah yang dikenal dengan sebutan "Kota Pelajar," masyarakat DIY ...	0	0.370539
Banyak masyarakat menyuarakan keprihatinan mereka terhadap politik uang dan kampanye ...	2	1
Seniman musik tradisional sering kali menghadapi keterbatasan dalam hal fasilitas ...	1	0.567193
Masyarakat berharap kebijakan pemerintah dapat mengintegrasikan seni musik tradisional ke ...	1	0.252879
Melibatkan PKK dalam kegiatan pelestarian budaya melalui museum, seperti pelatihan, ...	1	1
Agar dana keistimewaan DIY dapat disalurkan secara berkelanjutan dan merata ...	-1	0.051547
HIMPAUDI ingin mengajak masyarakat untuk lebih menyadari pentingnya pendidikan anak ...	0	0.393071
HIMPAUDI bercita-cita agar setiap anak usia dini, terutama di daerah ...	0	0.341086
Memohon dukungan untuk pengembangan kompetensi dan profesionalisme tenaga pendidik anak ...	0	0.27804
DPD RI diharapkan dapat membantu mengupayakan dukungan dari pemerintah pusat ...	7	1
Pemerintah diharapkan mampu mendorong penyelesaian Hambatan Regulasi. Membantu menemukan solusi ...	7	0.300278

Aspirasi	Topik	Probabilitas
Memastikan Perlindungan dan Keberlanjutan Usaha Lokal. Memastikan bahwa kepentingan usaha ...	-1	0.316139
Sebagai masyarakat, tentu kami berharap adanya transparansi dalam kontestasi PILKADA ...	2	0.774066
Kami berharap, di masa tenang nanti, bawaslu bisa secara netral ...	2	1
Agar proses kampanye dilakukan secara etis, tanpa adanya praktik politik ...	2	1
Dalam Pilkada, kami juga menginginkan pemilihan yang damai, adil, dan ...	2	1
Masyarakat menginginkan akses yang mudah ke informasi terkait pemilu, termasuk ...	2	0.317135
Masyarakat berharap agar penyelenggara pemilu seperti KPU (Komisi Pemilihan Umum) ...	2	0.17514
Pemerintah diharapkan tidak menggunakan fasilitas negara untuk kepentingan tertentu dalam ...	2	1
Mendorong penyiaran program atau acara diskusi yang memungkinkan publik mendapatkan ...	2	1
Mengenai aturan penyiaran kampanye dan iklan politik di media elektronik ...	2	1
Masyarakat berharap pemimpin baru lebih memperhatikan isu lingkungan, seperti pengelolaan ...	6	1
Program peningkatan ekonomi, terutama bagi UMKM lokal, menjadi perhatian masyarakat ...	-1	0.130829
Ada aspirasi kuat untuk menjaga kebudayaan lokal dan kearifan tradisional ...	7	1
Untuk meningkatkan daya saing dan kualitas pengusaha muda, HIPMI Yogyakarta ...	1	1
Pengusaha muda berharap agar pemerintah memberikan insentif bagi pengusaha ...	1	1
Pengusaha muda di Yogyakarta berharap agar proses perizinan usaha dapat ...	1	1
Salah satu tantangan utama bagi pengusaha muda adalah akses terhadap ...	1	1
Banyak petani berharap agar hasil pertanian mereka dapat meningkat, baik ...	8	1
Melibatkan lebih banyak petani perempuan dan anak muda dalam program ...	-1	0.090459
Masyarakat berharap agar Reforma Agraria dapat menyelesaikan masalah sengketa tanah ...	3	0.234508
Perbaikan infrastruktur pertanian, seperti jaringan irigasi, jalan desa, serta fasilitas ...	8	0.437574
Adanya penggunaan ruang yang ramah lingkungan, dengan penghijauan yang lebih ...	-1	0.151177
Masyarakat, khususnya pedagang UMKM, menginginkan adanya kemudahan dalam memperoleh izin ...	1	0.174718

Aspirasi	Topik	Probabilitas
Agar ada kerjasama yang lebih intens antara pemerintah, komunitas UMKM, ...	1	0.124354
Menyusun landasan hukum dan kebijakan yang kuat untuk penyelenggaraan pariwisata ...	7	0.353662
Meningkatkan koordinasi dan sinergi antara DPRD DIY, Anggota DPD, pemerintah ...	7	1
Membahas dan menetapkan indikator yang dapat dijadikan acuan untuk menilai ...	7	0.46788
Seniman berharap adanya dukungan berupa kebijakan perlindungan hak kekayaan intelektual, ...	1	1
Harapan untuk lebih banyak kolaborasi antara pemerintah, komunitas seni, dan ...	1	1
Seniman sketsa membutuhkan akses terhadap ruang-ruang berkarya seperti studio, galeri, ...	1	1
Warga menginginkan agar pengelolaan lahan untuk desa wisata tidak hanya ...	-1	0.201626
Masyarakat ingin memastikan bahwa pengembangan desa wisata tidak hanya meningkatkan ...	7	0.369095
Reforma agraria diharapkan dapat memberikan kemudahan akses tanah bagi masyarakat ...	3	0.210424
YKI berharap pasien kanker mendapatkan akses yang lebih baik terhadap ...	5	0.198414
Pemerintah membuat kebijakan yang lebih berpihak pada pasien kanker, misalnya ...	5	0.148631
Perbaikan dan pengembangan fasilitas di Pura Bhakti Widi Ngawen agar ...	-1	0.16954
emberdayaan Ekonomi Masyarakat Aspirasi : Program pelatihan keterampilan bagi masyarakat ...	1	0.08658
Masyarakat dan komunitas seni mengharapkan adanya alokasi anggaran yang lebih ...	1	0.5964
Diharapkan adanya program-program pembinaan, pelatihan, dan pendampingan bagi komunitas orkestra ...	1	0.369465
Kalau ada program beasiswa musik, itu akan jadi langkah besar. ...	1	0.287261
Fasilitas layanan kesehatan untuk anak dengan kanker, khususnya di daerah, ...	10	0.158881
Prosedur birokrasi yang panjang dan kompleks sering kali menghambat LKS ...	3	1
Banyak dari kami yang orang tuanya pensiunan, tidak mampu membiayai ...	0	0.385578
Agar pemerintah memperluas program keamanan berbasis masyarakat, seperti revitalisasi ronda ...	6	1
Banyak UMKM yang usahanya bagus, tapi tidak paham cara mengelola ...	1	1
Para lansia berharap adanya fasilitas layanan kesehatan khusus lansia yang ...	10	1

Aspirasi	Topik	Probabilitas
Mendorong pemerintah daerah untuk lebih aktif menyusun dan melaksanakan Perda ...	10	0.496958
Komunitas lansia berharap adanya fasilitas layanan kesehatan khusus lansia yang ...	10	1
Penguatan peran PKK dalam penanganan stunting dan gizi buruk, yang ...	2	0.166974
pengembangan wellness tourism melibatkan secara aktif masyarakat setempat, khususnya pelaku ...	12	0.583056
Akademisi UNU Yogyakarta mengusulkan pengembangan Pesantren AMAN sebagai model pesantren ...	0	0.135197
DPD RI DIY dapat menjadi jembatan antara aspirasi perguruan ...	0	0.215356
Lansia menyampaikan kebutuhan mendesak atas peningkatan akses terhadap layanan kesehatan ...	10	1
Banyak lansia mengungkapkan rasa kesepian, minimnya perhatian keluarga, dan ketidakpastian ...	10	1
Diharapkan pemerintah daerah memiliki layanan bagi siswa yang berhadapan dengan ...	0	0.403618
Saya sangat setuju dengan Tripusat pendidikan. Parenting sangat dibutuhkan karena ...	0	1
BK3S mendorong agar pemerintah daerah dan pusat memberikan dukungan yang ...	3	0.166983
BK3S mengusulkan agar pemerintah melibatkan organisasi sosial secara lebih aktif ...	-1	0.070407
DPD RI perlu mendorong percepatan pembangunan infrastruktur air bersih berbasis ...	6	1
Perlunya kejelasan regulasi dan dukungan anggaran rutin dari pemerintah bagi ...	14	1
Pengurus yayasan menyampaikan kebutuhan akan pelatihan berkelanjutan dan sertifikasi tenaga ...	14	1
Kurangnya sinkronisasi antara Dinas Sosial, Dinas Pendidikan, Dinas Kesehatan, dan ...	0	1
Membuat program literasi digital berbasis gender yang menysasar kelompok perempuan, ...	1	0.115178
Dibutuhkan mekanisme pengaduan dan pemulihan korban KBGS yang ramah perempuan ...	0	0.315735
Perempuan masih kurang terlibat dalam proses pengambilan keputusan di bidang ...	-1	0.210888
Ingin ada 1 tempat umat budha yang layak dan representatif. ...	4	0.176796
Dengan dicanangkannya candi prambanan sebagai tempat ibadah nusantara, kita sangat ...	1	0.484563
Pengurus dan anggota Nasyiatul Aisyiah mendorong partai politik untuk lebih ...	0	1

Aspirasi	Topik	Probabilitas
Pemerintah daerah diharapkan bisa meningkatkan akses perempuan terhadap pelatihan keterampilan ...	0	1
Banyak perempuan di DIY yang terlibat dalam usaha mikro, kecil, ...	0	1
Diharapkan kepada Gusti Kanjeng Ratu Hemas (B-53) Anggota DPD RI ...	-1	0.111853
Ketiadaan atau minimnya pelayanan rehabilitasi berbasis komunitas di tingkat desa/kecamatan ...	0	0.417079
Memperjuangkan pembentukan Pusat Layanan Rehabilitasi Ringan berbasis desa/kelurahan dengan dukungan ...	3	0.161278
Pendidikan anti-narkoba di sekolah dan perguruan tinggi masih bersifat sporadis, ...	0	0.439846
Mendorong pilot project "Sekolah Bersinar" di DIY sebagai bagian dari ...	0	0.220563
Masyarakat mengeluhkan masih adanya tumpang tindih alur birokrasi antara layanan ...	3	1
Masyarakat menyampaikan harapan agar layanan-layanan publik yang tersedia di Mal ...	3	1
Pihak kalurahan menyampaikan perlunya peningkatan kapasitas SDM dan kelembagaan pamong ...	3	0.33463
kurangnya jumlah tenaga medis yang bertugas di Puskesmas, terutama bila ...	5	0.139976
bahwa banyak tenaga honorer yang selama ini membantu operasional pelayanan ...	11	0.218242
tumpang tindihnya sistem aplikasi pelaporan dan input data pasien yang ...	5	0.443684
minimnya anggaran operasional non-medis, khususnya untuk kegiatan promotif dan preventif ...	3	0.130586
menurunnya jumlah tenaga pendidik di sekolah-sekolah swasta akibat migrasi besar-besaran ...	-1	0.278824
perlunya peningkatan kapasitas sistem digital dan integrasi layanan daring lintas ...	3	1
MPP Sleman mengusulkan dukungan untuk penguatan mobile service atau pelayanan ...	3	1
Banyak program top down dari pemerintah sehingga dari masyarakat harus ...	6	0.532804
Masyarakat membutuhkan embung untuk menampung air untuk mengairi lahan pertanian ...	8	0.32572
permintaan pembangunan atau revitalisasi saluran irigasi, embung desa, dan bantuan ...	6	0.562347
Masyarakat Desa Gerbosari menyampaikan kebutuhan mendesak untuk pelebaran jalan utama ...	6	0.251838
Perempuan Desa Gerbosari berharap adanya program pemberdayaan yang terintegrasi dengan ...	-1	0.125025

Aspirasi	Topik	Probabilitas
Pentingnya peran pemuda dalam merancang masa depan desa perbatasan melalui ...	3	0.355368
Pemerintah Pusat dan Daerah mengalokasikan secara khusus anggaran pengadaan komputer ...	9	1
percepatan rekrutmen guru dan tenaga kependidikan, insentif/bonus khusus untuk penempatan ...	0	0.393411
masyarakat berharap adanya intervensi dari pemerintah. Baik pemerintah daerah maupun ...	8	1
Masyarakat menginginkan lebih banyak aktivitas pelatihan seperti pelabelan halal, pengemasan ...	-1	0.12402
Pemuda Karang Taruna DIY meminta adanya fasilitasi pelatihan digital secara ...	1	1
Karang Taruna DIY menginginkan adanya program dukungan konkret terhadap inisiatif ...	1	1
Pemuda Karang Taruna DIY mengusulkan agar organisasi Karang Taruna diakui ...	1	0.379284
pemerintah didorong untuk menyediakan tenaga konselor atau psikolog anak di ...	10	0.24965
kebutuhan akan dukungan intervensi dari pemerintah pusat dalam pengadaan sarana ...	3	0.215762
Kasus gangguan mental yang terjadi pada anak dan remaja harus ...	0	1
Tindak kriminalitas yang berupa Klitih menjadi salah satu faktor utama ...	0	0.255431
Momen Sumpah Pemuda dapat dijadikan ajang bagi pemuda khususnya di ...	3	1
Harapan mahasiswa Katolik UGM, Akademisi, dan Tokoh Agama agar langkah ...	2	0.045962
Menitipkan agar kerukunan umat beragama tetap dijaga dan hindari polarisasi ...	2	0.109129
Sebuah pesan bahwa Indonesia merupakan rumah yang nyaman bagi seluruh ...	4	1
Dalam rangka menyambut Pilkada serentak pada tanggal 27 November 2024, ...	2	1
Di era berkembangnya digitalisasi yang semakin masif, penyebaran hoaks (berita ...	2	0.046087
Proses Revitalisasi Pasar Sentul realitasnya memberikan dampak negatif terhadap perolehan ...	-1	0.148142
Adanya penghasilan yang tidak sebanding dengan harga sewa tempat menyebabkan ...	8	0.293461
Advokasi anak-anak difabel untuk bisa bersekolah di sekolah inklusi. Perlunya ...	0	0.368981
Adanya harapan untuk tetap menyelenggarakan kelas gratis guna menstimulasi anak-anak ...	0	0.443912
Kementerian Pendidikan Dasar dan Menengah harus sesegera mungkin melakukan pengawasan ...	0	0.466625

Aspirasi	Topik	Probabilitas
Peningkatan softskill menjadi catatan bahwa akses terhadap bidang ekonomi kreatif ...	13	1
Untuk mengurangi persentase tingkat keterlibatan generasi muda di kabupaten Bantul ...	13	1
Relevansi kebudayaan untuk menjangkau generasi muda saat ini dilakukan dengan ...	1	1
terkait umat Katolik yang sudah lanjut usia namun terdapat keterbatasan ...	-1	0.182237
Sebagai seorang Anggota DPD DIY, ada harapan untuk tetap memiliki ...	0	0.382099
Partisipasi masyarakat dalam pengelolaan pemerintah desa masih kurang, terutama menyangkut ...	3	1
Hubungan antara lembaga pemerintah pusat, pemerintah daerah, dan pemerintah desa ...	1	0.12851
Para pemuda desa mendorong adanya program pemberdayaan khusus bagi anak ...	-1	0.29834
Pemerintah diharapkan memberi kelonggaran bagi mahasiswa yang bersikap kritis terkait ...	2	0.08493
Mencabut atau merevisi UU ITE yang sering digunakan oleh pemerintah ...	2	0.102934
Anggota legislatif, termasuk DPD RI harus turut serta mengkritik kebijakan ...	2	0.327135
Melakukan edukasi politik yang massif di semua lini yang dilakukan ...	2	0.084911
Masyarakat desa Bendo Gedhe II menyampaikan aspirasi untuk tetap melanjutkan ...	6	0.254287
potensi sektor pertanian harus dimaksimalkan oleh Kementerian Pertanian, Kementerian Perindustrian, ...	8	1
Adanya kepentingan dari kaum perempuan untuk bisa berdaya secara mandiri ...	0	1
kabupaten Bantul masih rentan terhadap kekerasan yang dialami oleh perempuan ...	-1	0.148936
untuk membantu meningkatkan UMKM di daerah pedesaan. UMKM ini nantinya ...	1	0.091049
Adanya isu gangguan kesehatan mental hingga berujung bunuh diri terjadi ...	1	0.08677
Regulasi pengelolaan dana hibah untuk KPU dan Bawaslu yang sering ...	3	0.110904
Anggaran disusun dengan gambaran situasi normal tanpa adanya Pemungutan Suara ...	2	0.16438
Masih terjadi perbedaan dalam penentuan besaran uang transportasi yang disesuaikan ...	1	0.09168
Adanya gap antara tingkat literasi dan inklusi keuangan ini menjadi ...	13	1
OJK telah melakukan upaya preventif dan upaya kuratif. Upaya preventif ...	1	0.226544

Aspirasi	Topik	Probabilitas
Platform Pinjaman Online Legal yang didapati menyalahgunakan data pribadi masyarakat ...	-1	0.18878
Dalam hal perusahaan asuransi yang masih banyak tidak mematuhi regulasi, ...	1	0.267475
Pelatihan pembuatan lilin aroma terapi memiliki beberapa manfaat seperti dapat ...	1	0.190267
Mengajak generasi muda untuk memanfaatkan minyak jelantah untuk digunakan sebagai ...	-1	0.342431
Anak-anak di Panti Asuhan Bina Siwi merasa senang telah dikunjungi ...	0	0.494843
Anak-anak yang memiliki kebutuhan khusus bukan hanya dilihat dari sisi ...	0	1
Masyarakat yang sudah berada pada lanjut usia memerlukan asupan gizi ...	10	1
Masyarakat yang sudah berada pada lanjut usia juga dianjurkan untuk ...	10	1
Perlunya peran posyandu balita yang ada di setiap Kemantren maupun ...	0	1
Aspirasi terkait Makan Bergizi Gratis. Pasar Imogiri akan menjadi salah ...	1	0.101987
Aspirasi terkait peminjaman ilegal. Adanya oknum yang menawarkan pinjaman ilegal ...	1	0.178675
Tempat pembuangan sampah atau TPA. TPA menjadi permasalahan di Kabupaten ...	-1	0.24755
Terlebih saat ini sudah masuk musim penghujan sehingga irigasi ...	8	1
Infrastruktur akses pariwisata ke Gunungkidul masih sangat minim, mulai dari ...	7	1
Terdapat kebocoran anggaran di sektor pariwisata karena sistem tiket masih ...	7	1
Rata-rata lama sekolah (RLS) penduduk Kabupaten Gunungkidul pada tahun 2023 ...	-1	0.067514
Embung diharapkan meningkatkan debit air, dapat menunjang sektor perkebunan dan ...	8	0.21737
Pariwisata Gunungkidul menjadi sektor yang menjanjikan untuk meningkatkan pemasukan asli ...	7	1
Saat kekeringan terjadi para petani yang menggunakan teknik mencangkul manual ...	8	1
Droping air perlu disertai dengan mencatat aspirasi masyarakat agar tercipta ...	-1	0.286648
Mendorong bantuan traktor bagi para petani yang masih menggunakan teknik ...	8	1
Mendorong pemerintah Kabupaten Gunungkidul memikirkan solusi kekeringan untuk jangka panjang. ...	6	0.40429
Kekeringan merupakan masalah klasik Kabupaten Gunungkidul sejak lama. Solusi jangka	6	0.259273

Aspirasi	Topik	Probabilitas
Aspirasi yang disampaikan diantaranya meningkatkan UMKM di level padukuhan dan ...	1	0.082778
Perlunya edukasi serta sosialisasi terkait maraknya pinjol dan judi online ...	13	1
Perlu susunan program kerja prioritas dengan menyesuaikan kapasitas anggaran yang ...	-1	0.128603
Membangun kolaborasi dengan stakeholder di DIY untuk lebih memperhatikan kelompok ...	14	1
Perlu adanya dorongan dari Pemerintah Daerah terkait pemerataan sekolah berbasis ...	0	1
Mendorong lembaga pendidikan untuk selalu memberikan kesempatan kepada anak-anak kelompok ...	0	1
Sumber pendapatan dari sisi APBN di daerah D.I. Yogyakarta berasal ...	3	0.139844
Target penerimaan PNBP pada tahun 2024 mencapai Rp2,14 triliun mengalami ...	3	0.138501
Terdapat perubahan data yang signifikan atas paparan yang disampaikan oleh ...	3	0.141706
Mendorong kepada Badan Pengelola Keuangan dan Aset DIY untuk meningkatkan ...	3	0.110557
Mendorong kepada Kanwil Dirjen Perbendaharaan DIY untuk selalu memantau kinerja ...	3	0.138412
Kekerasan seksual berbasis gender di Kota Yogyakarta masih tergolong cukup ...	0	1
Jumlah mental disorder atau gangguan jiwa saat ini semakin meningkat. ...	-1	0.087899
Mendorong implementasi UU Tindak Pidana Kekerasan Seksual lebih dioptimalkan untuk ...	0	0.46772
Meminta untuk seluruh aparat penegak hukum di DIY agar turut ...	0	0.419876
Peningkatan sosialisasi terkait pinjol dan judol. Polda DIY selama pelaksanaan ...	-1	0.292408
Berdasarkan hasil pemantauan yang dilakukan oleh Pemerintah Kota Yogyakarta tahun ...	6	1
Terkait pengelolaan sampah di Kota Yogyakarta. Pengelolaan sampah pada setiap ...	-1	0.158318
Mendorong optimalisasi Dinas Lingkungan Hidup untuk memperhatikan kondisi air yang ...	6	1
Pertama, terkait dengan dana desa, seluruh Kalurahan (khususnya di Kabupaten ...	3	1
Kedua, terkait dengan status kepegawaian dari Lurah atau Pamong, Kepala ...	-1	0.305742
Ketiga, terkait dengan mutasi dan rotasi jabatan. Mutasi dan rotasi ...	3	0.245637
Keempat, terkait dengan alokasi dana perhal penghasilan tetap dan kesejahteraan ...	-1	0.139393

Aspirasi	Topik	Probabilitas
Kelima, peningkatan SDM para perangkat desa perlu dilakukan dengan kerja ...	-1	0.05701
Mendorong ide bahwa negara hanya perlu memberi kekuatan pada provinsi ...	3	1
Dalam rangka untuk meningkatkan akses layanan keuangan bagi masyarakat untuk ...	-1	0.108533
Dinas Koperasi dan UKM DIY memiliki Klinik Koperasi yang digunakan ...	1	0.178398
Mendorong Himbara maupun Perbankan swasta lainnya untuk memberikan sosialisasi terkait ...	1	0.106552
Dalam kondisi struktur sosial masyarakat peran perempuan harus mendapatkan porsi ...	0	1
Di Kabupaten Sleman sendiri mulai muncul banyaknya inisiasi terhadap Kelompok ...	-1	0.331193
Kondisi di wilayah Sleman Barat pada hakikatnya adalah lumbung pangan ...	3	0.717829
Secara perhitungan luas wilayah lahan pertanian yang menganggur sekitar 250 ...	8	1
Salah satu yang menjadi sorotan adalah bagaimana tingkat komoditas pertanian ...	8	1
Mendorong Pemerintah Daerah untuk menjaga agar tidak lagi adanya alih ...	8	1
Tanah longsor menjadi salah satu ancaman untuk pemukiman di daerah ...	6	1
Dinas Pekerjaan Umum Perumahan dan Kawasan Permukiman Kota Yogyakarta harus ...	6	1
Mendorong Pemerintah Kota untuk memetakan kondisi wilayah yang rawan terkena ...	6	1
Pertama, keterlibatan anak muda dalam pemilu menjadi salah satu isu ...	2	0.309885
Pengaruh media sosial juga memiliki dampak soal preferensi memilih calon ...	2	1
Catatan selanjutnya adalah tentang netralitas penyelenggara pemilu dan ancaman terhadap ...	2	0.170498
Mendorong penyelenggara pemilu dan aparat untuk bersikap netral dan tidak ...	2	0.123217
menyediakan program pelatihan berkelanjutan untuk meningkatkan keterampilan. Hal ini tentu ...	-1	0.193526
Mendorong Dinas Koperasi dan UKM DIY dan Dinas terkait di ...	1	0.190034
Kasus Tindak Pidana Perdagangan Orang atau TPPO terbongkar di Kulon ...	0	0.337677
Adanya aspirasi yang disampaikan bahwa tingkat kemiskinan di Kulon Progo ...	1	0.08811
Kemiskinan yang ada di Kabupaten Kulon Progo harus mampu diatasi ...	1	0.088583

Aspirasi	Topik	Probabilitas
Melakukan pendampingan dan edukasi dengan menggandeng aktivis perempuan untuk mengkampanyekan ...	0	1
Mendorong agar pemerintah Kabupaten Kulon Progo memperbaiki akses pendidikan terhadap ...	1	0.090007
Masyarakat mengeluhkan menjelang natal dan tahun baru yang lalu beberapa ...	8	0.400021
Adanya libur natal dan tahun baru ini menyebabkan munculnya keluhan ...	-1	0.163449
Masyarakat menyoroti kembali munculnya aksi kejahatan jalanan khususnya klitih yang ...	0	0.289872
Dalam hal harga bahan pokok yang meningkat, maka mendorong Dinas ...	8	1
Dalam hal volume sampah yang meningkat, maka mendorong Dinas Lingkungan ...	6	0.302525
Dalam hal munculnya kembali aksi jalanan khususnya Klitih, maka mendorong ...	0	0.250127
Data yang dipaparkan dari Satu Data Sleman menunjukkan jumlah pelaku ...	-1	0.096439
Masyarakat yang memiliki usaha menginginkan untuk mengadakan pembuatan sertifikat halal ...	4	0.055354
Dalam hal aspirasi yang disampaikan terkait masih kurangnya program sertifikat ...	4	0.070486
Mendorong adanya kerjasama antara Badan Penyelenggara Jaminan Produk Halal dengan ...	4	0.078164
Adanya aspirasi yang disampaikan masyarakat agar kondisi lansia untuk lebih ...	10	1
Pemberian perhatian untuk masyarakat Lansia penting untuk dilakukan, terlebih dengan ...	10	1
Mendorong untuk Dinas Sosial agar memonitor kondisi lansia yang ada ...	10	1
Pentingnya Peningkatan Sarana dan Prasarana untuk Operasi Penyelamatan SAR DIY ...	3	0.169925
SAR DIY juga meminta perhatian terhadap regulasi dan dukungan anggaran ...	-1	0.113924
kami mengusulkan agar SAR DIY dilibatkan dalam program pengembangan destinasi ...	12	1
kolaborasi antara SAR DIY dan industri pariwisata dapat membuka peluang ...	12	1
Pembentukan tim tanggap darurat wisata yang terdiri dari relawan SAR ...	12	1
Penyediaan platform digital untuk mempromosikan destinasi wisata aman, seperti aplikasi ...	12	1
Penyusunan regulasi yang mewajibkan setiap destinasi wisata untuk memiliki standar ...	12	1

Aspirasi	Topik	Probabilitas
Pengembangan destinasi wisata unggulan yang mengintegrasikan keamanan dan keselamatan melalui ...	12	1
Masyarakat Sleman menyampaikan aspirasi pentingnya pengadaan Tempat Pemakaman Umum (TPU) ...	3	1
mengusulkan agar pemerintah tidak hanya fokus pada penyediaan lahan baru, ...	6	1
Kami berharap agar sekolah-sekolah Muhammadiyah di Yogyakarta dapat terus meningkatkan ...	0	0.537272
kami ingin agar pendidikan Muhammadiyah dapat lebih merata dan terjangkau ...	0	1
sekolah-sekolah Muhammadiyah diharapkan dapat menjadi pelopor dalam menyediakan pendidikan inklusif ...	0	1
Kami juga menginginkan adanya peningkatan kerja sama antara sekolah sekolah ...	0	1
Warga berharap agar Masjid Faridan M Noto dapat menjadi pusat ...	4	0.49533
Salah satu aspirasi utama warga adalah pelaksanaan penyuluhan keagamaan secara ...	4	0.545713
Warga juga mengusulkan pelaksanaan pelatihan keterampilan di masjid sebagai bagian ...	4	1
Khususnya dalam bidang ekonomi kreatif, warga berharap masjid dapat menjadi ...	4	0.181096
Masjid Faridan M Noto diharapkan menjadi wadah untuk meningkatkan solidaritas ...	4	0.443696
Sepak bola di DIY terus menunjukkan perkembangan positif, baik dari ...	-1	0.12032
Menurut data Kementerian Pemuda dan Olahraga (Kemenpora), Indonesia membutuhkan lebih ...	-1	0.133814
Infrastruktur sepak bola di DIY masih membutuhkan peningkatan agar memenuhi ...	-1	0.154605
Warga berharap agar pemerintah daerah, investor, dan supporter dapat bersinergi ...	-1	0.186089
Dengan meningkatkan kualitas pemain dan infrastruktur sepak bola di DIY, ...	-1	0.129896
Dewan Pendidikan DIY menekankan pentingnya peningkatan kualitas pendidikan yang merata ...	-1	0.095331
pemerintah perlu memperhatikan distribusi fasilitas sekolah, seperti laboratorium, perpustakaan, dan ...	0	1
Dewan Pendidikan mendorong peningkatan fasilitas sekolah yang ramah disabilitas, seperti ...	0	1
Dewan Pendidikan DIY menekankan perlunya peningkatan tunjangan guru, pelatihan profesionalisme, ...	0	1
Untuk mengatasi kesenjangan pendidikan bagi siswa dari keluarga kurang mampu, ...	0	0.385559

Aspirasi	Topik	Probabilitas
Dewan Pendidikan DIY juga menyoroti pentingnya pembelajaran berbasis teknologi dan ...	0	0.443837
Kolaborasi antara pemerintah, sekolah, dan masyarakat menjadi kunci untuk menciptakan ...	0	1
Warga Kadipaten, Kota Yogyakarta, berharap terciptanya kehidupan bermasyarakat yang harmonis ...	4	0.249307
Masyarakat Kadipaten juga menginginkan peningkatan peran tokoh agama dan lembaga ...	4	0.102706
warga Kadipaten juga menyampaikan aspirasi pentingnya perlindungan anak dari berbagai ...	0	1
warga meminta agar pemerintah dan masyarakat bekerja sama untuk: Meningkatkan ...	0	1
warga berharap agar program-program seperti Posyandu, PAUD inklusif, dan ruang ...	4	0.299426
nilai gotong royong yang sudah tertanam kuat di masyarakat dapat ...	0	0.386777
Masyarakat berharap adanya langkah konkret dalam mitigasi banjir di lingkungan ...	6	1
Mengusulkan penerapan program pengelolaan sampah berbasis komunitas, seperti bank sampah ...	-1	0.250716
Agar pemerintah dapat memperbaiki infrastruktur pendukung pariwisata, seperti jalan akses, ...	7	0.463669
Pemerintah dapat memberikan pelatihan dan dukungan kepada pelaku UMKM di ...	12	0.750773
Mitigasi banjir, seperti perbaikan drainase dan pengelolaan sampah, dapat diintegrasikan ...	6	0.222804
Untuk mewujudkan lingkungan yang aman dari banjir dan destinasi wisata ...	6	0.25162
Dalam pelaksanaan tugas di lapangan, kami sering menghadapi tantangan seperti ...	3	0.301578
Meningkatkan edukasi kepada orang tua dan masyarakat tentang hak hak ...	0	1
Penyediaan kendaraan ambulan yang dilengkapi dengan fasilitas ramah disabilitas, seperti ...	-1	0.183492
Pelatihan bagi relawan dan petugas layanan darurat untuk menangani kebutuhan ...	14	1
Peningkatan aksesibilitas di fasilitas umum, termasuk rumah sakit, puskesmas, dan ...	-1	0.162267
Untuk menciptakan masyarakat yang tangguh, inklusif, dan peduli, Korps AmbulanMu ...	14	1
Program-program kolaboratif seperti kampanye anti-kekerasan terhadap anak, pelatihan relawan untuk ...	14	1
Kami, jamaah Masjid Gedhe Yogyakarta, menyampaikan aspirasi kepada Anggota DPD ...	4	1
kami berharap adanya perhatian lebih dalam bentuk dukungan program, fasilitas, ...	4	1

Aspirasi	Topik	Probabilitas
kami berharap adanya sinergi antara pemerintah dan lembaga keagamaan dalam ...	4	1
Untuk menciptakan dampak yang lebih luas, kami mengusulkan integrasi antara ...	4	0.322625
Kami, para pegawai perusahaan konstruksi di Kota Yogyakarta, menyampaikan aspirasi ...	-1	0.061089
kami mendesak pemerintah untuk memastikan bahwa setiap pekerja konstruksi dilindungi ...	-1	0.056946
Kami juga mendorong adanya peningkatan pengawasan terhadap perusahaan agar mematuhi ...	-1	0.182223
Untuk menciptakan dampak yang lebih luas, kami mengusulkan sinergi antara ...	-1	0.056341
Dengan dukungan dan regulasi yang tepat dari pemerintah melalui DPD ...	-1	0.061006
Pimpinan Wilayah ~Aisyiyah DIY menyampaikan aspirasi kepada Anggota DPD RI ...	0	0.346336
~Aisyiyah menekankan pentingnya perlindungan anak dari berbagai bentuk kekerasan, penganiayaan, ...	0	0.382267
Peningkatan edukasi kepada orang tua dan masyarakat tentang hak hak ...	0	1
Penguatan lembaga perlindungan anak di tingkat desa/kelurahan untuk memberikan layanan ...	0	1
Penerapan sanksi hukum yang lebih tegas bagi pelaku kekerasan terhadap ...	-1	0.098022
~Aisyiyah siap menjadi mitra strategis pemerintah dalam menyelenggarakan program-program pendidikan ...	0	0.391394
kami berharap DPD RI dapat mendorong kebijakan yang berpihak pada ...	0	0.2367
~Aisyiyah sebagai gerakan perempuan Islam yang konsisten dalam isu-isu keperempuanan ...	0	0.324476
Penerbitan regulasi daerah yang mendukung perlindungan anak, seperti larangan kekerasan ...	0	0.645512
Penyediaan anggaran khusus untuk program pemberdayaan keluarga dan perlindungan anak ...	0	0.233518
Penguatan sinergi antara pemerintah, organisasi masyarakat, dan lembaga pendidikan dalam ...	0	0.303351
Kami, masyarakat Kota Yogyakarta, menyampaikan aspirasi kepada Anggota DPD RI ...	4	1
Kami berharap DPD RI dapat mendorong pemerintah daerah untuk menyediakan ...	-1	0.098967
Pengembangan Ekonomi Kreatif untuk Mendukung Pedagang Pasca Relokasi Selain penyediaan ...	1	0.184445
Para pedagang diberikan pelatihan keterampilan dalam bidang ekonomi kreatif, seperti ...	1	0.1127

Aspirasi	Topik	Probabilitas
Lokasi pengganti didesain sebagai pusat ekonomi kreatif yang menarik wisatawan, ...	1	1
Pemerintah menyediakan modal usaha dan program mentoring bagi pedagang untuk ...	1	0.286169
Kami mengusulkan agar masjid dan lokasi pengganti pedagang dapat dirancang ...	4	1
Untuk memastikan keberlanjutan usaha pedagang pasca relokasi, kami berharap pemerintah ...	1	0.55117
Dengan mengintegrasikan penataan kawasan masjid dan pengembangan ekonomi kreatif, kami ...	4	1
GIPI DIY menyampaikan aspirasi agar upaya peningkatan pariwisata di Daerah ...	7	1
pengembangan destinasi seperti desa wisata, situs budaya, dan atraksi seni ...	1	0.343393
GIPI DIY juga menekankan pentingnya pendekatan ramah lingkungan dalam pengelolaan ...	-1	0.21216
Untuk meningkatkan kenyamanan dan kualitas pengalaman wisatawan, GIPI DIY mendorong ...	7	1
Pembangunan jalan akses menuju destinasi wisata yang saat ini masih ...	7	0.194485
Penyediaan fasilitas umum yang representatif, seperti toilet bersih, area parkir ...	-1	0.143396
Digitalisasi layanan wisata melalui platform online untuk memesan tiket, reservasi ...	12	1
GIPI DIY mengusulkan program pelatihan rutin bagi pemandu wisata, pengelola ...	-1	0.149754
GIPI DIY juga berharap adanya kolaborasi yang lebih erat antara ...	1	1
Penyelenggaraan event budaya yang lebih sering dan bervariasi, termasuk festival ...	1	1
Promosi event tersebut secara intensif melalui media sosial, platform digital, ...	12	1
Kolaborasi dengan komunitas lokal untuk memastikan bahwa event budaya tersebut ...	1	1
GIPI DIY meminta dukungan Anggota DPD RI DIY untuk memperjuangkan ...	7	1
Program promosi terpadu melalui kampanye "Visit Jogja" di pasar domestik ...	12	1
Pemberian insentif pajak bagi pelaku usaha di sektor pariwisata, seperti ...	7	0.463302
Kerja sama dengan maskapai penerbangan untuk menyediakan paket wisata murah ...	12	1
Pengembangan pariwisata yang berkelanjutan tidak hanya meningkatkan jumlah kunjungan wisatawan, ...	7	1
Sinergi antara warga Muhammadiyah Kota Jogja dengan Pemerintah Kota Jogja ...	4	0.606406

Aspirasi	Topik	Probabilitas
Penyelenggaraan workshop rutin tentang manajemen usaha, pemasaran digital, dan branding ...	1	0.27273
Kolaborasi dengan perguruan tinggi dan lembaga pelatihan untuk memberikan pendampingan ...	1	0.584012
Penyediaan platform digital untuk mempromosikan produk-produk kreatif lokal kepada audiens ...	1	0.276815
Pengembangan ekonomi kreatif melalui sinergi antara warga Muhammadiyah dan pemerintah ...	4	0.808199
Penyelenggaraan festival ekonomi kreatif yang melibatkan seniman, pelaku UMKM, dan ...	1	1
Penyediaan insentif pajak dan bantuan modal bagi pelaku UMKM kreatif ...	1	0.55379
Pembentukan pusat pelatihan dan inkubator bisnis untuk mendukung inovasi dan ...	1	0.503357
sinergi antara warga Muhammadiyah dan pemerintah Kota Jogja dalam pengembangan ...	4	0.769536
Kami berharap DPD RI dapat memfasilitasi akses pembinaan, sarana latihan, ...	1	0.20457
Pembentukan pusat pelatihan Tapak Suci sebagai destinasi wisata edukasi, di ...	1	0.210911
Pelatihan bagi anggota Tapak Suci untuk menjadi duta budaya yang ...	1	0.458427
Penyusunan regulasi yang mendukung integrasi budaya tradisional dalam pengembangan pariwisata, ...	-1	0.287475
masyarakat seni berharap agar kegiatan ini mendapat dukungan berkelanjutan dari ...	1	1
Menyediakan ruang pameran permanen atau galeri komunal bagi seniman muda ...	1	0.618361
Mengadakan workshop dan pelatihan keterampilan kreatif, seperti seni lukis, kerajinan ...	1	1
Membuka peluang kolaborasi antara seniman, pelaku UMKM, dan komunitas budaya ...	1	1
Integrasi pameran seni rupa dengan destinasi wisata unggulan, seperti Malioboro, ...	12	0.210765
Penyediaan platform digital untuk mempromosikan karya seni dan produk kreatif ...	1	0.389933
Penyusunan regulasi yang mendukung pengembangan ekonomi kreatif, seperti insentif pajak ...	1	1
SPMB untuk program pendidikan vokasi belum sepenuhnya terhubung dengan kebutuhan ...	0	1
Revisi UU Sisdiknas diperlukan untuk memperkuat posisi pendidikan vokasi sebagai ...	0	1
Infrastruktur pendidikan vokasi di DIY masih terbatas, terutama di daerah ...	9	1
Meskipun DIY memiliki potensi besar dalam pengembangan pendidikan vokasi untuk ...	0	1

Aspirasi	Topik	Probabilitas
Infrastruktur pelatihan vokasi di DIY masih terbatas, terutama di daerah ...	9	1
Banyak mitra driver online di Komunitas PAGODJA yang mengkonsumsi minuman ...	-1	0.208567
Mitra driver online umumnya kurang mendapatkan edukasi kesehatan terkait gaya ...	-1	0.20497
Mitra driver online sering kali bekerja dalam kondisi lingkungan yang ...	-1	0.200741
Banyak mitra driver online memiliki anak-anak yang sedang menempuh pendidikan. ...	-1	0.179473
Saat ini, belum ada regulasi yang secara spesifik melindungi hak-hak ...	-1	0.175233
Mitra driver online sering kali menjadi korban kebijakan sepihak dari ...	-1	0.182466
Mitra driver online di DIY masih menghadapi tantangan infrastruktur digital, ...	-1	0.18045
Museum memiliki potensi besar untuk menjadi media edukasi kesehatan, termasuk ...	-1	0.070526
Museum dapat menjadi pusat transformasi pendidikan kesehatan melalui pameran interaktif ...	-1	0.070792
Museum dapat menjadi sarana kampanye kesehatan untuk mengedukasi masyarakat tentang ...	-1	0.070908
Kurikulum SPMB belum sepenuhnya adaptif terhadap kebutuhan pendidikan berbasis budaya ...	0	1
Museum belum dimanfaatkan secara optimal dalam sistem pendidikan nasional. Revisi ...	0	1
Infrastruktur museum di DIY masih terbatas, terutama di daerah pinggiran. ...	-1	0.01356
Penyandang disabilitas, terutama yang memiliki keterbatasan mobilitas atau gangguan penglihatan, ...	14	1
Pendidikan kesehatan bagi penyandang disabilitas belum sepenuhnya inklusif. Metode pembelajaran ...	0	0.458929
Penyandang disabilitas sering kali menghadapi hambatan dalam mendapatkan layanan kesehatan ...	14	1
Penyandang disabilitas masih menghadapi diskriminasi dalam sistem penerimaan murid baru ...	0	1
Kurikulum pendidikan nasional belum sepenuhnya inklusif untuk memenuhi kebutuhan penyandang ...	0	1
Infrastruktur pendidikan dan keagamaan untuk penyandang disabilitas di DIY masih ...	9	0.451811
Sinergi antara pemerintah, lembaga keagamaan, dan komunitas penyandang disabilitas masih ...	14	1
SPMB untuk SMK masih kurang diminati dibandingkan SMA, karena persepsi ...	0	1
Revisi UU Sisdiknas diperlukan untuk memperkuat posisi SMK sebagai inkubator ...	0	1

Aspirasi	Topik	Probabilitas
Anggaran untuk modernisasi peralatan dan bengkel praktik di SMK sangat ...	0	0.602544
Guru produktif di SMK sering kali kurang kompeten karena minimnya ...	-1	0.103829
Meskipun DIY memiliki potensi besar dalam pengembangan SMK, implementasi regulasi ...	0	0.411098
Infrastruktur SMK di DIY masih terbatas, terutama di daerah pedesaan. ...	9	1
Anak-anak usia dini sering kali menjadi sasaran pemasaran produk MBDK ...	-1	0.200028
Kurikulum PAUD saat ini belum sepenuhnya mencakup pendidikan kesehatan yang ...	0	0.364092
Fasilitas PAUD sering kali tidak memenuhi standar sanitasi dan keamanan ...	-1	0.149658
Proses penerimaan murid baru untuk PAUD sering kali tidak transparan ...	0	0.261108
Revisi UU Sisdiknas diperlukan untuk memperkuat posisi PAUD sebagai fondasi ...	0	0.408642
Infrastruktur PAUD di DIY masih terbatas, terutama di daerah pedesaan. ...	9	1
Meskipun DIY memiliki potensi besar dalam pengembangan pendidikan, implementasi regulasi ...	0	0.345621
Infrastruktur PAUD di DIY masih terbatas, terutama di daerah pedesaan. ...	9	1
Sinergi antara pemerintah daerah, masyarakat, dan lembaga pendidikan dalam pengembangan ...	0	0.192351
Dorong revisi regulasi terkait pendidikan usia dini untuk memperkuat sinergi ...	0	1
Industri makanan dan minuman dalam jaringan bisnis Muhammadiyah sering kali ...	4	0.106583
Jaringan bisnis Muhammadiyah memiliki potensi besar untuk menjadi media edukasi ...	4	0.189402
Fasilitas pariwisata dan hotel yang dikelola oleh Jaringan Wisata Muhammadiyah ...	-1	0.142535
Sinergi antara Majelis Ekonomi Bisnis dan Pariwisata (MEBP) dengan lembaga ...	-1	0.19346
Revisi UU Sisdiknas diperlukan untuk memperkuat sinergi antara pilar ekonomi ...	0	1
Infrastruktur pendidikan vokasi di DIY masih terbatas, terutama di daerah ...	9	1
Meskipun DIY memiliki potensi besar dalam pengembangan ekonomi berbasis komunitas, ...	-1	0.304868
Infrastruktur ekonomi di DIY masih terbatas, terutama di daerah pedesaan. ...	9	0.143742
Sinergi antara pemerintah daerah, Muhammadiyah, dan pelaku usaha dalam pengembangan ...	4	1

Aspirasi	Topik	Probabilitas
Dorong revisi regulasi terkait ekonomi berbasis komunitas untuk memperkuat sinergi ...	4	1
Program pendidikan kesehatan saat ini belum sepenuhnya mengintegrasikan nilai-nilai budaya ...	0	0.275978
Nilai-nilai budaya lokal, seperti kebersamaan dan gotong royong, dapat dimanfaatkan ...	0	0.144677
Kurikulum SPMB saat ini belum sepenuhnya mencakup materi tentang kebudayaan ...	0	1
Revisi UU Sisdiknas diperlukan untuk memperkuat posisi kebudayaan lokal sebagai ...	0	0.433431
Infrastruktur pendidikan budaya lokal di DIY masih terbatas, terutama di ...	9	1
Meskipun DIY memiliki otonomi khusus dalam pelestarian budaya, implementasi regulasi ...	-1	0.106265
Infrastruktur budaya di DIY masih terbatas, terutama di daerah pedesaan. ...	9	1
Sinergi antara pemerintah daerah, masyarakat, dan pelaku budaya dalam pengembangan ...	-1	0.081471
Kurikulum pendidikan kesehatan saat ini belum sepenuhnya mengintegrasikan nilai-nilai karakter ...	0	0.454945
SPMB saat ini cenderung menekankan capaian akademik semata, sehingga membatasi ...	0	1
Revisi UU Sisdiknas diperlukan untuk memperkuat posisi pendidikan karakter sebagai ...	0	0.421635
Infrastruktur pendidikan karakter di DIY masih terbatas, terutama di daerah ...	9	1
Meskipun DIY memiliki potensi besar dalam pengembangan pendidikan berbasis karakter, ...	0	0.441256
Rendahnya UMP di Yogyakarta berdampak langsung pada daya beli masyarakat. ...	13	1
Pentingnya gizi bagi produktivitas pekerja tidak bisa dipandang sebelah mata. ...	13	1
Untuk mengembangkan kebijakan microfinance dan UMKM secara efektif, diperlukan komunikasi ...	1	0.17885
Rendahnya UMP di Yogyakarta berdampak langsung pada daya beli masyarakat. ...	13	1
Program Makan Bergizi Gratis yang diluncurkan oleh pemerintah dan didukung ...	-1	0.101981
Pemerintah telah menetapkan kenaikan rata-rata Upah Minimum Provinsi (UMP) sebesar ...	13	1
Buruh di DIY telah menuntut agar UMP ditetapkan lebih tinggi, ...	13	1
Kebijakan upah murah berpotensi mengurangi daya beli masyarakat dan memperburuk ...	13	1
Pemerintah perlu melakukan peninjauan kembali terhadap UMP dengan mempertimbangkan garis ...	13	1

Aspirasi	Topik	Probabilitas
Program Makan Bergizi Gratis perlu dukungan semua pihak, termasuk Organisasi ...	-1	0.072601
Pemerintah perlu segera merevisi pasal tentang penyediaan alat kontrasepsi untuk ...	0	0.472829
Meningkatkan program edukasi kesehatan reproduksi di sekolah-sekolah untuk memberikan pemahaman ...	0	1
Peningkatan honor guru swasta dan tenaga pendidik di Daerah Istimewa ...	11	1
BPJS Kesehatan perlu ditingkatkan dalam hal aksesibilitas fasilitas yang dapat ...	5	1
nominal iuran BPJS cukup menyulitkan bagi masyarakat kelas menengah ke ...	5	0.122405
agar kru relawan AmbulanMu mendapatkan jaminan sosial, khususnya akses berobat ...	-1	0.152657
Perlu adanya pengawasan yang lebih ketat terhadap dana BPJS yang ...	5	0.084955
Iuran BPJS Kesehatan untuk guru dan karyawan di sekolah tidak ...	11	0.138532
Kesejahteraan guru honorer perlu mendapat perhatian khusus mengingat peran penting ...	11	1
Merevisi kebijakan pengangkatan dan penggajian guru honorer	11	1
Memberikan tunjangan tambahan untuk guru honorer	11	1
Memperluas akses guru honorer untuk mengikuti seleksi PPPK (Pegawai Pemerintah	11	1
Memberikan pelatihan dan pengembangan kompetensi untuk meningkatkan kualitas guru honorer ...	11	1
Adanya penguatan kegiatan pemuda di Daerah Istimewa Yogyakarta (DIY) dan	2	0.048383
Meningkatkan kolaborasi antara pemerintah, sektor swasta, dan organisasi kepemudaan untuk ...	1	1
Mengembangkan program mentoring yang menghubungkan pemuda dengan profesional berpengalaman di ...	-1	0.111368
Mendorong partisipasi aktif pemuda dalam pengambilan keputusan di tingkat desa ...	3	1
Memperluas akses pendidikan dan pelatihan keterampilan yang sesuai dengan kebutuhan ...	9	0.245889
Meningkatkan program kewirausahaan pemuda dengan dukungan modal dan pendampingan bisnis ...	1	1
Mengoptimalkan pemanfaatan teknologi digital untuk membuka peluang ekonomi bagi pemuda ...	1	1
Memperkuat jaringan perlindungan sosial bagi pemuda dari keluarga kurang mampu.	0	0.282064
Merevisi dan memperkuat undang-undang terkait kelautan dan perikanan untuk memastikan	-1	0.160023

Aspirasi	Topik	Probabilitas
Pemerintah perlu melakukan simplifikasi prosedur pendaftaran dan klaim layanan BPJS. ...	5	0.199061
Implementasi sistem antrian online yang sudah diterapkan di beberapa fasilitas ...	-1	0.214405
Pemanfaatan teknologi informasi dalam pengelolaan data peserta dan layanan kesehatan ...	5	0.549938
Meningkatkan kerjasama antara BPJS Kesehatan dengan berbagai fasilitas kesehatan, baik ...	5	1
Mengadakan sosialisasi dan pelatihan di komunitas pedagang untuk memberikan pemahaman ...	5	0.741802
Mengkaji ulang struktur iuran BPJS Kesehatan bagi pedagang kaki lima ...	5	0.145607
Memastikan bahwa manfaat yang diberikan oleh BPJS Kesehatan mencakup layanan ...	5	1
Meningkatkan kualitas pelayanan di fasilitas kesehatan yang bekerja sama dengan ...	5	1
Penting untuk mengembangkan sistem integrasi antara BPJS Kesehatan dan jaminan ...	5	1
BPJS Kesehatan perlu memperluas kerjasama dengan rumah sakit swasta untuk ...	5	1
BPJS Kesehatan harus terus berupaya meningkatkan kualitas pelayanan di seluruh ...	5	1
Melakukan sosialisasi secara aktif kepada masyarakat mengenai manfaat dan cara ...	5	1
Pemerintah perlu mempertimbangkan peningkatan subsidi untuk iuran BPJS Kesehatan guna ...	5	0.240096
Saat ini, banyak guru honorer dan inpassing belum mendapatkan perlindungan ...	11	1
Banyak guru honorer swasta dan pensiunan guru swasta yang tidak ...	11	1
Pensiunan guru swasta juga menghadapi tantangan serupa. Banyak dari mereka ...	11	0.963174
Pemerintah perlu mempertimbangkan pendaftaran otomatis bagi guru honorer swasta dan ...	11	1
Saat ini, meskipun BPJS Kesehatan mencakup banyak jenis penyakit, ada ...	5	1
Masyarakat perlu diberikan informasi yang jelas mengenai manfaat dan cakupan ...	5	1
Lansia sering kali menghadapi berbagai masalah kesehatan, seperti hipertensi, diabetes, ...	10	1
Kami melihat bahwa demokrasi hari ini seringkali hanya berhenti pada ...	2	0.101192
Kami merasa santri dan mahasiswa belum cukup mendapat ruang untuk ...	2	0.101718
Merasa prihatin melihat fenomena perilaku anak muda saat ini yang ...	2	0.100032

Aspirasi	Topik	Probabilitas
Peran perempuan di ruang publik, khususnya di pemerintahan, masih belum ...	0	1
Sebagai anggota Pemuda Pancasila DIY merasa prihatin atas situasi keamanan ...	2	0.096474
Banyak di antara kami yang merasa perlu wadah untuk mengasah ...	2	0.069301
Keberadaan madrasah diniyah yang selama ini berperan besar dalam mendidik ...	0	0.547613
Banyak di antara kami yang statusnya masih honorer dengan penghasilan ...	11	1
Di tengah gejolak politik tanah air yang kerap memunculkan polarisasi, ...	2	0.08725
Selama ini banyak kader Fatayat yang telah berusaha mandiri melalui ...	-1	0.070823
Kami merasa bahwa arus globalisasi dan produk impor sering kali ...	13	1
Banyak kaum muda yang ingin berkontribusi dalam ekonomi lokal, tetapi ...	13	1