

Pemodelan Topik Percakapan Konseling Kesehatan Mental Menggunakan Metode BERTopic

Clarissa Luna Maheswari

Departemen Teknik Informatika, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia

e-mail: 5025211003@mhs.its.ac.id

Abstrak— Permasalahan kesehatan mental semakin banyak diungkapkan melalui platform digital, termasuk dalam bentuk percakapan konseling daring. Percakapan konseling mengandung informasi penting namun tidak terstruktur dan emosional, sehingga sulit dianalisis secara sistematis. Penelitian ini menggunakan metode BERTopic untuk mengidentifikasi tema psikologis utama dalam percakapan konseling. Metode ini memungkinkan eksplorasi tematik berbasis unsupervised learning tanpa ketergantungan pada label manual. Fokus diberikan pada analisis topik menggunakan representasi semantik dan evaluasi kualitas hasil.

Tahapan implementasi mencakup preprocessing, pembuatan embedding dokumen menggunakan SentenceTransformer, reduksi dimensi dengan UMAP, klusterisasi melalui HDBSCAN, dan ekstraksi topik menggunakan c-TF-IDF. Pengujian dilakukan terhadap dua konfigurasi preprocessing, yaitu L-ON S-ON (lemmatisasi aktif) dan L-OFF S-ON (tanpa lemmatisasi). Evaluasi mencakup metrik kuantitatif seperti cosine similarity, semantic coherence, dan rasio outlier, serta validasi kualitatif melalui analisis manual dan spasial. Sebagai pembandingan, LDA diuji pada tiga konfigurasi topik (5, 10, 15) dengan evaluasi berbasis koherensi c_v .

Hasil evaluasi menunjukkan bahwa BERTopic menghasilkan trade-off antara granularitas topik dan stabilitas klusterisasi tergantung pada strategi preprocessing. L-ON S-ON menghasilkan 22 topik dengan coherence tertinggi (0,3412) namun rasio outlier tinggi (44,27%), sedangkan L-OFF S-ON lebih stabil dan mencatat cosine similarity tertinggi (0,4338) tetapi mengalami dominasi satu topik besar. Kombinasi SentenceTransformer, UMAP, dan HDBSCAN terbukti efektif dalam menangkap variasi semantik. Dibandingkan LDA, topik-topik BERTopic lebih tajam dan relevan secara psikologis. Penelitian menyimpulkan bahwa hasil pemodelan sangat dipengaruhi oleh strategi preprocessing, dan pendekatan berbasis embedding lebih sesuai untuk mengeksplorasi struktur tematik dalam percakapan konseling tidak terstruktur.

Kata Kunci—BERTopic, Embedding Semantik, Kesehatan Mental, Konseling Daring, Pemodelan Topik

I. PENDAHULUAN

Kesehatan mental telah menjadi isu global yang semakin mendesak. Sebelum pandemi COVID-19, gangguan depresi dan kecemasan sudah menjadi penyebab utama beban kesehatan mental di seluruh dunia. Pandemi memperparah situasi ini secara signifikan. Studi global memperkirakan adanya tambahan 53,2 juta kasus gangguan depresi mayor dan 76,2 juta kasus gangguan kecemasan selama tahun 2020, masing-masing mencerminkan peningkatan sebesar 27,6% dan 25,6% dibandingkan sebelum pandemi [1]. Di Indonesia,

situasi tercermin dalam hasil survei nasional I-NAMHS tahun 2023 yang menunjukkan bahwa 34,9% remaja mengalami masalah kesehatan mental dalam 12 bulan terakhir, setara dengan sekitar 15,5 juta individu usia 10–17 tahun [2]. Bersamaan dengan meningkatnya perhatian terhadap isu ini, kemajuan dalam bidang pemrosesan bahasa alami (Natural Language Processing/NLP) telah membuka peluang baru dalam memahami pola pikir dan kondisi psikologis melalui analisis teks, termasuk dalam konteks konseling daring.

Peningkatan kebutuhan akan layanan psikologis mendorong pertumbuhan layanan konseling digital, khususnya berbasis teks. Platform seperti Ibunda.id dan Bicarakan.id kini menyediakan konseling daring yang anonim, fleksibel, dan mudah diakses. Layanan ini menjadi semakin relevan di tengah pembatasan sosial serta meningkatnya kenyamanan masyarakat dalam berinteraksi secara digital. Konseling berbasis teks memungkinkan pengguna untuk mengekspresikan perasaan dan pengalaman secara lebih terbuka, tanpa tekanan sosial yang sering kali muncul dalam interaksi tatap muka. Data industri menunjukkan nilai pasar terapi daring global akan naik dari USD 3,8 miliar pada 2024 menjadi lebih dari USD 10 miliar pada 2031 [3]. Studi akademik juga mengkonfirmasi lonjakan kebutuhan terhadap layanan ini secara global [4]. Dengan meningkatnya volume data yang dihasilkan dari interaksi konseling daring, dibutuhkan pendekatan analitik yang efisien dan mampu menangkap makna tersembunyi di balik narasi pengguna.

Data percakapan konseling bersifat tidak terstruktur, emosional, dan kerap mengandung metafora, sehingga menyulitkan proses analisis, terutama ketika volume datanya sangat besar. Metode klasik seperti Latent Dirichlet Allocation (LDA) tidak mampu menangkap makna implisit atau mempertahankan konteks naratif dalam teks psikologis, sehingga kurang relevan untuk analisis topik percakapan konseling. Dalam berbagai studi, LDA terbukti menghasilkan topik yang kurang koheren dalam konteks data pendek dan emosional dibandingkan pendekatan modern seperti BERTopic [5], [6]. Tantangan lain dari LDA adalah keterbatasannya dalam menangkap struktur semantik dan konteks emosional dalam teks, akibat pendekatannya yang berbasis bag-of-words yang mengabaikan relasi makna antar kata [7]. Kebutuhan untuk memahami dinamika komunikasi konseling secara lebih mendalam mendorong penggunaan representasi semantik dan embedding kontekstual yang mampu mempertahankan makna dan konteks.

Untuk mengatasi keterbatasan ini, metode BERTopic dikembangkan menggunakan model transformer dan teknik clustering berbasis kepadatan. BERTopic mampu menghasilkan topik yang lebih koheren secara semantik dan mempertahankan konteks pada dokumen panjang, seperti percakapan konseling. Studi menunjukkan bahwa BERTopic menghasilkan kluster yang lebih padat dan interpretatif dibandingkan LDA, serta lebih unggul dalam menangkap variasi semantik dalam teks [8]. Keunggulan BERTopic dalam menangani teks dengan struktur naratif yang kompleks menjadikan metode ini kandidat ideal untuk mengeksplorasi dinamika percakapan psikologis secara lebih mendalam. Penelitian ini mengusulkan penerapan BERTopic untuk pemodelan topik otomatis dari data percakapan konseling kesehatan mental. Proses dilakukan mulai dari pengumpulan data teks, prapemrosesan, pembuatan representasi embedding, klusterisasi, hingga ekstraksi topik. Hasil akhir akan dievaluasi secara kuantitatif melalui metrik koherensi dan kualitas kluster, serta secara kualitatif melalui analisis isi.

II. TINJAUAN PUSTAKA

Tabel 1. Penelitian Terkait BERTopic dan Evaluasinya

Peneliti dan Tahun	Data	Metode	Library/ Tool	Hasil Pengujian
Abuza yed et al. (2021)	108.789 dokumen teks berbahasa Arab dari tiga portal berita	LDA, NMF, BERTopic (BERT Embeddings, UMAP, HDBSCAN)	Scikit-learn, Gensim, BERTopic, Sentence Transformers	BERTopic menghasilkan koherensi topik tertinggi berdasarkan NPMI. LDA menunjukkan nilai negatif. BERTopic dengan embedding AraVec memperlihatkan performa terbaik. Nilai estimasi NPMI dari grafik mencapai ~0,65.
Mutsaddi et al. (2025)	5.225 ulasan pendek berbahasa Hindi	BERTopic, Gensim, Scikit-learn, LDA, NMF, CTM, Top2Vec, dll.	BERTopic, Gensim, Scikit-learn, BigARTM, OCTIS, Sentence Transformers	BERTopic (mBERT-Uncased) menghasilkan nilai cv 0,76 dan cNPMI 0,070 pada 95 topik. LDA hanya mencapai cv 0,38 dan cNPMI -0,12 pada 165 topik.
De Groot et al. (2022)	62.522 tanggapan terbuka mahasiswa dari 5 fakultas	BERTopic (HDBSCAN, k-Means), LDA	BERTopic, Gensim, Scikit-learn	BERTopic (HDBSCAN) mencapai koherensi topik 0,091 dan diversity 0,880, mengungguli LDA yang hanya mencapai koherensi 0,031 dan diversity 0,718.
Mihajlov et al. (2024)	100 novel dalam bahasa Serbia dari korpus SrpELTeC	LDA, NMF, BERTopic	Sentence Transformer (paraphrase-mpnet-base-v2), Word2Vec, HDBSCAN, UMAP, RoBERTa	BERTopic (mpnet) menghasilkan skor c_v terendah 0.299, dibanding NMF 0.568 dan LDA 0.361. BERTopic (Jerteh-355) mencatat topic diversity tertinggi 0.970, menunjukkan c_v kurang representatif untuk model berbasis embedding.

		Ta, NLTK	
Rahimi et al. (2023)	20.000 dokumen berita dan 17.000 tweet Elon Musk	LDA, ETM, ATM, CTM, BERTopic, Top2Vec	BERTopic, CTM, CPML, CTC package, Sentence-BERT, ChatGPT
		Topik tidak bermakna dari Top2Vec mendapat skor c_v sangat tinggi (0.995), tapi menunjukkan nilai tinggi topik semantik yang buruk (0.009)	

Tabel 2.1 merangkum lima penelitian yang relevan, baik penelitian yang mendukung keunggulan BERTopic maupun penelitian yang mengkritisi validitas metrik evaluasi tertentu. Penelitian oleh Abuzayed et al. [9] menunjukkan bahwa BERTopic menghasilkan nilai koherensi tertinggi berdasarkan NPMI pada dokumen berita berbahasa Arab, mengungguli LDA yang menunjukkan skor negatif. Hasil yang didapatkan mendukung keandalan BERTopic ketika diterapkan pada data dengan struktur narasi eksplisit. Mutsaddi et al. [10] memperkuat temuan ini dengan membandingkan tujuh model berbeda pada data ulasan pendek berbahasa Hindi, di mana BERTopic dengan mBERT menghasilkan nilai c_V sebesar 0,76, jauh di atas LDA yang hanya mencapai 0,38.

De Groot et al. [11] menerapkan BERTopic untuk menganalisis tanggapan mahasiswa, dan membandingkannya dengan LDA dari sisi topic coherence dan topic diversity. Hasilnya menunjukkan bahwa BERTopic (HDBSCAN) unggul baik dalam koherensi (0,091 vs. 0,031) maupun keberagaman topik (0,880 vs. 0,718), mengindikasikan keunggulan model ini dalam menangani data naratif terbuka. Mihajlov et al. [12] menunjukkan bahwa BERTopic dengan embedding mpnet memperoleh nilai c_V terendah (0,299) dibanding NMF (0,568) dan LDA (0,361), walaupun mencatat topic diversity tertinggi (0,970). Temuan tersebut menyoroti keterbatasan c_V sebagai metrik evaluasi untuk model berbasis semantic embedding. Rahimi et al. [13] secara eksplisit menunjukkan kelemahan c_V, di mana topik tidak bermakna (trash topic) yang terdiri dari kode tim hoki mendapat skor c_V tinggi (0,995), namun skor semantic CPML-nya justru rendah (0,009). Tinjauan ini menjadi bukti bahwa c_V dapat memberikan sinyal positif palsu pada model berbasis embedding seperti BERTopic.

Berdasarkan pengamatan terhadap berbagai penelitian yang ditinjau, dapat disimpulkan bahwa penggunaan BERTopic cenderung menghasilkan topik yang lebih konsisten secara semantik dan lebih bervariasi dibandingkan metode tradisional. Efektivitas evaluasi sangat bergantung pada pemilihan metrik yang sesuai. Metrik koherensi tradisional seperti c_V, NPMI, dan UMass terbukti tidak selalu mampu menangkap kualitas semantik dari topik yang dihasilkan oleh model embedding seperti BERTopic. Penelitian juga menguji salah satu metode klasik yaitu LDA sebagai baseline. Evaluasi terhadap LDA menggunakan metrik c_V dengan tujuan perbandingan umum terhadap metode konvensional yang masih banyak digunakan dalam literatur. Langkah yang diambil bertujuan untuk menilai sejauh mana model berbasis probabilistik seperti LDA mampu mengungkap makna topik dalam domain data yang kompleks dan ekspresif seperti percakapan konseling daring. *Gap* yang

diisi dalam penelitian adalah evaluasi kualitatif berbasis makna psikologis serta strategi preprocessing yang disesuaikan dengan karakteristik data percakapan konseling.

III. METODOLOGI

A. Pengumpulan Dataset

Dataset yang digunakan berasal dari platform CounselChat.com, sebuah layanan konseling psikologi daring yang memungkinkan pengguna mengirimkan pertanyaan secara anonim kepada terapis profesional. Fokus analisis diarahkan pada kolom questionText yang memuat narasi pertanyaan pengguna dalam bahasa sehari-hari tanpa identitas personal. Data bersifat terbuka dan otentik, mencerminkan isu psikologis secara langsung dalam bentuk ekspresi emosional. Dataset diperoleh dari Kaggle dan merupakan hasil web scraping dari situs CounselChat.com. Dataset terdiri dari dua berkas terpisah, yaitu 20200325_counsel_chat.csv dan counselchat-data.csv. Setelah dilakukan proses deduplikasi dan pembersihan data, jumlah total entri berkurang dari 3.512 menjadi 994 dokumen unik. Panjang rata-rata teks meningkat dari 55,20 menjadi 69,52 kata, disertai peningkatan standar deviasi dari 48,27 menjadi 65,28 kata, yang menunjukkan tingginya variasi bentuk dan kedalaman narasi. Seluruh analisis difokuskan pada narasi penuh (questionText) tanpa menggabungkan title atau metadata lainnya. Karakteristik data yang panjang dan ekspresif memperkuat keputusan untuk menggunakan pendekatan unsupervised berbasis representasi semantik.

Dataset bersumber dari dua bekas terpisah, yaitu 20200325_counsel_chat.csv dan counselchat-data.csv. Setelah proses deduplikasi dan pembersihan, total entri berkurang dari 3.512 menjadi 994 dokumen unik. Panjang rata-rata teks meningkat dari 55,20 menjadi 69,52 kata, disertai peningkatan standar deviasi dari 48,27 menjadi 65,28 kata, menunjukkan tingginya variasi naratif. Seluruh analisis difokuskan pada narasi penuh (questionText) tanpa menggabungkan judul atau metadata lainnya. Karakteristik data yang bervariasi panjang dan kedalamannya memperkuat keputusan untuk menggunakan pendekatan unsupervised berbasis representasi semantik.

B. Pra-pemrosesan Teks

Pra-pemrosesan dilakukan untuk membersihkan dan menstandarkan teks naratif sebelum diproses lebih lanjut secara semantik. Proses mencakup normalisasi karakter non-ASCII, ekspansi kontraksi, konversi huruf kecil, dan koreksi ejaan menggunakan SymSpell. Seluruh karakter non-alfabet dihapus, dan teks disederhanakan menjadi bentuk standar agar stabil untuk representasi embedding.

Pemrosesan linguistik meliputi lemmatisasi menggunakan pustaka spaCy dan penghapusan stopword dari daftar bawaan scikit-learn. Dua konfigurasi pra-pemrosesan diuji, yaitu: (1) lemmatisasi aktif tanpa penghapusan stopword, dan (2) lemmatisasi + stopword removal. Konfigurasi pertama mempertahankan ekspresi emosional seperti "just" dan "not", sedangkan konfigurasi kedua menghasilkan teks yang lebih bersih secara statistik. Pemilihan kombinasi bertujuan

mengeksplorasi trade-off antara kebersihan linguistik dan pelestarian makna emosional dalam konteks data psikologis.

C. Pemodelan Topik dengan Metode BERTopic

Teks dianalisis menggunakan pendekatan pemodelan topik berbasis BERTopic, yang terdiri dari tahapan embedding, reduksi dimensi, klasterisasi, dan ekstraksi kata kunci. Representasi semantik diperoleh melalui model all-mpnet-base-v2 dari pustaka SentenceTransformers, yang menghasilkan embedding berdimensi 768 untuk setiap kalimat. Representasi ini direduksi menjadi lima dimensi menggunakan UMAP (n_neighbors=15) untuk mempertahankan struktur semantik dalam ruang rendah. Proyeksi hasil embedding kemudian dikelompokkan menggunakan algoritma HDBSCAN (min_cluster_size=10), yang memungkinkan pembentukan kluster topik secara otomatis tanpa perlu menetapkan jumlah topik sebelumnya. Topik yang terbentuk diekstraksi menggunakan teknik class-based TF-IDF (c-TF-IDF), yang menghitung bobot kata berdasarkan distribusinya dalam satu kluster dibandingkan seluruh korpus.

D. Pemodelan Topik dengan Metode LDA

Sebagai metode pembandingan, penelitian ini juga menerapkan Latent Dirichlet Allocation (LDA) menggunakan pendekatan Bag of Words (BoW). Metode LDA diterapkan dengan mengubah narasi konseling menjadi vektor frekuensi kata melalui pendekatan Bag of Words dan CountVectorizer. Parameter max_df = 0.95 digunakan untuk menyaring kata umum yang terlalu sering muncul, sementara min_df = 2 mengecualikan kata yang hanya muncul sekali. Eksperimen dilakukan pada tiga konfigurasi jumlah topik: 5, 10, dan 15, untuk menilai pengaruh granularitas terhadap struktur tematik. Rentang tersebut dipilih untuk menjaga keseimbangan antara generalisasi tema dan fragmentasi. Parameter random_state ditetapkan agar hasil reproduisibel, sedangkan learning_method diatur ke "online" untuk efisiensi komputasi. Batas max_iter = 10 menjadi kompromi antara konvergensi dan waktu komputasi. Model dilatih per konfigurasi, dan hasilnya diekstraksi dalam bentuk sepuluh kata kunci teratas berdasarkan bobot probabilitas tertinggi untuk merepresentasikan tiap topik.

E. Evaluasi Kuantitatif

Evaluasi pemodelan topik dalam penelitian ini menggunakan metrik statistik kuantitatif seperti c_V, NPMI, dan topic diversity. Metrik c_V dipilih sebagai indikator utama karena memiliki korelasi tinggi terhadap penilaian manusia dalam pengujian empiris [14]. NPMI digunakan sebagai pelengkap karena mengukur kekuatan asosiasi antar kata berdasarkan ko-okurensi [15]. Untuk menilai keberagaman kata kunci antar topik, digunakan metrik topic diversity sebagaimana diusulkan oleh pendekatan neural topic modeling [16]. Penghitungan seluruh metrik dilakukan secara otomatis menggunakan pustaka Gensim dan Scikit-learn tanpa campur tangan subjektif [13]. BERTopic secara eksplisit mendukung penggunaan metrik ini sebagai bagian dari evaluasi standar [18]. Namun, efektivitas metrik berbasis frekuensi ini masih diperdebatkan, terutama dalam domain psikologis. Studi sebelumnya

menunjukkan bahwa skor coherence tinggi tidak selalu mencerminkan interpretasi manusia terhadap makna topik, khususnya dalam data emosional dan informal [19][20]. Metrik kuantitatif dalam penelitian ini difungsikan sebagai pelengkap, bukan penentu utama kualitas topik. Evaluasi juga dilengkapi dengan outlier ratio yang mencerminkan efisiensi segmentasi HDBSCAN dalam membentuk kluster tematik [21]. Metrik yang digunakan memberikan indikasi tambahan terhadap keberhasilan model dalam menyusun struktur topik yang padat dan bermakna, khususnya dalam konteks data naratif konseling yang tidak terstruktur.

Evaluasi pemodelan topik dalam penelitian ini mengacu pada metrik kuantitatif berbasis statistik yang telah digunakan secara luas dalam studi topik modeling, seperti c_V , NPMI, dan topic diversity. Metrik c_V dipilih sebagai indikator utama karena terbukti memiliki korelasi tinggi terhadap penilaian manusia berdasarkan studi Röder et al. [15]. Sementara itu, NPMI digunakan sebagai pelengkap karena mengukur asosiasi antar kata berdasarkan kemunculan bersamaan dalam dokumen. Untuk mengukur keberagaman topik, digunakan metrik topic diversity seperti yang disarankan oleh Dieng et al. [16]. Seluruh metrik ini dihitung secara otomatis menggunakan pustaka Gensim dan Scikit-learn, tanpa campur tangan manusia. BERTopic sendiri mendukung penggunaan metrik tersebut sebagai bagian dari evaluasi standar [8]. Di sisi lain, efektivitas metrik berbasis frekuensi ini telah dikritisi oleh Syed dan Spruit [13] serta Hoyle et al. [14] yang menunjukkan bahwa coherence score tidak selalu sejalan dengan persepsi semantik manusia, khususnya pada data psikologis yang bersifat personal dan emosional. Penilaian kuantitatif dalam studi ini tidak dijadikan dasar utama dalam interpretasi tematik. Evaluasi juga dilengkapi dengan metrik outlier ratio yang mencerminkan efisiensi klusterisasi dokumen oleh HDBSCAN, sebagai bagian dari pipeline BERTopic. Nilai yang dihasilkan memberikan informasi tambahan tentang sejauh mana model mampu membentuk struktur topik yang bermakna dalam konteks data naratif tidak terstruktur.

F. Evaluasi Kualitatif

Evaluasi kualitatif dalam penelitian ini dirancang sebagai validasi teknis untuk menilai akurasi hasil pemodelan topik dalam konteks analisis teks komputasional. Fokus utama adalah koherensi semantik dan akurasi pengelompokan dokumen, bukan eksplorasi psikologis klinis. Validasi dilakukan untuk memastikan bahwa model Lematisasi Aktif mampu mengidentifikasi kata kunci yang bermakna secara semantik, mengelompokkan dokumen berdasarkan kemiripan, serta menghasilkan topik yang relevan dalam ruang semantik. Pendekatan ini mengadaptasi metode thematic analysis, narrative analysis, dan pattern recognition dari ranah linguistik dan psikologi komputasional (17)(18)(19).

Kerangka evaluasi disusun dengan mengadaptasi praktik sistematis dalam psikologi klinis, termasuk initial symptom screening (20), thematic analysis (17), narrative analysis (21), serta identifikasi pola dalam konteks klasifikasi ICD-11 (22). Evaluasi dilakukan bertahap, dimulai dari pemeriksaan daftar kata kunci, analisis kesesuaian makna dalam dokumen, hingga

visualisasi distribusi spasial. Triangulasi dilakukan melalui diskusi informal dengan seorang psikiater yang juga dosen di Program Studi Pendidikan Profesi Dokter ITS. Diskusi ini digunakan untuk menguji kredibilitas temuan dan mengaitkannya dengan pola asesmen psikologis klinis, yang melibatkan narasi pasien, identifikasi gejala, dan pencocokan terhadap sistem klasifikasi seperti ICD-11 (22). Pendekatan ini merujuk pada metode triangulation via researchers dalam penelitian kualitatif (23).

Strategi pengambilan sampel dilakukan secara purposif dengan memilih tiga topik yang mewakili variasi ukuran kluster dan kualitas semantik. Dari masing-masing topik dipilih tiga hingga empat dokumen secara acak sehingga terkumpul 20 dokumen yang dianalisis manual. Kategorisasi topik dilakukan berdasarkan kerangka ICD-11, dengan fokus pada Z-codes yang relevan dengan isu relasional dan psikososial (22). Distribusi topik dibandingkan terhadap struktur tematik dalam ICD-11 untuk menilai kesesuaian klasifikasi semantik.

Visualisasi hasil pemodelan dilakukan melalui fungsi `visualize_documents()` serta dua bentuk utama: word cloud dan histogram distribusi dokumen. Word cloud digunakan untuk menilai dominasi semantik dan kohesi kata kunci berbasis $c\text{-TF-IDF}$. Smith et al. menyebut word cloud mampu meningkatkan pemahaman pengguna terhadap struktur topik (24). Histogram distribusi membantu mengidentifikasi ketidakseimbangan kluster seperti supercluster. Constantinou et al. menekankan pentingnya visualisasi distribusi dalam evaluasi kestabilan model topik (25). Visualisasi lain seperti dendrogram dan bar chart disimpan di lampiran karena tidak memberikan informasi interpretatif tambahan yang signifikan.

Evaluasi ini merupakan bagian dari pendekatan mixed-methods yang menggabungkan computational modeling dan interpretasi manual. Kolaborasi lintas disiplin dilakukan melalui integrasi technical framework NLP dan domain knowledge psikologi klinis. Pendekatan ini sah dalam penelitian kualitatif sebagai bentuk triangulasi sumber dan perspektif (23). Tujuan akhir dari evaluasi ini adalah memastikan bahwa model topik yang digunakan mampu menghasilkan struktur tematik yang valid secara statistik dan bermakna secara kontekstual, serta layak digunakan sebagai alat bantu eksplorasi awal dalam ranah konseling psikologis berbasis teks.

IV. HASIL DAN PEMBAHASAN

Hasil pemodelan topik dievaluasi untuk menilai struktur semantik dan relevansi tematik dari keluaran BERTopic terhadap data konseling psikologis. Dua konfigurasi preprocessing diuji yaitu Lematisasi Aktif dan Lematisasi Nonaktif. Evaluasi dilakukan secara kuantitatif dan kualitatif untuk memperoleh gambaran menyeluruh terkait performa model.

Tabel 2. Hasil Evaluasi Kuantitatif Konfigurasi Processing

Dimensi Evaluasi	Lematisasi Aktif	Lematisasi Nonaktif
Jumlah Topik Terklasifikasi	22 topik	14 topik
Distribusi Dokumen	Merata (10 - 51 dokumen)	Dominan (10 – 328 dokumen)

c_V	0,4134	0,4378
NPMI	-0,1121	-0,1113
Topic Diversity	0,7478	0,8455

Secara kuantitatif, L-OFF S-ON menunjukkan nilai cosine similarity tertinggi (0,4338), semantic coherence lebih baik, dan rasio outlier terendah (29,28%). L-ON S-ON mencatat similarity lebih rendah (0,3907) dan rasio outlier lebih tinggi (44,27%), namun menghasilkan distribusi topik yang lebih merata dengan 22 topik (masing-masing 10–51 dokumen). Hal ini menunjukkan trade-off antara granularitas dan kestabilan kluster: lemmatisasi menghasilkan topik lebih spesifik namun meningkatkan jumlah dokumen yang tidak dapat diklasifikasikan stabil. Perbandingan karakteristik dan metrik kuantitatif kedua konfigurasi dirangkum pada Tabel 2.

Model LDA yang menjadi *baseline* diuji dengan konfigurasi 5, 10, dan 15 topik. Hasil evaluasi menunjukkan bahwa konfigurasi 5 topik memiliki c_V tertinggi (0,5222), diikuti oleh konfigurasi 10 (0,4582) dan 15 (0,4574). Skor koherensi yang menurun menunjukkan bahwa peningkatan jumlah topik dapat menurunkan stabilitas semantik. Model cenderung membentuk kluster kecil yang kurang bermakna, terutama dalam konteks narasi konseling yang tidak terstruktur. Hasil menunjukkan bahwa konfigurasi L-OFF S-ON memberikan nilai cosine similarity tertinggi (0,4338) dan rasio outlier terendah (29,28%), sedangkan L-ON S-ON mencatat similarity lebih rendah (0,3907) dan rasio outlier lebih tinggi (44,27%). Perbedaan yang ada menunjukkan bahwa lemmatisasi cenderung menghasilkan kluster lebih granular namun meningkatkan jumlah outlier.

A. Pemodelan BERTopic dengan Konfigurasi Lemmatisasi Aktif

Konfigurasi BERTopic dengan lemmatisasi aktif menghasilkan 22 topik, dengan distribusi dokumen antara 12 hingga 67 dokumen per topik. Sebaran angka menunjukkan bahwa model mampu menangkap keragaman tematik dalam data konseling daring tanpa membentuk kluster seragam. Evaluasi semantik menggunakan metrik c_V menunjukkan variasi kualitas topik yang signifikan, dengan skor tertinggi 0,5805 (Topik 19) dan terendah 0,2281 (Topik 10). Distribusi skor lengkap dan daftar kata kunci utama tiap topik disajikan tersedia di Lampiran A.

Tabel 3. Kategorisasi Topik Berdasarkan Domain Psikologis Konfigurasi Lemmatisasi Aktif

Tema Besar	Topik	Jumlah	Tema Spesifik
Kesehatan Mental Klinis	0	67	Gangguan kecemasan dan depresi klinis
	7	26	Episode depresi dengan masalah harga diri
	10	20	Pikiran obsesif dan ruminasi
	12	19	Kecemasan sosial dan fobia sekolah
	13	19	PTSD dan trauma kekerasan
	15	15	Depresi remaja dengan menyakiti diri
	16	15	Gangguan dukacita berkepanjangan
	21	12	Gangguan tidur dan

			mimpi buruk
Hubungan dan Intimasi	2	44	Masalah hubungan dan ketertarikan romantis
	3	39	Disfungsi seksual dan masalah intimasi
	4	28	Krisis pernikahan dan perselingkuhan
	5	27	Kepemilikan berlebihan dan cemburu dalam hubungan
	6	27	Kecemasan perpisahan dan ketergantungan
	17	13	Drama kehamilan dan kekacauan hubungan
	18	12	Manipulasi dan masalah kepercayaan
	20	12	Jarak emosional dalam pernikahan
Dinamika Keluarga	1	54	Manajemen kemarahan dan konflik keluarga
	8	25	Trauma keluarga dan masalah hak asuh
	11	19	Tantangan mengasuh
	14	15	Konflik mertua dan kekerasan verbal
	22	10	Stres mengasuh orang tua dengan alzheimer
Identitas dan Perkembangan	9	24	Eksplorasi identitas gender dan seksualitas
	19	12	Kebingunan identitas dan komunikasi

Pada saat validasi manual berlangsung, ditemukan sejumlah token hasil preprocessing yang mengalami distorsi bentuk (misal: exwife, selfharm, familyoriented) akibat penghapusan karakter khusus, lemmatisasi, dan koreksi ejaan. Token-token tersebut dipertahankan dengan interpretasi ulang agar makna semantik tetap konsisten dengan konteks naratif. Analisis lanjut terhadap kata kunci dan distribusi dokumen memungkinkan pengelompokan topik ke dalam empat domain psikologis utama, yaitu: (1) Kesehatan Mental Klinis, (2) Hubungan dan Intimasi, (3) Dinamika Keluarga, serta (4) Identitas dan Perkembangan. Rincian kategorisasi ini disajikan pada Tabel 2. Klasifikasi tersebut bersifat eksploratif dan bertujuan memberi struktur konseptual terhadap hasil pemodelan unsupervised untuk interpretasi tematik yang lebih sistematis.

Visualisasi word cloud untuk enam topik utama ditampilkan pada Lampiran C, distribusi jumlah dokumen per tema pada Lampiran E, dan pemetaan spasial dokumen berbasis UMAP ditampilkan pada Lampiran G. Evaluasi kualitatif dilakukan pada tiga topik representatif berdasarkan ukuran kluster, koherensi semantik, dan domain psikologis: Topik 0 (gangguan kecemasan), Topik 2 (hubungan romantis), dan Topik 19 (eksplorasi identitas). Temuan rinci pada masing-masing topik ini menunjukkan bahwa BERTopic mampu menghasilkan kluster tematik yang kohesif, meskipun pada beberapa kasus ditemukan indikasi kesalahan klasifikasi (false positive) seperti pada Topik 19.

B. Pemodelan BERTopic dengan Konfigurasi Lemmatisasi Nonaktif

Pengujian kedua dilakukan untuk menilai dampak dinonaktifkannya proses lemmatisasi terhadap kualitas dan

struktur hasil pemodelan topik dengan BERTopic. Berbeda dengan konfigurasi sebelumnya yang menggunakan normalisasi bentuk kata, konfigurasi ini mempertahankan bentuk asli kata dalam data, sementara penghapusan stopwords tetap aktif. Hasilnya menunjukkan perubahan signifikan dalam karakteristik distribusi dan kualitas semantik topik.

Model menghasilkan total 14 topik, dengan penyebaran dokumen yang sangat tidak merata. Topik 0 mendominasi hampir sepertiga dari keseluruhan data (328 dokumen dari 994), sementara beberapa topik lainnya seperti topik 12 dan 13 yang hanya terdiri dari 10 dokumen. Jumlah dokumen yang tidak dapat diklasifikasikan oleh algoritma HDBSCAN juga meningkat drastis, menghasilkan rasio outlier sebesar 29,28%. Hal tersebut menunjukkan bahwa ketika lemmatisasi tidak dilakukan, banyak dokumen kehilangan kesamaan semantik yang cukup untuk tergabung dalam satu kluster. Sebaliknya, dokumen-dokumen dengan ekspresi umum dan kata-kata frekwentatif cenderung terserap ke dalam topik besar yang lebih generik, seperti topik 0.

Evaluasi kuantitatif terhadap koherensi semantik menggunakan metrik c_v mengungkapkan kualitas topik yang bervariasi. Skor tertinggi dicapai oleh topik 9 (0,6918), yang menunjukkan hubungan semantik yang kuat antara kata kunci yang terbentuk. Lebih dari separuh topik memiliki skor c_v di bawah 0,40, yang merupakan batas minimum yang sering digunakan untuk menilai koherensi internal. Salah satu topik bahkan mencatat skor serendah 0,2504 (Topik 10), yang mengindikasikan bahwa kata kunci yang terkandung memiliki hubungan semantik yang lemah atau tidak konsisten.

Tabel 4. Kategorisasi Topik Berdasarkan Domain Psikologis Konfigurasi Lematisasi Nonaktif

Tema Besar	Topik	Jumlah	Tema Spesifik
Kesehatan Mental Klinis	0	328	Kecemasan dan depresi umum
	10	14	Dukacita dan trauma kematian
	13	10	Body image dan bullying
Hubungan dan Intimasi	2	57	Disfungsi seksual dan masalah intimasi
	3	47	Krisis pernikahan dan perselingkuhan
	4	38	Masalah hubungan
	8	16	Masalah persahabatan dan kedekatan
	11	12	Ketergantungan emosional dalam hubungan
	12	10	Ekspektasi pernikahan dan kekecewaan
Dinamika Keluarga	1	96	Konflik orangtua-anak
	5	29	Siklus kekerasan dan trauma masa kecil
	6	23	Konflik interpersonal dan kritik berlebihan
Identitas dan Perkembangan	7	21	Eksplorasi identitas gender dan seksualitas
Masalah Perilaku	9	16	Kebohongan kompulsif dan financial abuse

Salah satu penyebab utama rendahnya koherensi adalah

terpecahnya bentuk kata kerja dan lema. Contohnya, married, got, said, wanted, dan lies tetap dikenali sebagai token berbeda karena tidak dinormalisasi. Demikian pula kata-kata ekspresif seperti just, like, dan know yang tidak terhapus karena tidak termasuk dalam stopwords bawaan. Akumulasi bentuk linguistik semacam ini menghasilkan *noise* leksikal, yang berdampak pada keakuratan pemodelan semantik dan meningkatkan risiko terjadinya superklasterisasi atau misgruiping topik-topik spesifik ke dalam kategori umum.

Hasil struktur tematik dari konfigurasi ini tetap dapat dievaluasi dan ditafsirkan. Untuk memfasilitasi interpretasi hasil, seluruh topik yang dihasilkan dikelompokkan ke dalam lima domain psikologis besar, yaitu: (1) Kesehatan Mental Klinis, (2) Hubungan dan Intimasi, (3) Dinamika Keluarga, (4) Identitas dan Perkembangan, serta (5) Masalah Perilaku. Rincian pengelompokan tersebut, termasuk jumlah dokumen dan tema spesifik masing-masing topik, disajikan pada Tabel 3. Kerangka kategorisasi disusun secara konsisten dengan konfigurasi sebelumnya untuk mempermudah perbandingan antar pendekatan preprocessing.

Distribusi spasial hasil pemodelan divisualisasikan menggunakan algoritma UMAP untuk memproyeksikan dokumen ke dalam ruang dua dimensi. Visualisasi ini bertujuan untuk menilai penyebaran dokumen antar topik dan mendeteksi potensi tumpang tindih kluster. Hasil proyeksi menunjukkan bahwa dokumen dengan tema interpersonal dan keluarga cenderung membentuk kelompok yang cukup padat, sementara dokumen dengan ekspresi umum atau bentuk linguistik bervariasi cenderung tersebar. Visualisasi lengkap ditampilkan pada Lampiran H.

Representasi visual berbasis Word Cloud digunakan untuk menggambarkan dominasi kata kunci dalam setiap topik. Topik-topik dengan tema klinis dan hubungan menampilkan kata kunci yang lebih spesifik secara psikologis, sedangkan topik generik seperti Topik 0 memperlihatkan kata-kata umum yang sering muncul dalam ekspresi sehari-hari. Seluruh Word Cloud disajikan pada Lampiran D. Distribusi kuantitatif jumlah dokumen per tema divisualisasikan dalam bentuk histogram. Grafik ini menunjukkan bahwa topik 0 mendominasi dataset secara signifikan, sementara sebagian besar topik lainnya hanya terdiri dari 10–50 dokumen. Visualisasi histogram ditampilkan pada Lampiran F.

C. Pemodelan LDA dengan Tiga Konfigurasi

Evaluasi terhadap tiga konfigurasi LDA (5, 10, dan 15 topik) menunjukkan pola umum bahwa model ini cenderung membentuk kluster berdasarkan frekuensi kata, bukan keterhubungan semantik. Kecenderungan ini menyebabkan ketimpangan distribusi topik, dominasi kata-kata umum, dan lemahnya fokus tematik dalam banyak topik yang terbentuk.

Pada konfigurasi lima topik, topik 4 menyerap 63,9% dokumen, didominasi kata seperti *feel*, *want*, dan *like* yang bersifat umum. Sementara topik 2, satu-satunya topik dengan kata kunci eksplisit psikologis seperti *depression* dan *anxiety*, justru mencatat koherensi terendah ($c_v = 0,4101$). Meskipun topik 0 memiliki skor tertinggi ($c_v = 0,6554$), strukturnya tidak mencerminkan satu tema psikologis tertentu. Topik-topik

lainnya bersifat terlalu sempit atau terlalu netral untuk diinterpretasi secara klinis.

Konfigurasi sepuluh topik memperlihatkan distribusi yang tetap timpang. Topik 8 mencakup hampir separuh dataset (47,8%) dengan kata-kata seperti *want*, *say*, dan *like*. Enam topik mencakup kurang dari 50 dokumen. Topik 6 memperlihatkan arah tema klinis seperti *anxiety* dan *depression* ($c_v = 0,5216$), namun tetap bercampur dengan istilah umum. Topik 1 memiliki skor koherensi tinggi ($c_v = 0,5986$), tetapi hanya memuat enam dokumen, menyulitkan generalisasi.

Pada konfigurasi lima belas topik, muncul fragmentasi berlebihan. Topik 4, 8, dan 12 masing-masing mencakup lebih dari 290 dokumen dan tetap didominasi oleh kata-kata umum seperti *feel*, *want*, dan *know*. Sebagian besar topik kecil seperti topik 0, 2, dan 5 tidak memiliki konsistensi tematik, walaupun mengandung kata-kata yang menunjukkan arah ke isu psikologis tertentu. Satu-satunya topik dengan struktur semantik yang solid adalah topik 14 ($c_v = 0,5558$) dengan kata kunci seperti *depression*, *anxiety*, dan *stress*.

Secara keseluruhan, hasil evaluasi menunjukkan bahwa LDA tidak optimal untuk memodelkan tema dalam data percakapan konseling yang bersifat emosional dan tidak terstruktur. Struktur topik yang terbentuk lebih mencerminkan statistik frekuensi kata daripada hubungan makna, sehingga interpretasi psikologis menjadi terbatas. Distribusi lengkap topik dan skor koherensi per konfigurasi disajikan dalam Lampiran F.

D. Analisis Perbandingan dan Implikasi

Tabel 5. Perbandingan Rata-rata Skor Koherensi BERTopic dan LDA

Metode	Konfigurasi	c_V	NPMI	Topic Diversity
BERTopic	Lematisasi Aktif	0,4109	-0,1121	0,7478
LDA	5 Topik	0,5222	-0,0037	0,7800
LDA	10 Topik	0,4582	-0,1162	0,8300
LDA	15 Topik	0,4574	-0,1053	0,7467

Hasil evaluasi menunjukkan perbedaan karakteristik mendasar antara LDA dan BERTopic dalam pendekatan pemodelan topik, khususnya pada data tidak terstruktur seperti narasi konseling. LDA dengan konfigurasi lima topik mencatat skor koherensi semantik c_v tertinggi sebesar 0,5222 dan skor NPMI terbaik sebesar -0,0037, yang mencerminkan kemampuan metode probabilistik dalam mengidentifikasi pola ko-okurensi kata secara statistik. Skor c_v yang lebih tinggi pada LDA tidak secara otomatis merepresentasikan keunggulan semantik. Metrik c_v berbasis Pointwise Mutual Information (PMI) cenderung lebih menguntungkan model berbasis frekuensi kata seperti LDA, dibandingkan model berbasis representasi semantik kontekstual seperti BERTopic. Pada pengujian ini, BERTopic dengan konfigurasi lemmatisasi aktif mencatat skor c_v sebesar 0,4109 dan NPMI sebesar -0,1121. Meskipun secara numerik lebih rendah, nilai tersebut mencerminkan orientasi model terhadap kesamaan makna yang bersifat kontekstual antar dokumen.

Tabel 4 menyajikan perbandingan rata-rata skor koherensi, NPMI, dan topic diversity antara BERTopic dan LDA. LDA menunjukkan penurunan performa pada konfigurasi 10 dan 15

topik, dengan skor c_v masing-masing sebesar 0,4582 dan 0,4574. Penurunan ini menandakan keterbatasan LDA dalam mempertahankan kohesi topik saat jumlah topik meningkat, terutama pada data dengan struktur narasi yang bebas dan variasi linguistik tinggi. Meskipun konfigurasi LDA dengan 10 topik mencatat topic diversity tertinggi (0,8300), nilai tersebut mengindikasikan fragmentasi tema yang dapat menyulitkan interpretasi. BERTopic menunjukkan stabilitas dalam distribusi dan struktur topik. Topic diversity sebesar 0,7478 mengindikasikan bahwa model ini mampu menjaga keseimbangan antara keragaman kata dan konsistensi semantik dalam tiap topik. Pendekatan berbasis embedding yang digunakan BERTopic lebih sensitif terhadap keterkaitan makna implisit yang umum ditemukan dalam narasi konseling. Topik-topik yang terbentuk tidak hanya didasarkan pada kemunculan kata secara statistik, tetapi juga mempertimbangkan hubungan semantik kontekstual yang tidak eksplisit namun relevan secara psikologis dan klinis.

Dengan mempertimbangkan kompleksitas dan kekayaan ekspresi dalam data percakapan konseling, pendekatan BERTopic dinilai lebih unggul dalam menghasilkan topik-topik yang bermakna dan interpretatif. Walaupun skor koherensinya lebih rendah dalam metrik tradisional, hasil yang diperoleh lebih akurat dalam menangkap struktur semantik alami dari data. Angka yang didapat mendukung penggunaan BERTopic sebagai pendekatan yang lebih sesuai untuk eksplorasi tematik dalam domain kesehatan mental berbasis teks naratif.

V. KESIMPULAN

Berdasarkan hasil penelitian yang dilakukan, dapat disimpulkan bahwa: (1) Implementasi metode BERTopic berhasil dilakukan untuk mengekstraksi topik dari data konseling kesehatan mental. Metode diimplementasikan melalui pipeline terintegrasi yang mencakup tahapan prapemrosesan teks, representasi embedding, reduksi dimensi, klusterisasi, dan ekstraksi kata kunci. Proses ini diterapkan pada 994 dokumen teks tanpa pelabelan manual dan menunjukkan bahwa BERTopic mampu membentuk klaster semantik yang dapat divisualisasikan dan diinterpretasikan. (2) Dua konfigurasi preprocessing menghasilkan distribusi topik yang berbeda secara signifikan. Konfigurasi Lematisasi Aktif menghasilkan 22 topik dengan cakupan tema yang lebih granular dan distribusi dokumen yang merata. Konfigurasi Lematisasi Nonaktif hanya menghasilkan 14 topik dengan distribusi yang didominasi oleh satu klaster besar (328 dokumen), menunjukkan struktur topik yang lebih agregatif. (3) Evaluasi kuantitatif menunjukkan trade-off antara granularitas dan stabilitas. Konfigurasi L-ON S-ON mencatat skor *cosine similarity* sebesar 0,3907 dan rasio outlier 44,27%, sementara konfigurasi L-OFF S-ON memiliki similarity lebih tinggi sebesar 0,4338 dan rasio outlier lebih rendah sebesar 29,28%. Perbedaan tersebut menunjukkan bahwa lemmatisasi mendorong pemisahan topik yang lebih spesifik, namun dengan peningkatan jumlah dokumen yang tidak terklasifikasi. (4) Koherensi semantik pada topik-topik kecil cenderung lebih rendah. Rentang nilai *topic coherence* (c_v) pada kedua

konfigurasi berkisar antara 0,2504 hingga 0,6918. Beberapa topik kecil memiliki koherensi yang lemah, tetapi lima topik pada konfigurasi L-ON S-ON tercatat memiliki skor di atas 0,40, dimana hal ini menunjukkan bahwa granularitas tidak selalu mengurangi kualitas semantik secara signifikan. (5) Konfigurasi dengan lemmatisasi menghasilkan topik-topik yang lebih spesifik dan tajam, namun meningkatkan kemungkinan data tidak terklasifikasi. Tanpa lemmatisasi, model menjadi lebih inklusif terhadap variasi kata, namun cenderung membentuk super-klastr dengan interpretasi tematik yang kurang jelas. (6) Walaupun LDA mencatatkan skor c_V tertinggi (0,5222) dan NPMI terbaik (-0,0037) pada konfigurasi lima topik, topik-topik yang dihasilkan bersifat terlalu umum dan tidak terfokus secara tematik. Hal tersebut menyebabkan penurunan ketajaman dalam interpretasi psikologis. BERTopic, melalui pendekatan berbasis embedding semantik, lebih efektif dalam menangkap struktur tematik pada data konseling yang bersifat emosional, kontekstual, dan tidak terstruktur.

LAMPIRAN

LAMPIRAN A. Kata Kunci Lematisasi Aktif

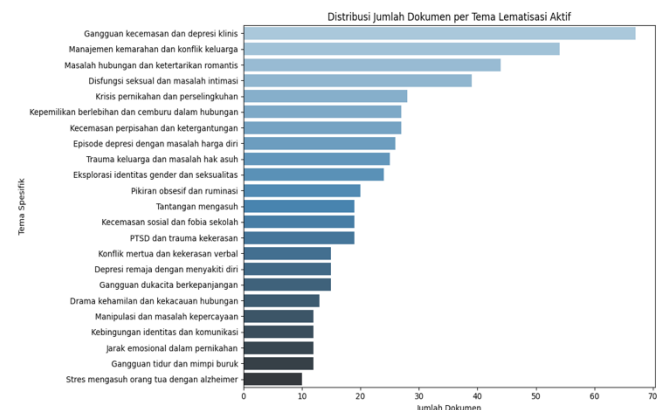
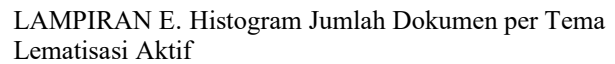
ID	Jumlah Dokumen	Kata Kunci	c_V
0	67	anxiety, depression, disorder, attack, diagnose, medication, stress, severe, therapist, panic	0,4011
1	54	anger, stepdaughter, make, angry, problem, thing, drama, exboyfriend, hurt, mood	0,2456
2	44	guy, date, love, year, know, think, exboyfriend, current, want, meet	0,383
3	39	sex, virgin, boyfriend, erection, want, wife, feel, partner, think, time	0,3214
4	28	wife, cheat, divorce, say, cheating, trust, want, love, marriage, work	0,2811
5	27	girlfriend, friend, relationship, trust, lot, boyfriend, really, thing, want, make	0,4995
6	27	break, love, boyfriend, dependent, breakup, busy, relationship, new, feel, talk	0,3761
7	26	like, feel, people, happy, make, anymore, song, enjoy, sad, talk	0,3212
8	25	child, year, infant, mother, exwife, relationship, daughter, leave, live, father	0,4566
9	24	girl, gay, like, gender, know, transgend, bisexual, look, male, feel	0,2582
10	20	thought, voice, think, hear, stop, remember, forever, heaven, head, live	0,2281
11	19	dad, mon, swear, brother, make, live, degree, tell, mother, sister	0,3795
12	19	people, school, big, popular, embarrass, social, uncomfortable, feel, class, public	0,3491
13	19	abuse, rape, emotionally, action, adult, attack, help, cope, dose, horrible	0,31
14	15	wife, mother, disagreement, usually, make, familyoriented, life, abuse, complain, wrong	0,3062
15	15	feel, start, like, sad, feeling, fear, store, selfharm, teenager, just	0,3029
16	15	job, pass, help, stop, grieve, stroke, awake, day, mother, need	0,2963

17	13	tell, want, confess, time, plan, love, later, money, woman, lunch	0,3477
18	12	wrong, boyfriend, texted, try, help, cat, girl, way, demeanour, shift	0,4389
19	12	message, answer, listen, people, understand, snapshot, instagram, respond, group, feel	0,5805
20	12	husband, listen, feel, marry, invest, drift, reconnected, uncured, apart, say	0,3629
21	12	nightmare, sleep, improve, asleep, solve, wake, lose, start, panic, fall	0,4288

LAMPIRAN B. Kata Kunci Lematisasi Nonaktif

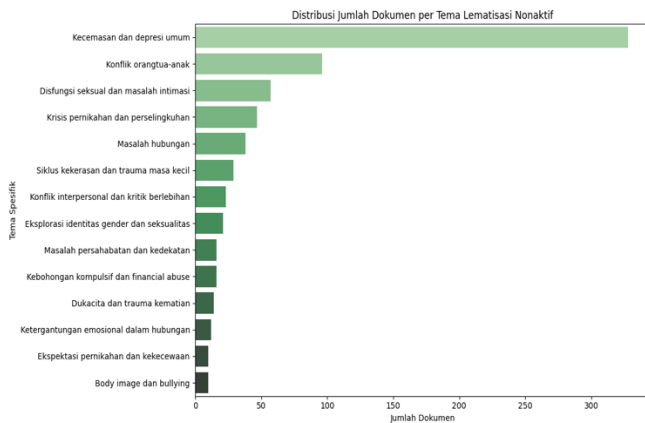
ID	Jumlah Dokumen	Kata Kunci	c_V
0	328	feel, like, anxiety, people, depression, know, just, want, life, think	0,4599
1	96	daughter, dad, mother, child, does, says, want, mon, just, house	0,3471
2	57	sex, men, drinking, told, want, husband, boyfriend, virgin, think, honest	0,3646
3	47	love, married, husband, cheated, time, said, wife, years, got, wants	0,5636
4	38	guy, love, want, relationship, dating, friend, years, feelings, know, did	0,4917
5	29	child, years, abused, having, children, remember, getting, abusive, daughter, louisiana	0,3013
6	23	comments, negative, frustrated, person, inconsiderate, alzheimer, character, sensitive, communication, fault	0,4513
7	21	girls, know, like, gender, transgender, gay, boy, tell, bisexual, accept	0,4842
8	16	girlfriend, friends, turned, really, friend, answer, month, says, kind, close	0,3961
9	16	lies, money, lying, friends, disappear, debt, borrow, boyfriend, just, games	0,6918
10	14	drugs, using, adopted, drug, overdosed, relative, birth, mother, old, father	0,2504
11	12	dependent, losing, does, better, sad, boyfriend, feelings, new, talking, love	0,3706
12	10	wanted, wedding, married, kids, proposal, know, cousin, disappointed, hair, jealous	0,3947
13	10	weight, parents, school, fat, bullied, skinny, think, girls, stop, invite	0,5621

LAMPIRAN C. Word Cloud BERTopic Lematisasi Aktif

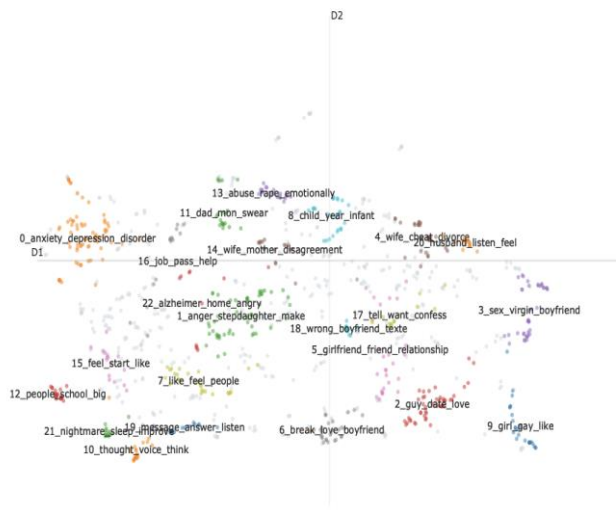


LAMPIRAN D. Word Cloud BERTopic Lematisasi Nonaktif

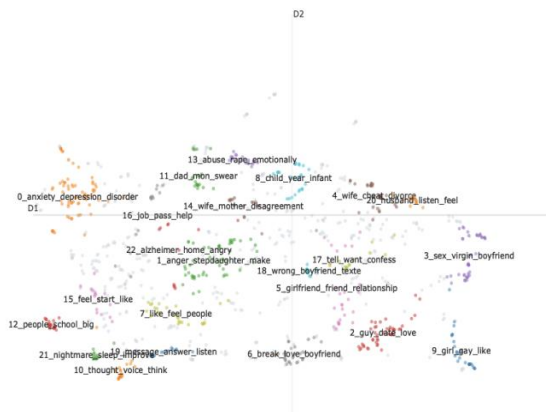
LAMPIRAN F. Histogram Jumlah Dokumen per Tema Lematisasi Nonaktif



LAMPIRAN G. Analisis Spasial Konfigurasi Lematisasi Aktif



LAMPIRAN H. Analisis Spasial Konfigurasi Lematisasi Nonaktif



UCAPAN TERIMA KASIH

Penulis juga diperkenankan menyampaikan ucapan terima kasih kepada sponsor penyedia dana penelitian. Penulis mengucapkan terima kasih kepada Agus Budi Raharjo, S.Kom., M.Kom., Ph.D., dan Prof. Dr. Diana Purwitasari, S.Kom., M.Sc., selaku dosen pembimbing yang telah memberikan bimbingan, arahan, serta masukan berharga selama proses penyusunan dan pelaksanaan penelitian ini. Penulis juga menyampaikan apresiasi kepada Departemen Teknik Informatika, Institut Teknologi Sepuluh Nopember, atas dukungan akademik dan fasilitas yang telah disediakan. Ucapan terima kasih turut disampaikan kepada rekan-rekan yang telah memberikan dukungan dan semangat selama proses pengerjaan artikel ini.

DAFTAR PUSTAKA

- [1] D. F. Santomauro *et al.*, “Global prevalence and burden of depressive and anxiety disorders in 204 countries and territories in 2020 due to the COVID-19 pandemic,” *The Lancet*, vol. 398, no. 10312, pp. 1700–1712, Nov. 2021, doi: 10.1016/S0140-6736(21)02143-7.
- [2] A. E. Wahdi, S. A. Wilopo, and H. E. Erskine, “122. The Prevalence of Adolescent Mental Disorders in Indonesia: An Analysis of Indonesia – National Mental Health Survey (I-NAMHS),” *Journal of Adolescent Health*, vol. 72, no. 3, p. S70, Mar. 2023, doi: 10.1016/j.jadohealth.2022.11.143.
- [3] Verified Market Research, “Global Online Therapy Services Market Size by Type (Cognitive Behavioral Therapy, Psychodynamic Therapy, and Personal Centered Therapy), by Application (Residential Use and Commercial), by Geographic Scope and Forecast,” 2024.
- [4] K. W. Lange, “Coronavirus disease 2019 (COVID-19) and global mental health,” *Global Health Journal*, vol. 5, no. 1, pp. 31–36, Mar. 2021, doi: 10.1016/j.glohj.2021.02.004.
- [5] O. Babalola, B. Ojokoh, and O. Boyinbode, “Comprehensive Evaluation of LDA, NMF, and BERTopic’s Performance on News Headline Topic Modeling,” *Journal of Computing Theories and Applications*, vol. 2, no. 2, pp. 268–289, Nov. 2024, doi: 10.62411/jcta.11635.
- [6] L. Ma *et al.*, “AI-powered topic modeling: comparing LDA and BERTopic in analyzing opioid-related cardiovascular risks in women,” *Exp Biol Med*, vol. 250, Feb. 2025, doi: 10.3389/ebm.2025.10389.
- [7] V. Narozhnyi and V. Kharchenko, “Semantic clustering method using integration of advanced LDA algorithm and BERT algorithm,” *INNOVATIVE TECHNOLOGIES AND SCIENTIFIC SOLUTIONS FOR INDUSTRIES*, no. 1 (27), pp. 140–153, Jul. 2024, doi: 10.30837/ITSSI.2024.27.140.
- [8] M. Grootendorst, “BERTopic: Neural topic modeling with a class-based TF–IDF procedure,” arXiv. Retrieved from <https://arxiv.org/abs/2203.05794>.

- [9] A. Abuzayed and H. Al-Khalifa, "BERT for Arabic Topic Modeling: An Experimental Study on BERTopic Technique," *Procedia Comput Sci*, vol. 189, pp. 191–194, 2021, doi: 10.1016/j.procs.2021.05.096.
- [10] A. Mutsaddi, A. Jamkhande, A. Thakre, and Y. Haribhakta, "BERTopic for Topic Modeling of Hindi Short Texts: A Comparative Study," Jan. 2025.
- [11] M. de Groot, M. Aliannejadi, and M. R. Haas, "Experiments on Generalizability of BERTopic on Multi-Domain Short Text," Dec. 2022
- [12] H. Rahimi, J. L. Hoover, D. Mimno, H. Naacke, C. Constantin, and B. Amann, "Contextualized Topic Coherence Metrics," May 2023
- [13] S. Syed and M. Spruit, "Full-Text or Abstract? Examining Topic Coherence Scores Using Latent Dirichlet Allocation," in *2017 IEEE International Conference on Data Science and Advanced Analytics (DSAA)*, IEEE, Oct. 2017, pp. 165–174. doi: 10.1109/DSAA.2017.61.
- [14] A. Hoyle, P. Goel, D. Peskov, A. Hian-Cheong, J. Boyd-Graber, and P. Resnik, "Is Automated Topic Model Evaluation Broken?: The Incoherence of Coherence," Jul. 2021.
- [15] M. Röder, A. Both, and A. Hinneburg, "Exploring the space of topic coherence measures," in *WSDM '15: Proceedings of the Eighth ACM International Conference on Web Search and Data Mining*, Feb. 2015.
- [16] Y. Dieng, F. Ruiz, and D. Blei, "Topic Modeling in Embedding Spaces," *Transactions of the Association for Computational Linguistics*, vol. 8, pp. 439–453, 2020.
- [17] Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative research in psychology*, 3(2), 77–101.
- [18] Guy-Evans, O. (2024). Narrative Analysis. *Simply Psychology*.
<https://www.simplypsychology.org/narrative-analysis.html>
- [19] Pattern Recognition (Psychology). (2025). Wikipedia.[https://en.wikipedia.org/wiki/Pattern_recognition_\(psychology\)](https://en.wikipedia.org/wiki/Pattern_recognition_(psychology))
- [20] Carlat, D. J. (1998). *The Psychiatric Interview: A Practical Guide*. Lippincott Williams & Wilkins.
- [21] Braun, V., & Clarke, V. (2013). *Successful Qualitative Research*. Sage.
- [22] WHO. (2021). *International Classification of Diseases 11th Revision (ICD-11)*. <https://icd.who.int>
- [23] Braun, V., & Clarke, V. (2013). *Successful Qualitative Research*. Sage.
- [24] Smith, A., et al. (2017). Evaluating visual representations for topic understanding and their effects on manually generated topic labels. *TACL*, 5, 1–16.
- [25] Constantinou, E., et al. (2023). Assessing topic stability and coherence with distribution-based visualizations. *Journal of Computational Social Science*.