

Skema Pendanaan: Madya

LAPORAN PENELITIAN



**EKSPLORASI TOPIK DAN FUNGSI KOMUNIKASI KESEHATAN
PUBLIK DI MEDIA SOSIAL MENGGUNAKAN PEMODELAN TOPIK
LDA**

TIM PENELITI

Ketua : Safitri Juanita, S.Kom, M.T.I (040043)
Anggota : M. Anif, S.Kom., M.Kom. (010012)

**FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS BUDI LUHUR
FEBRUARI 2026**

**HALAMAN PENGESAHAN
LAPORAN PENELITIAN**

Judul Penelitian : Eksplorasi Topik Dan Fungsi Komunikasi Kesehatan Publik Di Media Sosial Menggunakan Pemodelan Topik LDA

Bidang Penelitian : Data Mining

Ketua Peneliti

- a. Nama Lengkap : Safitri Juanita
- b. NIP/NIDN/ID-SINTA : 040043 / 0317098201 / 117415
- c. Jabatan Fungsional : Lektor
- d. Program Studi : Sistem Informasi
- e. Nomor HP : 083898928000
- f. Alamat e-mail : safitri.juanita@budiluhur.ac.id

Anggota Peneliti (1)

- a. Nama Lengkap : M. Anif
- b. NIP/NIDN/ID-SINTA : 010012/ 0321117001 / 6703012

Anggota Peneliti (2)

- a. Nama Lengkap : -
- b. NIP/NIDN/ID-SINTA : -

Mahasiswa (1)

- a. Nama Lengkap : Haniifaa Hafiizh
- b. NIM : 2212501098

Lama Penelitian : 6 bulan

Biaya Penelitian

- a. Sumber Universitas Budi Luhur : Rp. 15.000.000,-
- b. Sumber lain (sebutkan jika ada) : Rp. -


Jakarta, 27 Februari 2026

Mengetahui,
Dekan Fakultas Teknologi Informasi

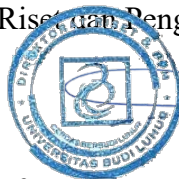


(Dr. Ir. Achmad Solichin, S.Kom., M.T.I)
NIP. 050023

Ketua Pelaksana


(Safitri Juanita, S.Kom, M.T.I)
NIP. 040043

Menyetujui,
Direktur Riset dan Pengabdian Kepada Masyarakat



(Prof. Dr. Ir. Prudensius Maring, M.A.)
NIP. 190043

RINGKASAN

Media sosial telah menjadi ruang utama bagi masyarakat untuk bertukar informasi dan pengalaman terkait kesehatan. Di platform seperti X (Twitter), pengguna non-medis sering mengajukan pertanyaan, berbagi pengalaman pribadi, memberi saran pengobatan, maupun menunjukkan empati terhadap sesama pengguna yang mengalami keluhan kesehatan. Aktivitas ini memperlihatkan terbentuknya wacana kesehatan daring yang kaya. Namun demikian, pola dan fungsi komunikasi masyarakat dalam berbagi informasi kesehatan belum teridentifikasi secara sistematis, sehingga diperlukan pendekatan komputasional untuk memetakan cara masyarakat berinteraksi dan berkomunikasi tentang isu kesehatan di media sosial. Penelitian ini bertujuan untuk mengidentifikasi topik-topik utama dalam komunikasi kesehatan yang dilakukan oleh pengguna non-medis di media sosial menggunakan metode Latent Dirichlet Allocation (LDA), menganalisis fungsi komunikasi yang muncul dalam percakapan kesehatan di media sosial, seperti *seeking information*, *providing information*, *sharing experience*, *giving advice*, *emotional support*, dan *referral to care*, mengkaji keterkaitan antara topik yang dihasilkan dari pemodelan topik dengan fungsi komunikasi yang muncul dalam setiap tema percakapan kesehatan, dan mengevaluasi pengaruh optimalisasi hiperparameter pada model LDA terhadap kualitas topik yang dihasilkan melalui pengukuran nilai *coherence score*. Tahapan penelitian meliputi pengumpulan data komentar publik bertema kesehatan dari platform X, prapemrosesan teks, pemodelan topik berbasis Latent Dirichlet Allocation (LDA), anotasi manual fungsi komunikasi berdasarkan kerangka teori komunikasi kesehatan, analisis proporsi fungsi, serta visualisasi hasil. Luaran wajib penelitian adalah satu artikel yang diterima pada jurnal Sinta 2 bernama Jurnal Nasional Pendidikan Teknik Informatika: JANAPATI, penerbit dari Universitas Pendidikan Ganesha untuk terbit di vol 15 no. 3 Desember 2026, dan luaran tambahan 1 sertifikat Hak Cipta (HKI).

Kata Kunci: media sosial; wacana kesehatan daring; analisis teks; latent dirichlet allocation (LDA); fungsi komunikasi

PRAKATA

Puji syukur dipanjatkan kehadirat Allah SWT, yang telah memberikan berkah dan hidayah-Nya sehingga kami memiliki kesempatan untuk melakukan penelitian dengan judul “Eksplorasi Topik Dan Fungsi Komunikasi Kesehatan Publik Di Media Sosial Menggunakan Pemodelan Topik LDA”.

Pada kesempatan ini tim penulis menyampaikan ucapan terima kasih kepada Universitas Budi Luhur yang telah mendanai penelitian ini melalui hibah internal perguruan tinggi. Ucapan terima kasih kepada Direktorat Penelitian dan Pengabdian Masyarakat dan Fakultas Teknologi Informasi yang telah memberikan bantuan dan dukungan terhadap penelitian ini sehingga peneliti dapat melakukan penelitian ini.

Tim penulis telah berusaha untuk menyempurnakan tulisan ini, namun sebagai manusia kami pun menyadari akan keterbatasan maupun kekhilafan dan kesalahan yang tanpa kami sadari. Oleh karena itu, saran dan kritik untuk perbaikan laporan penelitian ini akan sangat dinantikan.

Jakarta, Februari 2026

Penulis

DAFTAR ISI

HALAMAN PENGESAHAN	i
RINGKASAN.....	ii
PRAKATA	iii
DAFTAR ISI	iv
DAFTAR TABEL	v
DAFTAR GAMBAR.....	vi
DAFTAR LAMPIRAN	vii
BAB I PENDAHULUAN	1
1.1. Latar Belakang dan Rumusan Masalah	1
1.2. Pendekatan Pemecahan Masalah	2
1.3. <i>State-Of-The Art</i> dan Kebaruan	3
1.4. Peta Jalan Penelitian	4
BAB II METODE.....	6
BAB III HASIL PELAKSANAAN PENELITIAN	13
3.1. Tahap Pengumpulan Data.....	13
3.2. Tahap Persiapan Data	13
3.3. Tahap Proses Data	16
3.3.1. Membuat kamus kata.....	16
3.3.2. Implementasi Bags of Words (BOW)	17
3.3.3. Pemodelan LDA	17
1) Pendekatan LDA Tanpa Hiperparameter	17
2) Optimalisasi Model LDA Menggunakan Tuning Hiperparameter.....	25
3.3.4. Perbandingan LDA baseline dan LDA Optimasi	33
BAB IV KESIMPULAN DAN SARAN.....	34
4.1. Kesimpulan.....	34
4.2. Saran	35
DAFTAR PUSTAKA.....	36
LAMPIRAN	38

DAFTAR TABEL

Tabel 2.1. Contoh Dataset Penelitian	7
Tabel 2.2 Panduan kode untuk label kategori fungsi komunikasi.....	9
Tabel 2.3 Contoh Dokumen Validasi Ahli	11
Tabel 3.1. Kumpulan kata yang dinormalisasi	13
Tabel 3.2. Kumpulan kata yang tambahan Stopword.....	15
Tabel 3.3. Kamus Lemmatization.....	15
Tabel 3.4. Rekapitulasi Nilai Koherensi Berdasarkan Variasi Jumlah Topik pada model LDA tanpa Hiperparameter	18
Tabel 3.5. Interpretasi Topik menggunakan model LDA Berdasarkan Fungsi Komunikasi Pengguna Non-Medis	20
Tabel 3.6 Sepuluh Konfigurasi Hiperparameter Terbaik Berdasarkan Nilai Coherence pada Model LDA	26
Tabel 3.7 Interpretasi Topik LDA Berdasarkan Fungsi Komunikasi.....	28
Tabel 3.7 Perbandingan LDA dan Optimalisasi LDA.....	33

DAFTAR GAMBAR

Gambar 1.1. <i>State-of-the-art</i> Usulan Penelitian.....	3
Gambar 1.2. Peta Jalan Penelitian	4
Gambar 2.1 Tahapan Penelitian Eksplorasi Topik dan Fungsi Komunikasi Kesehatan Publik di Media Sosial Menggunakan Pemodelan Topik LDA	6
Gambar 2.2 Distribusi komentar dari pengguna non-medis pada aplikasi X per-tahun... 8	
Gambar 3.1 <i>Wordcloud</i> yang berisi kumpulan kata setelah pra-pemrosesan data	16
Gambar 3.2 Hubungan Jumlah Topik dengan Nilai Coherence Score pada Pemodelan LDA tanpa Hiperparameter	19
Gambar 3.3 Visualisasi Intertopic Distance Map dan Distribusi Kata pada Topik Menggunakan pyLDAvis	24
Gambar 3.4 Visualisasi Topik Menggunakan Word Cloud	25
Gambar 3.5 Perbandingan Nilai <i>Coherence</i> Tertinggi pada Setiap Jumlah Topik Hasil Tuning Hiperparameter LDA Menggunakan Grid Search	27
Gambar 3.6. Ringkasan Fungsi Komunikasi Menggunakan Optimasi LDA	30
Gambar 3.7. Distribusi Dokumen pada Setiap Topik Hasil Pemodelan LDA	31
Gambar 3.8. <i>Visualisasi Intertopic Distance Map</i> menggunakan pyLDAvis	32
Gambar 3.9. Visualisasi Word Cloud pada 8 Topik berdasarkan Optimalisasi LDA....	33

DAFTAR LAMPIRAN

Lampiran 1. Realisasi Penggunaan Anggaran	38
Lampiran 2. Surat Perjanjian Kontrak Penelitian	39
Lampiran 3. Catatan Harian.....	41
Lampiran 4. Artikel Ilmiah (draft).....	42
Lampiran 5. HKI (Terbit)	48

BAB I

PENDAHULUAN

1.1. Latar Belakang dan Rumusan Masalah

Media sosial kini menjadi ruang utama masyarakat untuk membahas isu kesehatan. Melalui platform seperti X (Twitter), Instagram, dan Facebook, pengguna dapat berbagi pengalaman, mengajukan pertanyaan, atau menanggapi informasi kesehatan. Penelitian menunjukkan bahwa media sosial berpotensi besar sebagai sarana promosi kesehatan yang efektif (1) karena mampu menjangkau populasi luas dengan biaya rendah dan penyebaran cepat. Selain itu, media sosial memfasilitasi akses informasi kesehatan, terutama bagi masyarakat yang sulit menjangkau layanan konvensional (2), serta menjadi wadah pertukaran pengetahuan, dukungan emosional, dan pembentukan opini publik.

Berbagai studi juga menegaskan bahwa media sosial meningkatkan kesadaran, mendorong perilaku hidup sehat, dan memperkuat partisipasi masyarakat dalam isu kesehatan (3). Di platform seperti X, masyarakat non-medis sering berbagi pengalaman, memberi saran, dan menunjukkan empati, membentuk ekosistem komunikasi kesehatan publik yang mencerminkan literasi digital masyarakat Indonesia. Dalam konteks ini, istilah kesehatan publik merujuk pada wacana kesehatan yang dibahas secara terbuka di ruang digital, sejalan dengan pendekatan public health communication yang menekankan partisipasi publik dalam membentuk persepsi dan perilaku kesehatan daring.

Komunikasi publik yang masif dan beragam sulit dipetakan secara manual, sehingga digunakan pendekatan komputasional berbasis Natural Language Processing (NLP). Metode Latent Dirichlet Allocation (LDA) mengekstraksi topik laten (4), sedangkan analisis fungsi komunikasi (5–7) menginterpretasikan makna sosial tiap topik. Dengan menggabungkan keduanya, penelitian ini bertujuan mengeksplorasi topik dan fungsi komunikasi kesehatan publik di media sosial, sebagaimana diterapkan dalam studi tentang pencegahan penyakit berisiko tinggi (8) dan wacana Covid-19 (9). Berdasarkan latar belakang tersebut, penelitian ini mengusulkan eksplorasi topik dan fungsi komunikasi kesehatan publik di media sosial menggunakan LDA. Berikut adalah rumusan masalah penelitian:

1. Bagaimana topik-topik utama dalam komunikasi kesehatan yang dilakukan oleh pengguna non-medis di media sosial dapat diidentifikasi menggunakan metode Latent Dirichlet Allocation (LDA)?
2. Fungsi komunikasi apa saja yang muncul dalam percakapan kesehatan di media sosial, seperti *seeking information*, *providing information*, *sharing experience*, *giving advice*, *emotional support*, dan *referral to care*?
3. Bagaimana keterkaitan antara topik yang dihasilkan oleh pemodelan topik LDA dengan fungsi komunikasi yang muncul dalam setiap tema percakapan kesehatan?
4. Bagaimana pengaruh optimalisasi hiperparameter pada model LDA terhadap kualitas topik yang dihasilkan berdasarkan nilai *coherence score*?

1.2. Pendekatan Pemecahan Masalah

Pendekatan penelitian ini terdiri atas dua tahap utama:

1. Pendekatan komputasional berbasis empiris

Pendekatan ini melibatkan pengumpulan dan analisis data teks dari media sosial X (Twitter) yang berisi percakapan publik non-medis tentang isu kesehatan. Data dikumpulkan melalui *web scraping* dengan kata kunci tertentu dan diproses melalui *preprocessing*. Hasilnya berupa dataset bersih yang dianalisis menggunakan *Latent Dirichlet Allocation* (LDA) untuk mengekstraksi topik laten dan mengidentifikasi tema utama komunikasi kesehatan di media sosial.

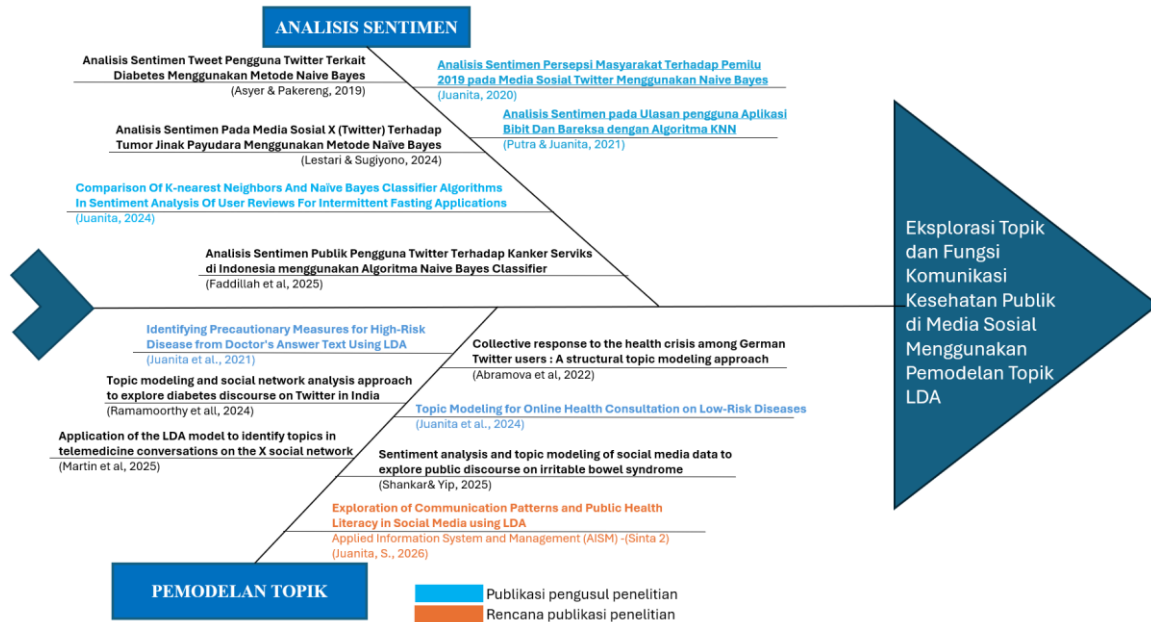
2. Pendekatan analitis *interpretative*

Pendekatan ini mengidentifikasi fungsi komunikasi dalam interaksi daring dengan menganotasi keluaran model LDA menggunakan panduan kode dari teori komunikasi kesehatan. Setiap topik dipetakan ke fungsi seperti informasi, pertanyaan, pengalaman, saran, empati, dan rujukan. Hasilnya memberikan gambaran pola komunikasi kesehatan publik berbasis data.

Kombinasi kedua pendekatan ini diharapkan memberikan gambaran empiris dan konseptual mengenai pola komunikasi kesehatan publik di media sosial serta menghasilkan kerangka analisis tematik-fungsional sebagai dasar penelitian lanjutan di bidang komunikasi kesehatan berbasis data.

1.3. State-Of-The Art dan Kebaruan

Gambar 1 menunjukkan *state-of-the-art* penelitian yang memetakan riwayat serta rencana publikasi pengusul.



Gambar 1.1. *State-of-the-art* Usulan Penelitian

Penjelasan pada Gambar 1.1 adalah sebagai berikut:

1. Analisis Sentimen

Berbagai penelitian terdahulu menunjukkan efektivitas machine learning dalam memahami opini publik di media sosial, khususnya di bidang kesehatan melalui analisis sentimen. Beberapa di antaranya menganalisis persepsi masyarakat terhadap diabetes di Twitter menggunakan Naïve Bayes (10), meneliti sentimen publik terhadap tumor jinak payudara dan pentingnya edukasi medis (11), serta menemukan peningkatan sentimen positif terhadap kanker serviks yang menunjukkan kesadaran publik meningkat (12).

Pengusul juga melakukan analisis sentimen di luar bidang kesehatan, seperti penelitian terhadap Pemilu 2019 menggunakan Naïve Bayes (13), ulasan aplikasi Bibit dan Bereksa dengan K-Nearest Neighbors (KNN) (14), serta studi pada aplikasi “Intermittent Fasting” menggunakan Naïve Bayes Classifier (NBC) dan KNN (15).

Keseluruhan penelitian tersebut menunjukkan bahwa analisis sentimen efektif untuk mengidentifikasi persepsi publik, namun masih terbatas pada polaritas (positif–negatif–netral). Penelitian ini mengusulkan pendekatan lebih komprehensif menggunakan

pemodelan topik LDA untuk mengeksplorasi tema laten dan fungsi komunikasi masyarakat non-medis di media sosial.

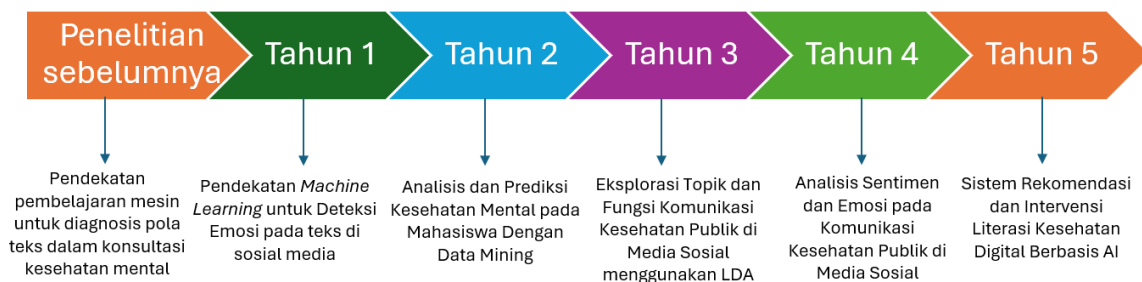
2. Pemodelan Topik

Selain analisis sentimen, penelitian berbasis pemodelan topik banyak digunakan untuk mengungkap pola komunikasi publik di bidang kesehatan. Abramova et al. (2022) (16) menggunakan Structural Topic Modeling (STM) pada lebih dari 800.000 tweet di Jerman terkait COVID-19 dan menemukan empat pola utama reaksi publik (understanding, action planning, hope, reassurance) serta pergeseran wacana dari individu ke komunitas. Ramamoorthy et al. (2024) (17) membandingkan algoritma LDA, NMF, BERTopic, dan Top2Vec pada percakapan diabetes di India dan mengidentifikasi delapan tema. Martin et al. (2025) (18) menerapkan LDA pada 150.000 komentar tentang telemedisin dan menemukan empat dimensi utama. Shankar & Yip (2025) (19) memadukan analisis sentimen dan pemodelan topik pada wacana Irritable Bowel Syndrome (IBS) di X.com dan menemukan delapan tema, termasuk gejala fisik, diet, dukungan sosial, dan kesehatan mental.

Dua penelitian pengusul sebelumnya juga menerapkan LDA pada komunikasi medis profesional, yakni Low-Risk Diseases (4) dan High-Risk Diseases (8), yang membuktikan efektivitas LDA dalam menemukan struktur semantik teks medis. Penelitian ini menjembatani konteks medis profesional dan komunikasi publik non-medis dengan menambahkan analisis fungsi komunikasi (pertanyaan, informasi, pengalaman, saran, empati, rujukan) untuk memahami pola wacana kesehatan digital secara komprehensif.

1.4. Peta Jalan Penelitian

Pada Gambar 2 merupakan peta jalan penelitian dari pengusul.



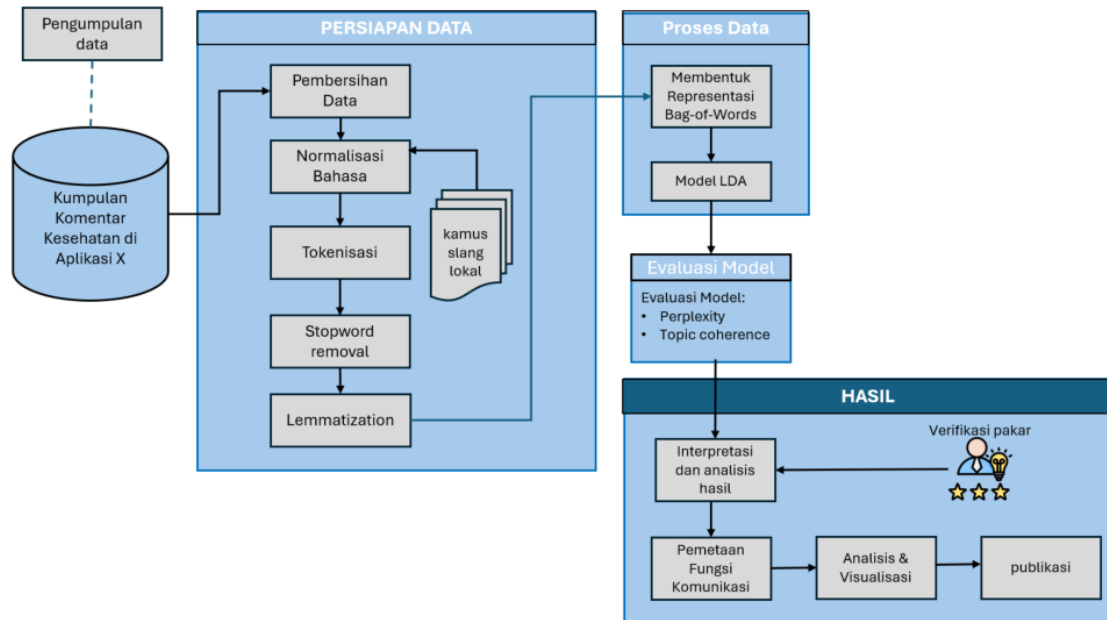
Gambar 1.2. Peta Jalan Penelitian

Gambar 1.2 menunjukkan peta jalan penelitian pengusul berdasarkan riwayat penelitian sebelumnya. Pada tahun 2024, penelitian difokuskan pada deteksi pola teks jawaban dokter terkait kesehatan mental dalam konsultasi daring menggunakan *machine learning* (20), serta analisis emosi (cemas dan depresi) di media sosial (21). Tahun berikutnya, penelitian diarahkan pada deteksi kesehatan mental mahasiswa menggunakan algoritma klasifikasi (22). Proposal tahun ini mengusulkan eksplorasi topik dan fungsi komunikasi kesehatan publik di media sosial menggunakan model LDA untuk memperluas konteks dari mahasiswa ke masyarakat umum. Tahun keempat direncanakan analisis sentimen dan emosi pada komunikasi kesehatan publik di media sosial, sedangkan tahun kelima pengembangan sistem rekomendasi literasi kesehatan digital berbasis *Artificial Intelligence* (AI)

BAB II METODE

2.1. Tahapan Penelitian

Pada Gambar 2.1 merupakan tahapan penelitian yang diusulkan.



Gambar 2.1 Tahapan Penelitian Eksplorasi Topik dan Fungsi Komunikasi Kesehatan Publik di Media Sosial Menggunakan Pemodelan Topik LDA

Berdasarkan Gambar 2.1 menampilkan tahapan penelitian yang terdiri dari lima tahap utama: (1) pengumpulan data, (2) persiapan data, (3) pemrosesan data, (4) evaluasi model, dan (5) interpretasi hasil. Setiap tahap dirancang sistematis untuk memastikan kualitas pemodelan topik yang merepresentasikan pola komunikasi publik di media sosial, sebagaimana ditunjukkan pada Gambar 3.

2.2. Pengumpulan data

Tahap ini mencakup pengumpulan data melalui *crawling* komentar pengguna non-medis di aplikasi X (Twitter) terkait topik kesehatan. Contoh dataset ditampilkan pada Tabel 2.1, terdiri atas unggahan pengguna yang dikumpulkan pada periode 2019–2025 dengan total 7,471 data.

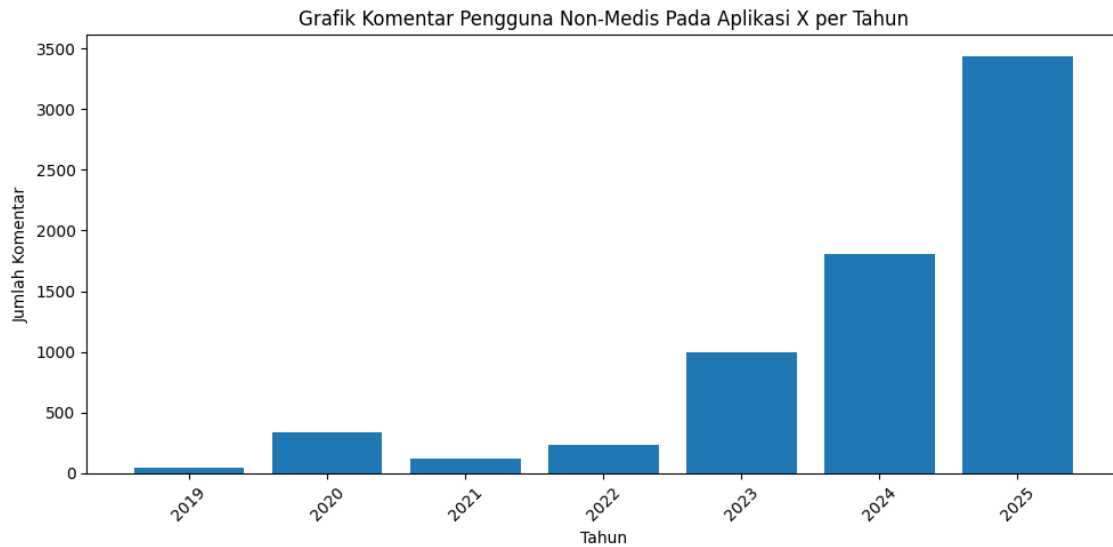
Tabel 2.1. Contoh Dataset Penelitian

No	Tanggal	Komentar
1	01 June 2019	MEMULAS Kadang makan rendang semalam atau leman tak masak,
2	20 July 2019	genetic yg menyebabkan dinding
3	16 September 2019	Terima kasih atas jawabannya sekarang aku mau meeting dulu sambil nangis.
	16 September 2019	Plis cek ke spog cari tau dulu sebabnya apa
5	16 September 2019	Tiduran di lantai nerima nasib sambil doa2 kalo2 mati (ini serius)
...
6977	30 December 2025	Ditambah Vit C, Vit D, Magnesium & Zinc jg bantu banget buat jaga imunitas tubuh.
6978	31 December 2025	Gongnya buah favorit aku ternyata juga salah satu sumber magnesium = Pisanggg Setelah tauu magensium ternyata memegang peranan penting, jadi makin senengg mam pisang wkwkw

2.3. Persiapan data

Tahap ini bertujuan untuk menghasilkan korpus teks yang bersih, terstandar, dan siap dianalisis oleh model *Latent Dirichlet Allocation (LDA)*. Pada tahap ini dibagi menjadi 2 langkah, meliputi:

- a. **Tahap 1** : menghapus data yang kosong dan duplikasi pada kolom komentar. Berdasarkan proses pada tahap 1 maka berhasil dihapus 490 baris, sehingga jumlah data menjadi 6981. Kemudian setelah dianalisis karena data tahun 2009, 2015, dan 2018 hanya memiliki 1 komentar maka dihapus. Sehingga total data penelitian adalah 6978 baris. Pada Gambar



Gambar 2.2 Distribusi komentar dari pengguna non-medis pada aplikasi X per-tahun

b. **Tahap 2** melakukan beberapa tahapan sebagai berikut:

- **Pembersihan Data:** Mengubah huruf besar menjadi huruf kecil, menghapus elemen non-teks seperti tanda baca, angka, URL, emoji, mention, dan hashtag untuk mengurangi noise.
- **Normalisasi Bahasa:** Mengubah penulisan informal ke bentuk baku (mis. *gk* → tidak, *bgt* → banget) menggunakan kamus slang lokal agar token konsisten.
- **Lematisasi:** Mengubah kata ke bentuk dasar, misalnya “memeriksa”, “diperiksa”, “pemeriksaan” menjadi “periksa”, agar kata seakar tidak dihitung ganda.
- **Penghapusan Stopword:** Menghapus kata umum seperti “yang”, “dan”, “untuk” guna memfokuskan analisis pada kata bermakna informatif.
- **Tokenisasi:** Memecah kalimat menjadi kata menggunakan word tokenizer Bahasa Indonesia agar batas kata dikenali akurat.

2.4. Proses data

Tahap ini berfokus pada transformasi teks menjadi representasi numerik dan penerapan model topik.

- a. **Bag-of-Words (BoW):** Setiap komentar direpresentasikan sebagai vektor frekuensi kata penting (*Bag-of-Words*) (20), yang menjadi input utama bagi algoritma LDA.
- b. **Pemodelan Topik LDA:** Model *Latent Dirichlet Allocation* mengekstraksi topik laten dari korpus dengan asumsi tiap dokumen memuat beberapa topik, dan tiap topik terdiri

atas distribusi kata. Parameter α dan β mengatur sebaran topik dan kata. Hasilnya berupa daftar topik serta kata kunci dominan sebagai dasar interpretasi fungsi komunikasi kesehatan publik.

2.5. Evaluasi Model

Tahap evaluasi dilakukan untuk menilai kualitas dan koherensi topik yang dihasilkan oleh model LDA menggunakan metode *Topic coherence*, metode ini digunakan untuk mengukur keterkaitan semantik antar kata dalam satu topik. Dalam penelitian ini digunakan metrik Cv atau NPMI (*Normalized Pointwise Mutual Information*) untuk menilai seberapa koheren kata-kata dalam setiap topik secara konseptual. Nilai *coherence* yang tinggi menunjukkan topik yang lebih mudah diinterpretasikan secara manusiawi.

2.6. Hasil

Pada bagian ini akan dibagi menjadi beberapa tahap yaitu:

a. Tahap Interpretasi dan Pemetaan Fungsi Komunikasi

Setelah model LDA menghasilkan sejumlah topik, dilakukan interpretasi manual terhadap setiap topik dengan meninjau kata kunci dominan dan komentar representatif. Setiap topik kemudian dipetakan ke dalam fungsi komunikasi kesehatan publik, seperti *pertanyaan, informasi, pengalaman, saran, empati, dan rujukan*.

b. Pemetaan Fungsi Komunikasi

Proses pemetaan fungsi komunikasi dilakukan secara otomatis menggunakan panduan kode (*codebook*) yang disusun berdasarkan adaptasi dari kerangka teori komunikasi kesehatan (5–7,23). Tabel 2.2 memuat panduan yang berisi definisi, batasan, dan contoh untuk setiap kategori fungsi komunikasi, seperti *pertanyaan, informasi, pengalaman, saran, empati, dan rujukan profesional*.

Tabel 2.2 Panduan kode untuk label kategori fungsi komunikasi

Kode	Fungsi Komunikasi	sumber	Definisi singkat	contoh
PERT	<i>Seeking Information</i> (Pertanyaan)	King & Hope (2013)	Meminta informasi, klarifikasi, atau solusi terkait gejala/pengobatan.	kenapa, gimana_cara, ada_yang_tahu, tolong_bantu

Kode	Fungsi Komunikasi	sumber	Definisi singkat	contoh
INFO	<i>Providing Information</i> (Informasi/Edukasi)	King & Hope (2013)	Memberi fakta, penjelasan, atau rujukan pengetahuan kesehatan.	menurut_kemenkes, gejala_umum, pencegahan, data_resmi
PENG	<i>Sharing Experience</i> (Pengalaman Pribadi)	Park et al. (2017); Hwang (2020)	Menceritakan pengalaman pribadi terkait kondisi atau terapi.	aku_pernah, dulu_kena, sembuh_setelah, efek_sampingku
SAR	<i>Giving Advice</i> (Saran/Rekomendasi)	Hwang (2020); Zhao (2014)	Memberi rekomendasi tindakan/produk untuk masalah kesehatan.	coba_pakai, oles_rutin, hindari_alergen, jaga_kebersihan
EMP	<i>Emotional Support</i> (Dukungan Emosional)	King & Hope (2013)	Menunjukkan empati dan dukungan emosional tanpa solusi teknis.	semoga_sembuh, sabar_ya, semangat, doa_terbaik
RUJ	<i>Referral to Care</i> (Rujukan Medis)	King & Hope (2013)	Menganjurkan pemeriksaan atau konsultasi ke tenaga medis.	periksa_ke_dokter, ke_puskesmas, spesialis_kulit, cek_lab

c. Analisis dan Evaluasi

Tahap analisis dan evaluasi dilakukan setelah proses anotasi fungsi komunikasi selesai. Tahap ini bertujuan untuk memahami pola tematik serta kecenderungan fungsi komunikasi dalam setiap topik yang dihasilkan oleh pemodelan *Latent Dirichlet Allocation* (LDA).

Analisis kuantitatif dilakukan dengan menghitung frekuensi dan proporsi kemunculan setiap fungsi komunikasi terhadap total komentar dalam satu topik. Pendekatan ini membantu mengidentifikasi fungsi komunikasi yang dominan serta membedakan pola interaksi pada berbagai tema kesehatan publik.

Selanjutnya, analisis kualitatif digunakan untuk menafsirkan konteks dan makna di balik sebaran fungsi tersebut. Komentar representatif dipilih untuk menjelaskan bagaimana masyarakat non-medis mengekspresikan pertanyaan, berbagi pengalaman,

memberikan saran, atau menunjukkan empati dalam diskusi kesehatan di media sosial. Evaluasi model dilakukan melalui dua aspek utama, yaitu:

- **Analisis dan evaluasi kualitas model LDA**

Kualitas model diukur menggunakan nilai *topic coherence* untuk menilai konsistensi struktur tematik yang terbentuk. Nilai ini digunakan sebagai dasar dalam menentukan jumlah topik yang paling representatif terhadap data.

- **Validitas anotasi fungsi komunikasi**

Validasi dilakukan dalam dua tahap, yaitu validasi otomatis dan validasi oleh pakar (*expert judgment*). Pada penelitian ini, proses validasi dilakukan oleh satu anotator ahli. Contoh dokumen yang digunakan dalam proses *expert judgment* ditampilkan pada Tabel 2.3.

Tabel 2.3 Contoh Dokumen Validasi Ahli

Komentar	Label Otomatis	Label Manual	Cocok?
junkfood tepung tepungan makanan manis berlebihann	<i>Giving Advice</i>	<i>Giving Advice</i>	Ya
ultra processed food macam sosis seblak frozen food lainnya cmiw ya	<i>Other</i>	<i>providing information</i>	Tidak
junk food pedes an	<i>Other</i>	<i>Other</i>	Ya
agak oot haid tidak teratur itu mkstdnya gmn ya soalnya saya sering haid tapi suka maju mundur seminggu dri kalender haid saya tapi slalu rutin haid tiap bulan kok	<i>Other</i>	<i>Seeking information</i>	Tidak

Tingkat kesesuaian antara label otomatis dan label manual diukur menggunakan *agreement score*, yang dihitung menggunakan rumus berikut:

$$Agreement\ Score = \frac{jumlah\ label\ yang\ cocok}{total\ label} \times 100\% \quad 2.1$$

Tahap analisis ini tidak hanya menghasilkan deskripsi statistik, tetapi juga memberikan pemahaman konseptual mengenai bagaimana masyarakat berinteraksi, bertukar informasi, dan membentuk wacana kesehatan di ruang digital.

2.7. Tahap *Deployment*

Membuat simpulan dari hasil eksperimen dan membuat 2 luaran yaitu:

- a. Luaran 1: Accepted di Jurnal Nasional Pendidikan Teknik Informatika: JANAPATI, penerbit dari Universitas Pendidikan Ganesha untuk terbit di vol 15 no. 3 Desember 2026 dengan judul “*Exploring the Topics and Functions of Public Health Communication on Social Media Using LDA Topic Modeling*”
- b. Luaran 2: HKI Granted (bersertifikat) Basis Data.

BAB III

HASIL PELAKSANAAN PENELITIAN

3.1. Tahap Pengumpulan Data

Tahap ini mencakup pengumpulan data melalui *crawling* komentar pengguna non-medis di aplikasi X (Twitter) terkait topik kesehatan. Pada tahap ini, kami berhasil mengumpulkan unggahan pengguna pada periode 1 Juni 2019 hingga 31 desember 2025 dengan total 7,471 data dengan atribut tanggal dan komentar. Contoh dataset ditampilkan pada Tabel 2.1

3.2. Tahap Persiapan Data

Pada tahap ini melakukan persiapan data agar dataset bersih sebelum proses berikutnya. Adapun tahapannya antara lain:

- a. Hapus baris yang kosong (tidak ada isinya). Pada tahap ini berhasil menghapus sebanyak 244 data, sehingga dataset pada tahap ini menjadi 7,227 baris.
- b. Hapus data yang duplikasi. Pada tahap ini, menghapus 490 baris yang memiliki data duplikasi, sehingga dataset pada tahap ini berkurang dan saat ini berjumlah sebanyak 6,981 baris.
- c. Filter komentar yang sedikit sebanyak 3 baris dan menghapus baris tersebut sehingga total dataset 6,978 baris.
- d. Case folding (merubah teks menjadi huruf kecil). Pada tahap ini merubah seluruh teks pada 6,978 baris menjadi huruf kecil.
- e. Bersihkan teks dengan menghapus beberapa hal berikut yaitu (URL, mention, hastags, karakter non-alfabet, tanda baca, spasi kosong, angka, superskrip dan subskrip), hapus baris yang kosong/tidak ada isinya).
- f. Normalisasi, pada tahap ini melakukan normalisasi pada kata-kata yang disingkat, atau istilah. Pada Tabel 3.1 menampilkan kumpulan kata yang dinormalisasi.

Tabel 3.1. Kumpulan kata yang dinormalisasi

bentuk "tidak"	"gk": "tidak", "ga": "tidak", "gak": "tidak", "ngga": "tidak", "nggak": "tidak", "engga": "tidak", "enggak": "tidak", "gabisa": "tidak bisa", "gabooleh": "tidak boleh", "gpp": "tidak apa apa",
----------------	--

kata umum	"btw": "by the way", "dll": "dan lain lain", "dsb": "dan sebagainya", "dr": "dari", "dgn": "dengan", "dg": "dengan", "dlm": "dalam", "org": "orang", "tp": "tapi", "tpi": "tapi", "trs": "terus", "trus": "terus", "hrs": "harus", "sm": "sama", "skrg": "sekarang", "skrng": "sekarang", "dlu": "dulu", "kalo": "kalau", "klo": "kalau", "apa2": "apa", "banget2": "sangat", "banget3": "sangat", "jg": "juga",
slang tindakan & fisik	"taro": "taruh", "teken": "tekan", "anget": "hangat", "geletakan": tiduran",
kata kerja slang	"ilang": "hilang", "bikin": "buat", "kasi": "kasih", "pake": "pakai", "pakein": "pakai", "olesin": "oles", "sampe": "sampai", "cuman": "cuma",
Kesehatan	"mempan": "efektif", "precaution": "tindakan pencegahan", "mens": "menstruasi", "menstruasiin": "menstruasi", "paracet": "paracetamol", "ranap": "rawat inap", "pusing2": "pusing", "tbtb": "tiba tiba",
kata ganti	"ak": "aku", "aq": "aku", "q": "aku", "gw": "saya", "gua": "saya", "elu": "kamu", "km": "kamu", "sy": "saya", "mrk": "mereka",
angka, waktu	"mg": "minggu", "brp": "berapa",
emosional / sosial	"plis": "tolong", "rekomen": "saran", "clear": "jelas", "evidencenya": "buktinya", "gws": "get well soon",
makan / minum	"mkn": "makan", "makan2": "makan", "minum2": "minum",
Tambahan	"kurangin": "kurangi", "ngurangin": "kurangi", "dikurangin": "kurangi", "tpi": "tapi", "pke": "pakai", "karna": "karena", "better": "lebih baik", "kayak": "mirip", "tau": "tahu", "trus": "kemudian", "terus", "anget": "hangat", "gatau": "tidak tahu", "gws": "semoga segera sembuh", "btw": "omong", "gabisa": "tidak bisa", "tetep": "tetap", "gtu": "begitu", "km": "kamu", "op": "operasi", "org": "orang", "hr": "hari", "kasi":

<p>"berikan", "jd": "jadi", "krna": "karena", "dlu": "dahulu", 'gatel': 'gatal', 'mending': 'lebih_baik', 'make': 'pakai', 'blm': 'belum', 'liat': 'lihat', 'dapat': 'dapat', 'sempet': 'sempat', 'ngerasain': 'merasakan', 'aslam': 'asam_lambung', 'kl': 'kalau', 'cepat': 'cepat',sg': 'langsung', 'emg': 'memang', 'kangker' : 'kanker',</p>
--

g. Hapus *Stopword* menggunakan *Library* NLTK dan menambah kata-kata yang tidak berhubungan dengan penelitian dan tidak mengandung arti. Pada Tabel 3.2 adalah kumpulan kata yang ditambahkan pada *Stopword*.

Tabel 3.2. Kumpulan kata yang tambahan *Stopword*

<p>'yang', 'dan', 'di', 'dari', 'dengan', 'untuk', 'pada', 'ke', 'dalam', 'ini', 'itu', 'ada', 'adalah', 'akan', 'telah', 'sudah', 'lagi', 'sekali', 'saja', 'sambil', 'atau', 'juga', 'bahwa', 'tersebut', 'oleh', 'agar', 'sehingga', 'namun', 'tetapi', 'melainkan', 'walau', 'meski', 'sementara', 'seraya', 'kali', 'ya', 'yaa', 'yaaa', 'yaah', 'deh', 'sih', 'tuh', 'eh', 'aja', 'doang', 'yg', 'ka', 'kak', 'kakk', 'lu', 'gitu', 'pake', 'kek', 'asa', 'way', 'mah', 'nya', 'x', 'gue', 'wkwk', 'nder', 'nderr', 'si', 'dll', 'tu', 'kah', 'dah', 'kl', 'ma', 'mom', 'nderrr'</p>
--

h. *Hybrid Stemming* dan *Lemmatization*.

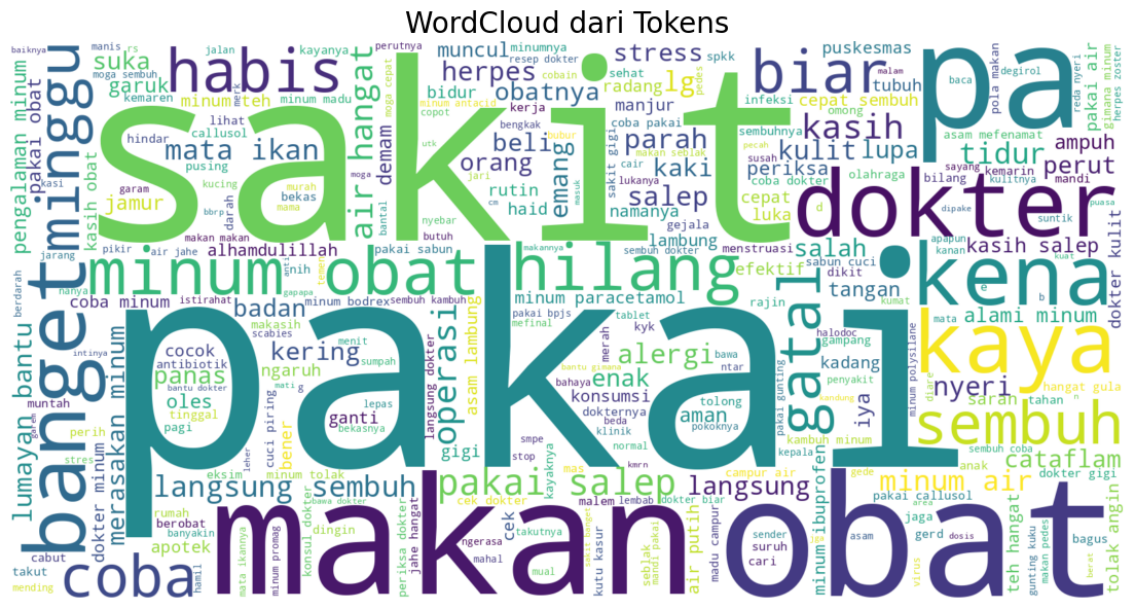
Keduanya menggunakan *Library* Sastrawi, *Stemming* untuk mengonversi kata berimbuhan Bahasa Indonesia ke bentuk dasarnya sebelum analisis topik, sedangkan *Lemmatization* untuk mengonversi kata berimbuhan Bahasa Indonesia ke bentuk dasarnya. Tahap ini bertujuan mengurangi variasi morfologis kata sehingga meningkatkan konsistensi representasi teks. Pada Tabel 3.3 menampilkan kamus tambahan untuk proses *Lemmatization*.

Tabel 3.3. Kamus *Lemmatization*

<p>"menjelaskan": "jelas", "dijelaskan": "jelas", "penjelasan": "jelas", "menanyakan": "tanya", "pertanyaan": "tanya", "ditanyakan": "tanya", "mengalami": "alami", "dialami": "alami", "penyebab": "sebab", "penyebabnya": "sebab", "dikurangi": "kurang", "mengurangi": "kurang", "pengurangan": "kurang", "kurangi": "kurang", "kurangin": "kurang", "ngurangi": "kurang",</p>

i. Setelah tahap hapus *stopword*, serta *hybrid stemming* dan *lemmatization*, maka dataset berkurang sehingga berjumlah 6,931 baris.

- j. Setelah itu dilakukan proses hapus duplikasi sehingga data berkurang kembali menjadi 6,894 baris.
- k. Tokenisasi, mengubah kumpulan kata yang sudah di Stemming dan lemmatization dengan cara memisahkan teks menjadi token.
1. Wordcloud hasil tokenisasi setelah pra-pemrosesan data terdapat pada Gambar 3.1.



Gambar 3.1 Wordcloud yang berisi kumpulan kata setelah pra-pemrosesan data

3.3. Tahap Proses Data

Setelah data menjadi lebih bersih, maka pada tahap ini, melakukan ekstraksi fitur menggunakan *Bags of Words* (BOW). Dengan tahapan sebagai berikut:

3.3.1. Membuat kamus kata

Pada tahap ini kamus kata dibentuk menggunakan filter maksimal muncul di 5 dokumen, dan maksimal proporsi dokumen adalah 0.5, berarti jika muncul di lebih dari 50% maka kata tersebut dibuang atau tidak masuk ke dalam kamus, Batasi jumlah kata 10,000. Filter tersebut diwakili dengan 1 baris perintah berikut :

```
dictionary_lemma.filter_extremes(no_below=5, no_above=0.5, keep_n=10000)
```

Pada tahap ini jumlah korpus adalah 6,894, sedangkan jumlah *term/vocab* setelah filter adalah 1,670.

3.3.2. Implementasi Bags of Words (BOW)

Pada tahap ini melakukan representasi kamus yang sudah terbentuk ke BOW. Dengan kode program sebagai berikut:

```
corpus = [dictionary.doc2bow(t) for t in texts_gensim]
```

contoh hasil dari proses BOW adalah sebagai berikut:

BoW doc-0 (id_kata, frekuensi): [(0, 1), (1, 1), (2, 1), (3, 1), (4, 1), (5, 1), (6, 1), (7, 1), (8, 1)]

3.3.3. Pemodelan LDA

Pemodelan topik pada penelitian ini dilakukan menggunakan metode Latent Dirichlet Allocation (LDA) dengan dua pendekatan, yaitu tanpa penyesuaian hiperparameter dan dengan penyesuaian hiperparameter. Pada tahap awal, model LDA dibangun menggunakan konfigurasi parameter standar dari pustaka Gensim dengan memanfaatkan Dictionary dari modul gensim.corpora untuk membangun kamus kata dan merepresentasikan dokumen dalam bentuk korpus *bag-of-words*.

Selanjutnya dilakukan proses tuning hiperparameter untuk memperoleh model dengan kualitas topik yang lebih baik. Parameter yang diuji meliputi alpha (symmetric, asymmetric), eta (auto, 0.3), passes (10, 20), dan iterations (200, 400). Setiap kombinasi parameter diuji pada beberapa jumlah topik menggunakan pendekatan grid search.

Kualitas model kemudian dievaluasi menggunakan nilai koherensi (*Coherence Score*) (*c_v*) melalui modul CoherenceModel. Nilai koherensi digunakan untuk mengukur tingkat keterkaitan semantik antar kata dalam suatu topik sehingga dapat membantu menentukan jumlah topik yang paling representatif terhadap kumpulan dokumen. Berikut ini adalah hasil pengujian dari kedua tahap.

1) Pendekatan LDA Tanpa Hiperparameter

Proses ini dilakukan dengan membangun beberapa model LDA menggunakan pustaka Gensim. Jumlah topik diuji secara bertahap mulai dari 2 hingga 10 topik dengan interval kenaikan satu topik. Untuk setiap jumlah topik yang diuji, model LDA dibangun menggunakan parameter *random_state* = 42 untuk menjaga konsistensi hasil eksperimen, *passes* = 10 untuk meningkatkan kualitas pelatihan model melalui iterasi berulang terhadap korpus, serta *alpha* = 'auto' agar model dapat menyesuaikan distribusi topik secara otomatis.

a) Penentuan Jumlah Topik Optimal

Setelah setiap model LDA tanpa hiperparameter terbentuk, dilakukan evaluasi menggunakan CoherenceModel dengan metrik C_v coherence. Metrik ini dipilih karena mampu mengukur tingkat koherensi topik berdasarkan kemiripan semantik kata-kata dalam topik menggunakan pendekatan berbasis *sliding window* dan *normalized pointwise mutual information* (NPMI).

Nilai koherensi dari setiap model kemudian disimpan untuk dibandingkan. Model dengan nilai *coherence* tertinggi dipilih sebagai model yang paling optimal karena menunjukkan bahwa topik-topik yang dihasilkan memiliki hubungan semantik yang lebih kuat dan lebih mudah diinterpretasikan.

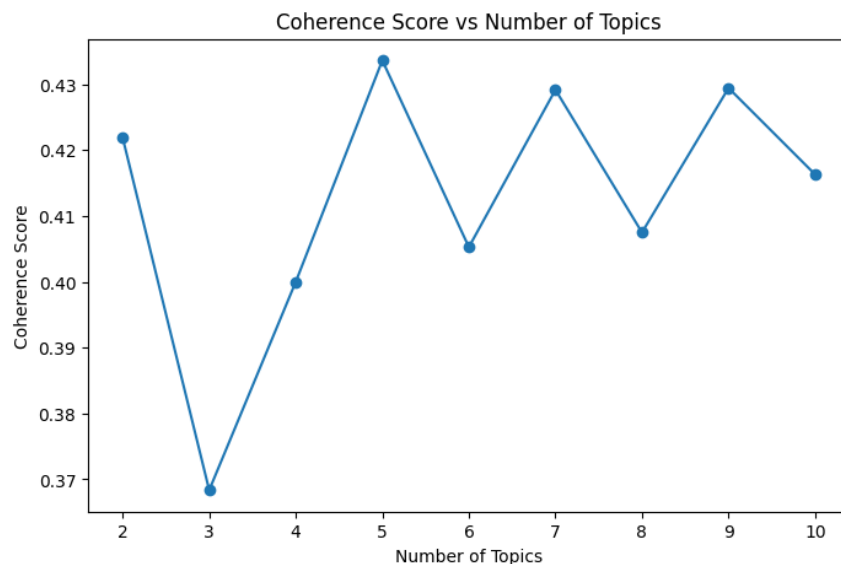
Tabel 3.4. Rekapitulasi Nilai Koherensi Berdasarkan Variasi Jumlah Topik pada model LDA tanpa Hiperparameter

No	Jumlah Topik	Nilai Koherensi	Interpretasi Kualitas Topik
1	2	0.4219	Koherensi cukup baik, namun jumlah topik masih terlalu sedikit untuk merepresentasikan variasi topik dalam korpus
2	3	0.3684	Nilai coherence menurun, menunjukkan topik yang dihasilkan kurang stabil
3	4	0.4000	Koherensi mulai meningkat namun masih belum optimal.
4	5	0.4336	Koherensi tertinggi, menunjukkan struktur topik paling representatif dan mudah diinterpretasikan.
5	6	0.4053	Koherensi menurun dibandingkan 5 topik.
6	7	0.4292	Koherensi kembali meningkat tetapi masih di bawah model optimal.
7	8	0.4075	Terjadi penurunan koherensi.
8	9	0.4294	Koherensi cukup tinggi namun masih sedikit lebih rendah dibandingkan 5 topik.
9	10	0.4163	Koherensi kembali menurun.

Berdasarkan hasil pengujian, nilai koherensi (*coherence*) tertinggi diperoleh pada jumlah topik 5 dengan nilai 0.4336. Hal ini menunjukkan bahwa model dengan lima topik menghasilkan struktur topik yang paling koheren dibandingkan dengan jumlah topik lainnya. Sementara itu, nilai *coherence* yang relatif tinggi juga diperoleh pada 7 topik (0.4292) dan 9 topik (0.4294), namun nilainya masih sedikit lebih rendah dibandingkan model dengan lima topik.

Secara umum, hasil ini menunjukkan bahwa peningkatan jumlah topik tidak selalu menghasilkan kualitas topik yang lebih baik. Setelah jumlah topik tertentu, nilai *coherence* cenderung mengalami fluktuasi atau bahkan menurun. Oleh karena itu, dalam penelitian ini model LDA dengan 5 topik dipilih sebagai model optimal untuk merepresentasikan struktur topik dalam korpus data.

Untuk memberikan gambaran yang lebih jelas mengenai pola perubahan nilai *coherence* terhadap jumlah topik, hasil pengujian tersebut juga divisualisasikan dalam bentuk grafik seperti yang ditunjukkan pada Gambar 3.2. Grafik tersebut memperlihatkan hubungan antara jumlah topik yang diuji dengan nilai *coherence* yang dihasilkan oleh model LDA.



Gambar 3.2 Hubungan Jumlah Topik dengan Nilai Coherence Score pada Pemodelan LDA tanpa Hiperparameter

Berdasarkan grafik pada Gambar 3.2, terlihat bahwa nilai *coherence* mengalami fluktuasi pada setiap variasi jumlah topik. Pada dua topik, nilai *coherence* berada pada

kisaran yang cukup baik, namun kemudian mengalami penurunan yang cukup signifikan pada tiga topik. Setelah itu, nilai *coherence* kembali meningkat pada empat topik dan mencapai nilai tertinggi pada lima topik. Kondisi ini menunjukkan bahwa model dengan lima topik menghasilkan struktur topik yang paling koheren dibandingkan dengan jumlah topik lainnya.

Setelah mencapai puncak pada lima topik, nilai *coherence* kembali mengalami penurunan pada enam topik, kemudian meningkat kembali pada tujuh topik dan sembilan topik. Meskipun demikian, nilai *coherence* pada variasi jumlah topik tersebut masih sedikit lebih rendah dibandingkan dengan nilai *coherence* pada lima topik. Hal ini menunjukkan bahwa penambahan jumlah topik tidak selalu meningkatkan kualitas topik yang dihasilkan oleh model, karena jumlah topik yang terlalu banyak dapat menyebabkan distribusi kata dalam topik menjadi kurang fokus.

Berdasarkan hasil evaluasi tersebut, dapat disimpulkan bahwa model LDA dengan lima topik memberikan nilai *coherence* tertinggi dan menghasilkan topik yang paling representatif dalam menggambarkan struktur tematik pada korpus data penelitian ini. Oleh karena itu, jumlah topik sebanyak lima dipilih sebagai jumlah topik optimal yang digunakan pada tahap analisis topik selanjutnya.

b) Hasil Topik dan Interpretasi menggunakan LDA

Hasil pemodelan LDA menghasilkan lima topik utama, dimana setiap topik direpresentasikan oleh sepuluh kata dengan bobot probabilitas tertinggi. Kata-kata tersebut menggambarkan tema atau pola komunikasi yang muncul dalam data teks konsultasi kesehatan. Pada Tabel 3.5 menampilkan interpretasi dari hasil pemodelan topik menggunakan LDA.

Tabel 3.5. Interpretasi Topik menggunakan model LDA Berdasarkan Fungsi Komunikasi Pengguna Non-Medis

Topik	Token dan Probabilitas	Interpretasi	Fungsi Komunikasi
Topik 0	makan (0.030), sakit (0.028), obat (0.024), operasi (0.017), stress (0.013), orang (0.013), pola	Keluhan kesehatan terkait pola makan	<i>Seeking Information</i> (Pertanyaan)

Topik	Token dan Probabilitas	Interpretasi	Fungsi Komunikasi
	(0.011), iya (0.010), lambung (0.010), radang (0.010)		
Topik 1	obat (0.042), pakai (0.039), dokter (0.030), salep (0.024), gigi (0.021), keluhannya (0.021), sakit (0.020), kasih (0.019), beli (0.017), minggu (0.014)	Informasi penggunaan obat	<i>Providing Information</i> (Informasi/Edukasi)
Topik 2	pakai (0.084), dokter (0.023), kena (0.021), salep (0.019), mata (0.018), gatal (0.017), hilang (0.016), herpes (0.015), kaya (0.015), kulit (0.013)	Pengalaman kesehatan pribadi	<i>Sharing Experience</i> (Pengalaman Pribadi)
Topik 3	minum (0.151), sembuh (0.056), langsung (0.048), dokter (0.046), bantu (0.035), kambuh (0.034), lumayan (0.033), coba (0.030), hangat (0.028), alami (0.027)	Rekomendasi pengobatan	<i>Giving Advice</i> (Saran/Rekomendasi)
Topik 4	minum (0.061), makan (0.046), air (0.036), jam (0.034), gimana (0.023), hangat (0.020), tidur (0.018), campur (0.013), darah (0.013), madu (0.012)	Dukungan kesehatan	<i>Emotional Support</i> (Dukungan Emosional)

Berdasarkan Tabel 3.5 menampilkan hasil pemodelan topik menggunakan LDA pada lima topik utama yang menggambarkan pola komunikasi pengguna. Setiap topik kemudian diinterpretasikan dengan mengacu pada kategori fungsi komunikasi kesehatan yang telah ditetapkan dalam penelitian ini, yaitu *Seeking Information*, *Providing Information*, *Sharing Experience*, *Giving Advice*, dan *Emotional Support*.

Topik pertama menunjukkan dominasi kata seperti makan, sakit, obat, dan lambung, yang mengindikasikan adanya diskusi mengenai keluhan kesehatan yang berkaitan

dengan pola makan dan gangguan pencernaan. Dalam konteks komunikasi, topik ini mencerminkan pengguna yang mengajukan pertanyaan atau mencari informasi terkait kondisi kesehatan yang mereka alami.

Topik kedua memperlihatkan kata-kata seperti obat, pakai, dokter, dan salep, yang menunjukkan adanya pengguna yang memberikan informasi mengenai penggunaan obat atau pengalaman berobat. Hal ini menunjukkan fungsi komunikasi *Providing Information*, dimana pengguna membagikan pengetahuan atau informasi yang mereka miliki kepada pengguna lain.

Pada topik ketiga, kata-kata seperti mata, gatal, herpes, dan kulit menunjukkan bahwa pengguna sering membagikan pengalaman pribadi terkait kondisi kesehatan tertentu. Topik ini menggambarkan fungsi komunikasi *Sharing Experience*, dimana pengguna menceritakan pengalaman mereka dalam menghadapi gejala atau pengobatan tertentu.

Topik keempat didominasi oleh kata-kata seperti minum, sembuh, coba, dan hangat, yang menunjukkan adanya rekomendasi atau saran mengenai cara mengatasi keluhan kesehatan berdasarkan pengalaman pribadi. Oleh karena itu, topik ini diinterpretasikan sebagai fungsi komunikasi *Giving Advice*.

Sementara itu, topik kelima menampilkan kata-kata seperti minum, makan, air, tidur, dan madu, yang berkaitan dengan anjuran menjaga pola hidup sehat. Dalam konteks komunikasi, topik ini mencerminkan bentuk dukungan sosial atau emosional, dimana pengguna memberikan dorongan atau perhatian kepada pengguna lain untuk menjaga kesehatan mereka.

Secara keseluruhan, hasil pemodelan topik menunjukkan bahwa komunikasi kesehatan yang terjadi pada pengguna aplikasi X tidak hanya berupa pertanyaan mengenai kondisi kesehatan, tetapi juga mencakup berbagai bentuk interaksi sosial seperti berbagi pengalaman, memberikan saran, serta memberikan dukungan emosional.

c) Diskusi

Berdasarkan hasil pemodelan topik menggunakan metode Latent Dirichlet Allocation (LDA), penelitian ini berhasil mengidentifikasi lima fungsi komunikasi yang muncul dalam percakapan pengguna aplikasi X terkait topik kesehatan, yaitu *Seeking Information*, *Providing Information*, *Sharing Experience*, *Giving Advice*, dan *Emotional Support*. Namun demikian, dari enam kategori fungsi komunikasi yang dirumuskan

dalam kerangka penelitian, fungsi komunikasi *Referral to Care* (Rujukan Medis) tidak muncul secara eksplisit dalam hasil topik yang dihasilkan oleh model.

Tidak munculnya fungsi komunikasi *Referral to Care* dapat dipahami dari karakteristik dataset yang digunakan dalam penelitian ini. Data yang dianalisis berasal dari komentar pengguna aplikasi X yang sebagian besar merupakan pengguna awam tanpa latar belakang medis. Dalam konteks komunikasi kesehatan di media sosial, pengguna awam cenderung lebih sering mengajukan pertanyaan, berbagi pengalaman pribadi, atau memberikan saran berdasarkan pengalaman pribadi, dibandingkan memberikan rujukan medis yang bersifat formal seperti merekomendasikan fasilitas kesehatan atau tenaga medis tertentu.

Selain itu, komunikasi pada platform media sosial sering kali bersifat informal dan berbasis pengalaman personal, sehingga rujukan untuk mencari layanan kesehatan profesional tidak selalu disampaikan secara eksplisit dalam bentuk rekomendasi langsung. Pengguna lebih sering memberikan saran umum seperti mencoba pengobatan tertentu atau menjaga pola hidup sehat, dibandingkan menyarankan konsultasi langsung dengan dokter atau fasilitas kesehatan tertentu.

Temuan ini menunjukkan bahwa pola komunikasi kesehatan di kalangan pengguna non-medis pada media sosial lebih didominasi oleh pertukaran pengalaman, informasi informal, serta dukungan sosial, daripada rujukan medis formal. Dengan demikian, hasil penelitian ini memberikan gambaran bahwa komunikasi kesehatan dalam komunitas daring lebih berperan sebagai ruang berbagi pengalaman dan dukungan sosial, sementara fungsi rujukan medis cenderung kurang menonjol dalam interaksi antar pengguna awam.

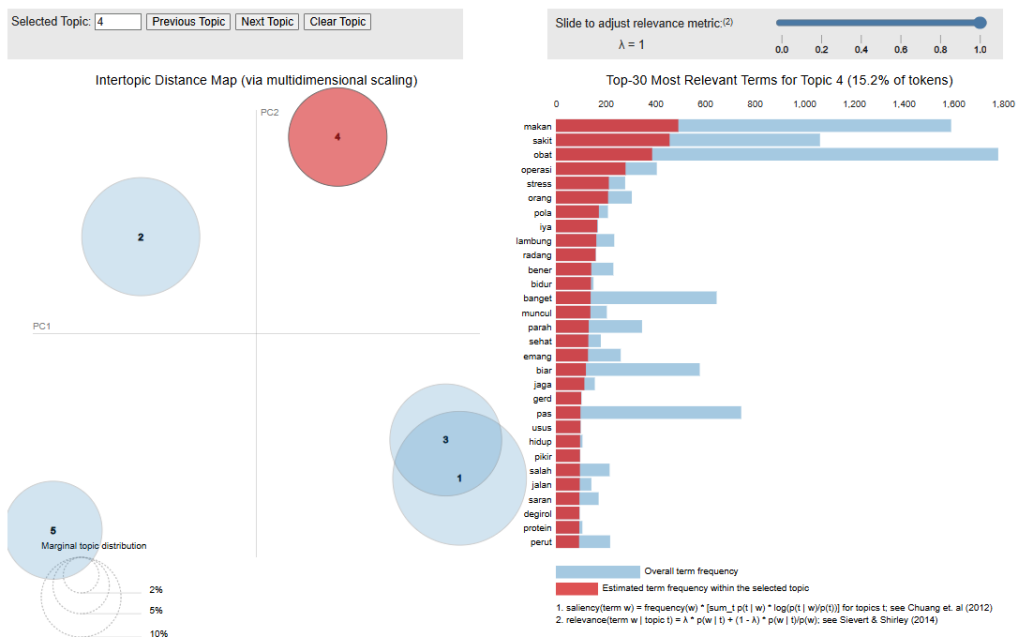
d) Visualisasi Distribusi Topik menggunakan PlyDAVIS

Untuk membantu memahami struktur topik yang dihasilkan oleh model Latent Dirichlet Allocation (LDA), dilakukan visualisasi menggunakan pyLDAvis seperti ditunjukkan pada Gambar 3.3. Visualisasi ini menampilkan dua komponen utama, yaitu peta jarak antar topik (*Intertopic Distance Map*) pada bagian kiri dan kata-kata paling relevan pada topik terpilih pada bagian kanan.

Pada peta jarak topik, setiap lingkaran merepresentasikan satu topik, dimana ukuran lingkaran menunjukkan proporsi topik dalam korpus, sedangkan jarak antar lingkaran menunjukkan tingkat kemiripan antar topik. Topik yang saling berjauhan menunjukkan

tema yang lebih berbeda, sementara topik yang berdekatan menunjukkan kemiripan konteks pembahasan.

Pada bagian kanan ditampilkan kata-kata yang paling relevan dalam topik yang dipilih. Kata-kata tersebut membantu mengidentifikasi tema utama dari topik yang terbentuk. Secara keseluruhan, visualisasi ini membantu memberikan gambaran yang lebih jelas mengenai distribusi topik serta kata-kata yang paling merepresentasikan setiap topik dalam dataset.



Gambar 3.3 Visualisasi Intertopic Distance Map dan Distribusi Kata pada Topik Menggunakan pyLDAvis

e) Visualisasi Kata Dominan pada Setiap Topik

Untuk memberikan gambaran visual mengenai kata-kata yang paling dominan pada setiap topik, penelitian ini juga menampilkan visualisasi word cloud seperti yang ditunjukkan pada Gambar 3.4. Pada visualisasi ini, ukuran kata mencerminkan tingkat frekuensi atau kontribusi kata dalam topik tertentu, dimana kata dengan ukuran lebih besar menunjukkan kemunculan yang lebih dominan dalam topik tersebut.

Word cloud ini membantu memperjelas interpretasi topik yang telah diidentifikasi sebelumnya, sehingga memudahkan dalam memahami konteks diskusi yang muncul dalam komentar pengguna aplikasi X terkait isu kesehatan. Dengan demikian, visualisasi ini berfungsi sebagai pelengkap dalam proses interpretasi hasil pemodelan topik.

Tabel 3.6 Sepuluh Konfigurasi Hiperparameter Terbaik Berdasarkan Nilai Coherence pada Model LDA

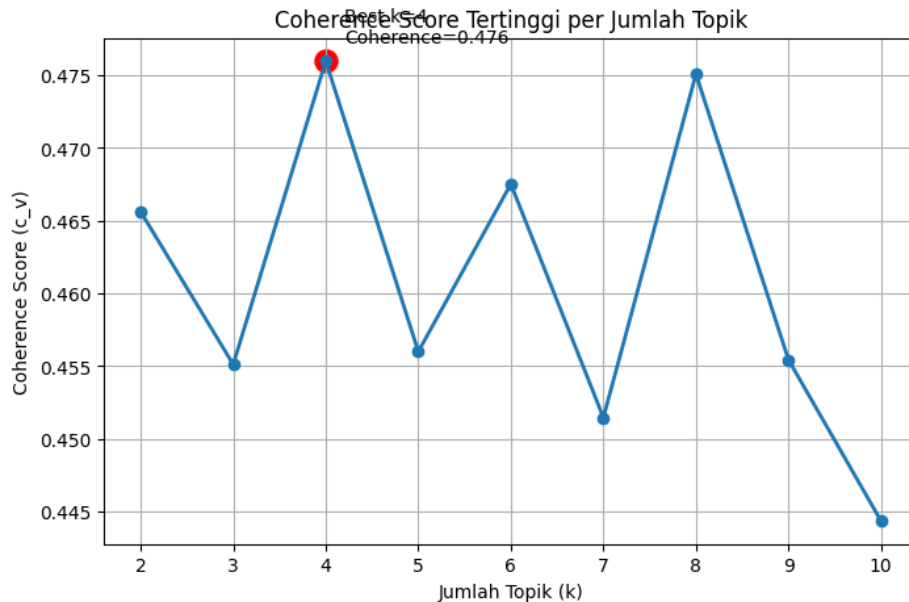
Peringkat	Jumlah Topik	Alpha	Eta	Passes	Iterations	Koherensi
1	4	asymmetric	0.3	10	200	0.4760
2	8	asymmetric	0.3	20	200	0.4751
3	8	asymmetric	0.3	10	400	0.4682
4	6	asymmetric	0.3	20	200	0.4675
5	6	asymmetric	0.3	20	400	0.4675
6	8	asymmetric	0.3	20	400	0.4660
7	6	asymmetric	0.3	10	200	0.4659
8	6	asymmetric	0.3	10	400	0.4659
9	2	asymmetric	auto	20	200	0.4656
10	2	asymmetric	auto	20	400	0.4656

Berdasarkan hasil tuning hiperparameter yang ditunjukkan pada Tabel 3.6, terlihat bahwa nilai *coherence score* yang dihasilkan oleh model LDA bervariasi tergantung pada kombinasi parameter yang digunakan. Berdasarkan hasil pengujian, nilai *coherence* tertinggi diperoleh pada konfigurasi jumlah topik 4 dengan parameter alpha = asymmetric, eta = 0.3, passes = 10, dan iterations = 200, yang menghasilkan nilai *coherence* sebesar 0.4760. Selain itu, konfigurasi dengan 8 topik juga menunjukkan nilai *coherence* yang relatif tinggi yaitu 0.4751, diikuti oleh konfigurasi 6 topik dengan nilai 0.4675.

Hasil ini menunjukkan bahwa pengaturan hiperparameter memiliki pengaruh terhadap kualitas topik yang dihasilkan oleh model LDA, dimana konfigurasi dengan alpha asymmetric dan eta bernilai 0.3 cenderung menghasilkan nilai *coherence* yang lebih tinggi dibandingkan konfigurasi lainnya.

Untuk memberikan gambaran yang lebih jelas mengenai pola perubahan nilai *coherence* terhadap jumlah topik, hasil pengujian tersebut juga divisualisasikan dalam bentuk grafik seperti yang ditunjukkan pada Gambar 3.5. Grafik tersebut memperlihatkan hubungan antara jumlah topik yang diuji dengan nilai *coherence* yang dihasilkan oleh model LDA.

Berdasarkan grafik pada Gambar 3.5, terlihat bahwa nilai *coherence* mengalami fluktuasi pada setiap variasi jumlah topik. Nilai *coherence* relatif tinggi diperoleh pada beberapa jumlah topik tertentu, terutama pada $k = 4$, $k = 6$, dan $k = 8$. Di antara seluruh konfigurasi yang diuji, nilai *coherence* tertinggi diperoleh pada jumlah topik 4 dengan nilai sekitar 0,476, yang ditandai pada grafik dengan titik merah.



Gambar 3.5 Perbandingan Nilai *Coherence* Tertinggi pada Setiap Jumlah Topik Hasil Tuning Hiperparameter LDA Menggunakan Grid Search

Temuan ini konsisten dengan hasil pengurutan konfigurasi hiperparameter sebelumnya, dimana kombinasi parameter $\alpha = \text{asymmetric}$, $\eta = 0.3$, $\text{passes} = 10$, dan $\text{iterations} = 200$ pada jumlah topik 4 menghasilkan nilai *coherence* tertinggi sebesar 0.4760. Berdasarkan hasil tuning hiperparameter, nilai *coherence* tertinggi diperoleh pada jumlah topik $k = 4$ dengan nilai 0.476. Namun demikian, nilai *coherence* pada $k = 6$ dan $k = 8$ juga menunjukkan nilai yang relatif tinggi dan tidak berbeda secara signifikan.

Mengingat tujuan penelitian ini adalah untuk mengidentifikasi fungsi komunikasi pengguna yang terdiri dari enam kategori, maka jumlah topik $k = 8$ menunjukkan nilai *coherence* yang sangat tinggi dan relatif mendekati nilai terbaik sehingga dipilih sebagai konfigurasi yang digunakan pada tahap analisis selanjutnya. Pemilihan ini mempertimbangkan tidak hanya nilai *coherence* yang dihasilkan oleh model, tetapi juga kesesuaian dengan kerangka konseptual penelitian sehingga interpretasi topik yang dihasilkan dapat merepresentasikan variasi fungsi komunikasi yang lebih komprehensif.

2. Hasil Topik dan Interpretasi menggunakan LDA

Tabel 3.7 menyajikan hasil pemodelan topik menggunakan metode Latent Dirichlet Allocation (LDA) dengan jumlah topik sebanyak delapan ($k = 8$). Setiap topik direpresentasikan oleh sejumlah token dengan nilai probabilitas tertentu yang menunjukkan tingkat kontribusi kata tersebut dalam membentuk topik. Setiap topik kemudian diinterpretasikan dan dipetakan ke dalam enam kategori fungsi komunikasi klinis, yaitu: *Seeking Information* (mencari informasi), *Giving Advice* (memberikan saran), *Providing Information* (menyediakan informasi), *Sharing Experience* (berbagi pengalaman), *Emotional Support* (dukungan emosional), dan *Referral to Care* (rujukan ke layanan perawatan). Berdasarkan analisis terhadap kata-kata dominan pada setiap topik, diperoleh gambaran mengenai pola komunikasi yang muncul dalam komentar pengguna aplikasi X terkait isu kesehatan.

Tabel 3.7 Interpretasi Topik LDA Berdasarkan Fungsi Komunikasi

Topik	Token dan Probabilitas	Interpretasi Topik	Fungsi Komunikasi
0	minum (0.032), makan (0.031), sakit (0.024), obat (0.023), tidur (0.018), asam (0.012), darah (0.011), nyeri (0.009), stress (0.009), lambung (0.008), perut (0.007), pola (0.007)	Diskusi mengenai keluhan kesehatan seperti sakit lambung, nyeri, atau masalah kesehatan yang berkaitan dengan pola makan dan kondisi tubuh. Pengguna biasanya menyampaikan keluhan dan kondisi yang dialami.	<i>Seeking Information</i>
1	minum (0.171), sembuh (0.051), hangat (0.050), air (0.050), langsung (0.042), lumayan (0.034), coba (0.032), kambuh (0.031), dokter (0.029), pengalaman (0.024), ibuprofen (0.020), paracetamol (0.018)	Percakapan yang menggambarkan pengalaman pribadi dalam mengatasi keluhan kesehatan, termasuk penggunaan obat atau minuman hangat yang dirasakan membantu pemulihan.	<i>Sharing Experience</i>

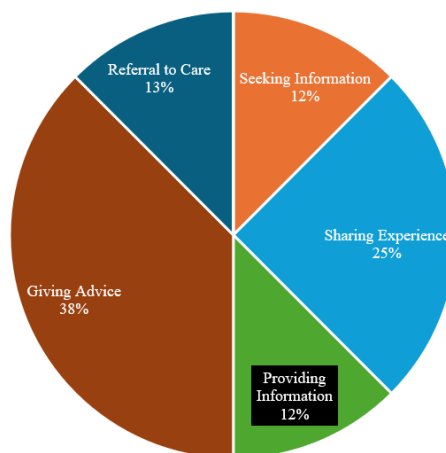
Topik	Token dan Probabilitas	Interpretasi Topik	Fungsi Komunikasi
2	dokter (0.088), obat (0.042), herpes (0.026), alergi (0.022), periksa (0.014), kulit (0.015), konsul (0.012), spkk (0.010), cek (0.010)	Diskusi mengenai penggunaan obat serta anjuran untuk memeriksakan diri atau berkonsultasi dengan dokter terkait kondisi kesehatan tertentu seperti alergi atau penyakit kulit.	<i>Referral to Care</i>
3	pakai (0.035), angin (0.029), gatal (0.024), cuci (0.019), panas (0.018), tangan (0.017), salep (0.013), sabun (0.012), hilang (0.010)	Pengguna memberikan saran mengenai cara mengatasi keluhan seperti gatal atau iritasi dengan penggunaan produk tertentu seperti sabun atau salep.	<i>Giving Advice</i>
4	pakai (0.113), salep (0.062), obat (0.027), sembuh (0.023), sabun (0.020), mandi (0.018), jamur (0.017), beli (0.017), apotek (0.013), oles (0.009)	Diskusi mengenai penggunaan obat atau salep untuk mengatasi masalah kulit seperti jamur, termasuk rekomendasi produk yang dapat digunakan.	<i>Giving Advice</i>
5	alami (0.091), makan (0.038), vit (0.020), kopi (0.019), kambuh (0.018), stres (0.012), pedes (0.012), bpjs (0.010), puskesmas (0.010)	Pembahasan mengenai faktor pemicu kondisi kesehatan seperti makanan, stres, atau gaya hidup serta pengalaman pengguna terkait gejala yang muncul.	<i>Sharing Experience</i>
6	makan (0.082), putih (0.047), buah (0.018), protein (0.014), sayur (0.013), hindar (0.013), daging (0.013), telur (0.011), anemia (0.011)	Diskusi mengenai pola makan sehat dan jenis makanan yang dianjurkan atau dihindari untuk menjaga kesehatan tubuh.	<i>Providing Information</i>

Topik	Token dan Probabilitas	Interpretasi Topik	Fungsi Komunikasi
7	pakai (0.106), mata (0.052), operasi (0.038), ikan (0.030), luka (0.014), bekas (0.013), kuku (0.011), plester (0.007)	Diskusi mengenai penanganan luka atau kondisi fisik tertentu serta penggunaan produk atau tindakan medis tertentu untuk mengatasi masalah tersebut.	<i>Giving Advice</i>

Hasil pemodelan topik menunjukkan bahwa percakapan kesehatan di antara pengguna aplikasi X didominasi oleh fungsi komunikasi berupa pemberian saran (*Giving Advice*) dan berbagi pengalaman pribadi (*Sharing Experience*). Hal ini menunjukkan bahwa pengguna cenderung berbagi pengalaman pribadi dan memberikan rekomendasi berdasarkan pengalaman tersebut dalam merespons keluhan kesehatan pengguna lain.

Selain itu, beberapa topik juga menunjukkan adanya fungsi *Providing Information*, dimana pengguna menyampaikan informasi terkait pola makan atau pengetahuan kesehatan dasar. Sementara itu, fungsi *Referral to Care* juga muncul melalui diskusi yang menyarankan pengguna untuk berkonsultasi dengan dokter atau memeriksakan diri ke fasilitas kesehatan tertentu.

Menariknya, fungsi *Emotional Support* tidak muncul secara dominan dalam hasil pemodelan topik, yang menunjukkan bahwa percakapan kesehatan pada dataset ini lebih berfokus pada pertukaran informasi praktis dan pengalaman terkait pengobatan dibandingkan dukungan emosional. Ringkasan distribusi fungsi komunikasi ditampilkan pada Gambar 3.6.

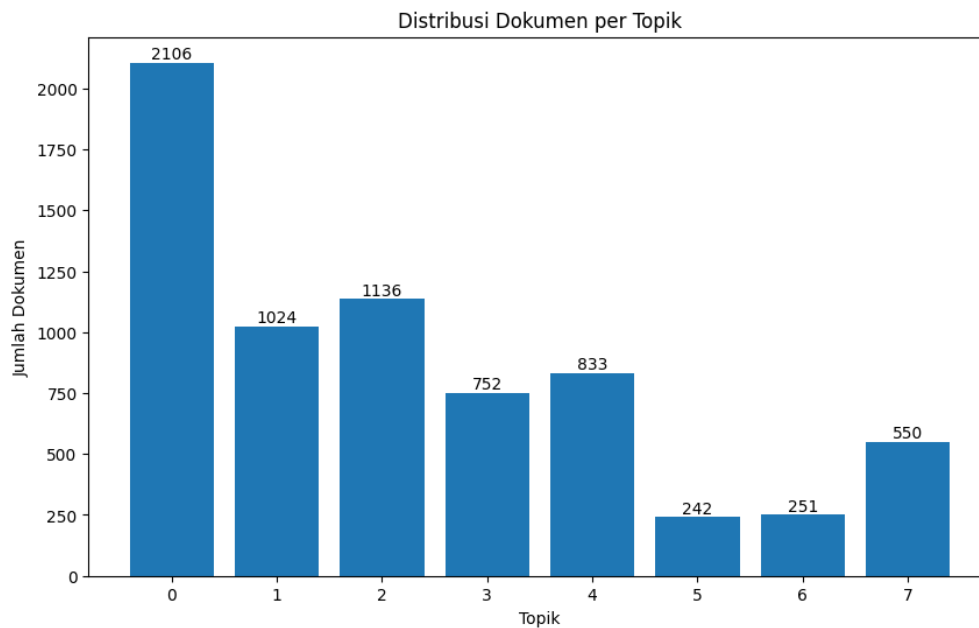


Gambar 3.6. Ringkasan Fungsi Komunikasi Menggunakan Optimasi LDA

3. Visualisasi data hasil Optimalisasi LDA

▪ Distribusi Dokumen pada Setiap Topik

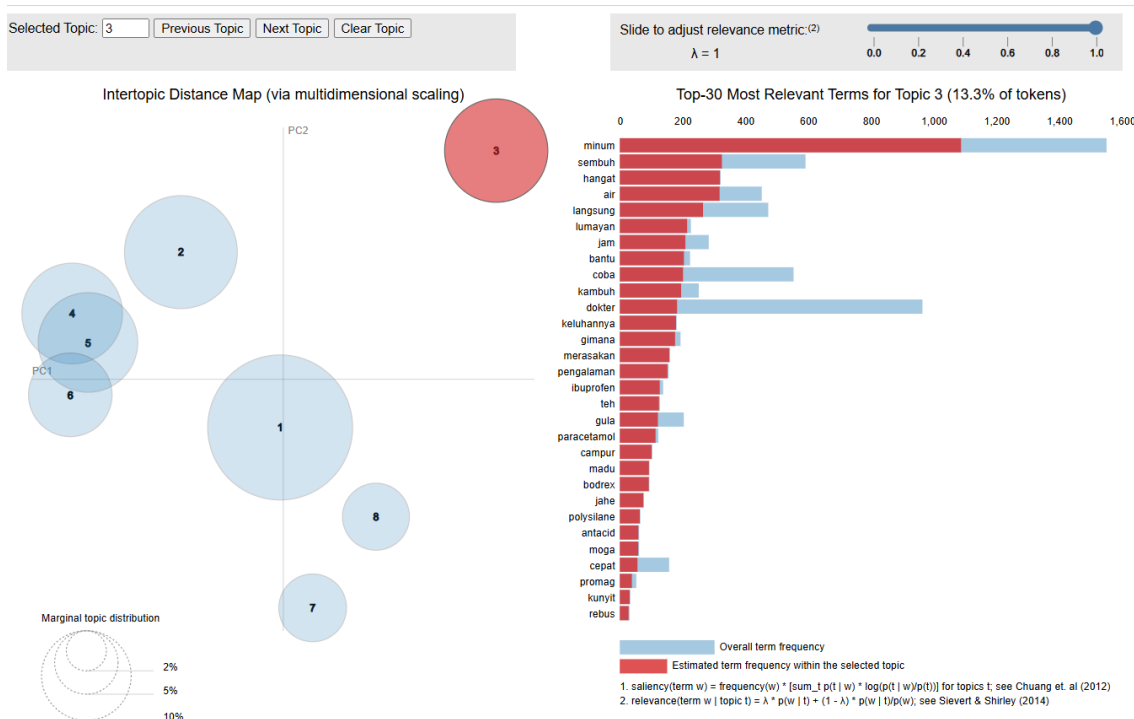
Gambar 3.7 menunjukkan distribusi jumlah dokumen pada setiap topik hasil pemodelan LDA yang telah dioptimalisasi. Hasil visualisasi menunjukkan bahwa beberapa topik memiliki jumlah dokumen yang lebih dominan dibandingkan topik lainnya. Hal ini menunjukkan bahwa tema tertentu lebih sering muncul dalam percakapan pengguna pada dataset penelitian. Distribusi dokumen ini juga memberikan gambaran mengenai tingkat kemunculan masing-masing topik dalam keseluruhan korpus data.



Gambar 3.7. Distribusi Dokumen pada Setiap Topik Hasil Pemodelan LDA

▪ Visualisasi Distribusi Topik menggunakan PlyDAVIS

Visualisasi intertopic distance menggunakan pyLDAvis digunakan untuk melihat hubungan antar topik yang dihasilkan oleh model LDA. Pada visualisasi ini, setiap lingkaran merepresentasikan satu topik, sedangkan ukuran lingkaran menunjukkan proporsi kemunculan topik dalam korpus data. Jarak antar lingkaran menunjukkan tingkat kemiripan antar topik, dimana topik yang lebih berjauhan menunjukkan perbedaan tema yang lebih jelas. Visualisasi ini membantu dalam memahami struktur topik serta hubungan antar topik yang muncul dalam percakapan pengguna. Pada Gambar 3.8 menampilkan *visualisasi intertopic distance map* menggunakan pyLDAvis.



Gambar 3.8. Visualisasi Intertopic Distance Map menggunakan pyLDAvis

Visualisasi Kata Dominan pada Setiap Topik

Gambar 3.9 menampilkan visualisasi *word cloud* untuk setiap topik yang dihasilkan oleh model Latent Dirichlet Allocation (LDA) dengan jumlah topik sebanyak delapan. Pada visualisasi ini, ukuran kata merepresentasikan bobot atau probabilitas kata dalam suatu topik, sehingga kata dengan ukuran lebih besar menunjukkan kontribusi yang lebih dominan dalam pembentukan topik tersebut.

Berdasarkan visualisasi tersebut, setiap topik memperlihatkan kumpulan kata yang merepresentasikan tema percakapan kesehatan yang berbeda. Beberapa kata dominan yang muncul antara lain *minum*, *obat*, *dokter*, *salep*, *makan*, dan *mata*, yang menunjukkan bahwa diskusi pengguna banyak berkaitan dengan pengalaman pengobatan, penggunaan obat atau produk kesehatan, serta pembahasan mengenai gejala dan penanganan penyakit.

Visualisasi ini membantu memperjelas struktur topik yang dihasilkan oleh model LDA dan mempermudah proses interpretasi tema percakapan dalam dataset, khususnya dalam mengidentifikasi pola komunikasi kesehatan yang dilakukan oleh pengguna non-medis pada platform X.

Visualisasi Word Cloud pada Setiap Topik LDA (K = 8)



Gambar 3.9. Visualisasi Word Cloud pada 8 Topik berdasarkan Optimalisasi LDA

3.3.4. Perbandingan LDA baseline dan LDA Optimasi

Model LDA dengan konfigurasi standar menghasilkan nilai *coherence* sebesar 0.4336 dengan jumlah topik optimal sebanyak lima topik. Selanjutnya dilakukan optimalisasi hiperparameter menggunakan pendekatan grid search untuk memperoleh konfigurasi model yang lebih optimal. Hasil eksperimen menunjukkan bahwa model LDA dengan hiperparameter optimal menghasilkan nilai *coherence* sebesar 0.4760, yang lebih tinggi dibandingkan model baseline. Hal ini menunjukkan bahwa proses optimalisasi hiperparameter mampu meningkatkan kualitas topik yang dihasilkan oleh model LDA. Tabel 3.7 menampilkan perbandingan kedua model.

Tabel 3.8 Perbandingan LDA dan Optimalisasi LDA

Model	Koherensi
LDA baseline	0.4336
LDA optimalisasi	0.4760

BAB IV

KESIMPULAN DAN SARAN

Berdasarkan analisis yang telah dilakukan terhadap permasalahan pada penelitian ini, maka dapat ditarik kesimpulan dan saran yang mungkin diperlukan untuk penelitian selanjutnya.

4.1. Kesimpulan

Berdasarkan hasil analisis yang telah dilakukan, penelitian ini menghasilkan beberapa kesimpulan sebagai berikut.

1. Identifikasi topik komunikasi kesehatan menggunakan LDA
Pemodelan topik menggunakan metode *Latent Dirichlet Allocation* (LDA) berhasil mengidentifikasi beberapa topik utama dalam percakapan kesehatan yang dilakukan oleh pengguna non-medis di media sosial. Topik-topik tersebut berkaitan dengan berbagai isu kesehatan, seperti keluhan penyakit, pengalaman penggunaan obat, pola makan dan gaya hidup sehat, serta penanganan gejala atau kondisi kesehatan tertentu. Hasil ini menunjukkan bahwa metode LDA efektif digunakan untuk mengekstraksi tema-tema utama dalam komunikasi kesehatan berbasis teks pada media sosial.
2. Fungsi komunikasi yang muncul dalam percakapan kesehatan
Analisis terhadap hasil pemodelan topik menunjukkan bahwa percakapan kesehatan di media sosial mencerminkan berbagai fungsi komunikasi, yaitu *seeking information, providing information, sharing experience, giving advice, dan referral to care*. Di antara fungsi komunikasi tersebut, percakapan didominasi oleh *sharing experience* dan *giving advice*, yang menunjukkan bahwa pengguna media sosial cenderung berbagi pengalaman pribadi serta memberikan saran terkait penanganan gejala atau penggunaan obat kepada pengguna lainnya.
3. Keterkaitan antara topik dan fungsi komunikasi
Hasil penelitian menunjukkan bahwa setiap topik yang dihasilkan oleh model LDA memiliki keterkaitan dengan fungsi komunikasi tertentu. Topik yang berkaitan dengan keluhan kesehatan cenderung menunjukkan fungsi *seeking information*, sementara topik yang membahas pengalaman penggunaan obat atau pengobatan lebih banyak mencerminkan fungsi *sharing experience*. Selain itu, topik yang berkaitan dengan penggunaan obat atau perawatan kesehatan sering kali

mencerminkan fungsi *giving advice*, sedangkan topik yang melibatkan konsultasi medis menunjukkan fungsi *referral to care*. Hal ini menunjukkan bahwa pemodelan topik dapat membantu mengungkap pola komunikasi kesehatan yang terjadi dalam diskusi pengguna media sosial.

4. Pengaruh optimalisasi hiperparameter terhadap kualitas model LDA
Hasil evaluasi model menunjukkan bahwa optimalisasi hiperparameter pada model LDA mampu meningkatkan kualitas topik yang dihasilkan. Model LDA *baseline* menghasilkan nilai *coherence* sebesar 0.4336, sedangkan model LDA yang dioptimalkan melalui proses tuning hiperparameter menghasilkan nilai *coherence* sebesar 0.476. Peningkatan nilai *coherence* ini menunjukkan bahwa optimalisasi hiperparameter berperan penting dalam menghasilkan topik yang lebih koheren dan mudah diinterpretasikan.

Secara keseluruhan, penelitian ini menunjukkan bahwa pendekatan pemodelan topik menggunakan LDA yang dikombinasikan dengan analisis fungsi komunikasi dapat memberikan pemahaman yang lebih komprehensif mengenai pola komunikasi kesehatan yang terjadi di media sosial, khususnya di antara pengguna yang tidak memiliki latar belakang medis.

4.2. Saran

Setelah ditarik kesimpulan, penulis juga memberikan beberapa saran-saran untuk penelitian selanjutnya sebagai berikut:

1. Mengembangkan pendekatan analisis komunikasi kesehatan di media sosial dengan metode pemodelan topik yang lebih mutakhir, seperti BERTopic atau neural topic modeling, yang mampu menangkap konteks semantik teks dengan lebih baik dibandingkan metode LDA.
2. Mengintegrasikan analisis sentimen atau analisis emosi untuk memahami bagaimana pengguna mengekspresikan kondisi emosional mereka ketika membahas masalah kesehatan di media sosial.
3. Memperluas sumber data dengan menggabungkan beberapa platform media sosial yang berbeda, sehingga dapat memberikan gambaran yang lebih komprehensif mengenai pola komunikasi kesehatan dalam ekosistem media sosial yang lebih luas.

DAFTAR PUSTAKA

1. University of Queensland C for RH, Johns Bloomberg Hopkins school of Public Health. Indonesia-National Adolescent Mental Health Survey (I-NAMHS): Laporan Penelitian [Internet]. Pusat Kesehatan Reproduksi. Yogyakarta; 2022. Available from: <https://qcmhr.org/outputs/reports/12-i-namhs-report-bahasa-indonesia>
2. Kohls E, Guenther L, Baldofski S, Brock T, Schuhr J, Rummel-Kluge C. Two years COVID-19 pandemic: Development of university students' mental health 2020–2022. *Front Psychiatry*. 2023;1122256(14):1–15.
3. Shafiee NSM, Mutalib S. Prediction of Mental Health Problems among Higher Education Student Using Machine Learning. *Int J Educ Manag Eng*. 2020;10(6):1–9.
4. Baba A, Bunji K. Prediction of Mental Health Problem Using Annual Student Health Survey: Machine Learning Approach. *JMIR Ment Heal*. 2023;10:1–23.
5. Maiyanti SI, Zayanti DA, Andriani Y, Suprihatin B, Desiani A, Salsabila A, et al. Perbandingan Klasifikasi Penyakit Kanker Paru-paru menggunakan Support Vector Machine dan K-Nearest Neighbor. *Process J Ilm Sist Informasi, Teknol Inf dan Sist Komput*. 2023;18(1):54–62.
6. Ramon E, Nazir A, Novriyanto N, Yusra Y, Oktavia L. Klasifikasi Status Gizi Bayi Posyandu Kecamatan Bangun Purba Menggunakan Algoritma Support Vector Machine (SVM). *J Sist Inf dan Inform*. 2022;5(2):143–50.
7. Ihsan TFF, Ramadhan I, Akbar DR, Ismanto E. Perbandingan Model Machine Learning Untuk Klasifikasi Deteksi Penyakit Jantung. *J Comput Sci Inf Technol*. 2025;6(2):199–205.
8. Ahuja R, Banga A. Mental stress detection in university students using machine learning algorithms. *Procedia Comput Sci* [Internet]. 2019;152:349–53. Available from: <https://doi.org/10.1016/j.procs.2019.05.007>
9. Sahlan F, Hamidi F, Misrat MZ, Adli MH, Wani S, Gulzar Y. Prediction of Mental Health Among University Students. *Int J Perceptive Cogn Comput* [Internet]. 2021;7(1):85–91. Available from: <https://journals.iium.edu.my/kict/index.php/IJPCC/article/view/225>
10. Malik SS. Anxiety , Depression and Stress prediction among College Students using Machine Learning Algorithms. In: *Second International Conference on Electrical,*

- Electronics, Information and Communication Technologies (ICEEICT). IEEE; 2023. p. 1–5.
11. Mahalakshmi G, Sujatha G. Predictions of College Students ' Mental Stress using Machine Learning Algorithms. In: 7th International Conference on Intelligent Computing and Control Systems (ICICCS). Madurai, India: IEEE; 2023. p. 288–95.
 12. Juanita S. Analisis Sentimen Persepsi Masyarakat Terhadap Pemilu 2019 Pada Media Sosial Twitter Menggunakan Naive Bayes. *J Media Inform Budidarma*. 2020;4(3):552–8.
 13. Putra ADA, Juanita S. Analisis Sentimen pada Ulasan pengguna Aplikasi Bibit Dan Bareksa dengan Algoritma KNN. *JATISI (Jurnal Tek Inform dan Sist Informasi)*. 2021;8(2):636–46.
 14. Juanita S, Adiyarta K, Syafrullah M. Sentiment analysis on E-Marketplace User Opinions Using Lexicon-Based and Naïve Bayes Model. In: International Conference on Electrical Engineering, Computer Science and Informatics (EECSI). Institute of Electrical and Electronics Engineers Inc.; 2022. p. 379–82.
 15. Sembiring ERB. Pengembangan Instrumen Pengukuran Kesehatan Mental [Internet]. Universitas Pendidikan Ganesha; 2023. Available from: <https://repo.undiksha.ac.id/14766/11/1911011024-LAMPIRAN.pdf>
 16. Li C. Preprocessing Methods and Pipelines of Data Mining: An Overview. In: Seminar Data Mining [Internet]. 2019. p. 1–7. Available from: <http://arxiv.org/abs/1906.08510>
 17. Wheeler EG. Age Periods of Human Life. *Adv Soc Sci Res J*. 2017;4(6):258–63.
 18. Sokolova M, Lapalme G. A systematic analysis of performance measures for classification tasks. *Inf Process Manag*. 2009;45:427–37.
 19. Sembiring ERB. Pengembangan Instrumen Pengukuran Kesehatan Mental (Mental Health Inventory) Setelah Pandemi Covid-19 pada Siswa SMP di Kota Singaraja [Internet]. Universitas Pendidikan Ganesha; 2023. Available from: <https://repo.undiksha.ac.id/14766/>

LAMPIRAN

Lampiran 1. Realisasi Penggunaan Anggaran

Dana disetujui: Rp 15.000.000,-

Jenis Pembelajaran	Komponen	Item	Kuantitas	Biaya Satuan	Total
Belanja Bahan	ATK	ATK	1 Paket	950.000	950.000
Belanja Bahan	Bahan penelitian (habis pakai)	Bahan penelitian (habis pakai)	1 Paket	1.500.000	1.500.000
Pengumpulan Data	Honor pembantu peneliti	Honor asisten peneliti (2 orang)	100 jam	25.000	2.500.000
Pengumpulan Data	FGD	FGD persiapan penelitian (2x3 orang)	6 paket	150.000	900.000
Pengumpulan Data	Transport	transport selama penelitian	6 paket	50.000	300.000
Pengumpulan Data	Konsumsi	konsumsi (2x3 orang)	6 paket	50.000	300.000
Analisis Data	Honor pengolah data	honor pengolah data (2 orang)	100 jam	25.000	2.500.000
Analisis Data	Honor narasumber	honor verifikasi pakar	1 orang	1.500.000	1.500.000
Sewa Peralatan	Peralatan penelitian	sewa collab google 3 bulan	1 paket	3.000.000	3.000.000
Pelaporan penelitian	Honor administrasi peneliti	honor administrasi peneliti	50 jam	25.000	1.250.000
Lainnya	Biaya pendaftaran HKI	Biaya pendaftaran HKI	1 paket	300.000	300.000

Lampiran 2. Surat Perjanjian Kontrak Penelitian



**UNIVERSITAS
BUDI LUHUR**

Kampus Pusat : Jl. Raya Ciledug - Petukangan Utara - Jakarta Selatan 12260
Telp : 021-5853753 (hunting), Fax : 021-5853489, <http://www.budiluhur.ac.id>

FAKULTAS TEKNOLOGI INFORMASI
FAKULTAS EKONOMI DAN BISNIS
FAKULTAS ILMU SOSIAL DAN STUDI GLOBAL
FAKULTAS TEKNIK
FAKULTAS KOMUNIKASI DAN DESAIN KREATIF

SURAT PERJANJIAN KONTRAK PENELITIAN Nomor A/UBL/DRPM/000/196/11/25

Pada hari ini, Rabu 05 November 2025 Semester Gasal Tahun Ajaran 2025/2026, kami yang bertandatangan di bawah ini:

1. **Prof. Dr. Ir. Prudensius Maring, M.A.**, selaku Direktur Riset dan Pengabdian Kepada Masyarakat Universitas Budi Luhur, selanjutnya disebut PIHAK PERTAMA.
2. **Safitri Juanita, S.Kom., M.T.I.**, selaku Peneliti selanjutnya disebut PIHAK KEDUA.

Kedua belah pihak menyatakan bersepakat untuk membuat perjanjian kontrak penelitian sebagai berikut:

Pasal 1 **Judul Penelitian**

PIHAK PERTAMA dalam jabatannya tersebut di atas, memberikan tugas kepada PIHAK KEDUA untuk melaksanakan penelitian yang berjudul: Eksplorasi Topik dan Fungsi Komunikasi Kesehatan Publik di Media Sosial Menggunakan Pemodelan Topik LDA

Pasal 2 **Personalia Penelitian**

Peneliti Utama : Safitri Juanita, S.Kom., M.T.I.
Anggota Peneliti : M. Anif, S.Kom., M.Kom.

Pasal 3 **Waktu dan Biaya Penelitian**

1. Waktu penelitian adalah 5 bulan, terhitung sejak tanggal 08 September 2025 sampai dengan 08 Februari 2026.
2. Biaya pelaksanaan penelitian ini dibebankan pada Yayasan Pendidikan Budi Luhur Cakti Tahun 2025 dengan nilai kontrak sebesar Rp 15,000,000.00 (lima belas juta rupiah)

Pasal 4 **Cara Pembayaran**

Pembayaran biaya penelitian diberikan secara bertahap, sebagai berikut:

1. Tahap pertama sebesar 50% dari nilai kontrak, setelah surat perjanjian kontrak penelitian ini ditandatangani oleh kedua belah pihak.
2. Tahap kedua sebesar 50% dari nilai kontrak, setelah PIHAK KEDUA menyerahkan Laporan Hasil Penelitian kepada PIHAK PERTAMA.

Pasal 5 **Keaslian Penelitian dan Ketidakterikatan dengan Pihak Lain**

1. PIHAK KEDUA bertanggungjawab atas keaslian judul penelitian sebagaimana disebutkan dalam Pasal 1 Surat Perjanjian Kontrak Penelitian ini (bukan duplikat/jiplakan/plagiat) dari penelitian orang lain.
2. PIHAK KEDUA menjamin bahwa judul penelitian tersebut bebas dari ikatan dengan pihak lain atau tidak sedang didanai oleh pihak lain.

KAMPUS ROXY : Pusat Niaga Roxy Mas Blok E.2 No. 38-39 Telp : 021-6328709 - 6328710, Fax : 021-6322872
KAMPUS SALEMBA : Sentra Salemba Mas Blok S-T, Telp : 021-3928688 - 3928689, Fax : 021-3161636

3. PIHAK KEDUA menjamin bahwa judul penelitian tersebut bukan merupakan penelitian yang SEDANG ATAU SUDAH selesai dikerjakan, baik didanai oleh pihak lain maupun oleh sendiri.
4. PIHAK PERTAMA tidak bertanggungjawab terhadap tindakan plagiat yang dilakukan oleh PIHAK KEDUA.
5. Apabila dikemudian hari diketahui ketidakbenaran pernyataan ini, maka kontrak penelitian DINYATAKAN BATAL, dan PIHAK KEDUA wajib mengembalikan dana yang telah diterima kepada Yayasan Pendidikan Budi Luhur Cakti sebagai pemberi dana.

Pasal 6 Monitoring Penelitian

1. PIHAK PERTAMA berhak untuk:
 - a. Melakukan pengawasan administrasi, monitoring, dan evaluasi terhadap pelaksanaan penelitian.
 - b. Memberikan sanksi jika dalam pelaksanaan penelitian terjadi pelanggaran terhadap isi perjanjian oleh peneliti.
 - c. Bentuk sanksi disesuaikan dengan tingkat pelanggaran yang dilakukan.
2. Pemantauan kemajuan penelitian dikoordinasikan oleh PIHAK PERTAMA.
3. Pelaksanaan kemajuan penelitian dilaksanakan pada tanggal 13 Desember 2025.
4. Format Laporan Kemajuan dan teknis pelaksanaannya diatur oleh PIHAK PERTAMA.

Pasal 7 Laporan Akhir Penelitian

PIHAK KEDUA wajib menyerahkan laporan akhir dalam bentuk softcopy, paling lambat tanggal 07 Februari 2026.

Pasal 8 Sanksi

Segala kelalaian baik disengaja maupun tidak, sehingga menyebabkan keterlambatan menyerahkan laporan hasil penelitian dengan batas waktu yang telah ditentukan akan mendapatkan sanksi sebagai berikut:

1. Tidak diperbolehkan mengajukan usulan penelitian pada semester berikutnya bagi ketua dan anggota peneliti.
2. PIHAK KEDUA diberikan kesempatan perpanjangan waktu penelitian selama 2 (dua) minggu sampai dengan tanggal 21 Februari 2026.
3. Jika setelah masa perpanjangan tersebut PIHAK KEDUA tidak dapat menyelesaikan penelitiannya, PIHAK KEDUA diwajibkan mengembalikan dana yang sudah diterima kepada Yayasan Pendidikan Budi Luhur Cakti dengan cara mengembalikan tunai kepada PIHAK PERTAMA.

Lampiran 3. Catatan Harian

No	Tanggal	Kegiatan
1.	8/09/2025	Catatan: <ul style="list-style-type: none"> • Studi Literatur • Membuat Proposal Penelitian
2.	9/9/2025	Catatan: <ul style="list-style-type: none"> • Crawling dataset
3.	10/11/2025	Catatan: <ul style="list-style-type: none"> • Mengolah dataset hasil pengumpulan data • Membersihkan dataset, menghapus atribut yang tidak berhubungan dengan penelitian • Melakukan visualisasi data
4.	7/12/2025	Catatan: <ul style="list-style-type: none"> • Membuat laporan kemajuan
5.	13/12/2025	Catatan: <ul style="list-style-type: none"> • Mengupload laporan kemajuan
6.	15/12/2025	Catatan: <ul style="list-style-type: none"> • Membersihkan dataset, menghapus atribut yang tidak berhubungan dengan penelitian • Membuat kode program untuk Implementasi metode LDA
7.	20/12/2025	<ul style="list-style-type: none"> • Kumpulkan dataset kembali • Proses dataset dengan metode LDA
8.	5/01/2025	Catatan: <ul style="list-style-type: none"> • Interpretasi hasil dari model LDA • Membuat kode program untuk anotasi otomatis fungsi komunikasi
9.	20/01/2025	Catatan: <ul style="list-style-type: none"> • Meminta validasi dari ahli untuk verifikasi hasil anotasi otomatis fungsi komunikasi
10.	25/01/2025	Catatan: <ul style="list-style-type: none"> • Analisis hasil validasi ahli dan hitung score analisis
11.	10/02/2025	Catatan: <ul style="list-style-type: none"> • Merapikan semua hasil eksperimen menggunakan Microsoft excel
12.	15/02/2025	Catatan: <ul style="list-style-type: none"> • Membuat draft dan submit ke JANAPATI
13.	19/02/2026	Catatan: Membuat draft HKI
14.	21/02/2026	Catatan: Membuat laporan penelitian

Topic Modeling and Communication Function Analysis of Non-Medical Users in Online Health Discussions

Safitri Juanita^{1*}, M. Anif²

^{1,2} Sistem Informasi, Fakultas Teknologi Informasi, Universitas Budi Luhur, Jakarta, Indonesia

E-mail: safitri.juanita@budiluhur.ac.id*

*corresponding author

Abstract— Online health discussions are increasingly used by the public as a space to seek information, share experiences, and exchange emotional support. However, the communication patterns of non-medical users in these discussions remain underexplored, particularly in the context of health literacy and public understanding of medical information. This study aims to identify thematic structures and communication functions expressed by non-medical users in online health conversations using *Latent Dirichlet Allocation* (LDA) and semi-automatic labeling. The research begins with text preprocessing and topic modeling using LDA to uncover dominant themes discussed by users. An automatic labeling mechanism was employed to assign communication functions to each topic, followed by expert judgment validation conducted by one clinical communication specialist. The quality of the LDA model was evaluated using topic coherence, while the validity of the automatic labels was assessed using an agreement score between machine-generated labels and expert annotations. Quantitative analysis was performed to examine the frequency and proportion of communication functions within each topic, whereas qualitative analysis interpreted representative user comments to provide deeper insights into how non-medical users ask questions, share experiences, give advice, and express emotions in digital health spaces. The findings reveal variations in communication patterns across health themes and highlight the dominance of certain communication functions, such as advice-giving, emotional expression, and experience sharing. The expert validation indicates that the automatic labeling approach achieves a high level of agreement, demonstrating its potential as an assistive method for annotating large-scale health communication data. This study contributes to a better understanding of how non-medical users construct and negotiate health-related meanings in online environments, offering implications for digital health literacy, community-based health communication, and the development of automated monitoring tools for public health discourse.

Key words — Topik Modeling, LDA, Non-Medical User, Online Health Discussion.

I. INTRODUCTION

Social media has become a primary arena for public engagement in health-related discussions. Through platforms such as X (Twitter), Instagram, and Facebook, users share personal experiences, ask health questions, exchange information, and respond to emerging health issues. Prior studies highlight the substantial potential of social media as an effective medium for health promotion (1), given its capacity to reach wide audiences at low cost and with rapid dissemination. It also facilitates health information access for individuals with limited contact with conventional healthcare services (2), and serves as a space for knowledge exchange, emotional support, and the formation of public opinion. Many studies further emphasize that social media increases public awareness, encourages healthy behaviors, and strengthens community participation in health issues (3). On platforms such as X, non-medical users frequently share experiences, offer advice, and express empathy, forming an ecosystem of public health communication that reflects the digital health literacy of Indonesian society. In this context, public health communication refers to openly accessible digital discussions on health-related topics, aligned with the public health communication paradigm that underscores public participation in shaping online health perceptions and behaviors.

The scale and diversity of communication on social media make manual analysis challenging; thus, computational approaches based on Natural Language Processing (NLP) are widely adopted to extract meaningful patterns from large volumes of text. Latent Dirichlet Allocation (LDA) supports the discovery of latent topics embedded in user-generated content (4), while communication function analysis (5–7) helps interpret the social meaning of these topics. Integrating both approaches enables a comprehensive exploration of public health discourse, as demonstrated in studies on high-risk disease prevention (8) and COVID-19 discussions (9). Building on this foundation, the present research aims to explore topics and communication functions expressed by non-medical users in online public health conversations using LDA. This study seeks to answer three key questions: (1) how LDA can be applied to identify topics in public health communication on social media; (2) how dominant communication functions—such as questions, information sharing, experience sharing, advice-giving, empathy, and referrals—can be identified; and (3) how topics and communication functions interact within each health theme.

To address these questions, the study employs a combination of computational and interpretative approaches. The computational empirical approach involves collecting and analyzing textual data from public conversations

on X using keyword-based web scraping, followed by a series of preprocessing steps. The resulting dataset is examined using LDA to extract latent topics and identify major themes. The interpretative analytical approach focuses on annotating LDA outputs using communication function coding derived from health communication theory, thereby mapping each topic to communicative categories such as information, questions, experiences, advice, empathy, and referrals. This combined approach is expected to produce both empirical and conceptual insights into public health communication patterns, facilitating the development of a thematic–functional analytical framework for subsequent research in data-driven health communication.

The state-of-the-art review further supports the significance of this research. Previous studies employing sentiment analysis demonstrate the effectiveness of machine learning in capturing public attitudes toward health issues, such as analyses of diabetes (10), benign breast tumors (11), and cervical cancer (12), as well as several sentiment studies conducted by the researcher on political and financial topics using Naïve Bayes, KNN, and NBC (13–15). However, sentiment analysis is limited to polarities and does not reveal deeper thematic patterns. Research utilizing topic modeling, including Structural Topic Modeling and LDA, has effectively uncovered public communication patterns on COVID-19 (16), diabetes (17), telemedicine (18), and Irritable Bowel Syndrome (19). The researcher’s previous work also applied LDA to professional medical communication in low-risk (4) and high-risk diseases (8). Building on these findings, the present research bridges professional medical discourse and public non-medical discourse by incorporating communication function analysis to construct a more comprehensive understanding of digital health interactions.

II. METHODOLOGY

This study used the Cross-Industry Standard Process for Data Mining (CRISP-DM) method to obtain information from student mental health data. CRISP-DM is a systematic methodology for data mining. This process involves several stages, starting from: business understanding, data understanding, data preparation, modeling, evaluation, and interpretation [14]. Figure 1 shows the stages in this study.

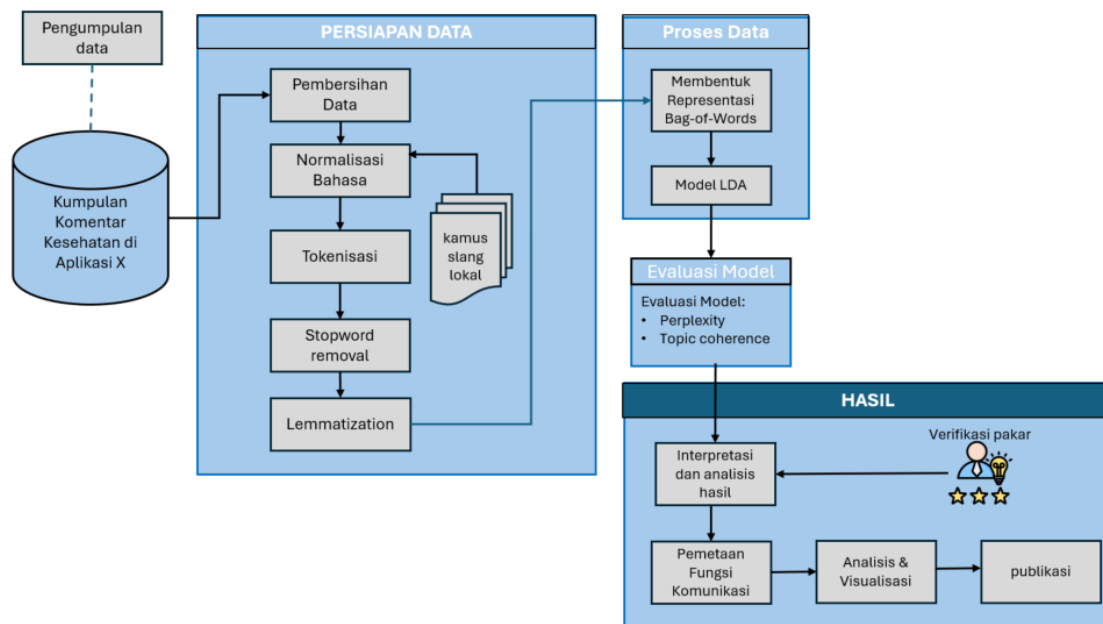


Figure 1. Kerangka Penelitian

This study adopts a mixed computational–interpretative methodology to analyze public health communication produced by non-medical users on social media. The research workflow begins with data collection from X (Twitter), focusing on public conversations related to health issues. Data were gathered through keyword-based web scraping to retrieve user-generated posts containing health-related terms. To ensure the quality and relevance of the dataset, a series of preprocessing steps were performed, including case folding, removal of URLs, emojis, and punctuation, elimination of stopwords, token normalization, and filtering of non-informative expressions. The resulting clean corpus serves as the input for topic modeling.

Latent Dirichlet Allocation (LDA) was employed to uncover latent thematic structures within the dataset. The number of topics was determined through experimental evaluation using topic coherence scores to ensure that the resulting topics exhibited semantic consistency. Each generated topic was represented through its top keywords and a distribution of contributing documents. The LDA output was then interpreted to identify major themes discussed by non-medical users in online public health conversations.

Following topic extraction, the study incorporated a communication function analysis to capture the pragmatic dimensions of health discourse. A semi-automatic labeling approach was used, where each LDA topic was initially assigned communication function labels—such as information sharing, questioning, experience sharing, advice-giving, empathy, and referral—based on keyword patterns and topic descriptors. These automatic labels were subsequently validated through expert judgment conducted by a clinical communication specialist. The expert reviewed each topic, confirmed or revised the communication function label, and provided interpretative notes when necessary. To assess the reliability of the automatic labeling, an agreement score was calculated by comparing the number of matching labels between the automatic and expert annotations relative to the total number of topics.

Both quantitative and qualitative analyses were conducted to interpret the final topics and communication functions. Quantitatively, the frequency and proportion of communication functions appearing within each topic were measured to identify dominant communicative patterns. Qualitatively, representative user comments were examined to provide contextual explanations of how non-medical users express questions, share personal experiences, provide advice, or convey emotional support in digital health spaces. This integrated methodological framework allows for a comprehensive thematic–functional interpretation of public health communication and offers insights into how health-related meanings are collaboratively constructed in online environments.

III. RESULT AND DISCUSSION

3.1 Topic Modeling Performance

Topic modeling was performed using the Latent Dirichlet Allocation (LDA) method to identify the main themes in health-related discussions posted by non-medical users on social media. In the initial experiment, the baseline LDA model produced a coherence score of 0.4336, with the optimal number of topics identified as five topics. To improve the quality of the topic model, hyperparameter optimization was conducted using a grid search approach by adjusting parameters such as alpha, eta, passes, and iterations. The optimized LDA model achieved a higher coherence score of 0.476, indicating an improvement in topic coherence and interpretability. This result demonstrates that hyperparameter tuning plays an important role in enhancing the performance of topic modeling for social media text data.

Although the highest coherence value was obtained at four topics, the model with eight topics ($K = 8$) was selected for further analysis because it provides a more detailed representation of the discussion themes and better supports the analysis of communication functions within health conversations.

Table 1. Ten Best Hyperparameter Configurations Based on Coherence Values in the LDA Model

Peringkat	Jumlah Topik	Alpha	Eta	Passes	Iterations	Koherensi
1	4	asymmetric	0.3	10	200	0.4760
2	8	asymmetric	0.3	20	200	0.4751
3	8	asymmetric	0.3	10	400	0.4682
4	6	asymmetric	0.3	20	200	0.4675
5	6	asymmetric	0.3	20	400	0.4675
6	8	asymmetric	0.3	20	400	0.4660
7	6	asymmetric	0.3	10	200	0.4659
8	6	asymmetric	0.3	10	400	0.4659
9	2	asymmetric	auto	20	200	0.4656
10	2	asymmetric	auto	20	400	0.4656

Table 2. Comparison of Baseline LDA and LDA Optimization

Model	Koherensi
LDA baseline	0.4336
LDA optimalisasi	0.4760

3.2 Visualization of Topic Distribution

To better understand the structure of the extracted topics, the intertopic distance visualization was generated using **pyLDAvis**. In this visualization, each circle represents a topic, and the size of the circle indicates the proportion of documents associated with that topic. The distance between circles reflects the similarity between topics, where topics located farther apart represent more distinct discussion themes. The visualization shows that most topics are relatively well separated, indicating that the LDA model successfully captures distinct themes within the dataset. Some topics appear closer to each other, suggesting partial semantic overlap in certain health discussions.

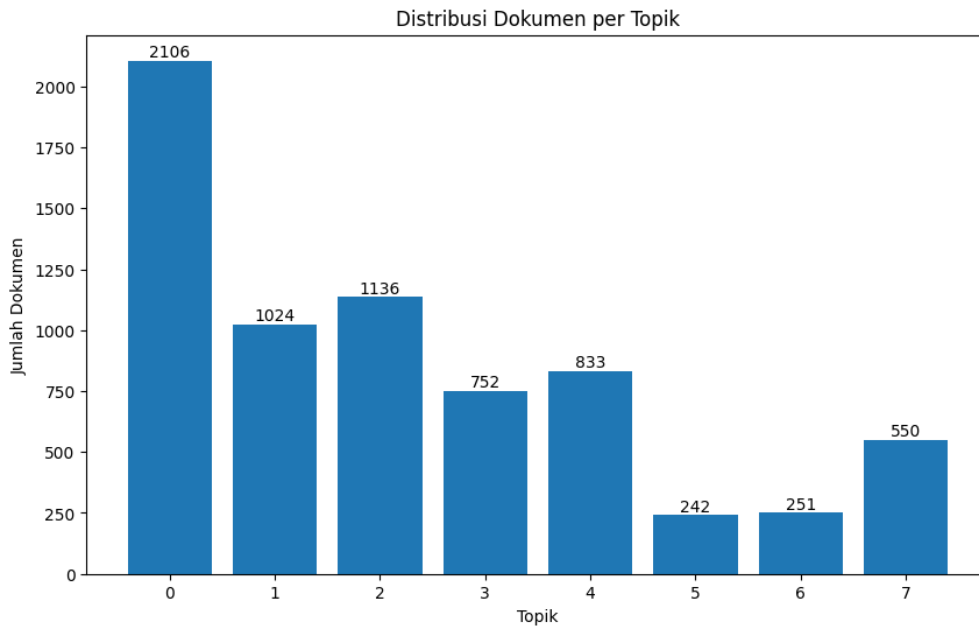


Figure 4. Distribution of Documents Across Topics

3.5 Communication Function Analysis

To understand the purpose of communication within the identified topics, the results were further analyzed using communication function categories, including seeking information, providing information, sharing experience, giving advice, emotional support, and referral to care. The analysis shows that health discussions on social media are mainly dominated by sharing experience and giving advice. Users frequently share personal experiences related to symptoms, treatments, or medication usage, and provide recommendations to other users facing similar conditions.

In addition, some conversations reflect information-seeking behavior, where users describe their symptoms and ask for possible explanations or solutions. Discussions involving healthcare professionals also indicate referral to care, where users suggest visiting medical professionals for further diagnosis or treatment.

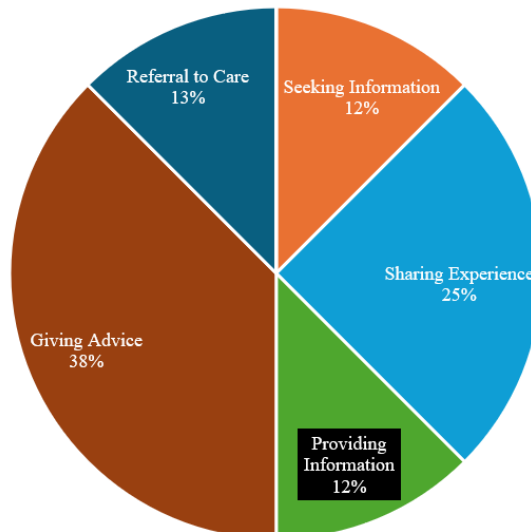


Figure 5. Distribution of Communication Functions

IV. CONCLUSION

This study contributes to the growing body of research on health communication in social media by integrating topic modeling and communication function analysis to better understand how non-medical users discuss health issues online. Unlike previous studies that primarily focus on identifying discussion topics, this research further interprets the identified topics through the lens of communication functions, enabling a deeper understanding of the purpose and dynamics of health-related conversations. From a methodological perspective, this study also

demonstrates the importance of hyperparameter optimization in LDA-based topic modeling, which significantly improves topic coherence and interpretability. The improved model performance suggests that optimized topic modeling can provide more reliable insights into large-scale social media discussions.

Practically, the findings highlight the role of social media as an informal health communication space, where users actively exchange experiences, advice, and information related to health conditions. Understanding these communication patterns can support public health organizations, health communicators, and policymakers in developing more effective strategies for disseminating accurate health information and addressing misinformation within online communities.

UCAPAN TERIMA KASIH

We would like to thank Budi Luhur University for the 2025 internal research grant for our study

DAFTAR PUSTAKA

- University of Queensland C for RH, Johns Bloomberg Hopkins school of Public Health. Indonesia-National Adolescent Mental Health Survey (I-NAMHS): Laporan Penelitian [Internet]. Pusat Kesehatan Reproduksi. Yogyakarta; 2022. Available from: <https://qcmhr.org/outputs/reports/12-i-namhs-report-bahasa-indonesia>
- Kohls E, Guenther L, Baldofski S, Brock T, Schuhr J, Rummel-Kluge C. Two years COVID-19 pandemic: Development of university students' mental health 2020–2022. *Front Psychiatry*. 2023;1122256(14):1–15.
- Shafiee NSM, Mutalib S. Prediction of Mental Health Problems among Higher Education Student Using Machine Learning. *Int J Educ Manag Eng*. 2020;10(6):1–9.
- Baba A, Bunji K. Prediction of Mental Health Problem Using Annual Student Health Survey: Machine Learning Approach. *JMIR Ment Heal*. 2023;10(6):1–23.
- Maiyanti SI, Zayanti DA, Andriani Y, Suprihatin B, Desiani A, Salsabila A, et al. Perbandingan Klasifikasi Penyakit Kanker Paru-paru menggunakan Support Vector Machine dan K-Nearest Neighbor. *Process J Ilm Sist Informasi, Teknol Inf dan Sist Komput*. 2023;18(1):54–62.
- Ramon E, Nazir A, Novriyanto N, Yusra Y, Oktavia L. Klasifikasi Status Gizi Bayi Posyandu Kecamatan Bangun Purba Menggunakan Algoritma Support Vector Machine (SVM). *J Sist Inf dan Inform*. 2022;5(2):143–50.
- Ihsan TFF, Ramadhan I, Akbar DR, Ismanto E. Perbandingan Model Machine Learning Untuk Klasifikasi Deteksi Penyakit Jantung. *J Comput Sci Inf Technol*. 2025;6(2):199–205.
- Ahuja R, Banga A. Mental stress detection in university students using machine learning algorithms. *Procedia Comput Sci* [Internet]. 2019;152:349–53. Available from: <https://doi.org/10.1016/j.procs.2019.05.007>
- Sahlan F, Hamidi F, Misrat MZ, Adli MH, Wani S, Gulzar Y. Prediction of Mental Health Among University Students. *Int J Perceptive Cogn Comput* [Internet]. 2021;7(1):85–91. Available from: <https://journals.iium.edu.my/kict/index.php/IJPC/article/view/225>
- Malik SS. Anxiety, Depression and Stress prediction among College Students using Machine Learning Algorithms. In: *Second International Conference on Electrical, Electronics, Information and Communication Technologies (ICEEICT)*. IEEE; 2023. p. 1–5.
- Mahalakshmi G, Sujatha G. Predictions of College Students' Mental Stress using Machine Learning Algorithms. In: *7th International Conference on Intelligent Computing and Control Systems (ICICCS)*. Madurai, India: IEEE; 2023. p. 288–95.
- Juanita S. Analisis Sentimen Persepsi Masyarakat Terhadap Pemilu 2019 Pada Media Sosial Twitter Menggunakan Naive Bayes. *J Media Inform Budidarma*. 2020;4(3):552–8.
- Putra ADA, Juanita S. Analisis Sentimen pada Ulasan pengguna Aplikasi Bibit Dan Bareksa dengan Algoritma KNN. *JATISI (Jurnal Tek Inform dan Sist Informasi)*. 2021;8(2):636–46.
- Juanita S, Adiyarta K, Syafrullah M. Sentiment analysis on E-Marketplace User Opinions Using Lexicon-Based and Naive Bayes Model. In: *International Conference on Electrical Engineering, Computer Science and Informatics (EECSI)*. Institute of Electrical and Electronics Engineers Inc.; 2022. p. 379–82.
- Sembiring ERB. Pengembangan Instrumen Pengukuran Kesehatan Mental [Internet]. Universitas Pendidikan Ganesha; 2023. Available from: <https://repo.undiksha.ac.id/14766/11/1911011024-LAMPIRAN.pdf>
- Li C. Preprocessing Methods and Pipelines of Data Mining: An Overview. In: *Seminar Data Mining* [Internet]. 2019. p. 1–7. Available from: <http://arxiv.org/abs/1906.08510>
- Wheeler EG. Age Periods of Human Life. *Adv Soc Sci Res J*. 2017;4(6):258–63.
- Sokolova M, Lapalme G. A systematic analysis of performance measures for classification tasks. *Inf Process Manag*. 2009;45:427–37.
- Sembiring ERB. Pengembangan Instrumen Pengukuran Kesehatan Mental (Mental Health Inventory) Setelah Pandemi Covid-19 pada Siswa SMP di Kota Singaraja [Internet]. Universitas Pendidikan Ganesha; 2023. Available from: <https://repo.undiksha.ac.id/14766/>

Lampiran 5. HKI (Terbit)


REPUBLIC INDONESIA
KEMENTERIAN HUKUM

SURAT PENCATATAN CIPTAAN

Dalam rangka perlindungan ciptaan di bidang ilmu pengetahuan, seni dan sastra berdasarkan Undang-Undang Nomor 28 Tahun 2014 tentang Hak Cipta, dengan ini menerangkan:

Nomor dan tanggal permohonan : EC002026036404, 6 Maret 2026

Pencipta

Nama : **Safitri Juanita dan M. Anif**

Alamat : Jl. Lembang Baru IIIA-RT.02 RW.09 No.62, Ciledug, Kota Tangerang, Banten, 15151

Kewarganegaraan : Indonesia

Pemegang Hak Cipta

Nama : **Direktorat Riset dan Pengabdian kepada Masyarakat Universitas Budi Luhur**

Alamat : Jl. Ciledug Raya, RT.10/RW.2, Pesanggrahan, Kota Adm. Jakarta Selatan, DKI Jakarta, 12260

Kewarganegaraan : Indonesia

Jenis Ciptaan : **Basis Data**

Judul Ciptaan : **Basis Data : Data Komentar Pengguna Non-Medis Pada Aplikasi X Terkait Topik Kesehatan**

Tanggal dan tempat diumumkan untuk pertama kali di wilayah Indonesia atau di luar wilayah Indonesia : 26 Februari 2026, di DKI Jakarta

Jangka waktu perlindungan : Berlaku selama 50 (lima puluh) tahun sejak Ciptaan tersebut pertama kali dilakukan Pengumuman.

Nomor Pencatatan : 001164211

adalah benar berdasarkan keterangan yang diberikan oleh Pemohon.
Surat Pencatatan Hak Cipta atau produk Hak terkait ini sesuai dengan Pasal 72 Undang-Undang Nomor 28 Tahun 2014 tentang Hak Cipta.

a.n. MENTERI HUKUM
DIREKTUR JENDERAL KEKAYAAN INTELEKTUAL
u.b
Direktur Hak Cipta dan Desain Industri


Agung Damarsongko,SH.,MH.
NIP. 196912261994031001





Disclaimer:

1. Dalam hal pemohon memberikan keterangan tidak sesuai dengan surat pernyataan, Menteri berwenang untuk mencabut surat pencatatan permohonan.
2. Surat Pencatatan ini telah disegel secara elektronik menggunakan segel elektronik yang diterbitkan oleh Balai Besar Sertifikasi Elektronik, Badan Siber dan Sandi Negara.
3. Surat Pencatatan ini dapat dibuktikan keasliannya dengan memindai kode QR pada dokumen ini dan informasi akan ditampilkan dalam browser.