

# Analisis Sentimen menggunakan IndoBERT dan Tren Topik Keluhan Pasien pada Ulasan Google Maps Rumah Sakit Menggunakan Latent Dirichlet Allocation

Naufal Muhammad Afif<sup>1</sup>, Ghuftron<sup>2</sup>

<sup>1</sup> Universitas Islam Sultan Agung; noval13131@gmail.com

<sup>2</sup> Universitas Islam Sultan Agung; ghuftron@unissula.ac.id

\*

## Info Artikel:

Dikirim: 12 Mei 2026

Direvisi: 30 Mei 2026

Diterima: 06 Juni 2026

**Abstract:** Patient satisfaction is a crucial indicator of hospital quality, yet management often focuses solely on star ratings that fail to explain the root causes of issues. This study develops a hybrid Natural Language Processing (NLP) model using IndoBERT for sentiment classification of Google Maps reviews. Reviews classified as negative sentiment are then filtered and processed using the Latent Dirichlet Allocation (LDA) method to uncover hidden themes within patient complaints. The test results show that the IndoBERT model achieves exceptionally high performance, with an accuracy of 95.23%, precision of 95.22%, recall of 95.23%, and an F1-score of 95.22%. The LDA analysis successfully identifies 10 optimal topics, which are categorized into five main complaint categories: time efficiency, medical services, facilities/parking, administrative procedures, and specialist services. The integration of IndoBERT and LDA proves effective in transforming raw digital reviews into strategic information for the automated evaluation of hospital service quality.

**Keywords:** Sentiment Analysis; IndoBERT; Topic Modeling; LDA; Google Maps; Patient Reviews

**Intisari:** Kepuasan pasien adalah indikator krusial mutu rumah sakit, namun manajemen sering kali hanya terpaku pada *rating* bintang yang tidak menjelaskan akar masalah. Penelitian ini mengembangkan model hibrida *Natural Language Processing* (NLP) menggunakan IndoBERT untuk klasifikasi sentimen ulasan Google Maps. Ulasan yang terklasifikasi sebagai sentimen negatif kemudian disaring dan diolah menggunakan metode *Latent Dirichlet Allocation* (LDA) untuk mengungkap tema tersembunyi pada keluhan pasien. Hasil pengujian menunjukkan model IndoBERT mencapai performa sangat tinggi dengan nilai akurasi 95,23%, presisi 95,22%, recall 95,23%, dan F1-score 95,22%. Analisis LDA berhasil mengidentifikasi 10 topik optimal yang terbagi ke dalam lima kategori keluhan utama: efisiensi waktu, layanan medis, fasilitas/parkir, prosedur administrasi, dan layanan spesialis. Integrasi IndoBERT dan LDA terbukti efektif mentransformasi ulasan digital mentah menjadi informasi strategis untuk evaluasi kualitas pelayanan rumah sakit secara otomatis.

**Kata Kunci:** Analisis Sentimen; IndoBERT; Pemodelan Topik; LDA; Google Maps; Ulasan Pasien.

## 1. Pendahuluan

Kualitas pelayanan kesehatan merupakan fondasi utama dalam sistem kesehatan global yang secara langsung berdampak pada keselamatan dan kesejahteraan masyarakat luas. Diperkirakan terdapat 8,4 juta kematian setiap tahunnya yang disebabkan oleh buruknya kualitas pelayanan di negara-negara berpenghasilan rendah dan menengah, yang mana angka ini menyumbang sekitar 15% dari total kematian di wilayah tersebut [1]. Dalam konteks persaingan industri kesehatan yang semakin ketat, kepuasan pasien bukan hanya sekadar indikator kinerja klinis, melainkan juga menjadi penentu keberlangsungan bisnis Rumah Sakit (RS) [2].

Meskipun standar operasional prosedur telah diterapkan secara ketat, keluhan atau komplain dari pasien merupakan hal yang tidak dapat dihindari dalam operasional harian rumah sakit. Berbagai aspek layanan sering menjadi sumber keluhan, mulai dari waktu tunggu yang lama hingga sikap petugas medis yang dianggap kurang empatik [3]. Seiring dengan perkembangan teknologi informasi, cara pasien menyampaikan aspirasi telah mengalami pergeseran drastis dari metode konvensional menuju platform digital terbuka seperti Google Maps [4].

Akan tetapi, ketergantungan manajemen rumah sakit pada indikator kuantitatif sederhana seperti rating bintang (star rating) memiliki kelemahan dasar karena tidak mampu menjelaskan akar permasalahan di balik ketidakpuasan pasien [4]. Tantangan utama dalam memanfaatkan data ulasan daring adalah volume data yang sangat besar dan formatnya yang tidak terstruktur, atau dikenal sebagai *Unstructured Patient Feedback* (UPF). Pemanfaatan teknik NLP, khususnya analisis sentimen dan pemodelan topik, memungkinkan manajemen rumah sakit untuk memantau persepsi publik secara cepat dan akurat [5].

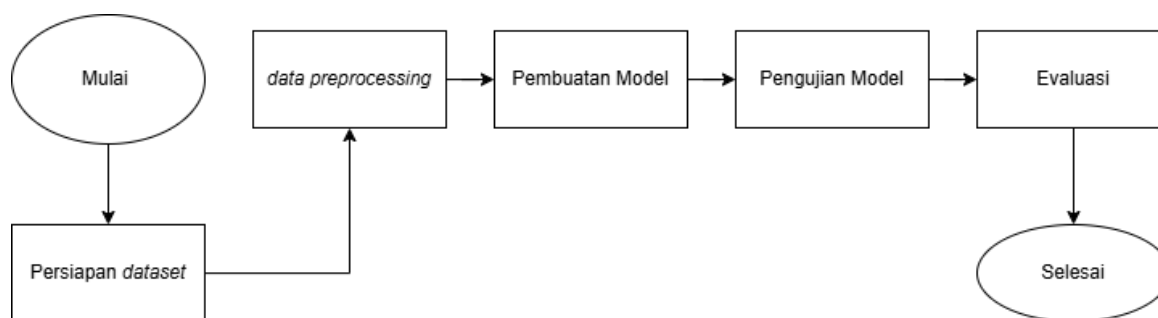
Dalam konteks penerapan NLP di Indonesia, masyarakat cenderung menggunakan ragam bahasa yang sangat tidak baku, penuh singkatan, dan istilah gaul (slang) yang membuat metode pembelajaran mesin klasik sering kesulitan menangkap makna sentimen [6]. IndoBERT muncul sebagai solusi karena arsitektur Transformer-nya memiliki kemampuan superior dalam memahami nuansa kontekstual bahasa Indonesia [7].

Namun, sekadar mengetahui sentimen belum cukup; diperlukan metode Latent Dirichlet Allocation (LDA) memegang peranan vital untuk mengekstraksi struktur topik tersembunyi (*latent topics*) secara otomatis tanpa perlu pelabelan manual sebelumnya [8]. Integrasi antara analisis sentimen menggunakan IndoBERT dan pemetaan topik menggunakan LDA menawarkan pendekatan hibrida yang sangat kuat sebagai sistem peringatan dini (*early detection system*) terhadap isu-isu krusial rumah sakit [8].

## 2. Metode Penelitian

. Dalam penelitian ini, terdapat rancangan tahapan dimulai dari proses *web scraping* untuk mendapatkan *dataset review Google Maps*. Data yang menjadi objek penelitian ini adalah ulasan tidak terstruktur (*Unstructured Patient Feedback*) dari platform Google Maps yang berhubungan dengan pelayanan beberapa Rumah Sakit Pemerintah di Semarang, seperti RSUP Dr. Kariadi, RSUD Adhyatma Tugurejo, dan RS Daerah KRMT Wongsonegoro. Selanjutnya akan dilakukan proses text preprocessing untuk menghasilkan dataset yang lebih terstruktur.

Selanjutnya, dilakukan proses klasifikasi sentimen menggunakan model *fine-tuned* IndoBERT untuk memisahkan ulasan ke dalam kategori sentimen positif dan negatif. Sistem kemudian melakukan proses filterisasi otomatis guna menyaring dan mengambil hanya ulasan yang diprediksi memiliki polaritas sentimen negatif. Hasil ulasan negatif (keluhan) tersebut kemudian akan dimasukkan ke dalam proses pemodelan topik dengan metode *Latent Dirichlet Allocation* (LDA) untuk mencari tren kelompok topik keluhan dari ulasan rumah sakit tersebut. Hasil dari rancangan ini dapat dilihat pada Gambar 3.1.



Gambar 1 Flowchart Penelitian

## 2.1 Persiapan Dataset

Dalam penelitian ini, pengumpulan data dilakukan dengan cara *web scraping*. *Web scraping* adalah sebuah cara pengambilan suatu data atau informasi tertentu dengan jumlah besar untuk nantinya digunakan dalam berbagai keperluan riset dan analisis. Penelitian ini mengambil data ulasan publik dari platform Google Maps pada objek tiga rumah sakit pemerintah di Semarang. Fokus penelitian ini terbagi menjadi dua bagian dataset fungsional:.

1. Dataset Penelitian: Terdiri dari ulasan campuran yang dikumpulkan dari RSUD Adhyatma Tugurejo dan RS Daerah KRMT Wongsonegoro dengan jumlah total 944 ulasan. Dataset ini dipecah menggunakan rasio 80% untuk data latih (*train\_df*) dan 20% untuk data validasi (*val\_df*) menggunakan parameter *random\_state=42* untuk konsistensi pengacakan serta *stratify* untuk menjaga keseimbangan proporsi label.
2. Dataset Uji Utama (Data Inti): Terdiri dari data khusus objek penelitian utama, yaitu RSUP Dr. Kariadi Semarang dengan jumlah total 1.447 ulasan. Data ini diposisikan murni sebagai data uji analitis dan tidak dilibatkan dalam fase pelatihan model.

## 2.2 Text Preprocessing

*Text Preprocessing* adalah sebuah proses yang dilakukan untuk mengubah bentuk data teks yang belum terstruktur menjadi data teks yang terstruktur sesuai dengan kebutuhan untuk proses analisis yang lebih lanjut [9], [10]. Dalam penelitian ini, dilakukan dua jalur tahapan *text preprocessing* untuk mengelola data ulasan rumah sakit yang telah dikumpulkan melalui *web scraping*:

1. Jalur *Preprocessing* Minimal (Untuk IndoBERT): Penerapan pembersihan karakter dilakukan secara minimal (*Light Preprocessing*) murni untuk menghapus komponen *emoji* dan karakter *non-ASCII*. Struktur kalimat asli dan tanda baca dipertahankan untuk menjaga keutuhan konteks kalimat yang dibutuhkan oleh mekanisme *attention* IndoBERT
2. Jalur *Preprocessing* Agresif (Untuk LDA): Diterapkan pada data ulasan negatif yang akan masuk ke tahap pemetaan topik (*Heavy Preprocessing*) guna menyisakan kata-kata substantif penting. Tahapan-tahapannya meliputi:
  - *Case Folding*: Mengubah penggunaan huruf besar dalam teks menjadi huruf kecil keseluruhan (*lowercase*).
  - *Cleaning*: Menghapus angka, tanda baca, tautan URL, dan nama pengguna (*username*) menggunakan pustaka *Regular Expression* (RE).
  - *Normalisasi*: Mengubah kata tidak baku menjadi kata baku berdasarkan kamus normalisasi
  - *Stemming*: Mengubah kata-kata berimbuhan menjadi kata dasar untuk mengurangi variasi kata menggunakan algoritma Sastrawi.
  - *Filtering/Stopword Removal*: Menghapus kata-kata fungsional umum yang tidak relevan dan tidak memiliki makna topik menggunakan pustaka Sastrawi.

## 2.3 Klasifikasi dan Filtering Sentimen IndoBERT

IndoBERT digunakan sebagai pengkodek kontekstual untuk melakukan analisis sentimen dua arah terhadap teks ulasan. Inti operasi model ini terletak pada fungsi *Scaled Dot-Product Attention* yang dirumuskan pada Persamaan (1):.

$$\text{Attention}(Q, K, V) = \text{softmax} \left( \frac{QK^T}{\sqrt{d_k}} \right) V \quad (1)$$

Ket:

Q = Matriks Query

K = Matriks Key

V = Matriks Value

$d_k$  = Dimensi dari kata key

Model merangkum representasi token ke dalam vektor utama [CLS] yang kemudian dimasukkan ke fungsi eksponensial *Softmax* untuk menghasilkan probabilitas klasifikasi akhir. Sistem melakukan pemindaian otomatis; ulasan yang terdeteksi sebagai sentimen positif diabaikan dari jalur pengolahan topik, sementara

ulasan yang diklasifikasikan sebagai sentimen negatif (keluhan) disaring untuk diteruskan ke pemodelan topik LDA

#### 2.4 Pemodelan Topik dengan Latent Dirichlet Allocation(LDA)

Metode LDA memproses data ulasan negatif hasil filterisasi untuk menemukan struktur semantik atau tema tersembunyi (*latent topics*) secara otomatis tanpa pelabelan manual. Hubungan parameter korpus, distribusi topik dalam dokumen, hingga pemilihan kata spesifik dihitung melalui model probabilitas bersama (*joint probability*) Bayesian pada Persamaan (2)

$$p(\theta, z, \omega|\alpha, \beta) = p(\theta|\alpha) \prod_{n=1}^N p(z_n|\theta)p(\omega_n|z_n, \beta) \tag{2}$$

Ket:

$\omega$  =Kata aktual yang diobservasi

$\alpha$  =Parameter *dirichlet prior* distribusi topik

$\theta$  =Vektor distribusi untuk sebuah dokumen

$\beta$  =Parameter *Dirichlet prior* distribusi kata per topik

$z$  =Topik untuk sebuah kata tertentu dalam dokumen

#### 2.5 Evaluasi

Teknik yang digunakan untuk melakukan evaluasi sistem pada penelitian ini menggunakan dua macam pendekatan metrik pengujian:

1. Evaluasi Klasifikasi (Confusion Matrix): Mengukur ketepatan model IndoBERT dalam memprediksi kelas sentimen dengan membandingkan label aktual (*ground truth*) dan label hasil prediksi model. Parameter evaluasi diturunkan menggunakan komponen *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN) melalui Persamaan (3) Dataset Uji Utama (Data Inti): Terdiri dari data khusus objek penelitian utama, yaitu RSUP Dr. Kariadi Semarang dengan jumlah total 1.447 ulasan. Data ini diposisikan murni sebagai data uji analitis dan tidak dilibatkan dalam fase pelatihan model. Tingkat ketepatan model dalam mengklasifikasikan seluruh data secara benar diuji menggunakan metrik *Accuracy* melalui perhitungan rasio total prediksi benar dibagi keseluruhan data pada Persamaan (3)

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{3}$$

Ket:

TP = (True Positive) Jumlah Data aktual Positif yang diprediksi benar sebagai positif oleh model

TN = (True Negative) Jumlah Data aktual Negatif yang diprediksi benar sebagai negatif oleh model

FP = (False Positive) Jumlah Data aktual Negatif yang diprediksi salah sebagai positif oleh model

FN = (False Negative) Jumlah Data aktual Positif yang diprediksi salah sebagai negatif oleh model

Selanjutnya, kualitas ketepatan model dalam memprediksi kelas positif diukur melalui metrik *Precision* pada Persamaan (4) dan sebaliknya jika *precision* kelas negatif  $TN / (TN+FN)$

$$Precision = \frac{TP}{TP + FP} \tag{4}$$

Sedangkan kemampuan model dalam menemukan kembali seluruh data yang termasuk dalam kelas positif diuji menggunakan metrik *Recall* pada Persamaan (5) dan sebaliknya jika *recall* kelas negatif  $TN / (TN+FP)$

$$Recall = \frac{TP}{TP + FN} \tag{5}$$

Selanjutnya, untuk menyeimbangkan nilai performa klasifikasi tersebut pada kondisi riil, dilakukan penghitungan nilai harmonik rata-rata menggunakan metrik *F1-Score* melalui rumus matematis pada Persamaan (6).

$$F1-Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{6}$$

2. Evaluasi Pemodelan Topik (Topic Coherence): Menggunakan skor koherensi metrik  $C_v$  untuk mengevaluasi kualitas kedekatan makna dan hubungan antar kata kunci di dalam satu kelompok topik LDA. Penghitungan nilai memanfaatkan Normalized Pointwise Mutual Information (NPMI) dan fungsi sliding window yang diintegrasikan melalui rumus Cosine Similarity pada Persamaan (7):

$$C_v = \cos(u \cdot v) = \frac{u \cdot v}{\|u\| \|v\|} \tag{7}$$

### 3. Hasil Penelitian

#### 3.1 Hasil Fine-Tuning IndoBERT

Proses pelatihan atau *fine-tuning* model IndoBERT dilakukan menggunakan dataset pelatihan yang berjumlah 944 ulasan. Dataset tersebut dibagi dengan komposisi rasio 80% untuk data latih (*train*) dan 20% untuk data validasi (*validation*). Eksperimen dijalankan selama 3 *epoch* menggunakan parameter tingkat pembelajaran (*learning rate*) sebesar  $2e-5$  dan ukuran *batch* sebesar 16 untuk melihat performa optimal model. Hasil pemantauan parameter performa model pengklasifikasi IndoBERT selama 3 siklus (*epoch*) terekam secara berkala pada Tabel 1

Tabel 1 Hasil Evaluasi Nilai Kerja per Epoch Pelatihan

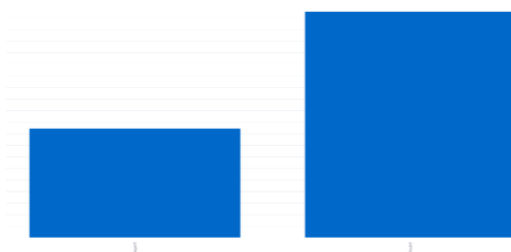
Epoch	Training Loss	Validation Loss	Accuracy	F1 Score	Recall	Precision
1	No Log	0.224628	0.925926	0.927083	0.927083	0.927083
2	No Log	0.271384	0.920635	0.922280	0.917526	0.927083
3	No Log	0.300721	0.925926	0.927835	0.918367	0.937500

Berdasarkan Tabel 1, pergerakan nilai validation loss merangkak naik secara bertahap mulai dari 0.224628 pada epoch pertama hingga menyentuh angka 0.300721 pada epoch ketiga. Kondisi ini mengindikasikan adanya gejala *overfitting* ringan akibat karakteristik variasi bahasa teks ulasan informal. Kendati demikian, metrik penentu utama keberhasilan model yang didasarkan pada nilai F1-Score justru mencatatkan puncak performa tertinggi pada epoch ketiga, yaitu sebesar 0.927835 (92,78%). Pada siklus akhir tersebut, nilai Accuracy stabil di angka 0.925926 (92,59%), nilai Recall mencatatkan angka sebesar 0.918367 (91,83%), dan nilai Precision berhasil menyentuh angka tertinggi

sebesar 0.937500 (93,75%). Dengan mengaktifkan parameter konfigurasi penyimpanan otomatis model terbaik pada akhir sesi pelatihan, maka bobot parameter dari hasil epoch ketiga ini resmi diekspor sebagai model pengklasifikasi sentimen utama dalam sistem.

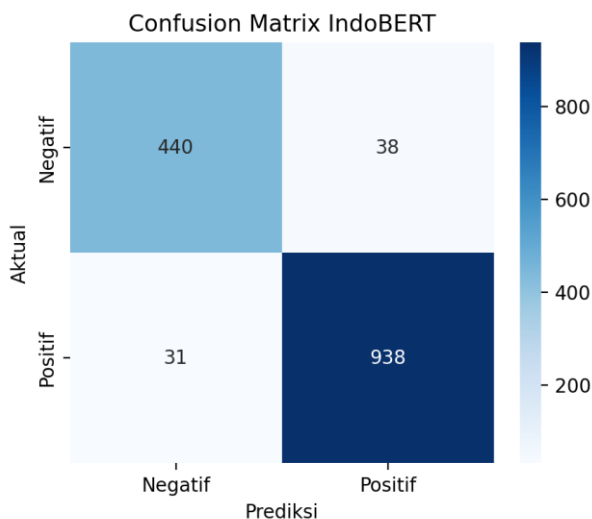
### 3.2 Hasil Evaluasi Klasifikasi Data Uji Utama

Model hasil pelatihan terbaik kemudian diuji kemampuannya menggunakan data uji utama (data inti) yang diambil secara independen dari RSUP Dr. Kariadi Semarang sebanyak 1.447 ulasan. Proses prediksi otomatis oleh model berhasil memisahkan data ke dalam dua polaritas sentimen utama. Distribusi kuantitas hasil klasifikasi antara ulasan sentimen positif dan sentimen negatif tersebut disajikan secara visual melalui grafik batang pada Gambar 2.



Gambar 2 Grafik Batang Jumlah Data Sentimen

Berdasarkan grafik pada Gambar 2, terlihat bahwa model mengklasifikasikan ulasan dengan sentimen positif dalam jumlah yang lebih dominan, yaitu sebanyak 976 ulasan, sedangkan untuk sentimen negatif (keluhan) tercatat sebanyak 471 ulasan. Selanjutnya, untuk melakukan pengujian performa ketepatan prediksi model terhadap data *ground truth* (label aktual asli), digunakan tabel kontingensi *confusion matrix* yang dipaparkan pada Gambar 3.



Gambar 3 Evaluasi *Confusion Matrix IndoBERT*

Berdasarkan matriks evaluasi pada Gambar 3, diperoleh rincian data performa berupa nilai True Negative (TN) sebesar 440, False Positive (FP) sebesar 31, False Negative (FN) sebesar 38, dan True Positive (TP) sebesar 938. Seluruh komponen kuantitatif tersebut kemudian diturunkan untuk menghitung laporan metrik kinerja menyeluruh yang dirangkum secara detail pada Tabel 2.

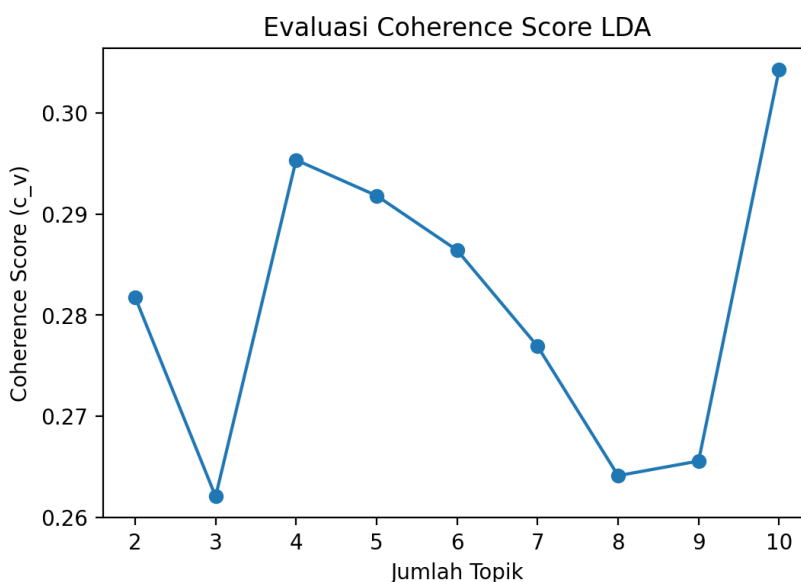
**Tabel 2** Hasil Detail Evaluasi Model IndoBERT pada Data Uji Utama

Kelas Sentimen	Precision	Recall	F1-Score	Jumlah Data
Negatif	0.9342	0.9205	0.9273	478
Positif	0.9611	0.9680	0.9645	969
Akurasi (Accuracy)			0.9523	1447
Macro Average	0.9476	0.9443	0.9459	1447
Weighted Average	0.9522	0.9523	0.9522	1447

Melalui hasil yang ditunjukkan pada Tabel 2, pengujian pada data riil membuktikan tingkat generalisasi model yang sangat kuat. Nilai akurasi menyeluruh (*Accuracy*) sistem berada pada angka 0.9523 atau sebesar 95,23%. Sesuai dengan fokus perancangan sistem evaluasi, performa pengujian pada kelas negatif (ulasan berisi keluhan) juga mencatatkan nilai yang sangat tinggi dengan perolehan skor *Precision* sebesar 0.9342 (93,42%) serta skor *Recall* sebesar 0.9205 (92,05%). Tingginya nilai parameter kuantitatif di atas 95% untuk nilai rata-rata tertimbang (*weighted average*) menegaskan bahwa pemanfaatan fitur kontekstual dua arah (*bidirectional embedding*) pada model IndoBERT sukses mengatasi kendala bias makna semantik akibat penggunaan bahasa gaul atau tidak baku masyarakat urban di platform Google Maps.

### 3.3 Hasil Evaluasi Klasifikasi Data Uji Utama

Sebanyak 471 ulasan yang berhasil disaring sebagai sentimen negatif (keluhan) oleh model IndoBERT kemudian diteruskan ke dalam modul algoritma LDA. Eksperimen pembentukan model LDA dijalankan secara iteratif dengan menguji variasi kuantitas kluster topik mulai dari nilai  $k = 2$  hingga  $k = 10$ . Pemantauan kualitas keterpisahan dan kedekatan makna kata di dalam topik dievaluasi menggunakan metrik skor koherensi (*Coherence Score  $C_v$* ). Pola perubahan nilai koherensi semantik dari hasil pengujian variasi jumlah topik tersebut disajikan secara visual melalui grafik pada Gambar 4



**Gambar 4** Grafik Evaluasi *Coherence Score* LDA

Berdasarkan grafik evaluasi pada Gambar 3.3, diperoleh nilai koherensi tertinggi saat jumlah kluster diatur sebanyak 10 topik  $k = 10$ , dengan pencapaian nilai berada di atas angka 0,30. Kenaikan nilai yang signifikan ini membuktikan secara kuantitatif bahwa konfigurasi 10 topik menghasilkan pembagian kluster kata yang paling optimal dan minim ambiguitas semantik. Selanjutnya, hasil ekstraksi kata kunci teratas (*Top Terms*) beserta nilai probabilitas kemunculan kata untuk masing-masing dari 10 topik laten yang terbentuk dipaparkan secara lengkap pada Tabel 3.

**Tabel 3** Tabel Topik Hasil LDA

Topik	Isi Topik
Topik 0	Dokter, layan, awat, anak, bagus
Topik 1	Kurang, layan, susah, kontrol, parkir
Topik 2	Parkir, Ramah, Satpam, Kartu, kurang
Topik 3	Periksa, Dokter, Parkir, Garuda, kontrol
Topik 4	Nunggu, Tunggu, Panggil, Lama, Tugas
Topik 5	Ruang, Proses, Tengah, Layan, Sama
Topik 6	Dokter, Daftar, Jadwal, Datang, lama
Topik 7	Rawat, Antri, Kurang, Mata, Layan
Topik 8	Layan, Ruang, Awat, Antri, Rawat
Topik 9	Layan, Lama, Nunggu, Dokter, banyak

Berdasarkan data kata kunci pada Tabel 3, kesepuluh klaster topik tersebut menunjukkan adanya keterkaitan makna semantik yang kuat satu sama lain. Guna mempermudah proses interpretasi analitis, peneliti melakukan reduksi dan pengelompokan fungsional dari 10 topik laten tersebut ke dalam 5 kategori area kritis keluhan utama pelayanan rumah sakit. Hasil pemetaan sektoral tersebut disajikan pada Tabel 4.

**Tabel 4** Hasil Pemetaan Kategori Keluhan Rumah Sakit

Kategori Keluhan	Topik Terkait	Deskripsi Masalah
Kategori 1 (Efisiensi Waktu dan Antrean)	Topic 4 dan topik 9	Berfokus pada keluhan durasi menunggu, proses pemanggilan antrean yang lama, dan banyaknya tumpukan pasien.
Kategori 2 (Layanan Medis dan Dokter)	Topik 0, Topik 3 dan Topik 6	Berkaitan dengan kualitas pemeriksaan dokter, spesialis anak, serta ketepatan jadwal kedatangan dokter.
Kategori 3 (Fasilitas Umum dan Parkir)	Topik 1 dan Topik2	Menyoroti kendala pada area parkir, akses masuk, serta interaksi dengan petugas keamanan (satpam).
Kategori 4 (Prosedur dan Administrasi)	Topik 5, dan Topik8	Mencakup alur pelayanan secara umum, proses rawat inap, serta prosedur administrasi ruangan.
Kategori 5 (Prosedur dan Administrasi pada poliklinik mata)	Topik 7	Keluhan spesifik yang berkaitan dengan unit rawat jalan tertentu (seperti poliklinik mata) dan antreannya.

Selain melalui pengelompokan secara fungsional konseptual, struktur keterpautan semantik serta pemisahan antar-klaster dari 10 topik keluhan yang terbentuk juga dianalisis secara spasial. Hasil visualisasi interaktif pemodelan topik menggunakan *library* pyLDAvis yang diproyeksikan melalui metode *Multidimensional Scaling* (MDS) disajikan pada Gambar 5



**Gambar 5** Hasil Pemodelan LDA Menggunakan Library pyLDAvis

Gambar 5 menyajikan visualisasi interaktif hasil pemodelan LDA menggunakan *library* pyLDAvis, yang terdiri dari dua bagian utama: *Intertopic Distance Map* di sebelah kiri dan *Top-30 Most Salient Terms* di sebelah kanan. Visualisasi ini bertujuan untuk memberikan gambaran komprehensif mengenai distribusi topik, ukuran relatif topik dalam korpus, serta hubungan antar topik berdasarkan kemiripan semantik melalui proyeksi *Multidimensional Scaling* (MDS).

Pada sisi kiri (*Intertopic Distance Map*), setiap lingkaran mewakili satu dari 10 topik yang terbentuk. Berbeda dengan model yang memiliki jumlah topik sedikit, visualisasi ini menunjukkan struktur yang lebih kompleks di mana terdapat beberapa kluster topik yang saling berdekatan bahkan tumpang tindih (*overlap*), seperti pada kelompok Topik 3, 4, 8, dan 9 di kuadran kiri, serta Topik 1 dan 7 di kuadran kanan. *Overlap* ini mengindikasikan adanya kesamaan kata kunci atau keterkaitan tema yang sangat erat di antara topik-topik tersebut. Namun, terdapat pula topik yang benar-benar terisolasi dan unik, seperti Topik 10, 5, 6, dan 2, yang menunjukkan adanya kategori keluhan yang sangat spesifik dan berbeda dari tema utama lainnya.

Meskipun model mampu mengisolasi topik-topik unik seperti infrastruktur parkir (Topik 2), visualisasi spasial ini juga menyingkap adanya keterbatasan internal dari algoritma LDA yang dibangun. Adanya fenomena tumpang tindih (*overlap*) yang masif pada Topik 3, 4, 8, dan 9 di kuadran kiri menandakan batas semantik antar-kluster keluhan tersebut masih sangat bias dan ambigu, di mana kata kunci penanda seperti "lama", "antri", dan "tunggu" saling bersilangan.

Selain itu, analisis pada komponen kata kunci teratas mendeteksi adanya kemunculan beberapa istilah yang bersifat terselubung dan kontradiktif di dalam korpus ulasan negatif, seperti hadirnya token berbobot positif berupa kata "bagus" pada Topik 1. Gejala anomali ini mengonfirmasi bahwa meskipun model IndoBERT sebelumnya telah menyaring ulasan ke dalam polaritas negatif, struktur teks ulasan masyarakat pada Google Maps sering kali mengandung kalimat sarkasme atau bentuk perbandingan komparatif (misalnya, "*pelayanan dokter bagus tapi antreanya lama*"). Munculnya bias kata kunci dan tingginya tingkat *overlap* spasial ini menjadi indikator penting bahwa pemodelan topik menggunakan pendekatan parameter LDA standar dalam penelitian ini masih memerlukan optimasi lebih lanjut, baik melalui penyesuaian nilai *hyperparameter*  $\alpha$  dan  $\beta$  secara lebih ketat maupun penerapan seleksi fitur kata (*feature selection*) yang lebih selektif pada tahap *preprocessing*.

#### 4. Kesimpulan

Berdasarkan hasil penelitian dan pengujian yang telah dilakukan pada sistem hibrida klasifikasi sentimen dan pemodelan topik ulasan pasien, dapat disimpulkan bahwa integrasi model *fine-tuned* IndoBERT mampu memberikan performa generalisasi konteks yang sangat kuat dalam memproses teks tidak terstruktur dari Google Maps. Model IndoBERT berhasil mencapai nilai akurasi menyeluruh (*Accuracy*) sebesar 0,9523 (95,23%) serta nilai rata-rata tertimbang *F1-Score* sebesar 0,9522 (95,22%). Melalui fungsi filterisasi sentimen negatif, model ini terbukti andal dalam mengisolasi ulasan berisi keluhan pasien secara otomatis untuk kemudian diteruskan ke dalam modul algoritma *Latent Dirichlet Allocation* (LDA).

Namun secara kritis, hasil pengujian pada pemodelan topik LDA menyingkap adanya keterbatasan internal yang cukup signifikan dalam menangkap batas semantik teks ulasan informal. Meskipun grafik koherensi

merekomendasikan jumlah 10 topik, visualisasi spasial pyLDAvis menunjukkan adanya fenomena tumpang tindih (*overlap*) yang sangat masif di antara beberapa topik utama. Keterbatasan ini diperparah dengan munculnya anomali kata kunci berbobot positif yang "terselubung" di dalam kluster keluhan negatif (seperti token "bagus").

Hal tersebut mengonfirmasi bahwa model LDA standar masih rentan mengalami bias makna akibat karakteristik bahasa ulasan masyarakat yang sering kali menggunakan kalimat perbandingan komparatif atau sarkasme. Oleh karena itu, bagi penelitian selanjutnya, sangat disarankan untuk melakukan optimasi penyesuaian *hyperparameter* secara lebih ketat, menerapkan seleksi fitur kata (*feature selection*) seperti *stopword* yang lebih selektif pada tahap *preprocessing*, atau mengintegrasikan metode pemodelan topik berbasis *deep learning embeddings* (seperti BERTopic) guna mereduksi *overlap* spasial dan menghasilkan kluster keluhan yang jauh lebih bersih serta spesifik.

## Daftar Pustaka

- [1] A. Harokan, A. D. Priyatno, A. Wahyudi, S. Tinggi, I. Kesehatan, and B. Husada, "Analisis Mutu Pelayanan Terhadap Kepuasan Pasien Rawat Inap Rumah Sakit Umum Daerah Tahun2024," vol. 9, no. 1, 2024.
- [2] D. T. Rindasiwi, P. H. P. Tan, U. Pelita, and H. Jakarta, "The Influence of Hospital Brand Image , Health Service Quality and Patient Satisfaction on Loyalty at Arosuka Regional Hospital," vol. 6, no. 4, pp. 247–267, 2024.
- [3] S. Nabila, T. D. Santi, and H. Hasnur, "Analisis Manajemen Komplain terhadap Kualitas Pelayanan Kesehatan di Puskesmas Banda Raya Kecamatan Banda Raya Kota Banda Aceh," vol. 4, no. 3, pp. 476–489, 2025, doi: 10.55123/insologi.v4i3.5231.
- [4] P. Permatasari and N. A. Rajebta, "Improving Quality of Care on Patient Satisfaction in Health Service Facilities by Rating on Google Maps : Literature Review," vol. 13, no. 2, pp. 276–285, 2025.
- [5] A. Feizollah et al., "The Use of Natural Language Processing to Interpret Unstructured Patient Feedback on Health Services : Scoping Review," vol. 27, pp. 1–13, 2025, doi: 10.2196/72853.
- [6] A. A. Chamid, R. Nindyasari, N. Azizah, and A. Hariyadi, "Analysis of public opinion on the governor candidate debate using LDA and IndoBERT," vol. 4, no. 3, 2025.
- [7] B. Wilie et al., "IndoNLU : Benchmark and Resources for Evaluating Indonesian Natural Language Understanding," pp. 843–857, 2020.
- [8] I. Z. Mustaqim and R. R. Suryono, "A Systematic Literature Review of Topic Modeling Techniques in User Reviews," vol. 11, no. 2, pp. 238–253, 2025.
- [9] D. Darmanto, K. Kharisma, A. R. Muhammad, and E. Wahyudi, "Perbandingan Kinerja Algoritma Jaro-Winkler dan Levenshtein Distance untuk Deteksi Kesalahan Penulisan Bahasa Indonesia pada Karya Ilmiah Mahasiswa," Jurnal Rekayasa Teknologi Informasi (JURTI), vol. 9, no. 2, pp. 112–121, 2025.
- [10] A. -R. Muhammad, A. E. Permasari and I. Hidayah, "Personalized Recommendation of Study Materials Based on Automatic Short Answer Scoring Results," 2022 8th International Conference on Education and Technology (ICET), Malang, Indonesia, 2022, pp. 70-75, doi: 10.1109/ICET56879.2022.9990778.