Institut Riset dan Publikasi Indonesia (IRPI)
MALCOM: Indonesian Journal of Machine Learning and Computer Science
Journal Homepage: https://journal.irpi.or.id/index.php/malcom
Vol. 6 Iss. 4 October 2026, pp: 2201-2210
ISSN(P): 2797-2313 | ISSN(E): 2775-8575

Topic Modeling of Google Reviews Using Latent Dirichlet Allocation for Natural Attractions

Pemodelan Topik Ulasan Google Menggunakan Latent Dirichlet Allocation untuk Objek Wisata Alam

Firmansyah[1*], Diqy Fakhrun Shiddieq[2], Dwi Nurhayati[3]

[1,2,3]Program Studi Bisnis Digital, Fakultas Ekonomi, Universitas Garut, Indonesia

E-Mail: [1]24025122008@fekon.uniga.ac.id, [2]diqy@uniga.ac.id, [3]dwinurhayati@uniga.ac.id

Received Jul 06nd 2026; Revised Aug 14th 2026; Accepted Oct 17th 2026; Available Online Oct 30th 2026
Corresponding Author: Firmansyah
Copyright © 2026 by Authors, Published by Institut Riset dan Publikasi Indonesia (IRPI)

Abstract

The growing volume of Google Reviews provides valuable insights into tourists' experiences, but analyzing unstructured textual data manually is time-consuming. This study aims to identify the main themes in visitor reviews of natural attractions in Garut Regency using Latent Dirichlet Allocation (LDA). The novelty lies in applying a preprocessing pipeline tailored to local review language, including slang normalization and a tourism-specific stopword dictionary, to natural tourism in Garut Regency, which has not previously been examined using LDA. A total of 12,861 reviews from various natural destinations were collected using Instant Data Scraper. The preprocessing stages included case folding, cleaning, normalization, tokenization, stopword removal, and stemming. Several topic configurations were evaluated using coherence scores, and the resulting topics were visualized with pyLDAvis. The optimal model contained five topics, achieving a coherence score of 0.5461. These topics concerned destination accessibility, visiting expenses, coastal tourism and environmental cleanliness, destination attractiveness and comfort, and mountain tourism experiences. The pyLDAvis visualization illustrated relationships among topics and the distribution of representative words. Overall, the findings demonstrate the usefulness of LDA for extracting meaningful insights from tourist reviews and provide practical guidance for destination managers and regional authorities to improve services, facilities, environmental cleanliness, and sustainable natural tourism development in Garut Regency.

Keywords: Coherence Score, Google Review, Latent Dirichlet Allocation (LDA), Natural Tourism, Topic Modeling.

Abstrak

Pertumbuhan jumlah ulasan wisatawan di Google Review menghasilkan data teks yang kaya informasi mengenai pengalaman pengunjung, tetapi sifatnya yang tidak terstruktur menyulitkan analisis manual. Penelitian ini bertujuan mengidentifikasi tema utama dalam ulasan wisata alam di Kabupaten Garut menggunakan metode Latent Dirichlet Allocation (LDA). Kebaruan penelitian terletak pada penerapan preprocessing yang disesuaikan dengan karakteristik bahasa ulasan lokal, meliputi normalisasi kata tidak baku dan penggunaan kamus stopword khusus pariwisata, pada konteks wisata alam Kabupaten Garut yang belum dikaji secara khusus menggunakan LDA. Sebanyak 12.861 ulasan dari berbagai destinasi wisata alam dikumpulkan menggunakan Instant Data Scraper. Data kemudian diproses melalui tahapan case folding, cleaning, normalization, tokenizing, stopword removal, dan stemming. Pemodelan dilakukan dengan menguji beberapa jumlah topik berdasarkan nilai coherence score, kemudian hasilnya divisualisasikan menggunakan pyLDAvis. Model terbaik menghasilkan lima topik dengan nilai coherence 0,5461. Kelima topik tersebut mencakup aksesibilitas destinasi, biaya kunjungan, wisata pantai dan kebersihan lingkungan, keindahan dan kenyamanan destinasi, serta pengalaman wisata pegunungan. Visualisasi pyLDAvis memperlihatkan hubungan antartopik dan distribusi kata dominan. Hasil penelitian menunjukkan bahwa LDA dapat mengekstraksi informasi bermakna dari ulasan wisatawan serta memberikan masukan bagi pengelola destinasi dan pemerintah daerah dalam meningkatkan kualitas layanan, fasilitas, kebersihan lingkungan, dan pengembangan wisata alam berkelanjutan di Kabupaten Garut.

Kata Kunci: *Coherence Score, Google Review, Latent Dirichlet Allocation (LDA), Topic Modeling*, Wisata Alam.

1. PENDAHULUAN

Industri pariwisata sekali lagi menegaskan posisinya sebagai salah satu penopang utama dalam mendorong laju perekonomian nasional seusai pandemi. Merujuk data perkembangan *Tourism Satellite Account* (TSA) tahun 2023, sumbangan sektor pariwisata pada Produk Domestik Bruto (PDB) nasional melesat cukup besar hingga menyentuh 4,67 persen, hampir menyamai kondisi sebelum pandemi [1]. Perkembangan positif ini tak lepas dari makin tingginya pergerakan masyarakat serta pelonggaran aturan perjalanan yang mulai dirasakan sejak 2022. Penguatan tren pemulihan pada 2023 tersebut menjadi pertanda penting bagi wilayah berpotensi wisata besar, salah satunya Kabupaten Garut, untuk memaksimalkan sektor pariwisata sebagai motor perekonomian daerah [2].

Studi tentang analisis ulasan wisata berbasis topic modeling terus tumbuh sebab metode ini sanggup mengolah ulasan berjumlah besar secara otomatis untuk menangkap tema-tema pokok yang menggambarkan pengalaman wisatawan. Jika dibandingkan dengan survei konvensional, penelaahan ulasan daring memberi peneliti informasi yang lebih kaya, apa adanya, dan bebas dari pengaruh susunan pertanyaan peneliti, sehingga persepsi wisatawan terhadap mutu destinasi tergambar secara lebih menyeluruh [3]. Kabupaten Garut memperlihatkan pemulihan sektor pariwisata yang cukup nyata seusai pandemi; angka kunjungan pada 2024 tercatat menembus 3,3 juta orang dan melewati target Pemerintah Daerah. Capaian ini menguatkan kedudukan strategis wilayah yang dikelilingi lanskap pegunungan sekaligus pesisir tersebut sebagai salah satu destinasi unggulan di Jawa Barat [4]. Kondisi ini dipicu oleh bergesernya selera masyarakat yang belakangan lebih menyukai kegiatan di ruang terbuka (*outdoor*) serta wisata bernuansa alam sebagai cara memulihkan kesehatan mental pascapandemi. Pergeseran perilaku ini selaras dengan pandangan literatur pariwisata mutakhir bahwa harapan wisatawan tidak lagi semata bertumpu pada manfaat fungsional destinasi, melainkan turut dipengaruhi pencarian manfaat emosional yang bersifat irasional, misalnya keinginan memperoleh ketenangan dan pengalaman yang utuh [5]. Beberapa penelitian terdahulu telah menerapkan penambangan teks pada ulasan destinasi wisata di Indonesia, seperti analisis sentimen objek wisata Kalimantan Barat berbasis ulasan Google Maps [6] dan pemodelan topik forum pariwisata nasional [7]. Namun, kajian-kajian tersebut umumnya berhenti pada klasifikasi polaritas sentimen atau menggunakan sumber data forum, dan belum banyak yang mengeksplorasi struktur topik laten dari ulasan Google Review objek wisata alam secara spesifik [8].

Perkembangan teknologi informasi telah menggeser kecenderungan komunikasi wisatawan menuju pemanfaatan *User Generated Content* (UGC) sebagai rujukan utama, di mana konten semacam itu terbukti nyata mampu menguatkan citra destinasi (*destination image*) sekaligus menjadi penentu utama dalam menumbuhkan kepuasan (*satisfaction*) dan niat kunjungan ulang (*revisit intention*) wisatawan [9]. Membanjirnya ulasan daring yang tidak terstruktur menjadikan proses penyaringan serta analisis informasi secara manual nyaris tidak mungkin ditempuh. Karena itu, dibutuhkan pendekatan komputasional berbasis *Natural Language Processing* (NLP) seperti pemodelan topik (*topic modeling*) untuk menggali struktur tema tersembunyi secara otomatis. Khusus di bidang pariwisata, penambangan teks atas ulasan daring ini terbukti memberi gambaran pengalaman konsumen yang jauh lebih kaya, apa adanya, dan menyeluruh ketimbang survei tradisional [10][11][12].

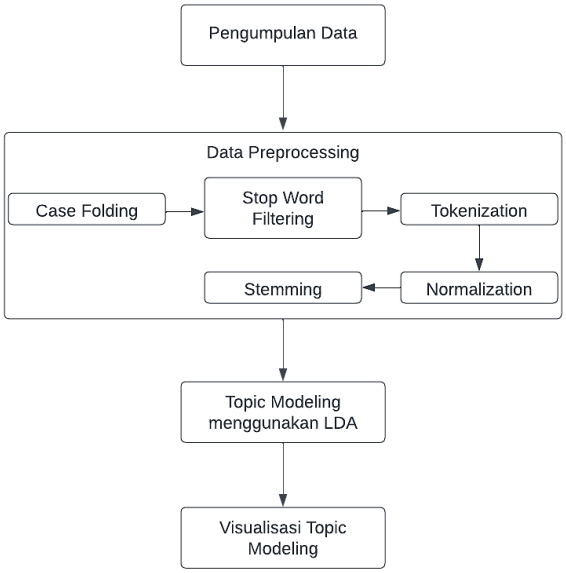
Latent Dirichlet Allocation (LDA) tergolong algoritma topic modeling yang hingga kini tetap luas dipakai karena sanggup mengenali struktur topik laten dari sekumpulan dokumen tanpa label secara efektif. Kendati muncul metode yang lebih baru seperti BERTopic maupun Top2Vec, LDA masih menjadi pilihan yang bersaing untuk analisis data teks sekaligus menghasilkan luaran yang gampang ditafsirkan di berbagai bidang kajian [13]. Kajian-kajian sebelumnya kerap bertumpu pada pendekatan deduktif lewat survei konvensional yang tidak jarang gagal merekam pengalaman konsumen secara utuh [5]. Menggali topik dari ulasan daring memunculkan tantangan teknis tersendiri karena wujudnya berupa teks pendek (*short-text*) yang penuh dengan data bising (*noisy*), bahasa *slang*, dan minimnya konteks [12]. Kajian ini hadir guna menjembatani celah itu dengan mengusung pendekatan induktif lewat algoritma *Latent Dirichlet Allocation* (LDA). Pendekatan tersebut ditempatkan secara khusus pada ekosistem wisata alam Kabupaten Garut guna merangkum gambaran ulasan pengunjung dengan cara yang lebih organik dan bertumpu pada data.

Kebaruan penelitian ini terletak pada perpaduan teknik *Natural Language Processing* (NLP) yang diselaraskan dengan ciri kebahasaan ulasan lokal, mencakup normalisasi kata serta pemakaian kamus stopword bertema pariwisata untuk mempertajam ketepatan ekstraksi topik [14]. Tidak seperti riset terdahulu yang cenderung umum, kajian ini menelaah dimensi khusus seperti mutu daya tarik alam serta faktor aksesibilitas yang menjadi penentu penting dalam pengembangan destinasi wisata [15].

Sasaran pokok kajian ini ialah menggali topik dominan seputar gambaran wisata alam Garut memakai pemodelan LDA. Lewat pendekatan induktif yang bertumpu pada data (*data-driven*) ini, temuan yang diperoleh diharapkan sanggup memperkokoh landasan teori informatika pariwisata sekaligus menyuguhkan bukti empiris langsung dari kacamata pengunjung [5]. Pemahaman yang mendalam atas preferensi nyata wisatawan ini amat penting bagi para *stakeholder* serta pengambil kebijakan dalam menyusun strategi pembangunan destinasi dan peningkatan mutu layanan yang lebih tepat, terarah, dan berorientasi pada konsumen [5] [16].

2. METODOLOGI PENELITIAN

Tahapan penelitian ini digambarkan secara visual pada Gambar 1. Rangkaiannya dimulai dengan mengumpulkan data ulasan pengunjung objek wisata alam di Kabupaten Garut dari platform Google Review. Begitu data terkumpul, langkah selanjutnya berupa *preprocessing* untuk memilah dan menyisihkan unsur kata yang tidak relevan. Proses lalu dilanjutkan dengan penggalian topik dengan metode *Latent Dirichlet Allocation* (LDA), yang ditutup dengan penyajian visualisasi luaran pemodelan melalui *pyLDAvis*.



Gambar 1. Alur Penelitian

2.1. Pengumpulan Data

Kajian ini bertumpu pada sumber data primer berupa ulasan pengguna yang dikumpulkan lewat platform Google *Reviews*. Objek yang ditetapkan mengacu pada tingkat interaksi serta kredibilitas ulasan agar keabsahan data yang diolah tetap terjaga. Pengambilan data ditempuh secara otomatis dengan bantuan *tool Instant Data Scraper*, yang memungkinkan penarikan teks dalam jumlah besar secara efisien. Peneliti menetapkan parameter pencarian yang sesuai supaya ketepatan informasi yang diambil terjamin. Seusai data berhasil ditarik, dijalankan verifikasi konten sebelum akhirnya seluruh dataset dirapikan dan disimpan dalam berkas berformat **.**xlsx. Proses scraping dilakukan pada periode November hingga Desember 2025. Objek wisata yang dianalisis berjumlah 25 destinasi wisata alam di Kabupaten Garut, yang dipilih berdasarkan kriteria inklusi berupa: (1) tergolong objek wisata alam; (2) memiliki halaman ulasan Google Review yang aktif; dan (3) memiliki jumlah ulasan yang memadai untuk dianalisis. Adapun kriteria eksklusi mencakup ulasan yang berupa teks kosong, duplikat, hanya berisi emoji/angka, atau bukan berbahasa Indonesia, yang dibuang pada tahap pembersihan data.

2.2. Preprocessing Data

Tahap preprocessing menempati posisi penting dalam alur topic modeling karena bermaksud menekan noise pada data teks lewat pembersihan, normalisasi, tokenisasi, pembuangan stopwords, serta stemming. Tahap ini turut menaikkan mutu topik yang dihasilkan sehingga penafsirannya menjadi lebih tepat dan mewakili muatan dokumen [11]. Keseluruhan tahapan ini dijalankan menggunakan bahasa pemrograman Python versi 3.x dengan dukungan sejumlah pustaka, yakni Sastrawi untuk stemming Bahasa Indonesia, NLTK untuk tokenizing dan stopword removal, pandas untuk manipulasi data, Gensim untuk pemodelan LDA dan penghitungan coherence score, serta pyLDAvis untuk visualisasi topik:

1. Case Folding & Cleaning: Langkah pembuka untuk menyeragamkan bentuk teks dengan mengubah seluruh huruf menjadi huruf kecil (*lowercase*). Proses ini dibarengi dengan pembersihan data (*cleaning*) guna membuang karakter yang tidak dibutuhkan seperti tanda baca, data kembar, serta keterangan yang tidak konsisten.
2. Normalization: Menyeragamkan kata-kata tidak baku atau menyimpang supaya selaras dengan kaidah bahasa yang berlaku.
3. Tokenizing: Memecah untaian teks menjadi kepingan-kepingan kata yang lebih kecil, yang lazim disebut token.
4. Filtering (Stopword Removal): Menyingkirkan kata-kata yang dinilai kurang informatif atau tidak memberi sumbangan berarti pada muatan dokumen, sehingga tersisa kata-kata yang berbobot informasi.

5. **S**temming: Perubahan akhir yang mengembalikan tiap kata ke bentuk dasarnya (akar kata) agar makna yang muncul lebih konsisten dan tepat.

2.3. Topic Modeling Menggunakan Latent Dirichlet Allocation

Topic modeling merupakan teknik *unsupervised machine learning* yang dipakai untuk menyingkap tema tersembunyi dalam himpunan dokumen teks berukuran besar. Pemodelan topik memandang tiap dokumen terbentuk dari paduan beragam topik dan kata [17]. Sejumlah algoritma topic modeling telah dikembangkan guna menggali informasi laten dari data teks, di antaranya *Latent Semantic Analysis* (LSA), *Probabilistic Latent Semantic Analysis* (PLSA), *Non-negative Matrix Factorization* (NMF), dan *Latent Dirichlet Allocation* (LDA). Dari berbagai metode itu, LDA menjadi salah satu pendekatan yang paling luas dipakai karena sanggup memunculkan topik yang koheren sekaligus luaran yang gampang ditafsirkan pada bermacam jenis data teks [12]. Salah satu metode yang paling luas dipakai dalam pemodelan topik ialah *Latent Dirichlet Allocation* (LDA) [18]. Istilah *latent* mengacu pada elemen yang tersembunyi dalam data, *Dirichlet* berkaitan dengan distribusi peluang yang menerangkan persebaran topik di dalam dokumen serta kata di dalam topik, sementara *allocation* bermakna proses penempatan topik atau kata secara lebih rinci ke dalam dokumen [18]. LDA menggambarkan dokumen sebagai paduan topik, dan tiap topik ditampilkan lewat distribusi kata dengan peluang tertentu [19]. Pada ranah pariwisata, topic modeling telah dipakai secara luas untuk menelusuri beragam tema riset berbasis data besar, mulai dari transformasi digital, pemasaran destinasi, pengalaman wisatawan, sampai pengelolaan destinasi pintar. Pendekatan ini memungkinkan pengenalan pola riset maupun persepsi wisatawan secara otomatis melalui penelaahan himpunan dokumen dalam jumlah besar [20].

Metode ini juga dipandang sanggup menghasilkan topik yang masuk akal secara logis, mudah ditafsirkan, sekaligus mendukung evaluasi model yang bertumpu pada koherensi dan penilaian manusia [18]. Pada langkah ini, penulis memakai dua parameter rujukan, yakni *number of topics* dan *words of topic*, guna memperoleh model dengan jumlah topik yang paling optimal [21]. Dalam kajian ini, tiap topik diwakili oleh sepuluh kata berprobabilitas tertinggi (*top words*) untuk mempermudah penafsiran atas ciri tiap topik. Penetapan jumlah kata tersebut bertujuan menyajikan gambaran yang cukup informatif tanpa menurunkan kejelasan keterkaitan antarkata pada setiap topik yang dihasilkan [12]. Sesudah itu, penulis menelaah number of topics atau jumlah topik lewat evaluasi memakai nilai *coherence*, yakni tolok ukur yang menakar himpunan kata luaran topic modeling menurut tingkat keterpahamannya oleh manusia [22].

Nilai *coherence* dipakai untuk menakar keterkaitan antarkata dalam pemodelan topik, dan topik terbaik ditetapkan berdasarkan nilai *coherence* tertinggi yang diperoleh [14]. Di samping itu, kajian ini turut menerapkan evaluasi dengan pendekatan human judgement guna menafsirkan persebaran kata dari luaran pemodelan topik; pada penilaian manusia, penulis memakai pendekatan intrusi kata untuk menilai topik yang bermakna serupa dan saling terkait sehingga dapat dikenali oleh manusia [18].

2.4. Visualisasi Topik Modeling

Luaran ekstraksi dari pemodelan topik berikutnya disajikan lewat antarmuka visualisasi interaktif berbasis web. Kehadiran perkakas visual ini dinilai amat penting bagi peneliti agar mampu menelaah ciri serta hubungan antartopik secara efisien [11]. Sebab itu, peneliti memakai pustaka pyLDAvis di ekosistem Python untuk memetakan persebaran kata beserta derajat keterkaitannya secara lebih jelas. Pendekatan analitik visual seperti ini secara luas terbukti manjur dalam menelusuri himpunan teks berskala besar guna memahami struktur tema di baliknya [12]. Demi memaksimalkan penelusuran kaitan antara topik dan istilah pembentuknya, *output* visualisasi ini lalu disimpan secara dinamis dalam format .html. [23].

3. HASIL DAN PEMBAHASAN

3.1. Hasil Pengumpulan Data

Penulis menghimpun data berupa ulasan pada tiap objek wisata alam di Kabupaten Garut yang diperoleh dari Google Review memakai Instant Data Scraper. Penarikan data ditempuh secara otomatis melalui penelusuran ulasan yang tersedia di masing-masing objek wisata alam. Berikutnya, data yang terkumpul disimpan dalam berkas berformat .xlsx. Adapun keseluruhan ulasan yang berhasil dihimpun dari seluruh objek wisata alam yang dikaji berjumlah 12.861 data, sebagaimana disajikan pada Tabel 1.

Tabel 1. Hasil pengumpulan data

Tempat Wisata	Jumlah Komentar
Bagendit	291
Cangkuang	157
Cikuray	985
Curug Cihanyawar	213
Curug Cisarua	138
Curug Jagapati	205
Curug Nyogog	137

Tempat Wisata	Jumlah Komentar
Curug Orok	112
-	-
Talaga Bodas	1046
Total	12861

3.2. Hasil Preprocessing Data

Pada tahap ini, penulis menjalankan preprocessing atas 12.861 data ulasan Google Review yang telah dihimpun. Proses preprocessing dikerjakan memakai bahasa pemrograman Python pada perangkat lunak berbasis Python. Langkah pertama ialah case folding, yakni mengubah seluruh huruf kapital menjadi huruf kecil. Setelah itu, dilakukan pembersihan data untuk menyingkirkan karakter yang tidak diperlukan, misalnya tanda baca, angka, simbol, garis miring, serta data yang berulang dan tidak konsisten. Langkah selanjutnya adalah normalisasi, yaitu menyulih kata-kata tidak baku menjadi kata baku. Berikutnya, dijalankan tokenisasi, yakni memilah teks menjadi satuan kata. Tahap sesudahnya berupa filtering, yaitu membuang kata-kata yang tidak memberi sumbangan berarti melalui stopword removal. Tahap penutup adalah stemming, yakni mengembalikan kata ke bentuk dasarnya. Selesai melewati seluruh langkah tersebut, data disimpan dalam format .xlsx. Adapun contoh luaran preprocessing data ulasan tersaji pada Tabel 2.

Tabel 2. Hasil *Preprocessing* Data

Proses	Hasil
Data colection	Berakit rakit disini, ditengah situ Cangkuang ke seberang lokasi yang ada Candi Cangkuang dan permukiman adat.
Case Folding	berakit rakit disini ditengah situ cangkuang ke seberang lokasi yang ada candi cangkuang dan pemukiman adat
Filtering	berakit rakit ditengah situ cangkuang seberang lokasi candi cangkuang pemukiman adat
Tokenization	['berakit', 'rakit', 'ditengah', 'situ', 'cangkuang', 'seberang', 'lokasi', 'candi', 'cangkuang', 'pemukiman', 'adat']
Normalization	['berakit', 'rakit', 'ditengah', 'situ', 'cangkuang', 'seberang', 'lokasi', 'candi', 'cangkuang', 'pemukiman', 'adat']
Stemming	['rakit', 'rakit', 'tengah', 'situ', 'cangkuang', 'seberang', 'lokasi', 'candi', 'cangkuang', 'mukim', 'adat']

3.3. Hasil Topic Modeling Menggunakan Latent Dirichlet Allocation

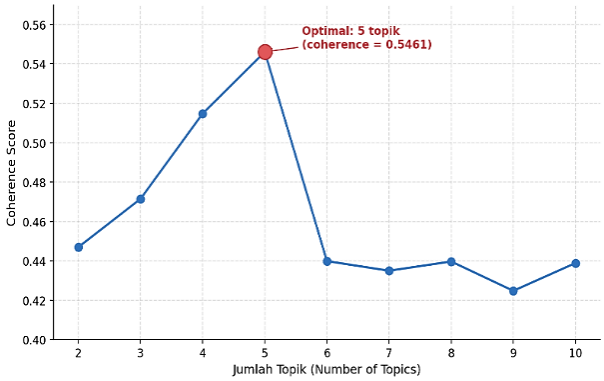
Pada tahap topic modeling dengan Latent Dirichlet Allocation (LDA), penulis memanfaatkan library Gensim yang berguna untuk menggali topik semantik secara otomatis dari himpunan dokumen secara efisien. Langkah pertama yang ditempuh ialah menetapkan jumlah kata pembentuk topik, atau words of topic, sebanyak 10 kata supaya topik yang muncul tidak saling tumpang tindih. Sesudah itu, model diuji memakai nilai coherence score sebagai tolok ukur utama, dengan jumlah topik (number of topics) sebagai dasar pengujian. Coherence score sendiri merupakan metrik untuk menakar mutu serta tingkat keterpahaman topik, sehingga kian tinggi nilainya, kian gampang topik itu dicerna manusia [24]. Guna memperoleh model terbaik, penulis menjalankan pengujian sebanyak 50 passes, dan hasil nilai coherence tersaji pada Tabel 3.

Tabel 3. Hasil Nilai Coherence

Topik	Nilai Coherence
2	0.4470
3	0.4715
4	0.5148
5	0.5461
6	0.4399
7	0.4350
8	0.4397
9	0.4248
10	0.4389

Merujuk keterangan pada Tabel 3, tampak bahwa tiap jumlah topik memunculkan nilai coherence yang beragam. Nilai coherence paling tinggi diraih pada susunan 5 topik, yakni sebesar 0.5461. Karena itu, model beranggotakan 5 topik ditetapkan sebagai model terbaik untuk mewakili muatan data ulasan.

Hubungan antara jumlah topik dan nilai coherence disajikan pada Gambar 2. Grafik memperlihatkan tren peningkatan nilai coherence dari 2 hingga 5 topik, mencapai puncak pada konfigurasi 5 topik (0,5461), lalu menurun konsisten pada jumlah topik yang lebih besar. Pola ini menegaskan bahwa 5 topik merupakan titik optimal, karena penambahan topik justru menurunkan keterpaduan antarkata.



Gambar 2. Hubungan Jumlah Topik terhadap Nilai *Coherence*

Luaran pemodelan lalu ditampilkan lewat sehimpunan kata yang menggambarkan ciri tiap topik. Masing-masing topik diberi label 0, 1, 2, 3, dan 4 sesuai peluang atau frekuensi kemunculan kata yang mewakili topik tertentu. Berikutnya, penulis menjalankan evaluasi dengan pendekatan penilaian manusia (human judgement) untuk menafsirkan tiap topik dari kata-kata yang paling kerap muncul pada model yang terbentuk. Luaran penafsiran pemodelan topik tersebut ditampilkan pada Tabel 4.

Tabel 4. Hasil Interpretasi Model 5 Topik

Topik	Probabilitas* Kata	Analisis Topik
0	0.200*"jalan" + 0.057*"akses" + 0.047*"lokasi" + 0.032*"motor" + 0.027*"talaga" + 0.024*"bodas" + 0.023*"mobil" + 0.021*"curug" + 0.019*"jam" + 0.017*"ojek"'	Kondisi akses jalan, lokasi destinasi, serta kemudahan transportasi seperti motor, mobil, dan ojek menuju objek wisata.
1	0.053*"masuk" + 0.039*"tiket" + 0.038*"enak" + 0.030*"bayar" + 0.029*"makan" + 0.028*"parkir" + 0.026*"ribu" + 0.026*"pas" + 0.026*"harga" + 0.026*"murah"'	Harga tiket masuk, biaya parkir, harga makanan, serta persepsi wisatawan terhadap keterjangkauan biaya selama berkunjung.
2	0.203*"pantai" + 0.046*"sampah" + 0.043*"sayang" + 0.028*"ombak" + 0.024*"laut" + 0.023*"pasir" + 0.023*"karang" + 0.022*"santolo" + 0.019*"renang" + 0.017*"keren"'	Pengalaman wisatawan terhadap wisata pantai, mencakup kebersihan lingkungan, kondisi ombak, pasir, laut, serta daya tarik Pantai Santolo.
3	0.069*"bagus" + 0.056*"indah" + 0.035*"pandang" + 0.030*"tempat" + 0.026*"wisata" + 0.025*"bersih" + 0.024*"air" + 0.023*"cocok" + 0.020*"nyaman" + 0.019*"lumayan"	Penilaian wisatawan terhadap keindahan pemandangan, kebersihan, kenyamanan, dan kelayakan destinasi sebagai tempat berwisata.
4	0.038*"gunung" + 0.029*"garut" + 0.026*"kawah" + 0.016*"salah" + 0.015*"kali" + 0.014*"puncak" + 0.013*"daki" + 0.013*"jawa" + 0.013*"jalur"'	Pengalaman wisatawan dalam menjelajahi kawasan pegunungan serta kawah di Garut, mencakup kegiatan pendakian, jalur, dan panorama puncak.

Berdasarkan hasil interpretasi terhadap Tabel 4, pemodelan *Latent Dirichlet Allocation* (LDA) beranggotakan 5 topik memunculkan lima kelompok tema yang menggambarkan fokus utama ulasan wisatawan atas destinasi wisata alam di Kabupaten Garut. Topik 0 didominasi kata jalan, akses, lokasi, motor, mobil, dan ojek, sehingga dapat dimaknai sebagai tema aksesibilitas destinasi, yang menyangkut kondisi jalan serta kemudahan wisatawan menjangkau lokasi wisata memakai beragam moda transportasi. Topik 1 memuat kata seperti tiket, masuk, bayar, parkir, harga, murah, dan makan, yang menandakan topik ini berkaitan dengan biaya berkunjung dan pengeluaran wisatawan selama di destinasi. Berikutnya, Topik 2 didominasi kata pantai, sampah, ombak, laut, pasir, karang, dan Santolo, sehingga mewakili keadaan wisata pantai, terutama yang menyangkut daya tarik alam dan kebersihan lingkungan. Topik 3 menghimpun kata bagus, indah, pandang, bersih, nyaman, dan cocok, yang menggambarkan persepsi positif wisatawan atas mutu destinasi, khususnya dari sisi keindahan, kebersihan, dan kenyamanan. Sementara itu, Topik 4 tersusun atas kata gunung, Garut, kawah, puncak, daki, dan jalur, yang menunjukkan tema wisata pegunungan, dengan penekanan pada aktivitas pendakian serta ciri kawasan pegunungan di Kabupaten Garut. Model lima topik dipilih sebagai model terbaik karena meraih coherence score sebesar 0,5461, yang menandakan keterkaitan antarkata dalam tiap topik

tergolong lebih ajek sehingga menghasilkan penafsiran yang lebih gampang dipahami dibanding susunan jumlah topik lainnya.

Kelima topik yang dihasilkan memberikan gambaran mengenai faktor-faktor yang paling sering menjadi perhatian wisatawan ketika memberikan ulasan terhadap destinasi wisata alam di Kabupaten Garut. Aspek aksesibilitas menunjukkan bahwa kemudahan menuju lokasi wisata masih menjadi salah satu unsur penting yang memengaruhi pengalaman berkunjung. Dari sisi ekonomi, biaya tiket masuk, parkir, dan konsumsi juga menjadi pertimbangan dalam membentuk penilaian wisatawan terhadap suatu destinasi. Pada kawasan wisata pantai, keberadaan kata sampah mengindikasikan bahwa kebersihan lingkungan masih menjadi isu yang perlu mendapat perhatian, meskipun keindahan pantai, ombak, dan hamparan pasir tetap menjadi daya tarik utama. Selain itu, dominasi kata-kata bernilai positif pada topik keindahan destinasi menunjukkan bahwa panorama alam, kebersihan, dan kenyamanan merupakan keunggulan yang banyak diapresiasi oleh pengunjung. Di sisi lain, munculnya tema wisata pegunungan memperlihatkan bahwa aktivitas pendakian, keberadaan kawah, serta panorama puncak merupakan karakteristik yang membedakan wisata alam Kabupaten Garut dengan destinasi lainnya. Secara keseluruhan, hasil pemodelan ini memberikan informasi yang bermanfaat bagi pengelola destinasi dalam menentukan prioritas pengembangan, terutama pada peningkatan akses menuju lokasi wisata, pengelolaan kebersihan lingkungan, penyediaan fasilitas yang memadai, serta pelestarian daya tarik alam sebagai faktor utama yang memengaruhi kepuasan wisatawan.

Tabel 5. Distribusi Topik Dominan per Kategori Objek Wisata

Kategori objek	Jumlah ulasan	Aksesibilitas (%)	Biaya (%)	Keindahan & kenyamanan (%)	Wisata pantai & kebersihan (%)	Wisata pegunungan (%)
Pantai	3908	1,4	7,2	66,5	23,3	1,7
Pegunungan	2555	2,4	7,4	65	5,6	19,7
Danau/Situ	1931	6,1	11,8	77,6	1,8	2,7
Curug (air terjun)	1958	12	3,6	77,9	2,7	3,9
Lainnya	942	2,5	12,7	69,4	13,3	2

Berdasarkan Tabel 5, distribusi topik dominan menunjukkan bahwa topik keindahan dan kenyamanan menjadi tema yang paling banyak dibahas pada seluruh kategori objek wisata. Persentase topik ini berkisar antara 65,0% hingga 77,9%, dengan nilai tertinggi terdapat pada kategori curug (77,9%), diikuti danau/situ (77,6%), lainnya (69,4%), pantai (66,5%), dan pegunungan (65,0%). Temuan tersebut menunjukkan bahwa pengalaman wisatawan pada berbagai jenis objek wisata alam di Kabupaten Garut lebih banyak dipengaruhi oleh persepsi terhadap panorama, kondisi lingkungan, dan kenyamanan selama berkunjung.

Meskipun demikian, setiap kategori objek wisata memiliki karakteristik pembahasan yang berbeda. Pada kategori pantai, topik wisata pantai dan kebersihan menempati proporsi terbesar kedua (23,3%), mengindikasikan bahwa kondisi kebersihan kawasan pesisir menjadi perhatian penting wisatawan. Pada kategori pegunungan, topik wisata pegunungan memiliki proporsi sebesar 19,7%, yang mencerminkan pembahasan mengenai karakteristik lingkungan pegunungan sebagai bagian dari pengalaman berwisata. Sementara itu, kategori curug menunjukkan proporsi topik aksesibilitas tertinggi dibandingkan kategori lainnya (12,0%), yang mengindikasikan bahwa kemudahan maupun kendala dalam mencapai lokasi wisata menjadi aspek yang cukup sering disampaikan dalam ulasan. Di sisi lain, kategori danau/situ dan lainnya memperlihatkan distribusi topik yang lebih terpusat pada aspek keindahan dan kenyamanan, sedangkan topik biaya, aksesibilitas, maupun tema spesifik lainnya muncul dengan proporsi yang relatif lebih rendah.

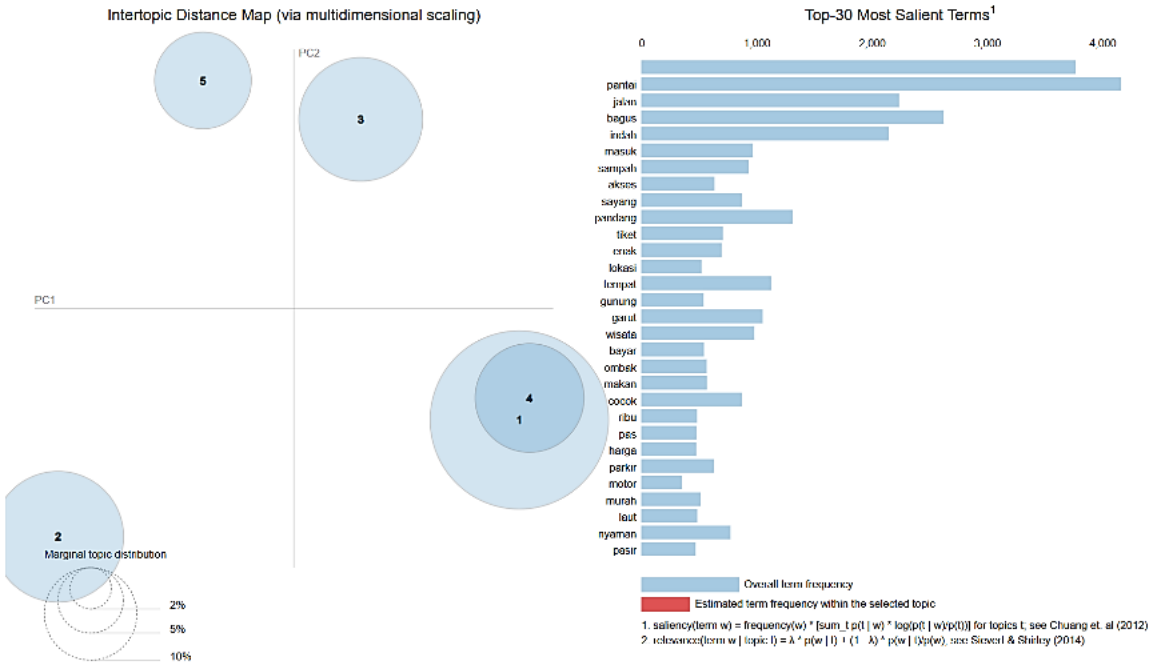
Secara keseluruhan, hasil tersebut menunjukkan bahwa persepsi wisatawan terhadap keindahan dan kenyamanan merupakan faktor utama yang membentuk ulasan pada seluruh kategori objek wisata alam. Namun, keberadaan topik-topik spesifik pada masing-masing kategori mengindikasikan bahwa karakteristik destinasi turut memengaruhi fokus pembahasan wisatawan dalam memberikan ulasan.

3.4. Hasil Visualisasi Pemodelan Topik

Selesai menjalankan pemodelan topik dengan Latent Dirichlet Allocation (LDA), penulis lalu memvisualisasikan susunan 5 topik memakai pyLDAvis guna mengetahui persebaran kata dari tiap topik yang terbentuk. Gambar 2 berikut menyajikan visualisasi dari pemodelan topik tersebut.

Visualisasi pyLDAvis pada Gambar 3 menggambarkan luaran ekstraksi pemodelan topik yang disusun menjadi lima klaster. Panel di sisi kiri menyajikan peta jarak antar-topik (*Intertopic Distance Map*) yang memetakan letak spasial kelima klaster tersebut lewat teknik multidimensional scaling (MDS). Ukuran tiap lingkaran menandai peluang marjinal masing-masing topik atas keseluruhan korpus; kian besar radius lingkaran, kian tinggi tingkat kemunculan topik itu dalam dokumen. Sementara itu, jarak antar-lingkaran menandakan tingkat kedekatan semantik. Dari pengamatan visual, terdapat tumpang tindih (*overlap*) yang cukup nyata antara Topik 1 dan Topik 4 pada wilayah kuadran kanan bawah. Gejala ini menandakan tingginya

irisan leksikal atau kemiripan tema bahasan di antara kedua klaster tersebut. Sebaliknya, tersebarnya Topik 2, 3, dan 5 pada kuadran yang berlainan menegaskan bahwa ketiga topik itu memiliki ciri bahasan yang lebih mandiri dan khas.



Gambar 3. Visualisasi Pemodelan Topik

Di panel sisi kanan, tersaji diagram batang mendatar yang memuat 30 term beserta nilai kepentingannya (*Top-30 Most Salient Terms*) yang mendominasi korpus data. Sejumlah term itu meliputi leksikon seperti pantai, jalan, bagus, indah, masuk, sampah, akses, sayang, pandang, tiket, enak, lokasi, tempat, gunung, garut, wisata, bayar, ombak, makan, cocok, ribu, pas, harga, parkir, motor, murah, laut, nyaman, dan pasir. Berdasarkan telaah atas persebaran term dominan tersebut, dapat ditarik simpulan bahwa fokus perbincangan dalam himpunan dokumen erat kaitannya dengan penilaian pengalaman pengunjung di destinasi wisata alam Kabupaten Garut. Lebih jauh, penggalian term ini mengelompokkan persepsi wisatawan ke dalam sejumlah dimensi pokok, yaitu: daya tarik wisata bahari (pantai, ombak, laut, pasir) serta pegunungan (gunung, indah, pandang), infrastruktur dan aksesibilitas (jalan, akses, lokasi, parkir, motor), aspek nilai ekonomi (tiket, bayar, harga, murah, ribu, pas), serta penilaian atas mutu kenyamanan, fasilitas, dan kebersihan destinasi (bagus, nyaman, enak, cocok, tempat, makan, sampah). Dimensi-dimensi inilah yang menjadi landasan pola pengelompokan topik secara menyeluruh dalam penelitian ini.

Konfigurasi lima topik yang dihasilkan penelitian ini sejalan dengan sejumlah studi pemodelan topik pariwisata lainnya. Ali dkk. [25] yang menerapkan LDA pada ulasan TripAdvisor kota Marrakech memperoleh empat topik sebagai dimensi optimal, sementara Putu dkk. [26] pada ulasan pariwisata Lombok berbahasa Indonesia turut menunjukkan bahwa LDA mampu merangkum pengalaman wisatawan ke dalam sejumlah topik yang ringkas dan interpretatif. Kedekatan jumlah topik ini menegaskan bahwa ulasan destinasi wisata cenderung terkelompok pada sejumlah dimensi pengalaman yang terbatas namun konsisten lintas konteks geografis dan bahasa.

4. KESIMPULAN

Penelitian ini berhasil menggali topik dominan dari 12.861 ulasan pengunjung objek wisata alam di Kabupaten Garut menggunakan pemodelan Latent Dirichlet Allocation (LDA). Konfigurasi lima topik terbukti paling optimal dengan coherence score tertinggi (0,5461), mencakup dimensi aksesibilitas, biaya berkunjung, wisata pantai dan kebersihan, keindahan-kenyamanan destinasi, serta wisata pegunungan. Kelima dimensi ini menjawab tujuan penelitian dalam mengidentifikasi tema utama yang menjadi perhatian wisatawan.

Secara teoretis, penelitian ini menambah khazanah literatur serta kajian tourism informatics dengan menawarkan pendekatan induktif (data-driven) untuk memahami perilaku wisatawan. Perpaduan teknik Natural Language Processing (NLP) dan algoritma LDA pada dataset ulasan lokal menunjukkan bahwa pemodelan topik komputasional sanggup menggali struktur tema dan pengalaman emosional wisatawan secara lebih organik dan menyeluruh dibandingkan dengan survei konvensional.

Secara praktis, temuan penelitian ini amat berguna untuk menopang pengelola wisata, pemerintah daerah, serta pemangku kepentingan terkait di Kabupaten Garut. Penggalian dimensi persepsi wisatawan ini dapat diterapkan sebagai pijakan dalam merancang strategi pengembangan destinasi yang berpusat pada konsumen, khususnya dalam mengutamakan perbaikan aksesibilitas jalan, penguatan pengelolaan kebersihan lingkungan di kawasan pesisir, serta penyempurnaan fasilitas demi menunjang kenyamanan pengunjung.

Kendati menyuguhkan wawasan empiris yang menyeluruh, kajian ini menyimpan sejumlah keterbatasan. Pertama, pengambilan data ulasan pengunjung hanya berasal dari satu platform, yakni Google Review. Kedua, cakupan pengamatan hanya dibatasi pada objek wisata alam, sehingga belum mewakili keseluruhan lanskap pariwisata Kabupaten Garut. Ketiga, korpus teks yang ditelaah hanya berpusat pada ulasan berbahasa Indonesia. Untuk riset mendatang, disarankan memadukan data dari beragam platform Online Travel Agent (OTA) lain, memperluas jangkauan pada kategori wisata buatan maupun budaya, serta merangkul analisis multibahasa agar perspektif wisatawan lintas demografi tertangkap lebih luas.

REFERENSI

[1] Kementerian Pariwisata Republik Indonesia, "No Title," Kementerian Pariwisata RI. [Online]. Available: https://kemenpar.go.id/direktori-statistik/perkembangan-tourism-satellite-account-indonesia-tahun-2023

[2] D. P. dan K. K. Garut, "No Title," Satu Data Kabupaten Garut. [Online]. Available: https://satudata.garutkab.go.id/datasets/jumlah-kunjungan-wisatawan-di-kabupaten-garut-4616/

[3] Z. Shang and J. M. Luo, "Current Issues in Tourism Topic modelling for wildlife tourism online reviews : analysis of quality factors," *Curr. Issues Tour.*, pp. 1–15, 2022, doi: 10.1080/13683500.2022.2086107.

[4] D. Maspupah, "The Analysis of Visitor Journey of Tourists in Garut Regency," vol. 3, pp. 139–149, 2023, doi: 10.35313/jtospolban.v3i3.103.

[5] I. Sutherland, Y. Sim, S. K. Lee, J. Byun, and K. Kiatkawsin, "Topic Modeling of Online Accommodation Reviews via Latent Dirichlet Allocation," 2020, doi: 10.3390/su12051821.

[6] A. Rifa'i, H. Sujaini, and D. Prawira, "Sentiment Analysis Objek Wisata Kalimantan Barat Pada Google Maps Menggunakan Metode Naive Bayes," *JEPIN (Jurnal Edukasi dan Penelit. Inform.*, vol. 7, no. 3, pp. 400–407, 2021, doi: 10.26418/jp.v7i3.48132.

[7] "Analisis Topik pada Forum Pariwisata di Indonesia Menggunakan Latent Dirichlet Allocation," *Teknika*, vol. 17, no. 1, pp. 8–14, 2023, doi: 10.35134/teknika.v17i1.6893.

[8] Y. A. Singgalen, "Analisis Sentimen dan Pemodelan Topik dalam Optimalisasi Pemasaran Destinasi Pariwisata Prioritas di Indonesia," *J. Inf. Syst. Informatics*, vol. 4, no. 1, pp. 459–470, 2021.

[9] V. Willyanto and H. K. Tunjungsari, "Peran User Generated Content dan Destination Image dalam Membentuk Satisfaction dan Revisit Intention ke Candi Borobudur," vol. 9, no. 1, pp. 261–274, 2025.

[10] T. Modelling, "Indonesian Journal of Computer Science," vol. 12, no. 1, pp. 3326–3339, 2023.

[11] D. Buenaño-fernandez, S. Luján-mora, M. González, and D. Gil, "Text Mining of Open-Ended Questions in Self-Assessment of University Teachers : An LDA Topic Modeling Approach," vol. 8, 2020, doi: 10.1109/ACCESS.2020.2974983.

[12] R. Albalawi, T. H. Yeap, and M. Benyoucef, "Using Topic Modeling Methods for Short-Text Data : A Comparative Analysis," vol. 3, no. July, pp. 1–14, 2020, doi: 10.3389/frai.2020.00042.

[13] R. Egger, J. Yu, and J. Yu, "A Topic Modeling Comparison Between LDA , NMF , Top2Vec , and BERTopic to Demystify Twitter Posts Making Sense of Social Media Using," vol. 7, no. May, pp. 1–16, 2022, doi: 10.3389/fsoc.2022.886498.

[14] U. N. Khadijah and N. Cahyono, "The Indonesian Journal of Computer Science," vol. 13, no. 4, pp. 6075–6086, 2024, doi: https://doi.org/10.33022/ijcs.v13i4.3816.

[15] M. S. Nugraha and M. I. Triguna, "Analisis Potensi Wisata Alam Situ Bagendit di Kabupaten Garut," pp. 24–32, 2024, doi: 10.33364/jidar.

[16] H. Zhang, T. Daim, and Y. Peggy, "Technological Forecasting & Social Change Integrating patent analysis into technology roadmapping : A latent dirichlet allocation based technology assessment and roadmapping in the field of Blockchain," *Technol. Forecast. Soc. Chang.*, vol. 167, no. August 2019, p. 120729, 2021, doi: 10.1016/j.techfore.2021.120729.

[17] U. Pahlawan, T. Tambusai, M. H. Ababil, G. Joshua, and B. Setiawan, "Jurnal Pendidikan dan Konseling Dirichlet Allocation (LDA)," vol. 4, pp. 6185–6196, 2022.

[18] E. Puspita, D. F. Shiddieq, and F. F. Roji, "Topic Modeling on Online News Media Using Latent Diriclet Allocation (Case Study Somethinc Brand) Pemodelan Topik pada Media Berita Online Menggunakan Latent Dirichlet Allocation (Studi Kasus Merek Somethinc)," vol. 4, no. April, pp. 481–489, 2024.

[19] J. Chang, S. Gerrish, C. Wang, and D. M. Blei, "Reading Tea Leaves : How Humans Interpret Topic Models," pp. 1–9.

[20] P. Madz and M. Valeri, "Digital transformation in tourism : bibliometric literature review based on machine learning approach s," vol. 26, no. 7, pp. 177–205, 2026, doi: 10.1108/EJIM-09-2022-0531.

[21] J. Penerapan, T. Informasi, W. Yohanes, Y. Mramra, and Y. Nataliani, "IT-EXPLORE Penentuan bidang unggulan akademik universitas melalui metode topik Linear Discriminant Analysis (LDA)," vol. 04, pp. 93–105, 2025, doi: 10.24246/itexplore.v4i1.2025.pp93-105.

[22] A. Hoyle, A. Hian-cheong, J. Boyd-graber, and P. Resnik, "Is Automated Topic Model Evaluation Broken ?: The Incoherence of Coherence," no. NeurIPS, 2021.

[23] C. Sievert, K. E. Shirley, and N. York, "LDAvis : A method for visualizing and interpreting topics," pp. 63–70, 2014.

[24] N. Cahyono, N. Nurcahyo, A. Fauzan, and R. Agung, "Analisa Perbandingan Latent Semantic Indexing (LSI) dan Latent Dirichlet Allocation (LDA) untuk Topic Modelling Aplikasi Identitas Kependudukan Digital (IKD)," vol. 6, no. 3, pp. 1638–1647, 2024, doi: 10.47065/bits.v6i3.5970.

[25] T. Ali, B. Omar, and K. Soulaimane, "Analyzing tourism reviews using an LDA topic-based sentiment analysis approach," *MethodsX*, vol. 9, p. 101894, 2022, doi: 10.1016/j.mex.2022.101894.

[26] N. L. P. M. Putu, A. Z. Amrullah, and Ismarmiaty, "Analisis Sentimen dan Pemodelan Topik Pariwisata Lombok Menggunakan Algoritma Naive Bayes dan Latent Dirichlet Allocation," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 1, pp. 123–131, 2021, doi: 10.29207/resti.v5i1.2587.