JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 5, Oktober 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v15i5.10101
Hal 1362-1374
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Pemodelan Topik Ulasan Tidak Terstruktur TikTok Shop Seller Center Menggunakan LDA dan BERTopic Berbasis Coherence Score

Salsabila Savitri, Nur Cahyo Hendro Wibowo*, Maya Rini Handayani

Fakultas Sains dan Teknologi, Teknologi Informasi, UIN Walisongo, Semarang, Indonesia
Email: [1]23080960004@student.walisongo.ac.id, [2,*]nur_cahyo@walisongo.ac.id, [3]maya@walisongo.ac.id
Email Penulis Korespondensi: nurcahyo@walisongo.ac.id
Submitted **30-09-2026**; Accepted **15-10-2026**; Published **31-10-2026**

Abstrak

Ulasan aplikasi *social commerce* di Google Play Store merupakan teks tidak terstruktur yang sulit dianalisis secara manual. Penelitian ini memodelkan topik ulasan aplikasi TikTok Shop Seller Center dengan *Latent Dirichlet Allocation* (LDA) dan BERTopic, lalu membandingkannya dengan tiga ukuran *coherence score*, yaitu C_V, UMass, dan C_NPMI. Kontribusi penelitian ini adalah perbandingan kedua metode dengan kriteria pemilihan parameter, data masukan, dan prosedur evaluasi yang sama. Sebanyak 2.480 ulasan diambil dengan google-play-scraper pada 14 Juli 2026 (tanggal ulasan 13 Juni 2022 sampai 13 Juli 2026). Setelah *case folding*, *cleaning*, tokenisasi, penghapusan *stopword*, dan *stemming* dengan Sastrawi, tersisa 2.234 dokumen. Pencarian jumlah topik LDA pada k = 2 sampai 15 memilih k = 3 dengan C_V 0,6107, sedangkan pengujian min_cluster_size BERTopic pada 10, 15, 20, 25, dan 30 memilih nilai 15 dengan lima topik dan C_V 0,6011. Pada evaluasi akhir dengan 10 kata teratas per topik, LDA memperoleh C_V 0,6107, UMass -2,0906, dan C_NPMI 0,0133, sedangkan BERTopic memperoleh 0,6011, -4,1687, dan 0,0040. LDA lebih tinggi pada ketiga ukuran, tetapi selisih C_V dan C_NPMI kecil, sedangkan selisih UMass lebih jelas. Topik yang muncul berkaitan dengan jual-beli dan pengelolaan toko, pesanan, pengiriman, potongan biaya, serta akun dan kemudahan penggunaan. Interpretasi topik belum divalidasi oleh pakar atau penilai manusia dan perlu dilakukan pada penelitian selanjutnya.

Kata Kunci: BERTopic; *coherence score*; LDA; pemodelan topik; TikTok Shop Seller Center

Abstract

Reviews of social commerce applications on Google Play Store are unstructured texts that are difficult to analyze manually. This study models the topics of TikTok Shop Seller Center reviews using Latent Dirichlet Allocation (LDA) and BERTopic, then compares them with three coherence scores: C_V, UMass, and C_NPMI. Its contribution is a comparison of both methods using the same parameter selection criterion, input data, and evaluation procedure. A total of 2,480 reviews were collected with google-play-scraper on 14 July 2026 (review dates 13 June 2022 to 13 July 2026). After case folding, cleaning, tokenization, stopword removal, and stemming with Sastrawi, 2,234 documents remained. The LDA topic number search (k = 2 to 15) selected k = 3 (C_V 0.6107), and testing the BERTopic min_cluster_size (10, 15, 20, 25, 30) selected 15, giving five topics (C_V 0.6011). In the final evaluation using the top 10 words per topic, LDA obtained a C_V of 0.6107, a UMass of -2.0906, and a C_NPMI of 0.0133, while BERTopic obtained 0.6011, -4.1687, and 0.0040. LDA scored higher on all three measures, although the C_V and C_NPMI differences are small and the UMass difference is clearer. The topics concern buying and selling, orders, shipping, fee deductions, and accounts. The topic interpretations have not been validated by experts or human raters, which remains for future work.

Keywords: BERTopic; coherence score; LDA; topic modeling; TikTok Shop Seller Center

1. PENDAHULUAN

Penggunaan media sosial di Indonesia terus menjadi dasar bagi tumbuhnya platform *e-commerce* berbasis video pendek. Laporan Digital 2025: Indonesia dari DataReportal mencatat 143 juta identitas pengguna media sosial pada awal tahun 2025, setara dengan 50,2% dari total populasi [1]. TikTok Shop Seller Center adalah aplikasi yang disediakan TikTok Shop bagi penjual untuk mengelola produk, pesanan, pengiriman, promosi, dan operasional toko melalui perangkat seluler. Pengguna aplikasi ini dapat menulis ulasan di Google Play Store. Ulasan tersebut berupa teks bebas yang tidak terstruktur, berisi pendapat dan pengalaman pengguna, dan belum tertata secara sistematis. Data teks seperti ini kaya informasi, tetapi sulit dianalisis secara manual karena jumlah dan ragam isinya [2]. Karena itu, diperlukan cara yang otomatis dan terukur untuk menemukan topik-topik utama yang dibicarakan dalam ribuan ulasan tersebut.

Salah satu pendekatan yang dapat dilakukan adalah pemodelan topik (*topic modeling*), yaitu teknik pembelajaran mesin tanpa pengawasan untuk menemukan struktur tema tersembunyi dalam kumpulan dokumen yang besar [3]. Latent Dirichlet Allocation (LDA) adalah model probabilistik generatif yang menggambarkan setiap dokumen sebagai campuran dari beberapa topik laten [3], dan telah banyak dipakai untuk menganalisis ulasan aplikasi [4]. BERTopic, yang diperkenalkan oleh Grootendorst [5], bekerja dengan cara berbeda. Metode ini mengubah dokumen menjadi *embedding* dari model bahasa yang sudah dilatih, mengelompokkannya dengan HDBSCAN, lalu memilih kata perciri tiap kelompok dengan *class-based TF-IDF* (c-TF-IDF). Berbeda dengan LDA yang bertumpu pada kemunculan bersama kata, BERTopic memakai kemiripan makna dari embedding sehingga berpotensi mengelompokkan ulasan yang maknanya serupa walaupun kata-katanya berbeda [5].

Sejumlah penelitian telah membandingkan LDA dan BERTopic, dan ringkasannya disajikan pada Tabel 1. Pada ulasan aplikasi mitra pengemudi GoPartner, BERTopic memperoleh koherensi tertinggi sebesar 0,6905 pada data tertokenisasi, sedangkan LDA terbaik pada data yang di-*stemming* dan NMF terendah [6]. Pada sekitar 150.000 judul berita dengan 40 topik, BERTopic memiliki koherensi sedikit di atas NMF dan konsistensi lebih baik pada penilaian oleh manusia dan model bahasa besar, sementara LDA tertinggal cukup jauh [7]. Pada data Twitter, penilaian berdasarkan

Copyright © 2026 The **Salsabila Savitri**, Page 1362
This Journal is licensed under a Creative Commons Attribution 4.0 International License

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 5, Oktober 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v15i5.10101
Hal 1362-1374
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

interpretasi manusia menempatkan BERTopic dan NMF di atas Top2Vec dan LDA [8]. Pada abstrak biomedis, akurasi LDA dan BERTopic setara menurut penilaian pakar, tetapi klaster BERTopic lebih padat dan terpisah pada visualisasi t-SNE [9]. Sebaliknya, pada perbandingan lima metode non-*embedding* untuk teks pendek, LDA dan NMF memberikan topik yang lebih bermakna [10].

Tabel 1. Ringkasan penelitian terdahulu yang relevan dan posisi penelitian ini

Ref.	Objek dan data	Metode	Metrik evaluasi	Hasil utama
[4]	Ulasan aplikasi Netflix di Google Play Store (3.028 ulasan)	LDA	Coherence score	Dua belas topik dipilih berdasarkan koherensi tertinggi
[6]	Ulasan aplikasi GoPartner di Google Play Store	BERTopic, LDA, NMF	Coherence score	BERTopic 0,6905 tertinggi pada data tertokenisasi; LDA terbaik pada data stemming; NMF terendah
[7]	Sekitar 150.000 judul berita	LDA, NMF, BERTopic (40 topik)	Koherensi berbasis Word2Vec, penilai manusia, dua model bahasa besar	BERTopic sedikit di atas NMF pada koherensi dan konsistensi lebih baik pada penilaian oleh manusia dan model bahasa; LDA tertinggal jauh
[8]	31.800 tweet unik (tagar #covid dan #travel)	LDA, NMF, Top2Vec, BERTopic	Interpretasi manusia	BERTopic dan NMF paling unggul, diikuti Top2Vec dan LDA
[9]	1.837 abstrak PubMed	LDA (MALLET), BERTopic (BioBERT)	Coherence, penilaian pakar, visualisasi t-SNE	Akurasi setara; klaster BERTopic lebih padat dan terpisah
[10]	Dua dataset teks (20-newsgroup dan percakapan Facebook)	LSA, LDA, NMF, RP, PCA	Recall, precision, F-score, koherensi topik	LDA dan NMF memberi topik yang lebih bermakna
Penelitian ini	2.234 ulasan TikTok Shop Seller Center	LDA dan BERTopic dengan kriteria pemilihan parameter yang sama	C_V, UMass, C_NPMI (10 kata teratas, teks dan kamus sama)	Dibahas pada Bagian 3

LDA sendiri sering diterapkan pada ulasan berbahasa Indonesia, antara lain pada ulasan aplikasi Netflix yang menghasilkan dua belas topik berdasarkan koherensi tertinggi [4], ulasan PeduliLindungi dengan nilai koherensi 0,3963 untuk lima topik [11], ulasan Lazada Indonesia dengan tema kualitas produk, layanan pengiriman, dan harga [12], serta komentar produk perawatan kulit di Female Daily yang jumlah topiknya dipilih dengan koherensi [13]. Pada ulasan Alfagift, LDA dipakai setelah analisis sentimen dan menunjukkan dua isu dominan pada ulasan negatif, yaitu masalah login dan keterlambatan pengiriman [14]. LDA juga dibandingkan dengan LSA pada data bencana alam berbahasa Indonesia dan menghasilkan koherensi yang lebih tinggi, yaitu 0,4715 berbanding 0,4467 [15]. Ulasan aplikasi di Google Play Store juga dianalisis untuk tugas lain, misalnya klasifikasi sentimen ulasan aplikasi MyICON+ dengan Naïve Bayes, SVM, dan Random Forest [16]. Untuk BERTopic dengan model multibahasa pada teks Indonesia, penerapan yang ditemukan dalam rujukan penelitian ini adalah klasifikasi berita hoaks [17].

Dari penelitian-penelitian tersebut terlihat dua hal. Pertama, hasil perbandingan LDA dan BERTopic bergantung pada data dan cara evaluasinya, sehingga tidak ada metode yang selalu lebih baik. Kedua, dalam rujukan yang kami telaah, belum ditemukan penelitian yang membandingkan LDA dan BERTopic pada ulasan aplikasi TikTok Shop Seller Center dengan kriteria pemilihan parameter yang sama dan tiga ukuran *coherence score* yang dihitung dengan prosedur yang sama untuk kedua metode. Penelitian ini mengisi celah tersebut.

Copyright © 2026 The **Salsabila Savitri**, Page 1363
This Journal is licensed under a Creative Commons Attribution 4.0 International License

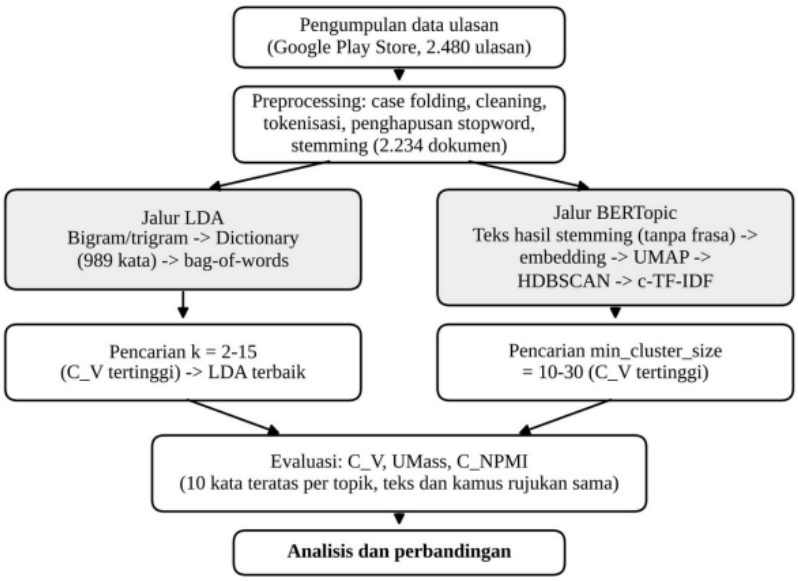
JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 5, Oktober 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v15i5.10101
Hal 1362-1374
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Kedua metode dibandingkan dengan kriteria yang sama, bukan dengan parameter yang sama. Jumlah topik pada LDA dan min_cluster_size pada BERTopic adalah parameter yang berbeda sifatnya. Jumlah topik ditetapkan langsung oleh peneliti, sedangkan min_cluster_size hanya membatasi ukuran minimum klaster sehingga banyaknya topik ditentukan oleh data. Kedua parameter tidak dapat disamakan nilainya. Yang disamakan adalah kriterianya: untuk masing-masing metode, sejumlah nilai kandidat diuji pada data yang sama, lalu dipilih nilai yang memberi C_V tertinggi. Setelah itu, kedua model terpilih dievaluasi dengan 10 kata teratas per topik, kamus (*dictionary*), dan teks yang sama. Jumlah kandidatnya berbeda (14 nilai k pada LDA dan 5 nilai min_cluster_size pada BERTopic), yang juga menjadi salah satu keterbatasan penelitian ini.

Tujuan penelitian ini adalah mengidentifikasi topik-topik dominan pada ulasan pengguna TikTok Shop Seller Center dan membandingkan LDA dan BERTopic berdasarkan C_V, UMass, dan C_NPMI. Kontribusi penelitian ini ada dua. Pertama, perbandingan kedua metode dilakukan dengan kriteria pemilihan parameter, data masukan, dan prosedur evaluasi yang sama, disertai versi perangkat lunak dan parameter agar dapat direplikasi. Kedua, hasilnya diperiksa dengan membandingkan nilai koherensi per topik, hubungan topik dengan rating, dan kata-kata topik dari kedua metode, tidak hanya nilai rata-rata koherensi.

2. METODOLOGI PENELITIAN

Penelitian dilakukan dalam tujuh tahap seperti ditunjukkan pada Gambar 1. Tahap pertama adalah pengumpulan ulasan TikTok Shop Seller Center dari Google Play Store. Tahap kedua adalah *preprocessing* teks, yang meliputi *case folding*, *cleaning*, tokenisasi, penghapusan *stopword*, dan *stemming*. Setelah itu, alur terbagi menjadi dua jalur. Pada tahap ketiga (jalur LDA), teks dibentuk menjadi bigram dan trigram, lalu disusun menjadi *dictionary* dan representasi *bag-of-words*. Pada tahap keempat (jalur BERTopic), teks hasil *stemming* tanpa pembentukan frasa diubah menjadi *embedding*, direduksi dimensinya dengan UMAP, dikelompokkan dengan HDBSCAN, dan diberi kata penciri dengan c-TF-IDF. Tahap kelima adalah pencarian parameter, yaitu jumlah topik k = 2 sampai 15 pada LDA dan min_cluster_size = 10, 15, 20, 25, dan 30 pada BERTopic, yang masing-masing dipilih berdasarkan C_V tertinggi. Tahap keenam adalah evaluasi kedua model terpilih dengan C_V, UMass, dan C_NPMI menggunakan 10 kata teratas per topik serta teks dan kamus rujukan yang sama. Tahap ketujuh adalah analisis dan perbandingan hasil.



Gambar 1. Alur penelitian

2.1 Pengumpulan Data

Ulasan diambil dari aplikasi TikTok Shop Seller Center (package id: com.tiktokshop.seller) di Google Play Store menggunakan *library* google-play-scraper. Pengambilan data dilakukan pada 14 Juli 2026 dengan dua metode pengurutan, yaitu ulasan terbaru (NEWEST) dan ulasan paling relevan (MOST_RELEVANT), masing-masing dengan batas 1.500 ulasan. Kedua kumpulan digabung dan duplikat dihapus berdasarkan reviewId, sehingga diperoleh 2.480 ulasan mentah. Tanggal ulasan pada data ini berkisar dari 13 Juni 2022 sampai 13 Juli 2026. Sebagian besar ulasan relatif baru: 1.455 dokumen (65,1%) berasal dari Mei sampai Juli 2026 (masing-masing 474, 704, dan 277 dokumen) dan 1.973 dokumen (88,3%) berasal dari Januari sampai Juli 2026. Dengan demikian, data ini merupakan potret ulasan pada saat pengambilan dan bukan sampel acak dari seluruh masa hidup aplikasi. Data hasil pengambilan disimpan ke Google Drive dan dimuat kembali pada eksekusi berikutnya tanpa *scraping* ulang, sehingga data tidak berubah antarpercobaan.

Copyright © 2026 The **Salsabila Savitri**, Page 1364
This Journal is licensed under a Creative Commons Attribution 4.0 International License

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 5, Oktober 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v15i5.10101
Hal 1362-1374
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

Dari data yang diambil, kolom yang dipakai adalah reviewId, score (rating), at (tanggal ulasan), dan content (isi ulasan). Data bersumber dari ulasan yang tampil publik di Google Play Store. Kolom userName tidak digunakan dalam analisis dan tidak disertakan dalam *dataset* yang disimpan untuk pemodelan. Hasil dilaporkan sebagai topik dan jumlah dokumen secara agregat, tanpa menampilkan ulasan individual atau identitas pengguna.

2.2 Preprocessing Teks

Preprocessing bertujuan membersihkan dan menormalisasi teks mentah sebelum pemodelan [18]. Tahapnya adalah sebagai berikut. (1) *Case folding*, yaitu mengubah semua huruf menjadi huruf kecil. (2) *Cleaning*, yaitu menghapus URL, alamat berawalan www, mention (@), tagar (#), angka, karakter selain huruf, dan spasi berlebih. (3) Tokenisasi dengan word_tokenize dari NLTK. (4) Penghapusan *stopword* memakai daftar *stopword* berbahasa Indonesia dari NLTK yang ditambah 35 *stopword* khusus. (5) *Stemming* dengan StemmerFactory dari Sastrawi, yang mengembalikan kata ke bentuk dasarnya. Ulasan yang tidak menyisakan kata setelah tahap ini dikeluarkan, sehingga 2.480 ulasan menjadi 2.234 dokumen dengan total 33.367 token (rata-rata sekitar 14,9 token per dokumen).

Ke-35 *stopword* khusus ditambahkan karena tidak membedakan topik dalam ulasan aplikasi. Kelompok pertama adalah kata yang menyebut aplikasinya sendiri atau sapaan kepada pengembang: apk, aplikasi, admin, min, dan kak. Kelompok kedua adalah singkatan dan kata percakapan: yg, nya, aja, ga, gak, nggak, ngga, iah, dong, banget, bgt, udah, dan udahh. Kelompok ketiga adalah ungkapan pujian dan penutup yang umum: ok, oke, okee, good, best, mantap, mantaap, mantasseep, bagus, bgus, puas, terimakasih, thanks, thank, recommended, recommend, dan rekomendasi. Daftar lengkapnya adalah: apk, aplikasi, yg, nya, aja, ga, gak, nggak, ngga, iah, dong, min, admin, banget, bgt, udah, udahh, kak, ok, oke, okee, good, best, mantap, bgus, bagus, terimakasih, thanks, thank, puas, mantaap, mantasseep, recommended, recommend, dan rekomendasi. Kata seperti “bagus”, “mantap”, dan “puas” memuat informasi sentimen. Kata-kata itu dihapus karena penelitian ini memodelkan tema ulasan, bukan sentimennya, dan karena kata tersebut dapat muncul pada tema apa pun. Pilihan ini dapat mengurangi informasi pada ulasan yang sangat pendek.

Hasil *stemming* tersebut digunakan sebagai masukan kedua metode, tetapi dengan bentuk yang berbeda. Pada LDA, token hasil *stemming* dibentuk menjadi *bigram* dan *trigram* dengan Gensim Phrases (*bigram*: min_count = 5 dan threshold = 10; *trigram*: Phrases dibangun dari keluaran *bigram* dengan threshold = 10 dan min_count bawaan 5), misalnya “tiktok_shop” dan “cair_dana”. Pada BERTopic, masukannya adalah teks hasil *stemming* yang digabung kembali menjadi kalimat tanpa pembentukan frasa. Frasa seperti “bantu usaha” pada hasil BERTopic berasal dari CountVectorizer dengan ngram_range (1,2), bukan dari Phrases. *Stemming* dan penghapusan *stopword* yang agresif dapat memengaruhi *embedding* semantik: pada penelitian deteksi emosi dengan IndoBERT dan IndoBERTweet, model tanpa penghapusan *stopword* dan *stemming* justru memberi akurasi lebih tinggi [18]. Penelitian ini tidak menguji BERTopic dengan teks tanpa *stemming*, sehingga pengaruh tersebut belum dapat diukur.

2.3 Pemodelan Topik dengan LDA

LDA dijalankan dengan Gensim 4.4.0. *Dictionary* disusun dari teks *bigram* dan *trigram* dengan corpora.*Dictionary*, lalu kata yang terlalu jarang atau terlalu umum disaring dengan no_below = 5 dan no_above = 0,5, sehingga dari 4.103 kata unik tersisa 989 kata. Sebanyak 67 dokumen (3,0%) tidak memiliki kata tersisa setelah penyaringan ini, tetapi tetap ikut dihitung dalam distribusi topik. *Corpus* direpresentasikan sebagai *bag-of-words*. Jumlah topik dicari untuk k = 2 sampai 15 dengan nilai C_V dari CoherenceModel Gensim. Rentang ini ditetapkan oleh penulis dan tidak berasal dari rujukan tertentu. Nilai k dengan C_V tertinggi dipilih tanpa syarat tambahan. Model dilatih dengan random_state = 42, update_every = 1, chunksize = 100, passes = 20, dan alpha = ‘auto’. Parameter lain tidak diubah dari nilai bawaan Gensim 4.4.0, yaitu eta tidak ditentukan (None), decay = 0,5, offset = 1,0, iterations = 50, eval_every = 10, gamma_threshold = 0,001, dan minimum_probability = 0,01. Topik divisualisasikan dengan pyLDAvis. Versi seluruh perangkat lunak yang dipakai pada kedua metode disajikan pada Tabel 2.

Tabel 2. Versi perangkat lunak yang digunakan

Komponen	Versi	Komponen	Versi
Python	3.13.15	umap-learn	0.5.12
gensim	4.4.0	hdbscan	0.8.44
bertopic	0.17.4	scikit-learn	1.6.1
sentence-transformers	5.7.0	numpy / pandas	2.1.3 / 2.2.3
google-play-scraper	1.2.7	Sastrawi / NLTK	1.0.1 / 3.9.1

2.4 Pemodelan Topik dengan BERTopic

BERTopic dijalankan dengan *library* BERTopic versi 0.17.4 dan model *embedding* paraphrase-multilingual-MiniLM-L12-v2 dari sentence-transformers versi 5.7.0. Model ini merupakan model multibahasa yang dibuat berdasarkan pendekatan Sentence-BERT [19]. Reduksi dimensi memakai UMAP (umap-learn 0.5.12) dengan n_neighbors = 15, n_components = 5, min_dist = 0,0, metric = ‘cosine’, dan random_state = 42. Pengelompokan memakai HDBSCAN (hdbscan 0.8.44) dengan metric = ‘euclidean’, cluster_selection_method = ‘eom’, dan prediction_data = True. Kata topik

Copyright © 2026 The **Salsabila Savitri**, Page 1365
This Journal is licensed under a Creative Commons Attribution 4.0 International License

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 5, Oktober 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v15i5.10101
Hal 1362-1374
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

dibentuk dengan CountVectorizer (scikit-learn 1.6.1) dengan ngram_range = (1,2) dan min_df = 2. Model dilatih dalam mode multibahasa dengan calculate_probabilities = False dan tanpa pembatasan jumlah topik (nr_topics tidak diisi), serta mengeluarkan 10 kata per topik (nilai bawaan top_n_words). Penggunaan model multibahasa untuk teks Indonesia sejalan dengan penerapan BERTopic pada klasifikasi berita hoaks berbahasa Indonesia [17].

Embedding dokumen dihitung satu kali dan dipakai ulang pada setiap percobaan. Kemudian lima nilai min_cluster_size (10, 15, 20, 25, dan 30) diuji, dan nilai dengan C_V tertinggi dipilih sebagai model akhir. Rentang ini ditetapkan oleh penulis: dimulai dari nilai bawaan BERTopic (10), lalu dinaikkan per 5 sampai 30. Jumlah topik mengikuti hasil pengelompokan HDBSCAN. Dokumen yang tidak masuk klaster mana pun diberi label -1 (*outlier*). Dokumen *outlier* tidak dipindahkan ke topik lain.

2.5 Evaluasi Model

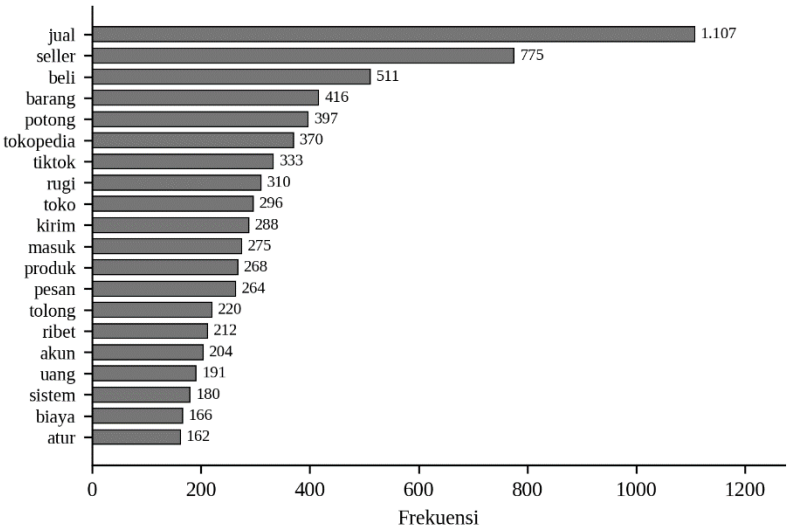
Evaluasi memakai tiga ukuran koherensi dari CoherenceModel Gensim. C_V menilai kemiripan antarkata teratas topik dari kemunculan bersama dalam jendela geser (*sliding window*), dengan nilai yang lebih tinggi berarti lebih koheren [20]. UMass menilai probabilitas bersyarat pasangan kata berdasarkan kemunculan bersama dalam satu dokumen, dan nilai yang lebih dekat ke nol berarti lebih koheren [20]. C_NPMI memakai normalized pointwise mutual information dengan rentang -1 sampai +1, dan nilai yang lebih tinggi berarti lebih koheren [21].

Agar nilai kedua model dapat dibandingkan, prosedur evaluasinya disamakan. Kedua model dinilai dengan 10 kata teratas per topik (topn = 10), yaitu jumlah kata yang secara bawaan dikeluarkan BERTopic. Kedua model memakai teks (*texts*), *dictionary*, dan *corpus* yang sama, yaitu teks *bigram-trigram* dan kamus 989 kata. Kata topik yang tidak ada di kamus tidak dihitung oleh Gensim, termasuk frasa dua kata dari BERTopic yang tidak terbentuk pada teks LDA. Jumlah kata yang terpakai pada setiap topik dilaporkan pada Bagian 3.4. Dokumen *outlier* BERTopic (topik -1) dikeluarkan dari perhitungan karena topik ini mewakili dokumen yang tidak masuk klaster mana pun, bukan satu tema. Perlakuan yang sama berlaku saat memilih min_cluster_size dan saat evaluasi akhir. Dengan perlakuan ini BERTopic dinilai pada 2.203 dokumen berklaster (98,6%), sedangkan LDA memberi topik dominan pada setiap dokumen. Nilai akhir per model adalah rata-rata nilai per topik.

3. HASIL DAN PEMBAHASAN

3.1 Karakteristik Dataset dan Preprocessing

Dataset terdiri dari 2.480 ulasan mentah, dan setelah *preprocessing* tersisa 2.234 dokumen dengan 33.367 token. Pada *dictionary* LDA terdapat 4.103 kata unik, yang menjadi 989 kata setelah penyaringan. Distribusi rating menunjukkan bahwa ulasan bernilai satu bintang paling banyak, yaitu 1.374 dokumen (61,5%), diikuti rating lima sebanyak 493 dokumen (22,1%), rating dua sebanyak 149 dokumen (6,7%), rating tiga sebanyak 125 dokumen (5,6%), dan rating empat sebanyak 93 dokumen (4,2%). Frekuensi kata pada Gambar 2 menunjukkan bahwa “jual” (1.107), “seller” (775), “beli” (511), “barang” (416), dan “potong” (397) adalah kata yang paling sering muncul. Dokumen yang dimodelkan juga sangat pendek: rata-rata 14,9 token per dokumen dengan median 12 token, kuartil pertama 4 token, kuartil ketiga 23 token, dan paling panjang 69 token. Panjang ini penting untuk menafsirkan hasil berikutnya, karena pada dokumen yang pendek sedikit kata yang dapat muncul bersama.



Gambar 2. Dua puluh kata dengan frekuensi tertinggi

Kata-kata yang sering muncul pada Gambar 2 berkaitan dengan aktivitas jual-beli, pengiriman, dan potongan biaya. Pada tahap pembentukan frasa, *trigram* yang paling sering muncul adalah “tolong_baik” (73), “tiktok_shop” (64), “cair_dana” (59), “kembali_barang” (49), dan “gabung_tiktok” (48). Frasa ini memberi konteks tambahan dibandingkan

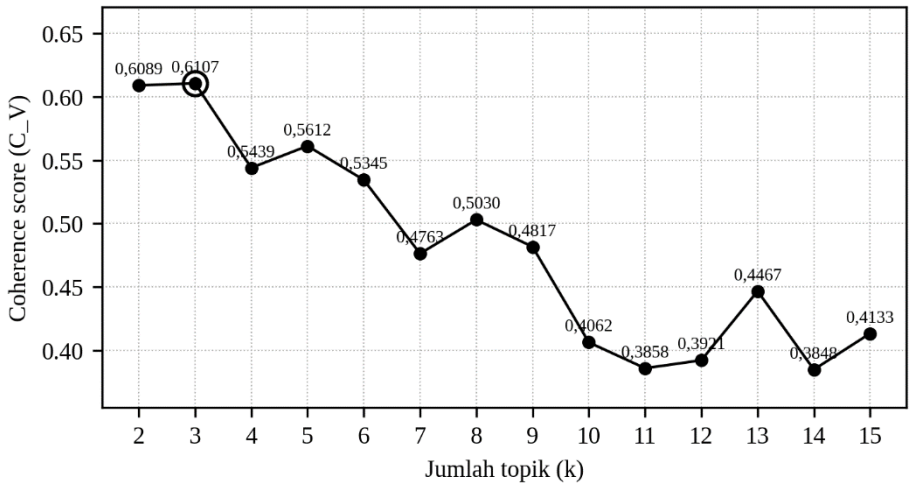
Copyright © 2026 The **Salsabila Savitri**, Page 1366
This Journal is licensed under a Creative Commons Attribution 4.0 International License

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 5, Oktober 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v15i5.10101
Hal 1362-1374
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

kata tunggal, misalnya pada pencairan dana dan pengembalian barang, dan frasa ini menjadi bagian dari teks yang dipakai pada LDA.

3.2 Hasil Pemodelan Topik LDA

Pencarian jumlah topik LDA dilakukan pada k = 2 sampai 15 dan hasilnya ditampilkan pada Gambar 3. Nilai C_V tertinggi diperoleh pada k = 3 sebesar 0,6107, hanya sedikit di atas k = 2 sebesar 0,6089. Setelah itu nilai C_V turun menjadi 0,5439 pada k = 4 dan berkisar antara 0,3848 dan 0,5612 pada k = 5 sampai 15. Karena aturan pemilihan yang dipakai adalah nilai C_V tertinggi, k = 3 dipilih sebagai model akhir. Selisih yang sangat kecil antara k = 2 dan k = 3 menunjukkan bahwa penetapan jumlah topik terbaik pada data ini tidak tajam. Memilih jumlah topik berdasarkan C_V tertinggi adalah pendekatan yang lazim dalam pemodelan topik [20].



Gambar 3. Nilai C_V LDA untuk setiap jumlah topik

Kata kunci ketiga topik LDA disajikan pada Tabel 3. Topik 0 berisi kata “seller”, “beli”, “produk”, “barang”, “toko”, dan “sistem”, yang merujuk pada aktivitas penjual, produk, dan pengaturan toko melalui sistem, dengan kata “rugi” dan “atur” juga termasuk sepuluh kata teratas. Topik 1 didominasi “jual” (0,102), disertai “potongan”, “pesan”, “uang”, “kirim”, “biaya”, dan “dana”, sehingga berkaitan dengan pesanan, pengiriman, potongan biaya, dan dana. Kata berikutnya pada topik ini, seperti “cair” dan “retur”, memperkuat kesan tersebut. Topik 2 berisi “tokopedia” (0,045), “masuk”, “ribet”, “akun”, “susah”, dan “tokped”, sehingga berkaitan dengan akun, proses masuk, dan kesulitan penggunaan yang disebut bersama nama Tokopedia. Bobot “jual” pada Topik 1 (0,102) adalah yang tertinggi di antara semua kata pada ketiga topik, sehingga topik ini paling kuat dicirikan oleh satu kata, sedangkan bobot pada dua topik lainnya lebih merata. Label topik pada Tabel 3 adalah interpretasi penulis berdasarkan kata kunci dan belum divalidasi oleh pihak lain.

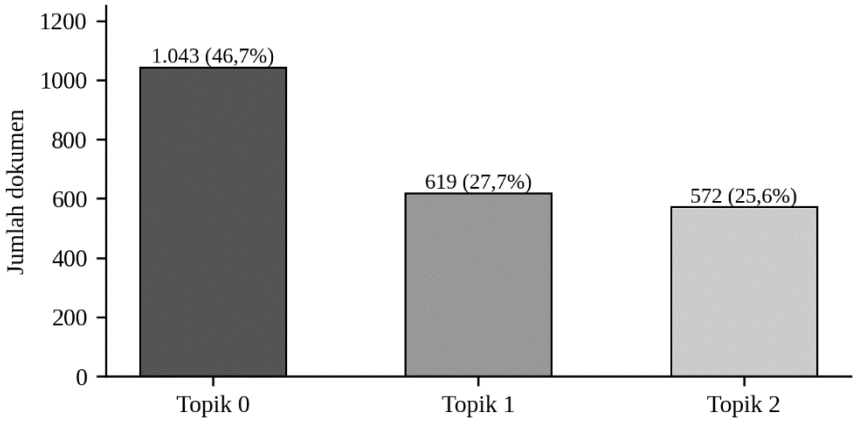
Tabel 3. Kata kunci tiga topik LDA terpilih (k =3) beserta bobotnya

Topik	Sepuluh kata teratas (bobot)	Interpretasi penulis
Topik 0	seller (0,065), beli (0,043), produk (0,029), barang (0,025), rugi (0,024), sistem (0,017), toko (0,016), tiktok (0,016), langsung (0,014), atur (0,014)	Aktivitas penjual, produk, dan pengaturan toko lewat sistem
Topik 1	jual (0,102), potong (0,038), pesan (0,035), uang (0,024), kirim (0,020), kalo (0,018), biaya (0,017), tolong (0,016), dana (0,016), barang (0,016)	Pesanan, pengiriman, potongan biaya, dan dana
Topik 2	tokopedia (0,045), masuk (0,033), ribet (0,027), akun (0,021), susah (0,017), bikin (0,016), tokped (0,014), tiktok (0,012), pakai (0,012), toko (0,011)	Akun, proses masuk, dan kesulitan penggunaan

Distribusi dokumen pada Gambar 4 menunjukkan bahwa Topik 0 menjadi topik dominan bagi 1.043 dokumen (46,7%), Topik 1 bagi 619 dokumen (27,7%), dan Topik 2 bagi 572 dokumen (25,6%). Proporsi ini sedikit dipengaruhi oleh 67 dokumen tanpa kata tersisa setelah penyaringan kamus, yang tetap dihitung dalam distribusi.

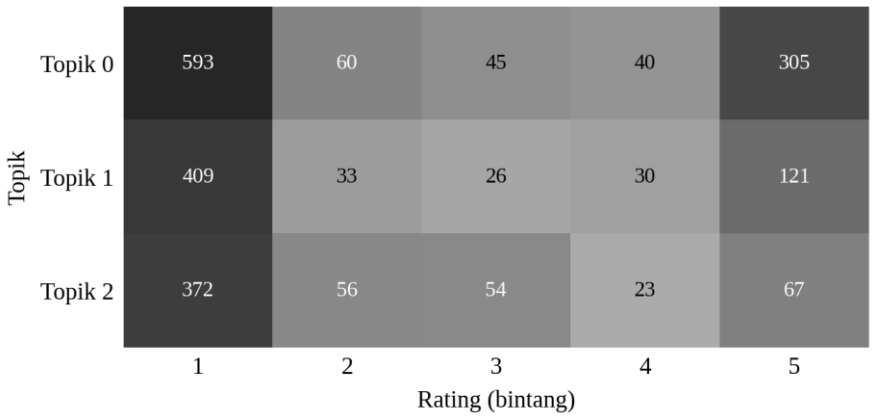
Copyright © 2026 The **Salsabila Savitri**, Page 1367
This Journal is licensed under a Creative Commons Attribution 4.0 International License

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 5, Oktober 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v15i5.10101
Hal 1362-1374
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom



Gambar 4. Distribusi dokumen pada topik LDA

Hubungan antara topik LDA dan rating ditampilkan pada Gambar 5. Rating satu merupakan rating terbanyak pada ketiga topik, yaitu 593 dokumen (56,9%) pada Topik 0, 409 dokumen (66,1%) pada Topik 1, dan 372 dokumen (65,0%) pada Topik 2, dibandingkan 61,5% pada seluruh data. Rating lima paling besar proporsinya pada Topik 0 (305 dokumen, 29,2%), diikuti Topik 1 (121 dokumen, 19,5%) dan Topik 2 (67 dokumen, 11,7%). Artinya, topik yang sama dapat muncul pada ulasan bernada baik maupun buruk, dan Topik 2 paling jarang berasal dari ulasan bintang lima.



Gambar 5. Jumlah dokumen pada setiap topik LDA menurut rating

3.3 Hasil Pemodelan Topik BERTopic

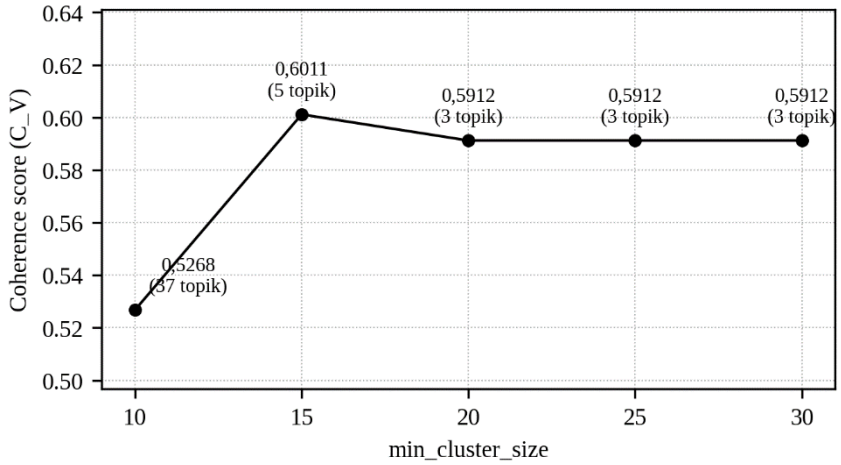
Hasil pengujian lima nilai min_cluster_size disajikan pada Tabel 4 dan Gambar 6. Nilai 10 menghasilkan 37 topik dengan C_V terendah (0,5268). Nilai 15 menghasilkan lima topik dengan C_V tertinggi (0,6011). Nilai 20, 25, dan 30 menghasilkan hasil yang sama, yaitu tiga topik dengan C_V 0,5912. Karena itu min_cluster_size = 15 dipilih sebagai model akhir. Keseragaman hasil pada nilai 20 sampai 30 menunjukkan bahwa pengelompokan HDBSCAN pada data ini tidak berubah lagi saat ambang ukuran klaster dinaikkan melewati titik tertentu.

Tabel 4. Hasil optimasi min_cluster_size BERTopic

min_cluster_size	Jumlah Topik	Coherence Score (C_V)
10	37	0,5268
15	5	0,6011
20	3	0,5912
25	3	0,5912
30	3	0,5912

Copyright © 2026 The **Salsabila Savitri**, Page 1368
This Journal is licensed under a Creative Commons Attribution 4.0 International License

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 5, Oktober 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v15i5.10101
Hal 1362-1374
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

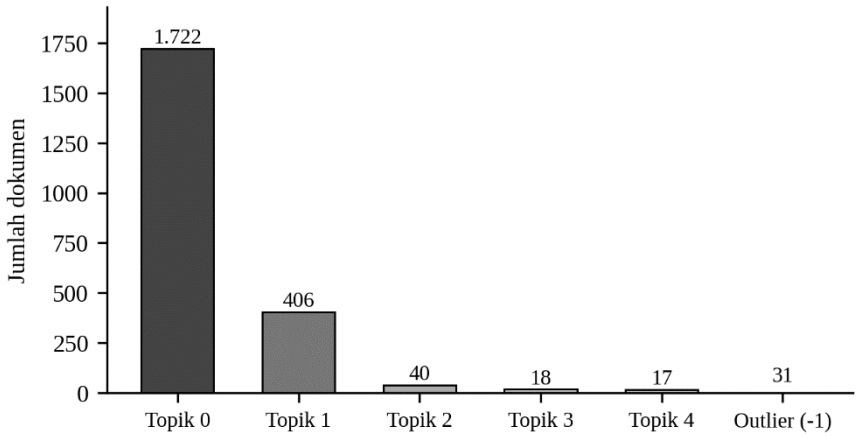


Gambar 6. Hasil optimasi min_cluster_size BERTopic

Model akhir membentuk lima topik dan satu kelompok *outlier* (topik -1) dari 2.234 dokumen. Kata kunci dan label setiap topik disajikan pada Tabel 5, dan jumlah dokumennya pada Gambar 7.

Tabel 5. Kata kunci dan interpretasi topik BERTopic (min_cluster_size =15)

Topik	Dokumen	Kata Representasi	Interpretasi penulis
0	1.722	jual, seller, beli, barang, tiktok, toko, potong, tokopedia, masuk, pesan	Jual-beli dan operasional toko
1	406	mudah, bantu, moga, suka, potong, ribet, paham, keren, cepat, umkm	Kesan penggunaan dan dukungan untuk UMKM
2	40	bantu bantu, bantu, usaha, bantu usaha	Ulasan singkat berulang dengan kata “bantu” dan “usaha”
3	18	bisnis, bantu ekonomi, ekonomi, bantu, usaha, umkm	Bisnis, ekonomi, dan UMKM
4	17	bintang, kasih bintang, kasih, bintang biar, selera	Ulasan yang menyebut pemberian bintang
-1 (outlier)	31	terima kasih, pokok, kasih, terima, trimakasi	Tidak terklasifikasi



Gambar 7. Distribusi dokumen pada topik BERTopic

Topik 0 memuat 1.722 dokumen (77,1%) dan topik 1 memuat 406 dokumen (18,2%), sedangkan Topik 2, 3, dan 4 masing-masing hanya memuat 40, 18, dan 17 dokumen. Sebanyak 31 dokumen (1,4%) menjadi *outlier*. Dominasi Topik 0 dapat berarti dua hal: struktur data yang memang didominasi pembahasan jual-beli, atau topik BERTopic yang masih terlalu umum. Data penelitian ini belum cukup untuk memisahkan keduanya. Alasan pertama didukung oleh kata teratas Topik 0 (“jual”, “seller”, “beli”, “barang”) yang juga merupakan kata paling sering pada seluruh korpus (Gambar 2) dan muncul

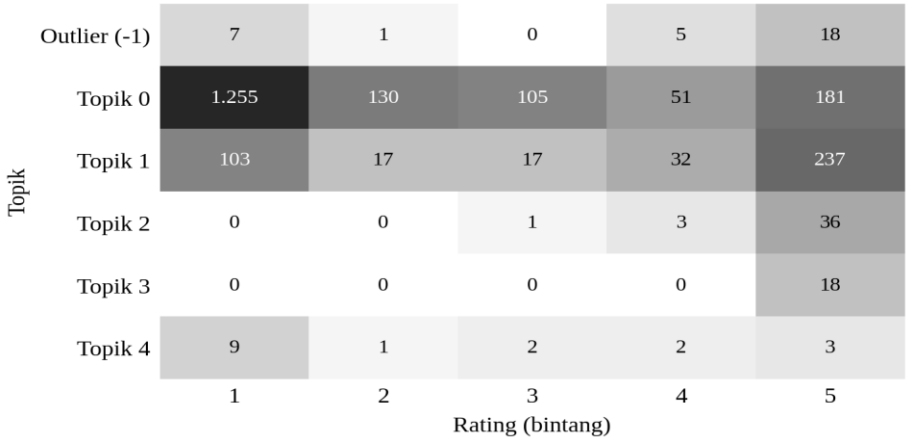
Copyright © 2026 The **Salsabila Savitri**, Page 1369
This Journal is licensed under a Creative Commons Attribution 4.0 International License

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 5, Oktober 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v15i5.10101
Hal 1362-1374
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

pada topik LDA. Alasan kedua didukung oleh min_cluster_size = 10, yang membagi data menjadi 37 topik. Artinya kelompok besar tersebut dapat dipecah lebih halus, walaupun C_V pada pembagian itu turun. Karena itu Topik 0 sebaiknya dibaca sebagai kelompok umum tentang jual-beli dan operasional toko, dan pembagian yang lebih rinci memerlukan penilaian manusia.

Topik 1 (406 dokumen) memuat kata “mudah”, “bantu”, “moga”, “suka”, “keren”, “cepat”, dan “umkm”, tetapi juga “potong” dan “ribet”, sehingga isinya campuran antara kesan positif dan kata “potong” yang juga ada di Topik 0. Topik 2 (40 dokumen) didominasi frasa “bantu bantu”, “bantu”, dan “usaha”. Topik 3 (18 dokumen) memuat “bisnis”, “bantu ekonomi”, “ekonomi”, “usaha”, dan “umkm”. Topik 4 (17 dokumen) memuat “bintang”, “kasih bintang”, dan “selera”. Kata kunci ketiga topik kecil ini banyak berupa frasa yang berulang. Frasa dua kata pada BERTopic tidak ada di kamus LDA, sehingga sebagian di antaranya tidak terpakai saat koherensi dihitung (lihat Bagian 3.4).

Hubungan topik BERTopic dan rating ditampilkan pada Gambar 8. Topik 0 didominasi rating satu (1.255 dokumen, 72,9%), sedangkan Topik 1 didominasi rating lima (237 dokumen, 58,4%). Topik 2 (36 dari 40 dokumen) dan Topik 3 (18 dari 18 dokumen) hampir atau seluruhnya berating lima. Topik 4 adalah ulasan yang menyebut pemberian bintang, tetapi data ini tidak menunjukkan apakah penyebutan itu berupa permintaan, pujian, atau keluhan.



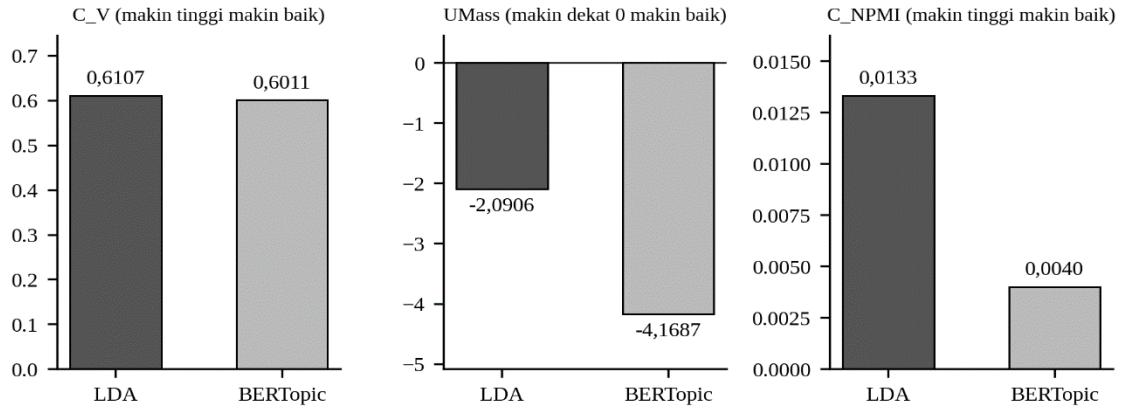
Gambar 8. Jumlah dokumen pada setiap topik BERTopic menurut rating

3.4 Perbandingan Kinerja LDA dan BERTopic

Perbandingan LDA dan BERTopic berdasarkan tiga ukuran koherensi disajikan pada Tabel 6 dan Gambar 9. Kedua model dinilai dengan prosedur yang sama (Bagian 2.5).

Tabel 6. Perbandingan coherence score LDA dan BERTopic (10 kata teratas per topik)

Metode	Jumlah topik	C_V	UMass	C_NPMI
LDA	3	0,6107	-2,0906	0,0133
BERTopic	5	0,6011	-4,1687	0,0040
selisih (LDA – BERTopic)		0,0096	2,0781	0,0093



Gambar 9. Perbandingan C_V, UMass, dan C_NPMI antara LDA dan BERTopic

Copyright © 2026 The **Salsabila Savitri**, Page 1370
This Journal is licensed under a Creative Commons Attribution 4.0 International License

JURIKOM (Jurnal Riset Komputer), Vol. 13 No. 5, Oktober 2026
e-ISSN 2715-7393 (Media Online), p-ISSN 2407-389X (Media Cetak)
DOI 10.30865/jurikom.v15i5.10101
Hal 1362-1374
https://ejurnal.stmik-budidarma.ac.id/index.php/jurikom

LDA memperoleh nilai lebih baik pada ketiga ukuran. Selisih C_V hanya 0,0096 (0,6107 berbanding 0,6011) dan selisih C_NPMI 0,0093 (0,0133 berbanding 0,0040). Selisih UMass lebih jelas, yaitu -2,0906 berbanding -4,1687. Nilai ini berasal dari satu kali pelatihan dengan random_state tetap dan tidak diuji secara statistik, sehingga selisih C_V yang kecil tidak cukup untuk menyatakan satu metode lebih unggul. Jumlah kata teratas yang dinilai memengaruhi nilai koherensi, sehingga perbandingan ini hanya berlaku untuk 10 kata teratas per topik.

Rata-rata di atas lebih mudah dipahami jika dilihat per topik (Tabel 7). Ketiga topik LDA memiliki C_V yang berdekatan, yaitu 0,5976 sampai 0,6264. Pada BERTopic, nilainya sangat bervariasi. Topik 1 hanya memperoleh C_V 0,2971 dan UMass -7,5023, yang sejalan dengan isinya yang bercampur antara kata bernada positif dan kata transaksi. Sebaliknya, Topik 2 memperoleh C_V tertinggi (0,8831), tetapi hanya dua kata yang terpakai ("bantu" dan "usaha"), dan Topik 3 serta Topik 4 hanya memakai empat dan tiga kata. Nilai yang tinggi ini dihitung dari sedikit pasangan kata, sehingga belum tentu mencerminkan kualitas topik secara keseluruhan. Jadi, rata-rata C_V BERTopic (0,6011) dihasilkan oleh dua topik besar dan tiga topik kecil dengan kondisi yang sangat berbeda.

Tabel 7. Nilai koherensi per topik

Model	Topik	Kata terpakai	C_V	UMass
LDA	0	10	0,6264	-1,9848
LDA	1	10	0,5976	-2,0982
LDA	2	10	0,6081	-2,1888
BERTopic	0	10	0,5893	-1,8218
BERTopic	1	10	0,2971	-7,5023
BERTopic	2	2	0,8831	-2,3308
BERTopic	3	4	0,5596	-5,9491
BERTopic	4	3	0,6766	-3,2395

Kata terpakai adalah jumlah kata topik yang ada di kamus 989 kata dan dihitung oleh Gensim. Rata-rata per model menghasilkan nilai pada Tabel 6.

Nilai C_NPMI kedua model sangat kecil, yaitu 0,0133 untuk LDA dan 0,0040 untuk BERTopic. C_NPMI bernilai nol jika dua kata muncul bersama sebanyak yang diharapkan bila keduanya saling bebas. Nilai yang mendekati nol berarti pasangan kata teratas pada topik hanya sedikit lebih sering muncul bersama daripada yang diharapkan secara kebetulan. Pada ulasan yang hanya sekitar 14,9 token, peluang dua kata berada dalam dokumen yang sama memang kecil, dan hal ini dapat ikut menjelaskan nilai yang rendah, walaupun penyebabnya tidak diuji di sini. Karena itu, C_NPMI dibaca sebagai pelengkap dan bukan pembeda utama: kedua model sama-sama tidak menunjukkan keterkaitan antarkata yang kuat menurut ukuran ini, dan selisih 0,0093 tidak cukup untuk membedakan keduanya.

Dari sisi isi topik, kedua metode menghasilkan kata yang beririsan: "jual", "seller", "beli", "barang", "tiktok", "toko", "potong", "Tokopedia", "masuk", dan "pesan" muncul pada sepuluh kata teratas Topik 0 BERTopic, dan sebagian besar juga muncul pada ketiga topik LDA. Keselarasan ini mendukung konsistensi temuan bahwa ulasan banyak membahas jual-beli, pesanan, pengiriman, potongan biaya, serta akun dan proses masuk. Namun, hal ini belum dianggap sebagai validasi isi topik, karena kedua metode memakai data yang sama dan kata-kata tersebut juga yang paling sering muncul pada seluruh korpus. Validasi yang lebih kuat memerlukan penilaian oleh pakar atau penilai manusia terhadap interpretasi topik, yang belum dilakukan dalam penelitian ini.

Temuan ini dibandingkan dengan penelitian terdahulu pada Tabel 8. Nilai koherensi antarpenelitian tidak dapat dibandingkan langsung karena data, jumlah topik, dan ukuran koherensinya berbeda, sehingga perbandingan dilakukan pada arah dan karakteristik hasilnya. Hasil penelitian ini searah dengan [6] yang juga menemukan LDA terbaik pada data hasil *stemming*, dan dengan [10] yang menemukan LDA memberi topik bermakna pada teks pendek. Hasil ini berbeda dari [7] dan [8], yang menempatkan BERTopic di atas LDA. Perbedaannya mungkin terkait ukuran koherensi dan penilaian yang dipakai (koherensi berbasis Word2Vec dan penilaian oleh manusia pada [7]; interpretasi manusia pada [8]) serta data yang jauh lebih panjang dan beragam. Dugaan ini tidak diuji dalam penelitian ini. Pada [9], akurasi LDA dan BERTopic setara, yang tidak bertentangan dengan selisih C_V yang kecil pada penelitian ini.

Tabel 8. Perbandingan hasil penelitian ini dengan penelitian terdahulu

Ref.	Temuan penelitian terdahulu	Hubungan dengan penelitian ini
[6]	BERTopic terbaik pada data tertokenisasi (0,6905); LDA terbaik pada data stemming	Searah: data penelitian ini di-stemming dan LDA lebih tinggi
[7]	BERTopic sedikit di atas NMF pada koherensi dan lebih baik pada penilaian oleh manusia dan model bahasa; LDA tertinggal jauh	Berbeda: LDA tidak tertinggal. Ukuran koherensi, jumlah topik, dan data berbeda

Copyright © 2026 The **Salsabila Savitri**, Page 1371
This Journal is licensed under a Creative Commons Attribution 4.0 International License

[8]	BERTopic dan NMF paling unggul pada interpretasi manusia; LDA terendah	Berbeda: penelitian ini belum memakai penilaian oleh manusia
[9]	Akurasi LDA dan BERTopic setara; klaster BERTopic lebih padat	Tidak bertentangan: selisih C_V kecil
[10]	LDA dan NMF memberi topik lebih bermakna pada teks pendek (metode non-embedding)	Searah: LDA kompetitif pada ulasan pendek
[11]	LDA pada ulasan PeduliLindungi: C_V 0,3963 untuk lima topik	Nilai tidak dapat dibandingkan langsung karena data dan prosedur berbeda

Hasil ini juga dapat dilihat dari sisi kegunaan topik. LDA memberi topik dominan pada semua 2.234 dokumen dengan pembagian yang cukup merata (46,7%, 27,7%, dan 25,6%). BERTopic menyisihkan 31 dokumen sebagai *outlier* dan menempatkan 77,1% dokumen pada satu topik, sedangkan tiga topik lainnya hanya memuat 75 dokumen (3,4%). Untuk korpus ini, ringkasan tiga topik dari LDA lebih langsung dapat dibaca daripada lima topik BERTopic, walaupun selisih koherensinya kecil. Sebaliknya, BERTopic memisahkan sekelompok ulasan dengan kata "umkm", "bisnis", dan "ekonomi" yang tidak muncul sebagai topik tersendiri pada LDA. Pada kedua metode, kata "potong" muncul pada topik yang sebagian besar berasal dari ulasan berating satu (66,1% pada Topik 1 LDA dan 72,9% pada Topik 0 BERTopic). Temuan ini dapat menjadi titik awal bagi pengembang untuk menelusuri ulasan terkait potongan biaya secara manual. Penyebab potongan tersebut tidak dapat disimpulkan dari data ini. Hasil penelitian ini memiliki beberapa keterbatasan. Pertama, data merupakan potret satu waktu dengan mayoritas ulasan berasal dari beberapa bulan terakhir sebelum pengambilan, sehingga topiknya dapat berbeda pada periode lain. Kedua, label topik adalah interpretasi penulis dan belum divalidasi oleh pakar. Ketiga, BERTopic hanya diuji dengan teks hasil *stemming*, dan hanya lima nilai min_cluster_size yang dicoba. Keempat, hasil berasal dari satu kali pelatihan dan tanpa uji statistik. Kelima, stopword seperti "bagus" dan "mantap" dihapus sehingga informasi sentimen tidak ikut dimodelkan. Keenam, jumlah kata teratas yang dinilai tetap 10. Pada pemeriksaan tambahan terhadap LDA, k terbaik tetap 3 untuk 5, 10, dan 15 kata, tetapi menjadi 2 untuk 20 kata (C_V 0,5733), sehingga pemilihan k bergantung pada jumlah kata yang dinilai. Pemeriksaan serupa untuk BERTopic belum dilakukan.

4. KESIMPULAN

Penelitian ini bertujuan memodelkan topik ulasan TikTok Shop Seller Center dari Google Play Store dan membandingkan LDA dan BERTopic dengan *coherence score*. Dari 2.234 dokumen hasil preprocessing, pencarian jumlah topik pada k = 2 sampai 15 memilih LDA dengan tiga topik (C_V 0,6107), dan pengujian min_cluster_size pada 10, 15, 20, 25, dan 30 memilih BERTopic dengan nilai 15 dan lima topik (C_V 0,6011). Dengan 10 kata teratas per topik serta teks dan kamus yang sama, LDA memperoleh C_V 0,6107, UMass -2,0906, dan C_NPMI 0,0133, sedangkan BERTopic memperoleh 0,6011, -4,1687, dan 0,0040. LDA lebih tinggi pada ketiga ukuran, tetapi selisih C_V dan C_NPMI kecil sehingga kedua metode belum dapat dibedakan secara tegas, sementara selisih UMass lebih jelas. Kedua metode menghasilkan topik yang beririsan, yaitu jual-beli dan pengelolaan toko, pesanan, pengiriman, potongan biaya, serta akun dan kemudahan penggunaan. Kontribusi penelitian ini adalah perbandingan dengan kriteria pemilihan parameter, data masukan, dan prosedur evaluasi yang sama, disertai analisis koherensi per topik dan penjelasan bahwa nilai rata-rata BERTopic dipengaruhi oleh topik kecil yang hanya memakai sedikit kata. Keterbatasannya adalah data berupa potret satu waktu dari satu aplikasi, label topik belum divalidasi pakar atau penilai manusia, BERTopic hanya diuji dengan teks hasil *stemming*, jumlah kata topik yang dinilai tetap 10, dan hasil berasal dari satu kali pelatihan tanpa uji statistik. Penelitian selanjutnya dapat memvalidasi topik dengan penilai manusia, menguji BERTopic pada teks tanpa *stemming*, memperluas rentang waktu data, dan menjalankan pelatihan berulang dengan uji statistik.

UCAPAN TERIMAKASIH

Terima kasih disampaikan kepada Program Studi Teknologi Informasi, Fakultas Sains dan Teknologi, Universitas Islam Negeri Walisongo Semarang, serta seluruh pihak yang telah memberikan dukungan, arahan, dan bimbingan selama proses penelitian ini berlangsung.

REFERENCES

[1] S. Kemp, "Digital 2025: Indonesia," DataReportal. Accessed: Oct. 10, 2026. [Online]. Available: https://datareportal.com/reports/digital-2025-indonesia

This Journal is licensed under a Creative Commons Attribution 4.0 International License

[2] W. L. Roldan-Baluis, N. A. Zapata, and M. S. M. Vásquez, "The effect of natural language processing on the analysis of unstructured text: A systematic review," *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 5, pp. 43–51, 2022, doi: 10.14569/IJACSA.2022.0130507.

[3] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet allocation," *Journal of Machine Learning Research*, vol. 3, pp. 993–1022, 2003, doi: 10.1162/jmlr.2003.3.4-5.993.

[4] G. Rosalinda, R. Santoso, and P. Kartikasari, "Pemodelan topik ulasan aplikasi Netflix pada Google Play Store menggunakan Latent Dirichlet Allocation," *Jurnal Gaussian*, vol. 11, no. 4, pp. 554–561, Feb. 2023, doi: 10.14710/j.gauss.11.4.554-561.

[5] M. Grootendorst, "BERTopic: Neural topic modeling with a class-based TF-IDF procedure," Mar. 11, 2022, *arXiv*: 2203.05794. doi: 10.48550/arXiv.2203.05794.

[6] M. D. Pratiwi and K. D. Tania, "Knowledge discovery through topic modeling on GoPartner user reviews using BERTopic, LDA, and NMF," *Journal of Applied Informatics and Computing*, vol. 9, no. 1, pp. 1–7, 2025, doi: 10.30871/jaic.v9i1.8782.

[7] O. Babalola, B. Ojokoh, and O. Boyinbode, "Comprehensive evaluation of LDA, NMF, and BERTopic's performance on news headline topic modeling," *Journal of Computing Theories and Applications*, vol. 2, no. 2, pp. 268–289, Nov. 2024, doi: 10.62411/jcta.11635.

[8] R. Egger and J. Yu, "A topic modeling comparison between LDA, NMF, Top2Vec, and BERTopic to demystify Twitter posts," *Frontiers in Sociology*, vol. 7, p. 886498, May 2022, doi: 10.3389/fsoc.2022.886498.

[9] L. Ma *et al.*, "AI-powered topic modeling: Comparing LDA and BERTopic in analyzing opioid-related cardiovascular risks in women," *Experimental Biology and Medicine*, vol. 250, p. 10389, Feb. 2025, doi: 10.3389/ebm.2025.10389.

[10] R. Albalawi, T. H. Yeap, and M. Benyoucef, "Using topic modeling methods for short-text data: A comparative analysis," *Frontiers in Artificial Intelligence*, vol. 3, p. 42, Jul. 2020, doi: 10.3389/frai.2020.00042.

[11] D. L. C. Pardede and M. A. I. Waskita, "Analisis pemodelan topik untuk ulasan tentang Peduli Lindungi," *Jurnal Ilmiah Informatika Komputer*, vol. 28, no. 1, pp. 17–26, Apr. 2023, doi: 10.35760/ik.2023.v28i1.7925.

[12] H. T. Saputra, A. Damanik, M. M. Shaquille, M. A. Rusydi, M. H. Riziq, and N. S. Zulva, "Analisis modelling pada reviewes Lazada Indonesia menggunakan Latent Dirichlet Allocation (LDA) untuk optimalisasi strategi bisnis," *JEKIN - Jurnal Teknik Informatika*, vol. 5, no. 1, pp. 361–371, Feb. 2025, doi: 10.58794/jekin.v5i1.1325.

[13] Nabila and C. I. Ratnasari, "Topic modeling of skincare comments from Female Daily," *Journal of Applied Informatics and Computing*, vol. 9, no. 4, pp. 1394–1405, Aug. 2025, doi: 10.30871/jaic.v9i4.9625.

[14] T. Z. Ramadhani, N. Y. Setiawan, and B. Rahayudi, "Evaluasi aplikasi layanan Alfagift berdasarkan topik ulasan pengguna menggunakan Latent Dirichled Allocation (LDA)," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, vol. 9, no. 6, Jun. 2025.

[15] M. G. Supiadin and A. D. Laksito, "Evaluating LDA and LSA for topic modeling in the Indonesian natural disaster," *Indonesian Journal of Computer Science*, vol. 12, no. 6, pp. 3326–3339, Dec. 2023, doi: 10.33022/ijcs.v12i6.3478.

[16] D. S. A. Rizki, M. S. Khabib, N. Rahmayuna, and V. G. Utomo, "Klasifikasi sentimen ulasan pengguna aplikasi layanan publik Google Play Store menggunakan NLP dan ML," *Jurnal Tekno Kompak*, vol. 20, no. 1, pp. 51–64, Feb. 2026, doi: 10.33365/jtk.v20i1.586.

[17] L. B. Hutama and D. Suhartono, "Indonesian hoax news classification with multilingual transformer model and BERTopic," *Informatica*, vol. 46, no. 8, pp. 81–90, Nov. 2022, doi: 10.31449/inf.v46i8.4336.

[18] U. Khairani, V. Mutiawani, and H. Ahmadian, "Pengaruh tahapan preprocessing terhadap model IndoBERT dan IndoBERTweet untuk mendeteksi emosi pada komentar akun berita Instagram," *Jurnal Teknologi Informasi dan Ilmu Komputer*, vol. 11, no. 4, pp. 887–894, Aug. 2024, doi: 10.25126/jtiik.1148315.

[19] N. Reimers and I. Gurevych, "Sentence-BERT: Sentence embeddings using Siamese BERT-networks," in *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, Hong Kong, China, 2019, pp. 3980–3990. doi: 10.18653/v1/D19-1410.

[20] M. Röder, A. Both, and A. Hinneburg, "Exploring the space of topic coherence measures," in *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining (WSDM)*, Shanghai, China, Feb. 2015, pp. 399–408. doi: 10.1145/2684822.2685324.

[21] G. Bouma, "Normalized (pointwise) mutual information in collocation extraction," in *Proceedings of the Biennial GSCL Conference 2009*, Tübingen, Germany, 2009, pp. 31–40.

This Journal is licensed under a Creative Commons Attribution 4.0 International License