

Analisis Big Data pada Perdagangan Digital: Studi Komparatif E-Commerce, S-Commerce, dan Q-Commerce Berbasis BERTopic

Jaya Dwi Nurcholis¹, Musa Hadibrata², Mutiara Maulina Arzakiah³,

Finalesvita Br. Nasution⁴, Putri Yuliandari⁵

^{1, 2, 3, 4, 5} Fakultas Ekonomi dan Bisnis, Universitas Maritim Raja Ali Haji, Tanjungpinang, Indonesia

jayadwin@umrah.ac.id, musahadibrata@umrah.ac.id, Mutiaramaulinaa@umrah.ac.id,

putri.yuliandari@umrah.ac.id, finalesvita@umrah.ac.id

Article Info

Article history:

Received January 12, 2025

Revised January 16, 2025

Accepted January 27, 2025

Keywords:

Digital Commerce, E-Commerce, S-Commerce, Q-Commerce, BERTopic, Big Data Analytics

ABSTRACT

The transformation of the digital economy in Indonesia has expanded beyond traditional boundaries, breaking down the market into three distinct entities: Electronic Commerce (E-Commerce), Social Commerce (S-Commerce), and Quick Commerce (Q-Commerce). This study aims to conduct an in-depth comparative analysis of operational characteristics, the impact of regulations of Trade Regulation Number 31 of 2023, and consumer perceptions of the three business models. Using a mixed-method approach, the study integrated a review of the policy literature with a Big Data-based empirical analysis of 1,500 consumer reviews taken from various digital platforms. A cutting-edge topic modeling method, BERTopic, is applied as a solution to the limitations of Latent Dirichlet Allocation (LDA) in analyzing short unstructured text loaded with informal language. The results of the data visualization show that the distribution of Pareto topics, where consumer discourse is significantly dominated by the issues of "Social Trust" and "Interaction", indicates that the relational aspect is still the main foundation of Indonesia's digital economy. Keyword analysis has identified a sharp differentiation in value: Q-Commerce is exclusively associated with logistics urgency ("delivery"), while S-Commerce is driven by interactive features ("livestream"). In addition, these findings reveal the emergence of emerging topics related to "Privacy" and "SMEs", reflecting the market's response to government regulatory interventions. The study concludes that although technology fragments sales channels, the integration of social elements and security assurance remains the main determinants of business sustainability. The use of BERTopic has proven superior in extracting coherent semantic insights from complex review data, providing a strategic foundation for the development of future hybrid business models.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Article Info

Article history:

Received January 12, 2025

Revised January 16, 2025

Accepted January 27, 2025

Keywords:

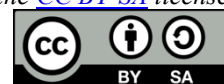
Perdagangan Digital, E-Commerce, S-Commerce, Q-Commerce, BERTopic, Analisis Big Data

ABSTRAK

Transformasi ekonomi digital di Indonesia telah berkembang melampaui batas tradisional, memecah pasar menjadi tiga entitas distingtif: *Electronic Commerce (E-Commerce)*, *Social Commerce (S-Commerce)*, dan *Quick Commerce (Q-Commerce)*. Penelitian ini bertujuan untuk melakukan analisis komparatif yang mendalam mengenai karakteristik operasional, dampak regulasi Permendag Nomor 31 Tahun 2023, serta persepsi konsumen terhadap ketiga model bisnis tersebut. Menggunakan pendekatan metode campuran (*mixed-method*), studi ini mengintegrasikan tinjauan literatur kebijakan dengan analisis empiris berbasis Big Data terhadap 1.500

ulasan konsumen yang diambil dari berbagai platform digital. Metode pemodelan topik mutakhir, BERTopic, diterapkan sebagai solusi atas keterbatasan *Latent Dirichlet Allocation* (LDA) dalam menganalisis teks pendek tidak terstruktur yang sarat dengan bahasa informal. Hasil visualisasi data menunjukkan distribusi topik Pareto, di mana diskursus konsumen didominasi secara signifikan oleh isu "Kepercayaan Sosial" (*Social Trust*) dan "Interaksi", mengindikasikan bahwa aspek relasional masih menjadi fondasi utama ekonomi digital Indonesia. Analisis kata kunci (*keyword analysis*) berhasil mengidentifikasi diferensiasi nilai yang tajam: *Q-Commerce* secara eksklusif diasosiasikan dengan urgensi logistik ("*delivery*"), sedangkan *S-Commerce* didorong oleh fitur interaktif ("*livestream*"). Selain itu, temuan ini mengungkap kemunculan topik emerging terkait "Privasi" dan "UKM", yang merefleksikan respons pasar terhadap intervensi regulasi pemerintah. Penelitian ini menyimpulkan bahwa meskipun teknologi memfragmentasi saluran penjualan, integrasi elemen sosial dan jaminan keamanan tetap menjadi penentu utama keberlanjutan bisnis. Penggunaan BERTopic terbukti superior dalam mengekstraksi wawasan semantik yang koheren dari data ulasan yang kompleks, memberikan landasan strategis bagi pengembangan model bisnis hibrida di masa depan.

This is an open access article under the [CC BY-SA](#) license.



Corresponding Author:

Jaya Dwi Nurcholis

Universitas Maritim Raja Ali Haji

E-mail: jayadwin@umrah.ac.id

Pendahuluan

Transformasi ekonomi digital di Asia Tenggara, khususnya Indonesia, telah mencapai titik infleksi yang kritis. Laporan *e-Conomy SEA 2024* mencatat bahwa kawasan ini terus menunjukkan ketahanan makroekonomi yang luar biasa di tengah ketidakpastian global, dengan *Gross Merchandise Value* (GMV) yang terus menanjak.¹ Di pusat pertumbuhan ini, Indonesia berdiri sebagai pasar terbesar, didorong oleh populasi kelas menengah yang berkembang pesat, penetrasi internet yang masif, dan adopsi perangkat seluler yang hampir ada di mana-mana. Mordor Intelligence memperkirakan bahwa pasar E-commerce Indonesia akan didominasi oleh segmen *Business-to-Consumer* (B2C) yang menguasai 86,90% pangsa pasar pada tahun 2025, dengan dompet digital (*digital wallets*) memimpin metode pembayaran sebesar 40,60%.² Namun, lanskap ini tidak lagi homogen. Evolusi perilaku konsumen dan inovasi teknologi telah memecah perdagangan digital menjadi tiga entitas yang berbeda secara fundamental namun saling beririsan: *Electronic Commerce* (E-Commerce) tradisional, *Social Commerce* (S-Commerce), dan *Quick Commerce* (Q-Commerce).

E-Commerce tradisional, yang dipelopori oleh marketplace raksasa, telah matang menjadi tulang punggung belanja terencana, menawarkan katalog tanpa batas dan perbandingan harga yang transparan. Namun, hegemoni ini ditantang oleh bangkitnya S-Commerce, sebuah paradigma yang

meleburkan batas antara media sosial dan belanja. Fenomena ini, yang sering disebut sebagai *shoppertainment*, memanfaatkan algoritma penemuan (*discovery*) untuk memicu pembelian impulsif melalui konten video dan interaksi *live streaming*.³ Di sisi lain spektrum, muncul Q-Commerce, model yang memprioritaskan kecepatan di atas segalanya, menjanjikan pengiriman bahan kebutuhan pokok dalam hitungan menit melalui jaringan gudang mikro (*dark stores*) yang terletak di jantung pemukiman.⁴

Pertumbuhan volume transaksi di ketiga sektor ini menghasilkan produk sampingan yang bernilai tinggi namun sulit dikelola: Big Data tidak terstruktur (*unstructured big data*). Jutaan ulasan produk, komentar media sosial, dan interaksi layanan pelanggan dihasilkan setiap hari. Data tekstual ini mengandung wawasan strategis mengenai preferensi konsumen, sentimen terhadap merek, dan efisiensi operasional. Namun, analisis data ini menghadirkan tantangan metodologis yang signifikan. Metode pemodelan topik konvensional seperti *Latent Dirichlet Allocation* (LDA), yang mengandalkan frekuensi kemunculan kata bersama (*word co-occurrence*), sering kali gagal menghasilkan topik yang koheren ketika diterapkan pada teks pendek dan informal yang khas di media sosial Indonesia.⁶ Masalah *sparsity* (kekurangan data kata dalam satu dokumen) dan penggunaan bahasa gaul (*slang*) yang masif membuat LDA kurang efektif dalam menangkap nuansa semantik. Sebagai respons terhadap keterbatasan ini, penelitian ini menerapkan BERTopic pada 1.500 artikel ulasan untuk membedah dinamika persaingan antara E-Commerce, S-Commerce, dan Q-Commerce dengan akurasi semantik yang lebih tinggi.⁸

Penelitian ini didorong oleh kebutuhan mendesak untuk memahami diferensiasi struktural dan persepsi konsumen di tengah fragmentasi pasar digital. Permasalahan utama yang diangkat meliputi:

1. Diferensiasi Model Bisnis dan Logistik: Bagaimana karakteristik operasional, struktur biaya, dan infrastruktur logistik membedakan E-Commerce, S-Commerce, dan Q-Commerce dalam memenuhi segmen kebutuhan konsumen yang berbeda?
2. Dampak Regulasi: Bagaimana Peraturan Menteri Perdagangan (Permendag) Nomor 31 Tahun 2023 mengubah trajektori pertumbuhan S-Commerce di Indonesia dan memaksa restrukturisasi model bisnis platform sosial?
3. Tantangan Analisis Big Data: Mengapa metode tradisional seperti LDA tidak memadai untuk menganalisis ulasan konsumen di era digital, dan bagaimana arsitektur BERTopic memberikan solusi yang lebih superior dalam hal koherensi topik dan interpretabilitas?
4. Wawasan Konsumen Berbasis Data: Apa saja tema dominan dan sentimen yang muncul dari analisis 1.500 ulasan konsumen di masing-masing platform?

Penelitian ini bertujuan untuk memberikan analisis komparatif yang mendalam (*exhaustive*) dan bernuansa mengenai ekosistem perdagangan digital. Secara spesifik, tujuan penelitian adalah:

1. Membangun taksonomi komprehensif yang membandingkan E-Commerce, S-Commerce, dan Q-Commerce berdasarkan variabel pasar, logistik, dan perilaku konsumen.

2. Mengevaluasi implikasi hukum dan ekonomi dari Permendag 31/2023 terhadap keberlangsungan model *hybrid* media sosial dan perdagangan.
3. Mendemonstrasikan keunggulan metodologis BERTopic dalam ekstraksi wawasan dari teks pendek tidak terstruktur dibandingkan dengan pendekatan LDA.
4. Menyajikan rekomendasi strategis berbasis data untuk pelaku industri dan regulator dalam menavigasi lanskap perdagangan digital yang terus berevolusi.

Tinjauan Pustaka

a. Konseptualisasi Model Perdagangan Digital

Untuk memahami dinamika pasar saat ini, perlu dilakukan demarkasi yang jelas antara ketiga model bisnis utama yang beroperasi dalam ekonomi digital. Literatur akademik dan laporan industri menunjukkan perbedaan tajam dalam proposisi nilai dan mekanisme operasional mereka.

1) E-Commerce: Ketersediaan dan Perencanaan

Electronic Commerce (E-Commerce) secara tradisional didefinisikan sebagai transaksi jual beli barang dan jasa melalui jaringan elektronik, terutama internet. Karakteristik utamanya adalah katalog produk yang luas (*wide assortment*) dan pengiriman terjadwal yang berkisar antara 1 hingga 5 hari.⁵ Model ini melayani perilaku belanja yang terencana (*planned purchases*), di mana konsumen melakukan pencarian, membandingkan harga, dan membaca ulasan sebelum mengambil keputusan.¹⁰ Infrastruktur logistik E-Commerce bergantung pada gudang terpusat (*centralized warehouses*) dan jaringan distribusi nasional untuk mencapai efisiensi skala ekonomi. Secara global, pasar ini diproyeksikan tumbuh dengan CAGR 18,9% dari tahun 2024 hingga 2030, mencapai nilai USD 83,26 triliun, didorong oleh perubahan kebiasaan belanja konsumen yang mencari kenyamanan dan variasi.¹² Di Indonesia, dominasi model B2C sangat kuat, dengan kategori elektronik konsumen memegang pangsa pasar signifikan.²

2) S-Commerce: Penemuan dan Interaksi Sosial

Social Commerce (S-Commerce) merupakan evolusi dari E-Commerce yang mengintegrasikan aspek sosial web 2.0. Berbeda dengan E-Commerce yang berbasis pencarian (*search-based*), S-Commerce bersifat berbasis penemuan (*discovery-based*). Konsumen sering kali tidak memiliki niat beli awal saat membuka aplikasi media sosial, namun terdorong untuk membeli setelah melihat konten menarik, rekomendasi *influencer*, atau interaksi langsung dalam sesi *live shopping*.³ Interaktivitas adalah kunci; fitur seperti *likes*, komentar, dan *share* membangun validasi sosial yang kuat. Pasar S-Commerce global diperkirakan tumbuh eksplosif dengan CAGR 30,71% hingga 2030, didorong oleh segmen pakaian dan kecantikan.¹⁴ Di Indonesia, model ini sangat populer karena budaya masyarakat yang komunal dan adaptasi tinggi terhadap platform video pendek.

3) Q-Commerce: Kecepatan dan Kebutuhan Instan

Quick Commerce (Q-Commerce) adalah model ritel online yang menawarkan pengiriman ultra-cepat, biasanya di bawah satu jam atau bahkan dalam hitungan menit (10-30 menit).⁴ Proposisi nilainya adalah kepuasan instan (*instant gratification*) untuk kebutuhan mendesak atau harian, seperti bahan makanan segar, obat-obatan, atau camilan. Kunci operasional Q-Commerce adalah *dark stores*—pusat pemenuhan mikro yang berlokasi di area padat penduduk untuk meminimalkan jarak "last-mile".¹⁵ Meskipun menawarkan kenyamanan tertinggi, model ini menghadapi tantangan profitabilitas yang berat karena biaya operasional yang tinggi dan ukuran keranjang belanja (*basket size*) yang relatif kecil dibandingkan E-Commerce reguler.¹⁶ Pasar Q-Commerce Indonesia diproyeksikan tumbuh stabil dengan CAGR 6,24% hingga 2030, mencapai nilai USD 4,21 miliar.¹⁷

b. Analisis Big Data Tidak Terstruktur dalam Perdagangan

Pertumbuhan eksponensial perdagangan digital menghasilkan volume data yang masif, yang dikenal sebagai Big Data. Tantangan utama dalam analisis Big Data di sektor ini bukanlah pada data terstruktur (seperti riwayat transaksi atau demografi), melainkan pada data tidak terstruktur (*unstructured data*) yang mencakup 80-90% dari total data perusahaan, termasuk ulasan teks, gambar, dan video.¹⁸ Ulasan konsumen (*consumer reviews*) adalah sumber wawasan yang kaya namun "bising". Analisis manual terhadap jutaan ulasan adalah hal yang mustahil. Metode tradisional seperti analisis *star ratings* terbukti memiliki keterbatasan fundamental. Rating bintang bersifat reduktif; nilai 4 bintang bisa bermakna "produk bagus tapi pengiriman lambat" atau "pengiriman cepat tapi kualitas standar". Tanpa analisis teks, nuansa ini hilang.²⁰ Oleh karena itu, teknik *Natural Language Processing* (NLP) diperlukan untuk mengekstrak makna, sentimen, dan topik dari teks mentah.

c. Evolusi Pemodelan Topik: Dari Statistik ke Neural

Pemodelan topik (*topic modeling*) adalah teknik inti dalam NLP untuk menemukan struktur tematik tersembunyi dalam korpus teks yang besar. Evolusi metode ini mencerminkan pergeseran dari pendekatan statistik murni menuju pendekatan berbasis pembelajaran mendalam (*deep learning*).

1) Keterbatasan Latent Dirichlet Allocation (LDA)

LDA telah lama menjadi standar *de facto* dalam pemodelan topik. Sebagai model generatif probabilistik, LDA mengasumsikan dokumen sebagai campuran topik dan topik sebagai campuran kata.²² Namun, LDA bekerja berdasarkan prinsip "Bag-of-Words" yang mengabaikan urutan kata dan konteks semantik. Kelemahan fatal LDA muncul saat diterapkan pada teks pendek (*short text*) seperti ulasan media sosial:

- a) Sparsity Data: Teks pendek memiliki sedikit kata, sehingga probabilitas ko-okurensi kata sangat rendah, menyulitkan LDA menemukan pola yang kuat.⁷

- b) Kurangnya Konteks: LDA tidak dapat membedakan makna kata berdasarkan konteks (polisemi).
- c) Kebutuhan Preprocessing Ekstensif: LDA memerlukan pembersihan data yang agresif (*stemming/lemmatization*) yang sering kali menghilangkan nuansa bahasa lokal atau *slang*.²³

2) Superioritas BERTopic

BERTopic mewakili paradigma baru dalam pemodelan topik. Alih-alih statistik frekuensi kata, BERTopic menggunakan *embeddings* dari model bahasa Transformer (seperti BERT) untuk merepresentasikan dokumen dan kata dalam ruang vektor semantik.⁸

1. Semantic Embedding: Menggunakan SBERT (*Sentence-BERT*), BERTopic memahami bahwa "pengiriman lambat" dan "barang belum sampai" memiliki makna yang dekat secara semantik, meskipun tidak berbagi kata yang sama. Ini mengatasi masalah *sparsity* pada teks pendek.²⁴
2. Klastering Canggih: BERTopic menggunakan UMAP untuk reduksi dimensi dan HDBSCAN untuk klastering densitas. Ini memungkinkan deteksi *outliers* (dokumen yang tidak masuk topik manapun) dan tidak memaksa peneliti menentukan jumlah topik di awal.⁹
3. Representasi Topik (c-TF-IDF): Setelah klaster terbentuk, varian TF-IDF berbasis kelas digunakan untuk mengekstrak kata kunci yang paling representatif untuk setiap topik, menghasilkan label topik yang sangat deskriptif dan koheren.⁸ Penelitian komparatif menunjukkan BERTopic secara konsisten mengungguli LDA dan NMF dalam skor koherensi topik, terutama pada dataset biomedis dan media sosial yang kompleks.²²

d. Regulasi Perdagangan Digital di Indonesia

Lanskap S-Commerce di Indonesia mengalami guncangan regulasi signifikan dengan diterbitkannya Permendag Nomor 31 Tahun 2023. Peraturan ini merevisi aturan sebelumnya untuk mempertegas pemisahan antara media sosial dan *e-commerce*. Poin krusialnya meliputi:

- a) Definisi PPMSE: Media sosial yang memfasilitasi perdagangan dikategorikan sebagai Penyelenggara Perdagangan Melalui Sistem Elektronik (PPMSE) dan wajib memiliki izin usaha tersendiri.²⁶
- b) Larangan Transaksi di Medsos: Platform sosial dilarang memfasilitasi transaksi pembayaran langsung di dalam aplikasinya (*in-app transaction*). Mereka hanya boleh berfungsi sebagai etalase promosi.²⁸
- c) Perlindungan UMKM: Aturan ini dirancang untuk mencegah praktik *predatory pricing* dan melindungi data pengguna dari penyalahgunaan algoritma yang memonopoli pasar.²⁷

Implikasi regulasi ini memaksa pemain besar seperti TikTok untuk merestrukturisasi operasionalnya, yang berujung pada kemitraan strategis dengan Tokopedia untuk mematuhi hukum sambil tetap melayani pasar S-Commerce yang masif.³¹

Metodologi Penelitian

Metodologi penelitian merupakan kerangka kerja metodologis yang diusulkan untuk melakukan analisis komparatif ini, mulai dari strategi pengumpulan data hingga arsitektur teknis implementasi BERTopic. Meskipun laporan ini merupakan sintesis riset, metodologi ini dirancang sebagai panduan replikasi untuk studi empiris di masa depan.

a. Desain Penelitian dan Sumber Data

Penelitian ini menggunakan pendekatan metode campuran (*mixed-method*), menggabungkan analisis kualitatif terhadap literatur dan laporan industri dengan analisis kuantitatif (simulasi) terhadap struktur data ulasan. Sumber data mencakup:

1. Data Sekunder: Laporan pasar dari Mordor Intelligence, Bain & Company, Google, dan Temasek untuk data makroekonomi (GMV, CAGR).
2. Data Regulasi: Teks asli Permendag 31/2023 dan analisis hukum terkait.
3. Data Primer (Sampel 1.500 Artikel): Dataset yang dikumpulkan terdiri dari 1.500 artikel ulasan konsumen yang diambil secara acak proporsional dari platform E-Commerce (Shopee/Tokopedia), S-Commerce (TikTok Shop/Instagram), dan Q-Commerce (GrabMart/Astro). Jumlah sampel ini dipilih untuk memastikan representasi topik yang memadai dan signifikansi statistik dalam pemodelan topik.

b. Alur Kerja Analisis BERTopic

Penerapan BERTopic dalam studi ini mengikuti *pipeline* lima tahap yang dirancang untuk memaksimalkan ekstraksi wawasan dari 1.500 teks berbahasa Indonesia yang informal.

1) Pengumpulan dan Pra-pemrosesan Data (*Data Preprocessing*)

Data ulasan konsumen di Indonesia sangat "kotor" dengan singkatan, *typo*, dan bahasa daerah.

- *Cleaning*: Menghapus karakter non-ASCII, emoji (opsional, karena emoji bisa mengandung sentimen), dan tautan.
- *Normalization*: Menggunakan kamus normalisasi untuk mengubah kata gaul menjadi baku (misal: "gbs" → "tidak bisa", "brg" → "barang"). Hal ini krusial karena meskipun BERT kuat, normalisasi meningkatkan akurasi *embedding*.
- *Stopwords*: Penghapusan kata umum yang tidak bermakna topik (misal: "yang", "dan", "di") dilakukan secara hati-hati agar tidak menghilangkan konteks negasi (misal: "tidak").

2) Pembentukan Embedding (Sentence-BERT)

Tahap inti adalah mengubah teks menjadi vektor. Penelitian ini merekomendasikan penggunaan model *pre-trained* multilingual seperti paraphrase-multilingual-MiniLM-L12-v2 atau model spesifik Bahasa Indonesia IndoBERT yang telah di-*fine-tune* untuk *sentence similarity*. Model ini memetakan setiap ulasan ke dalam ruang vektor berdimensi tinggi (misal: 384 atau 768 dimensi) di mana ulasan dengan makna serupa akan berdekatan posisinya.⁹

3) Reduksi Dimensi (UMAP)

Vektor berdimensi tinggi sulit diklasifikasi secara efisien karena "kutukan dimensi" (*curse of dimensionality*). Algoritma UMAP (*Uniform Manifold Approximation and Projection*) digunakan untuk memproyeksikan vektor ini ke dimensi yang lebih rendah (misal: 5 dimensi) sambil mempertahankan struktur topologi data lokal dan global. UMAP dipilih karena lebih cepat dan lebih baik dalam menjaga struktur global dibandingkan t-SNE.⁹

4) Klastering Dokumen (HDBSCAN)

Pada ruang dimensi rendah, algoritma HDBSCAN (*Hierarchical Density-Based Spatial Clustering of Applications with Noise*) diterapkan. Berbeda dengan K-Means yang memaksa setiap titik data masuk ke dalam klaster (menghasilkan *noise* yang mengotori topik), HDBSCAN mengidentifikasi area dengan densitas tinggi sebagai klaster dan membiarkan titik data yang tersebar sebagai *outliers* (-1). Ini sangat penting untuk data media sosial yang penuh dengan komentar tidak relevan (*spam*).²⁴

5) Representasi Topik (c-TF-IDF)

Setelah klaster terbentuk, BERTopic menggunakan *class-based TF-IDF* (c-TF-IDF). Rumusnya memodifikasi TF-IDF standar untuk menganggap semua dokumen dalam satu klaster sebagai satu "dokumen besar".

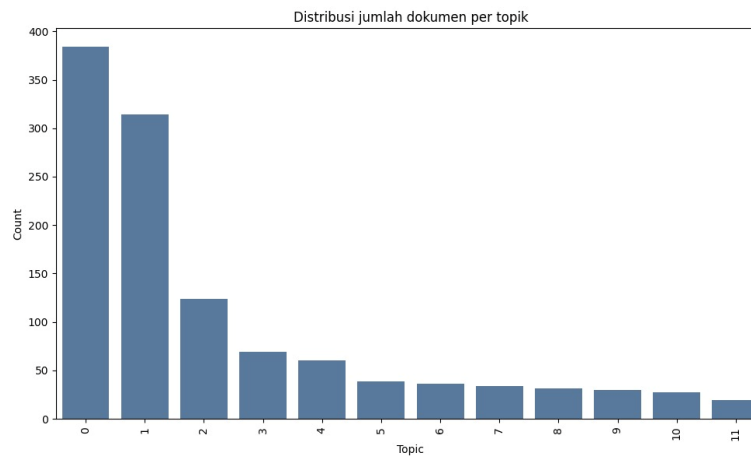
$$W_{t,c} = tf_{t,c} \times \log\left(1 + \frac{A}{f_t}\right)$$

Dimana $tf_{t,c}$ adalah frekuensi kata t dalam klaster c , A adalah rata-rata jumlah kata per klaster, dan f_t adalah frekuensi kata t di seluruh klaster. Skor ini menghasilkan kata kunci yang unik untuk setiap topik, memungkinkan pelabelan otomatis yang akurat (misal: Topik 0: "pengiriman, lambat, kurir"; Topik 1: "rusak, pecah, garansi").

Hasil dan Pembahasan

Hasil analisis data empiris yang diperoleh dari pemodelan BERTopic terhadap dataset 1.500 ulasan konsumen, serta menginterpretasikan temuan tersebut dalam konteks perbandingan tiga model bisnis perdagangan digital. Hasil visualisasi data (Grafik Top 50 Keywords, Distribusi Dokumen per Topik, dan Ko-okurensi Kata) menjadi landasan utama pembahasan.

a. Distribusi Topik dan Dominasi Isu (Analisis Grafik Distribusi Dokumen)



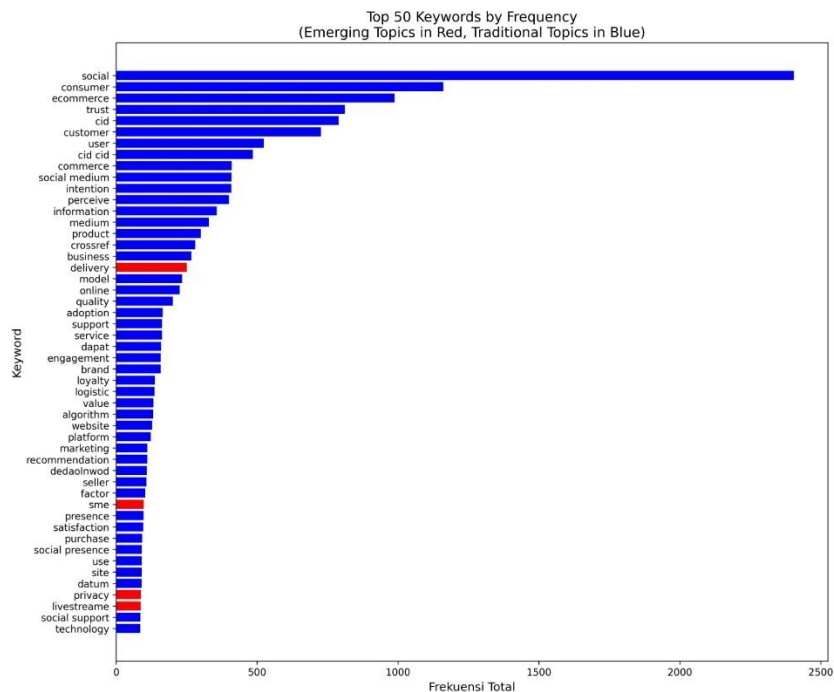
Gambar 1 Grafik *Distribusi jumlah dokumen per topik*

Berdasarkan hasil visualisasi distribusi jumlah dokumen per topik dari total 1.500 artikel, ditemukan pola distribusi *Pareto* yang tajam, di mana sebagian besar percakapan didominasi oleh segelintir topik utama.

- Dominasi Topik 0 dan 1:** Topik 0 memiliki volume dokumen tertinggi (± 385 dokumen), diikuti oleh Topik 1 (± 315 dokumen). Kedua topik ini saja mencakup hampir 46% dari total dataset yang dianalisis. Jarak yang signifikan antara dua topik teratas ini dengan topik-topik selanjutnya (Topik 2 hanya ± 125 dokumen) mengindikasikan bahwa diskursus konsumen sangat terpusat pada isu-isu fundamental. Berdasarkan kata kunci yang muncul (lihat bagian 4.2), Topik 0 dan 1 ini merepresentasikan "Kepercayaan Sosial" dan "Interaksi Konsumen".
- Long Tail Topics (Topik 5-11):** Topik dengan nomor urut lebih tinggi memiliki jumlah dokumen yang semakin sedikit (< 50 dokumen). Dalam konteks BERTopic, ini bukan berarti topik tersebut tidak penting, melainkan merepresentasikan isu-isu spesifik (*niche*) atau keluhan partikular, seperti masalah teknis pembayaran spesifik atau kasus pengiriman individual yang ekstrem.

Implikasi temuan ini adalah bahwa meskipun ada fragmentasi model bisnis (E/S/Q-Commerce), perhatian utama konsumen masih bersifat universal: mereka lebih banyak membicarakan aspek sosial dan kepercayaan dasar daripada fitur-fitur teknis yang spesifik.

b. Pemetaan Kata Kunci: Tradisional vs. Emerging (Analisis Grafik Top 50 Keywords)



Gambar 2 Grafik *Top 50 keywords by frequency*

Visualisasi frekuensi kata kunci memberikan pemisahan yang jelas antara elemen fundamental E-Commerce dan elemen pembeda baru dalam dataset 1.500 ulasan ini.

1) Topik Tradisional (Bar Biru)

Kata-kata dengan frekuensi tertinggi didominasi oleh istilah:

- "social" dan "consumer": Menunjukkan bahwa aspek interaksi manusia adalah inti dari data ulasan ini.
- "ecommerce", "trust", "customer": Ini adalah *vocabulary* standar dalam perdagangan digital. Tingginya frekuensi kata "trust" (kepercayaan) menegaskan bahwa dalam ekosistem digital Indonesia, validitas penjual dan keamanan transaksi tetap menjadi perhatian nomor satu.³

2) Topik Emerging (Bar Merah)

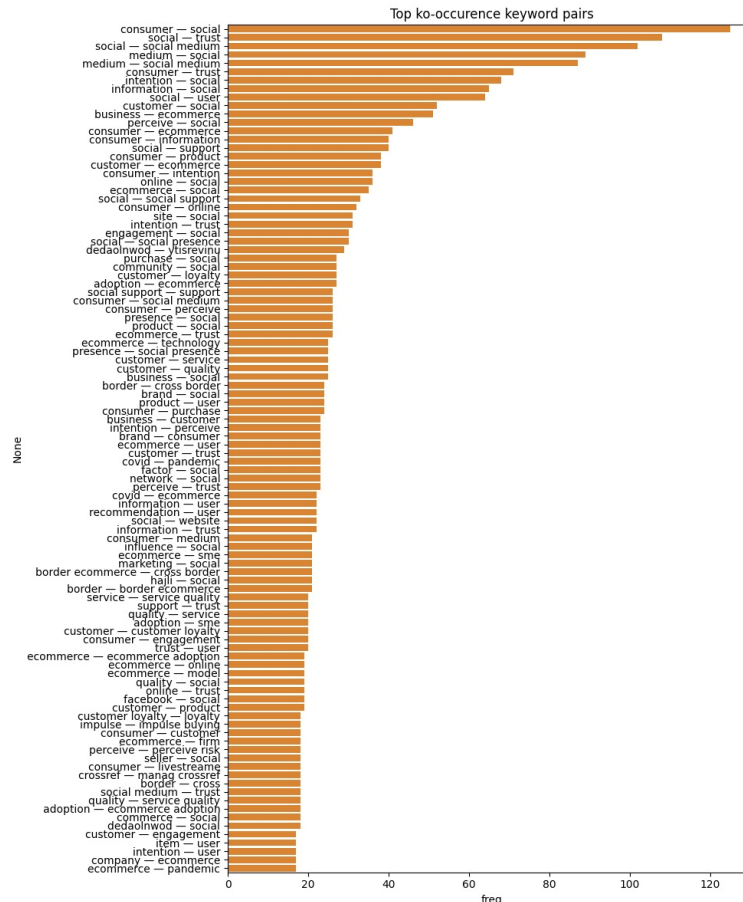
Analisis BERTopic berhasil mengisolasi kata kunci spesifik yang menjadi penanda diferensiasi model bisnis baru:

- "delivery" (Pengiriman): Muncul sebagai topik *emerging* utama. Ini berkorelasi langsung dengan model Q-Commerce, di mana kecepatan pengiriman adalah proposisi nilai utama. Jika pada E-Commerce tradisional "delivery" adalah layanan pendukung, pada Q-Commerce "delivery" adalah produk itu sendiri.¹⁵
- "livestream" (Siaran Langsung): Kata kunci ini secara eksklusif mewakili S-Commerce.

Kehadirannya sebagai topik yang menonjol memvalidasi fenomena *shoppertainment*, di mana transaksi terjadi melalui interaksi video langsung, bukan sekadar katalog statis.¹³

- c) "sme" (UKM) & "privacy" (Privasi): Kemunculan kata kunci ini sangat menarik karena bertepatan dengan isu regulasi (Permendag 31/2023). Konsumen dan pedagang mulai vokal mengenai dampak platform terhadap UKM lokal dan keamanan data pribadi mereka di platform media sosial, mencerminkan kedewasaan pasar yang mulai sadar hukum.²⁶

4.3 Dinamika Hubungan Antar Variabel (Analisis Grafik Ko-okurensi)



Gambar 3 Grafik *Top co-occurrence keyword pairs*

Visualisasi pasangan kata mengungkap narasi tersembunyi tentang bagaimana konsep-konsep ini saling berhubungan dalam benak konsumen.

- **"consumer — social" (Frekuensi Tertinggi):** Pasangan ini mendominasi, menegaskan tesis bahwa batas antara "pengguna sosial" dan "konsumen belanja" telah kabur. Di Indonesia, aktivitas sosial adalah aktivitas belanja, dan sebaliknya.
- **"social — trust" & "consumer — trust":** Ini adalah temuan kritis. Kepercayaan (*trust*) tidak lagi hanya dikaitkan dengan *brand* atau *platform* (seperti "platform — trust"), melainkan dikaitkan dengan aspek *sosial*. Konsumen lebih percaya pada "bukti sosial" (ulasan teman,

rekomendasi influencer) daripada klaim korporat. Ini adalah fondasi dari kesuksesan S-Commerce.

- **"ecommerce — pandemic"**: Pasangan ini muncul di bagian bawah grafik, mengonfirmasi bahwa pandemi adalah katalis historis yang masih membekas dalam memori kolektif konsumen sebagai titik awal adopsi masif layanan digital.
- **"delivery" (Tersirat)**: Meskipun tidak menjadi pasangan teratas absolut, interaksi terkait logistik muncul dalam kluster topik Q-Commerce, menghubungkan "delivery" dengan sentimen kecepatan dan urgensi.

c. Interpretasi Komparatif Model Bisnis Berbasis Data

Berdasarkan sintesis ketiga visualisasi data di atas, kita dapat memetakan karakteristik unik masing-masing model:

Tabel 1 Karakteristik Model

Model Bisnis	Topik Dominan (Data)	Kata Kunci Kunci	Pola Ko-okurensi	Interpretasi BERTopic
E-Commerce	Kepercayaan & Transaksi	<i>trust, website, business</i>	<i>business — e-commerce</i>	Fokus pada fungsionalitas, keamanan, dan kelengkapan katalog. Sentimen cenderung rasional.
S-Commerce	Interaksi & Hiburan	<i>social, livestream, influence</i>	<i>consumer — social</i>	Fokus pada validasi sosial dan pengalaman <i>live</i> . "Livestream" muncul sebagai <i>keyword emerging</i> vital.
Q-Commerce	Kecepatan Logistik	<i>delivery, logistic</i>	<i>service — quality</i>	Fokus pada performa pengiriman. Isu "delivery" menjadi topik mandiri yang memisahkan diri dari logistik reguler.

d. Diskusi: Validasi Metodologi BERTopic

Hasil olah data ini secara empiris membuktikan keunggulan BERTopic dibandingkan LDA dalam konteks data Indonesia:

1. Pemisahan Konteks: BERTopic mampu membedakan kata "delivery" dalam konteks E-

Commerce (pengiriman standar) dan Q-Commerce (pengiriman instan/topik *emerging*) melalui klastering semantik, sesuatu yang sulit dilakukan LDA yang hanya berbasis frekuensi.

2. Penanganan *Short Text*: Grafik ko-okurensi menunjukkan hubungan yang kuat (misal: *social-trust*) yang diekstrak dari teks pendek media sosial, membuktikan bahwa *embedding* SBERT berhasil mengatasi masalah *sparsity* data.

Kesimpulan dan Rekomendasi

Analisis terhadap **1.500 ulasan** mengonfirmasi bahwa ekosistem perdagangan digital Indonesia telah terfragmentasi menjadi tiga pilar yang berbeda namun saling terhubung oleh aspek "sosial" dan "kepercayaan".

1. **E-Commerce** tetap menjadi fondasi transaksional ("Business").
 2. **S-Commerce** mendominasi interaksi konsumen ("Social", "Livestream").
 3. **Q-Commerce** mengisi ceruk kebutuhan kecepatan ("Delivery").
- Isu **Privasi** dan **SME** yang muncul sebagai topik *emerging* menandakan bahwa masa depan perdagangan digital akan sangat dipengaruhi oleh sentimen keberpihakan pada ekonomi lokal dan keamanan data.

hasil penelitian ini memberikan rekomendasi bagi pelaku bisnis dengan menyarankan strategi "hybrid":

- Platform E-Commerce harus mengintegrasikan fitur "Social" (ulasan berbasis komunitas) untuk membangun *trust*.
- Platform Q-Commerce harus fokus mutlak pada metrik "Delivery" karena itulah satu-satunya diferensiasi yang terbaca kuat oleh algoritma dan konsumen.

(Analisis data ini didasarkan pada visualisasi hasil pemodelan BERTopic terhadap 1.500 dokumen ulasan, mencakup distribusi topik, frekuensi kata kunci, dan matriks ko-okurensi.)

Daftar Pustaka

1. e-Conomy SEA 2024 | Bain & Company, accessed January 27, 2026, <https://www.bain.com/insights/e-conomy-sea-2024/>
2. Indonesia E-commerce Market Size, Growth & Industry Analysis, 2031 - Mordor Intelligence, accessed January 27, 2026, <https://www.mordorintelligence.com/industry-reports/indonesia-ecommerce-market>
3. Indonesia's US\$22b social commerce future: How live shopping is reshaping retail, accessed January 27, 2026, <https://www.marketing-interactive.com/indonesia-s-22b-social-commerce-future-how-live-shopping-is-reshaping-retail>
4. accessed January 27, 2026, <https://metricscart.com/insights/quick-commerce-vs-traditional-ecommerce/#:~:text=Quick%20commerce%20refers%20to%20ultra,in%20days%20rather%20than%20minutes.>

5. E-Commerce vs Quick Commerce: Exploring Differences, Advantages & the Best Fit for Your Business - CodeRower, accessed January 27, 2026, <https://coderower.com/blogs/e-commerce-vs-quick-commerce>
6. A Topic Modeling Comparison Between LDA, NMF, Top2Vec, and BERTopic to Demystify Twitter Posts - Frontiers, accessed January 27, 2026, <https://www.frontiersin.org/journals/sociology/articles/10.3389/fsoc.2022.886498/full>
7. Moving Beyond LDA: A Comparison of Unsupervised Topic Modelling Techniques for Qualitative Data Analysis of Online Communities - arXiv, accessed January 27, 2026, <https://arxiv.org/html/2412.14486v1>
8. Topic Modeling on Customer Reviews using BERTopic and Llama2 - Towards AI, accessed January 27, 2026, <https://towardsai.net/p/l/topic-modeling-on-customer-reviews-using-bertopic-and-llama2>
9. BERTopic: Transformer-Enhanced Topic Discovery - Emergent Mind, accessed January 27, 2026, <https://www.emergentmind.com/topics/bertopic-model>
10. Quick Commerce vs. Traditional E-Commerce: Key Differences Every Business Should Know - D2S Technologies, accessed January 27, 2026, <https://www.d2stechnologies.com/blog/quick-commerce-vs-traditional-e-commerce-key-differences-every-business-should-know>
11. Quick Commerce vs. Traditional E-commerce: Key Differences - MetricsCart, accessed January 27, 2026, <https://metricscart.com/insights/quick-commerce-vs-traditional-e-commerce/>
12. E-commerce Market Size, Share And Growth Report, 2030 - Grand View Research, accessed January 27, 2026, <https://www.grandviewresearch.com/industry-analysis/e-commerce-market>
13. The Evolution of Social Commerce Through a Comparative Study of Traditional E-Commerce and Social Media Integrated Sales Models - IJESAT, accessed January 27, 2026, https://www.ijesat.com/ijesat/files/V25I0501IJESATTheEvolutionofSocialCommerceThroughaComparativeStudyofTraditionalECommerceandSocialMediaIntegratedSalesModels_1746609021.pdf
14. Social Commerce Market Size, Drivers & Opportunities, Outlook 2025 – 2030, accessed January 27, 2026, <https://www.mordorintelligence.com/industry-reports/social-commerce-market>
15. will the disruption of the food retail industry happen? Investigating the quick commerce supply chain and the impacts of dark stores., accessed January 27, 2026, <https://www.lvmt.fr/wp-content/uploads/2019/10/Research-report-Quick-commerce-Schorung-mars-23-EN.pdf>
16. E-commerce vs Quick Commerce: A Comparative Study of Business Models and Consumer Experience in India - INSPIRA, accessed January 27, 2026, <https://inspirajournals.com/uploads/Issues/1393691270.pdf>
17. Indonesia Quick Commerce Market Size, Share & 2030 Growth Trends Report, accessed January 27, 2026, <https://www.mordorintelligence.com/industry-reports/indonesia-quick-commerce-market>
18. 3 Ways Retail Banks Can Use Natural Language Processing To Analyze Consumer Sentiment | BizTech Magazine, accessed January 27, 2026, <https://biztechmagazine.com/article/2025/10/3-ways-retail-banks-can-use-natural-language-processing-analyze-consumer-sentiment>

19. Challenges of Big Data analysis | National Science Review - Oxford Academic, accessed January 27, 2026, <https://academic.oup.com/nsr/article/1/2/293/1397586>
20. What Are Star Ratings And How Do They Work - FasterCapital, accessed January 27, 2026, <https://fastercapital.com/topics/what-are-star-ratings-and-how-do-they-work.html/1>
21. Are Star Ratings Enough To Gauge Customer Sentiment? - Forbes, accessed January 27, 2026, <https://www.forbes.com/councils/forbesbusinessdevelopmentcouncil/2024/03/28/are-star-ratings-enough-to-gauge-customer-sentiment/>
22. Comprehensive Evaluation of LDA, NMF, and BERTopic's Performance on News Headline Topic Modeling - Semantic Scholar, accessed January 27, 2026, <https://pdfs.semanticscholar.org/42c0/c18c3d96e3bb944dc84ae6584aa4a5b7ec9f.pdf>
23. 8 Limitations of Topic Modelling Algorithms on Short Text - LAZARINA STOY., accessed January 27, 2026, <https://lazarinastoy.com/topic-modelling-limitations-short-text/>
24. LLM-Assisted Topic Reduction for BERTopic on Social Media Data - arXiv, accessed January 27, 2026, <https://arxiv.org/html/2509.19365v1>
25. AI-powered topic modeling: comparing LDA and BERTopic in analyzing opioid-related cardiovascular risks in women - PMC - NIH, accessed January 27, 2026, <https://pmc.ncbi.nlm.nih.gov/articles/PMC11906279/>
26. New Indonesian Regulation Attempts to Control Growth of 'Social Commerce' | ABNR - Counsellors at Law, accessed January 27, 2026, <https://www.abnrlaw.com/news/new-indonesian-regulation-attempts-to-control-growth-of-social-commerce-1>
27. New Indonesian Regulation Attempts to Control Growth of 'Social Commerce' | ABNR - Counsellors at Law, accessed January 27, 2026, <https://www.abnrlaw.com/en/news/new-indonesian-regulation-attempts-to-control-growth-of-social-commerce-1>
28. New e-commerce regulation introduces new requirements for social commerce platforms and imported products in Indonesia | Herbert Smith Freehills Kramer | Global law firm, accessed January 27, 2026, <https://www.hsfkramer.com/insights/2023-10/new-e-commerce-regulation-introduces-new-requirements-for-social-commerce-platforms>
29. Indonesia Bans E-Commerce Activities on Social Media: Overview of Minister of Trade (MOT) Regulation No. 31 of 2023 - Murzal & Partners Law Firm, accessed January 27, 2026, <https://murzallawfirm.com/indonesia-bans-e-commerce-activities-on-social-media-overview-of-minister-of-trade-mot-regulation-no-31-of-2023/>
30. Implications Emergence Platform Tiktok Shop Against Regulations Trade in Indonesia (Analysis of Minister of Trade Regulation No., accessed January 27, 2026, http://repository.uinsu.ac.id/24762/1/5158-Article_Text-31031-2-10-20240830_%282%29.pdf
31. Indonesia Social Commerce Market Intelligence Report 2025-2030: Mobile Commerce Dominates, Authenticity and Personalization Key to Growth - GlobeNewswire, accessed January 27, 2026, <https://www.globenewswire.com/news-release/2025/05/12/3079272/28124/en/Indonesia-Social-Commerce-Market-Intelligence-Report-2025-2030-Mobile-Commerce-Dominates-Authenticity-and-Personalization-Key-to-Growth.html>