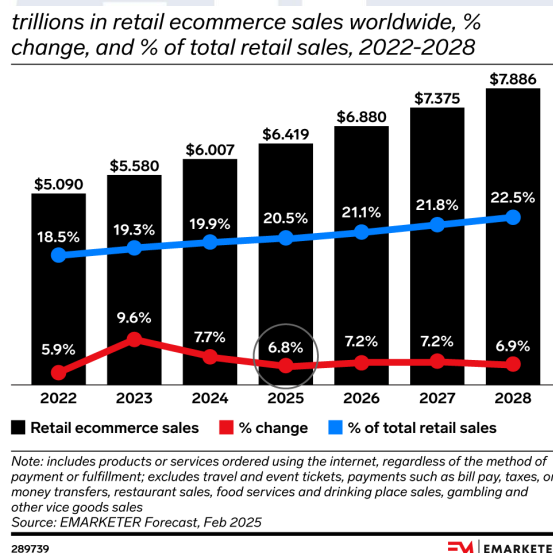


BAB I

PENDAHULUAN

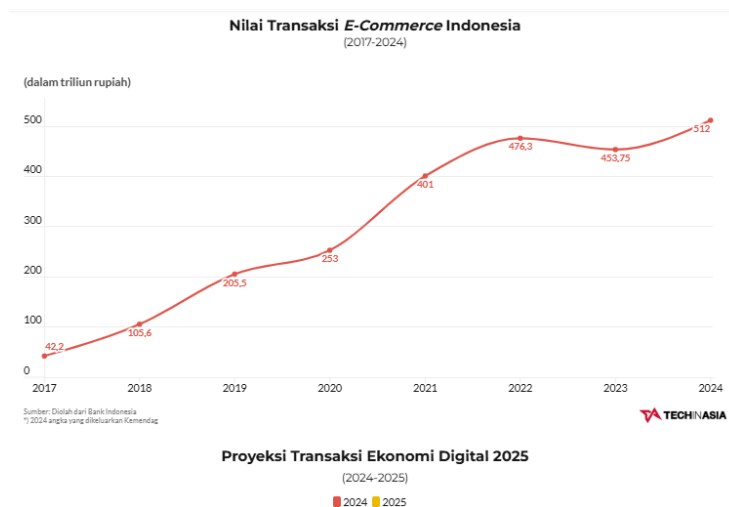
1.1 Latar Belakang

E-commerce merupakan salah satu sektor yang mengalami pertumbuhan pesat di tahun 2024 secara global[1]. Pada tahun 2024, penjualan *e-commerce* global mencapai angka penjualan \$6,007 triliun dan penjualan global pada *e-commerce* diprediksi meningkat sebesar 22,5% pada tahun 2028[2]. Kemudahan metode pembayaran dan akses untuk membeli berbagai produk melalui *smartphone* menjadi faktor belanja di *e-commerce* disukai oleh banyak orang[3].



Gambar 1.1 Pertumbuhan *e-commerce* global sales

Di Indonesia, potensi pertumbuhan transaksi *e-commerce* disebabkan pada semakin banyaknya pengguna *e-commerce*[4]. Berdasarkan Kementerian Perdagangan, nilai pembelian melalui *e-commerce* mencapai Rp 512 triliun sedangkan jumlah orang yang menggunakan *e-commerce* diperkirakan mencapai 65,65 juta orang pada tahun 2024[5][6]. Tren belanja melalui media sosial semakin berkembang seiring meluasnya internet dan pengguna media sosial[7].



Gambar 1.2 Nilai transaksi *e-commerce* di Indonesia

Perkembangan media sosial memberikan dampak pada perilaku konsumen yang mendasarkan pembelian pada rekomendasi atau ulasan untuk membeli barang [8]. *Social commerce* memungkinkan pengguna media sosial untuk membeli dan menjual barang atau jasa secara langsung melalui platform media sosial sehingga media sosial tidak hanya berfungsi untuk menemukan barang tetapi juga mendukung proses belanja[9]. Pengguna media sosial bisa mencari hingga membeli barang tanpa perlu keluar dari aplikasi media sosial. Fenomena tersebut terlihat jelas pada berbagai negara saat aplikasi media sosial seperti TikTok, Instagram, dan Facebook menawarkan fitur belanja langsung di dalam platform[10].

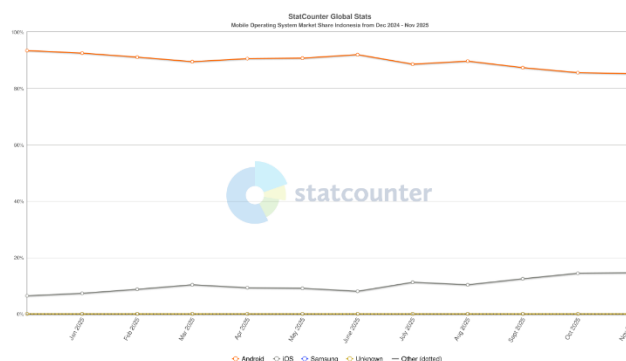
Sejak masuk di tahun 2021, TikTok Shop menjadi salah satu platform yang sukses menarik perhatian masyarakat Indonesia yang menggabungkan interaksi pengguna melalui video pendek dan live streaming dengan platform *e-commerce* [11]. Indonesia menjadi pasar terbesar Tiktok Shop di Asia Tenggara dengan nilai transaksi sebesar \$6,198 juta dolar di tahun 2024[12]. Akan tetapi, Tiktok Shop terpaksa menghentikan operasional pada 4 Oktober 2023 akibat Permendag Nomor 31 Tahun 2023 tentang perizinan Berusaha, Periklanan, Pembinaan, dan Pengawasan Pelaku Usaha dalam Perdagangan melalui Sistem Elektronik [13]. Peraturan tersebut mengakibatkan platform media sosial tidak diperbolehkan menyediakan fitur transaksi perdagangan pada sistem elektroniknya dan hanya memfasilitasi aktivitas promosi[14]. Tiktok kembali beroperasi di Indonesia pada

tanggal 12 Desember 2023 setelah menjalin Kerjasama dengan Tokopedia dan Goto[15]. Kerja sama tersebut meletakkan Tokopedia dan TikTok Shop Indonesia menjadi Shop Tokopedia yang berada di bawah entitas PT Tokopedia[16]. Salah satu syarat utama agar Tiktop Shop dapat beroperasi kembali adalah pemisahan platform media sosial TikTok dengan platform jual beli. Tiktok yang hanya memiliki izin PSE (Penyedia Sistem Elektronik) dapat beroperasi menggunakan izin PMSE dalam negeri yang sudah dimiliki oleh Tokopedia[17].

Integrasi Tiktok Shop dengan platform *e-commerce* Tokopedia memberikan dampak (*disruption*) pada pengguna TikTok Shop [18] khususnya pada ekosistem penjual sebagai salah satu *stakeholder* platform *e-commerce* yang sebelumnya menggunakan aplikasi *Seller Center*. Mitra penjual yang terdampak baik di Tiktok Shop dan Tokopedia mencapai 21 juta penjual lebih dengan pasar potensial pasca-integrasi Tiktok Shop ke Tokopedia adalah 100 juta orang pengguna aktif bulanan di Tokopedia dan 125 juta orang Tiktok di Indonesia di tahun 2024[19]. Oleh karena itu, Tokopedia dan TikTok Shop meluncurkan Tokopedia & TikTok Shop Seller Center sebagai pusat pengelola toko baru untuk membantu penjual dari kedua *comerce* dalam mengelola operasional bisnis di kedua platform tersebut dengan lebih mudah dan efisien [20]. Langkah tersebut dilakukan sebagai strategi Tiktok dan Tokopedia untuk mendorong pertumbuhan dan inovasi yang berkelanjutan dengan memanfaatkan sumber daya dari kedua belah pihak[21].

Perubahan ini menimbulkan reaksi dari sisi penjual yang menggunakan aplikasi Seller Center[22]. Salah satu cara penjual atau pedagang untuk memberikan reaksi atau keluhan mengenai peluncuran aplikasi *Seller Center* tersebut adalah melalui komentar yang terekam di fitur ulasan pengguna di platform Google Play Store sebagai tempat akses aplikasi Tokopedia & TikTok Shop Seller dan Tokopedia Seller[23]. Aplikasi *seller center* Tokopedia dan Tiktok Shop Seller telah diunduh lebih dari 10 juta kali di Google Play Store sebagai platform distribusi aplikasi dengan *operating system* Android [24]. Data *market share* sistem operasi perangkat seluler di Indonesia menunjukkan Android menguasai 85% pangsa pasar di Indonesia [25] sehingga ulasan pada Google Play Store atau aplikasi berbasis Android dapat menggambarkan persepsi mayoritas pengguna aplikasi seller center.

Postingan ulasan di Google Play Store dapat digunakan sebagai sumber data yang potensial untuk dianalisis karena memuat pengalaman pengguna [26] sebelum dan sesudah aplikasi Tokopedia & Tiktok Shop Seller resmi dapat diakses untuk migrasi penjual.



Gambar 1.3 Data market share sistem operasi perangkat seluler di Indonesia Desember 2024 - November 2025 oleh StatCounter[25]

Analisis ulasan secara manual memakan waktu yang lama sehingga salah satu pendekatan yang dapat digunakan untuk menganalisis ulasan adalah teknik analisis sentimen (*sentiment analysis*) yang dapat mengidentifikasi dan mengklasifikasikan opini dalam format data teks menjadi kategori positif, negatif, atau netral[27]. Penelitian sebelumnya [28][29] masih berfokus pada klasifikasi sentimen ulasan positif, negatif, atau netral. Jumlah ulasan aplikasi akan semakin meningkat seiring meningkatnya pengguna aplikasi sehingga menggunakan ABSA (*Aspect-based Sentiment Analysis*) dapat memberikan analisis sentimen terhadap topik atau tema tertentu[30]. Pendekatan *Topic Modeling* mampu mengelompokkan tema yang merepresentasikan *corpus* dokumen bergantung pada distribusi kata[31]. Distribusi kata sebagai representasi teks yang digunakan pada pemodelan topik diperoleh dengan teknik *word-embedding*, pembobotan kata atau frekuensi pada dokumen serta *corpus*[32]. Melalui pendekatan ini [33], kumpulan dokumen atau ulasan yang besar dapat diuraikan untuk mengali tema atau isu yang muncul pada ulasan aplikasi layanan seller center. PoS (*part-of-speech*) dapat digunakan untuk seleksi fitur pada aspek ekstraksi ABSA [34] pada penelitian terdahulu.

Metode *topic modeling* seperti LDA dan NMF mampu mengidentifikasi kelompok tema atau topik secara probabilistik dan non probabilistik serta

menghasilkan kata kunci berdasarkan topik [35]. Latent Dirichlet Allocation (LDA) bekerja dengan mengasumsikan bahwa setiap dokumen tersusun dari campuran beberapa topik dan setiap topik direpresentasikan oleh distribusi kata tertentu. NMF bekerja dengan melakukan dekomposisi matriks representasi teks menjadi dua matriks non-negatif yang masing-masing menggambarkan hubungan antara kata dan topik serta antara dokumen dan topik [36]. Teknik *feature extraction* TFIDF dan Count Vectorizer diterapkan pada data ulasan untuk membentuk representasi data dalam bentuk numerik [37]. TFIDF (Term Frequency — Inverse Document Frequency) menunjukan kata-kata penting dalam satu dokumen terhadap *corpus* dan Count Vectorizer yang membentuk vektor dari CBOW (*Continuous Bag-of-Words*)

Penelitian ini membandingkan hasil *topic modeling* antara LDA dan NMF untuk merepresentasikan tema ulasan pengguna. Hasil matriks representasi topik dari kedua metode tersebut dievaluasi berdasarkan nilai *coherence* dan digunakan sebagai fitur representatif numerik pada model klasifikasi teks Support Vector Machine (SVM) sebagai metode *aspect classification*[38]. Metode klasifikasi teks yang digunakan termasuk dalam kategori *supervised learning* yang memerlukan data label untuk melatih model agar mampu memprediksi sentimen pada data baru[39]. Hasil klasifikasi ini dapat diukur dengan matrik evaluasi seperti akurasi, presisi, recall, dan F1-score untuk menilai kinerja model[40] yang diterapkan pada evaluasi validasi silang untuk menguji stabilitas model[41].

Beberapa penelitian terdahulu memberikan hasil dari kinerja algoritma *supervised* dan *unsupervised machine learning*. Penelitian terdahulu menerapkan teknik analisis sentimen untuk mengukur sentimen terkait penutupan TikTok Shop di Indonesia menggunakan data Twitter dengan algoritma Multinomial Naive Bayes, Bernoulli Naive Bayes, dan Complement Naive Bayes[28]. Akurasi yang dihasilkan dengan Count Vectorizer dan TextBlob menghasilkan akurasi tertinggi sebesar 86,60% dalam klasifikasi sentimen.

Pada penelitian terdahulu, LDA menghasilkan 3 topik dominan dari data judul berita saat pandemi COVID-19 selama satu tahu masa pandemi seperti *corona cases*, *positive corona/ positive COVID*, dan *COVID-19* dengan akurasi prediksi

klasifikasi 82,4%[42]. Penelitian terdahulu [43] meneliti *sentiment analysis* dan *topic modeling* terhadap opini masyarakat umum di media sosial Twitter terhadap aktivitas Regsosek oleh Badan Pusat Statistik menggunakan algoritma Naïve Bayes, Nearest Centroid, K-Nearest Neighbors, dan Support Vector Machine untuk klasifikasi serta Latent Semantic Analysis (LSA) dan Latent Dirichlet Allocation (LDA) untuk topic modeling. Hasil penelitian menunjukkan bahwa metode Nearest Centroid dan LDA memberikan kinerja lebih baik dengan nilai F1-score masing-masing 59% (positif) dan 66% (negatif). Penelitian terdahulu menggunakan data media sosial Twitter untuk data dengan tagar #covid dan #travel mendapatkan hasil NMF performa yang lebih baik daripada LDA pada teks pendek dengan masing-masing menggunakan pembobotan TF-IDF [44].

Penelitian terdahulu menggunakan *aspect-based sentiment analysis* yang terbagi menjadi ekstraksi aspek (AE) dan klasifikasi sentimen aspek (ASC) untuk memperoleh wawasan (*insight*) tambahan tentang kualitas suatu produk atau layanan dengan menilai sentimen ulasan pengguna dari beberapa situs web dan platform e-commerce dan menemukan 14 tema atau aspek serta model *machine learning* RoBERTa menghasilkan akurasi tertinggi di atas 90%[33].

Penelitian terdahulu yang menerapkan ABSA pada data pandemi COVID 19 menggunakan teknik *word embedding* TFIDF dan Count Vectorizer pada LDA untuk mengekstrak topik dan menguji performa model klasifikasi pada aspek memberikan hasil akurasi tertinggi dengan SVM pada 85%[32].

Penelitian terdahulu menunjukkan penggunaan PoS tagging untuk meningkatkan ABSA [34] dan mampu memberikan akurasi prediksi 92% menggunakan Bi-LSTM ketika menangani 1 juta data ulasan pelanggan [30].

Berdasarkan penelitian terdahulu, penelitian tugas akhir ini mengkombinasikan pemodelan topik menggunakan NMF dan LDA dengan algoritma klasifikasi SVM pada data ulasan aplikasi berbahasa Indonesia untuk melihat tema dan sentimen pengguna terhadap integrasi aplikasi *seller center*.

Kontribusi pada penelitian, seperti: 1) Menerapkan kombinasi *topic modeling* dan *classification text* untuk *Aspect-based sentiment analysis*, 2) Menggunakan data ulasan pengguna aplikasi *seller center* Tokopedia dan Tiktok Shop yang diambil

pada periode sebelum dan sesudah aplikasi *seller center* terintegrasi dapat diakses, 3) Mengintegrasikan algoritma *unsupervised learning* LDA dan NMF dengan algoritma *supervised learning* SVM untuk memperoleh *Aspect-based sentiment analysis*, 4) Membandingkan kinerja *topic modeling* dengan teknik *feature extraction* dan *part-of-speech tagging*.

1.2 Rumusan Masalah

Rumusan masalah pada penelitian ini, yaitu:

1. Bagaimana mengolah data teks ulasan aplikasi untuk melatih algoritma *topic modeling* dan *sentiment analysis*?
2. Bagaimana cara membangun model untuk mengidentifikasi topik pada data ulasan dengan *topic modeling* LDA dan NMF pada *framework* CRISP-DM?
3. Bagaimana model yang telah dibangun dapat diimplementasikan pada ulasan baru?

1.3 Batasan Masalah

Berdasarkan rumusan masalah maka batasan yang ditetapkan adalah sebagai berikut:

1. Data yang digunakan adalah ulasan pengguna dari aplikasi Tokopedia & TikTok Shop Seller dan Tokopedia Seller yang diambil dari Google Play Store.
2. Periode pengambilan data dibatasi pada dua rentang waktu yaitu enam bulan sebelum aplikasi Tokopedia & TikTok Shop Seller yang resmi dapat diakses pada tanggal 8 April 2025 (7 November 2024 hingga 7 April 2025) dan enam bulan sesudah penjual di dorong untuk bermigrasi ke aplikasi *seller center* tersebut (periode 8 April 2025 hingga 7 Oktober 2025).
3. Jumlah ulasan yang diperoleh adalah 21 ribu ulasan. Namun, turun menjadi 15 ribu ulasan setelah *preprocessing* dengan *pos tagging* dan 17 ribu tanpa *pos tagging*.
4. Ulasan spam atau palsu dikurasi ketika menggunakan parameter pembobotan kata dengan hanya menampilkan kata yang muncul pada dua dokumen atau lebih.

5. Bahasa ulasan yang dianalisis adalah Bahasa Indonesia dan bahasa pemrograman yang digunakan adalah python.
6. Proses *topic modeling* hanya menggunakan metode Latent Dirichlet Allocation (LDA) dan Non-negative Matrix Factorization (NMF).
7. Latent Dirichlet Allocation (LDA) yang digunakan adalah algoritma *unsupervised machine learning* untuk pemodelan topik.
8. Proses klasifikasi sentimen hanya menggunakan Support Vector Machine (SVM).

1.4 Tujuan dan Manfaat Penelitian

1.4.1 Tujuan Penelitian

Tujuan penelitian ini adalah:

1. Membangun model *web scrapping* untuk memperoleh teks ulasan aplikasi serta menerapkan *text preprocessing* NLP (*Natural Language Processing*) dan *feature extraction* pada data teks ulasan aplikasi *seller center*.
2. Menggunakan *framework* CRISP-DM untuk membangun model *topic modeling* dengan membandingkan *coherence score* LDA dan NMF terhadap representasi fitur TFIDF dan Count Vectorizer serta membuat model *sentiment analysis* dengan SVM pada tiap aspek yang ditemukan.
3. Membangun model klasifikasi dalam bentuk *web* sederhana melalui Streamlit dan dapat diakses secara lokal.

1.4.2 Manfaat Penelitian

a) Teoritis

Manfaat teoritis pada penelitian ini:

1. Menyediakan hasil analisis sentimen dan *topic modeling* terhadap konteks integrasi platform *e-commerce*.
2. Memberikan perbandingan antara LDA dan NMF untuk pemodelan topik pada ulasan aplikasi berbahasa Indonesia.

b) Praktis

Manfaat praktis pada penelitian ini:

1. Penelitian ini diharapkan dapat memberikan wawasan bagi pelaku bisnis *e-commerce* mengenai persepsi pengguna sebelum dan sesudah integrasi layanan.
2. Penelitian ini diharapkan dapat menjadi acuan pengembangan model analisis sentimen dengan menerapkan *topic modeling*.

1.5 Sistematika Penulisan

BAB I Pendahuluan

Bab I memberikan latar belakang penelitian, rumusan masalah, batasan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan skripsi.

BAB II Landasan Teori

Bab II membahas teori-teori yang mendukung penelitian seperti penelitian terdahulu, konsep analisis sentimen, *topic modeling*, serta kerangka pemikiran penelitian.

BAB III Metodologi Penelitian

Bab III menjelaskan metode penelitian yang digunakan, meliputi kerangka kerja CRISP-DM, sumber data, teknik pengumpulan data, teknik analisis data, serta evaluasi model.

BAB IV – Analisis dan Hasil Penelitian

Bab IV menyajikan hasil pengolahan data, evaluasi topic modeling, hasil klasifikasi sentimen, perbandingan metode, dan interpretasi hasil.

BAB V Simpulan dan Saran

Bab V menyajikan kesimpulan dari penelitian serta saran untuk penelitian selanjutnya atau penerapan praktis hasil penelitian.