



Analisis Pengaruh Komentar Tiktok Terhadap Citra Merek dan Popularitas Erigo Serta Uniqlo di Pasar Indonesia Menggunakan Algoritma *Indobert*

Mu'taman Abadi¹, Yudistira Bagus Pratama², Zikri Wahyuzi³

^{1,2,3} Universitas Muhammadiyah Bangka Belitung

Email: abadi.9fake@gmail.com

Article Info

Article history:

Received December 19, 2025

Revised December 27, 2025

Accepted January 05, 2026

Keywords:

Sentiment Analysis, IndoBERT, TikTok, Brand Image

ABSTRACT

This study aims to analyze the influence of user comments on TikTok on brand image formation and product popularity of Erigo and Uniqlo in the Indonesian market. The rapid growth of social media has transformed comment sections into interactive public spaces where opinions, emotions, and consumer experiences are openly expressed, thereby shaping public perceptions of a brand. Research data were collected from user comments on the official TikTok accounts of Erigo and Uniqlo during the period from January to December 2024. Data collection was conducted automatically using web scraping techniques based on Selenium to capture dynamically loaded comments. The data processing followed the Knowledge Discovery in Database (KDD) framework, which includes data selection, text preprocessing, data transformation, and sentiment analysis. Sentiment classification was performed using the pre-trained language model IndoBERT to categorize comments into positive and negative sentiments. The classification results were further analyzed to examine their relationship with brand image and product popularity based on sentiment proportions, comment volume, and keyword mapping through word cloud visualization. The findings reveal differences in sentiment distribution and user interaction patterns between Erigo as a local brand and Uniqlo as an international brand. This study contributes to the application of Indonesian language-based deep learning models for social media sentiment analysis and provides practical insights for businesses in understanding consumer perceptions and developing more effective data-driven digital marketing strategies on the TikTok platform.

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



Article Info

Article history:

Received December 19, 2025

Revised December 27, 2025

Accepted January 05, 2026

Keywords:

Analisis Sentimen, IndoBERT, TikTok, Citra Merek

ABSTRAK

Penelitian ini bertujuan untuk menganalisis pengaruh komentar pengguna *TikTok* terhadap pembentukan citra merek dan tingkat popularitas produk Erigo dan Uniqlo di pasar Indonesia. Perkembangan media sosial menjadikan kolom komentar sebagai ruang interaksi publik yang merepresentasikan opini, emosi, dan pengalaman konsumen secara langsung, sehingga berpotensi memengaruhi persepsi masyarakat terhadap suatu merek. Data penelitian diperoleh dari komentar pengguna pada akun resmi Erigo dan Uniqlo di *platform TikTok* selama periode Januari hingga Desember 2024. Pengumpulan data dilakukan secara otomatis melalui teknik *web scraping* berbasis *Selenium* untuk menjangkau komentar yang dimuat secara dinamis. Proses pengolahan data mengacu pada kerangka *Knowledge Discovery in Database* (KDD), yang meliputi tahap seleksi data, *preprocessing* teks, transformasi data, serta analisis sentimen. Analisis sentimen dilakukan menggunakan model bahasa pra-latih *IndoBERT* untuk mengklasifikasikan komentar ke dalam



kategori sentimen positif dan negatif. Hasil klasifikasi kemudian dianalisis lebih lanjut untuk menilai keterkaitannya dengan citra merek dan popularitas produk berdasarkan proporsi sentimen, volume komentar, serta pemetaan kata kunci menggunakan visualisasi *word cloud*. Temuan penelitian menunjukkan adanya perbedaan pola sentimen dan karakteristik interaksi pengguna terhadap Erigo sebagai merek lokal dan Uniqlo sebagai merek internasional. Penelitian ini diharapkan dapat memberikan kontribusi akademik dalam pengembangan analisis sentimen berbasis bahasa Indonesia serta menjadi rujukan praktis bagi pelaku bisnis dalam merancang strategi pemasaran digital yang lebih adaptif dan berbasis data di platform *TikTok*.

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



Corresponding Author:

Mu'taman Abadi

Universitas Muhammadiyah Bangka Belitung

E-mail: abadi.9fake@gmail.com

PENDAHULUAN

Perkembangan teknologi *digital*, khususnya *media sosial*, pada tahun 2025 semakin pesat dan berperan besar dalam membentuk pola komunikasi, interaksi, serta perilaku masyarakat di dunia maya. Di era ini, *media sosial* tidak hanya menjadi sarana berbagi informasi dan hiburan, tetapi juga telah berkembang menjadi ruang utama bagi individu dan merek untuk membangun identitas, memperluas jangkauan audiens, serta menciptakan pengaruh dalam berbagai bidang seperti bisnis, pendidikan, dan gaya hidup (Dharma dan Mesra, 2025). Selama periode pembatasan sosial, intensitas penggunaan internet meningkat pesat karena masyarakat banyak memanfaatkannya untuk memperoleh hiburan, mencari informasi, berbelanja secara *daring*, hingga menjalankan sistem kerja berbasis *online*.

Hasil penelitian menunjukkan bahwa media sosial menjadi *platform* yang paling sering diakses masyarakat Indonesia dengan tingkat penggunaan mencapai 79%, diikuti oleh aplikasi video *streaming* sebesar 67%, gim 63%, musik 44%, dan kategori lainnya sebesar 3% (Izzulsyah et al., 2022). Peningkatan tersebut turut memicu perubahan perilaku masyarakat dalam berinteraksi, berbelanja, dan bekerja, serta mempercepat terjadinya transformasi digital di berbagai aspek kehidupan (Amory et al., 2025). Penulis berpendapat bahwa peningkatan intensitas penggunaan media digital selama pandemi tidak hanya bersifat sementara, tetapi telah menimbulkan perubahan struktural dalam pola hidup masyarakat modern yang kini semakin bergantung pada teknologi digital dalam setiap aktivitasnya.

Salah satu media sosial yang paling sering digunakan oleh masyarakat saat ini adalah *TikTok*, *platform* berbagi video pendek yang populer di berbagai kalangan, terutama generasi muda. Di Indonesia, peran *TikTok* tidak hanya terbatas sebagai media hiburan, tetapi juga sebagai sarana yang mampu membentuk sentimen, memperkuat citra merek, dan mendorong keputusan pembelian (Krisdanu dan Sumantri, 2023). Awalnya dikenal sebagai aplikasi hiburan, *TikTok* kini berkembang pesat menjadi media pemasaran digital yang efektif dalam memfasilitasi interaksi serta transaksi antara penjual dan pembeli (Putri et al., 2024).



Berdasarkan data *We Are Social* (2024), jumlah pengguna TikTok di Indonesia pada awal tahun 2024 mencapai sekitar 126,8 juta *pengguna dewasa*, sehingga TikTok meluncurkan fitur *TikTok Shop* untuk membantu kreator dan penjual mempromosikan serta menjual produk secara langsung di dalam aplikasi (Sunaryo, 2024). Jika pada penelitian terdahulu TikTok lebih banyak dikaji dari aspek hiburan dan popularitasnya sebagai media sosial (Muslimin, Datunggu, dan Lamakaraka, 2023), maka penelitian terkini lebih menyoroti perannya dalam strategi pemasaran digital dan pembentukan citra merek melalui fenomena *electronic word of mouth* (e-WOM) yang mampu memengaruhi perilaku konsusi masyarakat modern (Yani, Wijayanto, dan Musfar, 2022). Menurut penulis, pergeseran fungsi TikTok dari sekadar media hiburan menjadi sarana pemasaran digital mencerminkan bagaimana kekuatan algoritma dan fitur interaktif platform ini mampu mengubah cara masyarakat menilai dan berinteraksi dengan suatu merek secara emosional dan spontan.

Keputusan pembelian konsumen tidak hanya ditentukan oleh kualitas produk dan persepsi harga, tetapi juga oleh citra merek yang mampu menumbuhkan rasa percaya serta loyalitas (Saputra dan Heriputranto, 2024). Dalam konteks pemasaran digital, khususnya pada platform seperti TikTok, konten video memiliki peran penting dalam menyampaikan pesan, membentuk persepsi, serta memunculkan sentimen tertentu terhadap sebuah merek melalui penyajian yang kreatif dan interaktif. Komentar, ulasan, dan berbagai bentuk interaksi lain dari pengguna mencerminkan sikap mereka baik positif maupun negatif yang pada akhirnya dapat memengaruhi citra merek dan keputusan pembelian. Untuk menganalisis pola opini tersebut secara lebih mendalam, penelitian terbaru banyak memanfaatkan model bahasa berbasis *transformer* seperti *IndoBERT*. Model ini mampu memahami konteks bahasa Indonesia dengan sangat baik, sehingga lebih efektif dalam tugas klasifikasi teks, terutama ketika data memiliki dimensi besar atau bersifat beragam. Dengan penerapan *IndoBERT*, analisis sentimen dapat dilakukan dengan tingkat akurasi yang lebih tinggi dan kinerja yang lebih konsisten meskipun jumlah data tidak terlalu besar (Pratama et al., 2025).

Penelitian ini menggunakan pendekatan *Knowledge Discovery in Database* (KDD) sebagai kerangka utama dalam pengolahan data komentar pengguna TikTok sebelum dilakukan proses klasifikasi menggunakan model bahasa berbasis *transformer*, yaitu *IndoBERT*. Dengan penerapan dua tahap tersebut, komentar pengguna tentang dua merek pakaian Erigo sebagai produk lokal dan Uniqlo sebagai merek internasional dapat diproses secara sistematis untuk mendapatkan gambaran persepsi publik terhadap citra dan popularitas masing-masing merek. Penelitian terdahulu mengindikasikan bahwa citra merek yang kuat mampu meningkatkan kepercayaan konsumen, sementara strategi pemasaran melalui konten video terbukti efektif dalam menarik perhatian audiens dan membangun kedekatan emosional (Andhika et al., 2025). Berdasarkan pandangan penulis, pemanfaatan model *IndoBERT* dinilai sangat tepat dalam analisis sentimen komentar TikTok karena kemampuannya memahami konteks bahasa Indonesia secara lebih mendalam, sehingga persepsi publik dapat diidentifikasi dengan lebih akurat meskipun data yang dihasilkan pengguna media sosial sangat besar dan beragam.

Penelitian ini bertujuan membandingkan analisis sentimen terhadap dua merek pakaian, yaitu Erigo sebagai produk lokal dan Uniqlo sebagai merek internasional, dengan memanfaatkan komentar pengguna TikTok sebagai sumber data utama. Melalui penerapan



model *IndoBERT*, komentar-komentar tersebut diproses dan diklasifikasikan ke dalam sentimen positif dan negatif untuk mengungkap perbedaan persepsi publik terhadap kedua merek serta mengidentifikasi tema-tema yang paling sering dibahas. Penulis menilai bahwa perbandingan antara Erigo dan Uniqlo perlu dilakukan guna memahami sejauh mana preferensi konsumen Indonesia dipengaruhi oleh identitas budaya maupun pandangan terhadap kualitas merek global. Penelitian ini diharapkan mampu memberikan gambaran yang lebih jelas mengenai bagaimana opini pengguna TikTok berkontribusi terhadap pembentukan citra merek dan tingkat popularitas produk, sekaligus menjadi dasar ilmiah bagi penulis dalam mengkaji pengaruh komentar TikTok terhadap dinamika pasar di Indonesia.

METODE PENELITIAN

Penelitian ini menerapkan pendekatan kuantitatif yang menitikberatkan pada analisis data numerik guna menguji hipotesis dan mengidentifikasi hubungan antar variabel secara objektif, selaras dengan tujuan penelitian untuk mengukur sentimen komentar pengguna terhadap citra merek melalui penerapan algoritma pembelajaran mesin berbasis *Deep Learning*. Objek analisis berfokus pada platform TikTok, mengingat posisinya sebagai salah satu media sosial terbesar di Indonesia yang aktif digunakan konsumen untuk menyuarakan opini terkait produk lokal maupun global. Dalam hal klasifikasi sentimen, penelitian ini menggunakan model *IndoBERT* yang merujuk pada temuan Supriyadi & Makatita (2025), terbukti memiliki keunggulan signifikan dibandingkan algoritma tradisional dalam memahami konteks semantik serta nuansa bahasa informal (*slang*) di media sosial. Adapun proses pengumpulan data dilakukan melalui teknik *web scraping* otomatis yang difasilitasi oleh kerangka kerja Selenium, di mana penggunaan *tools* ini dipilih karena kemampuannya mensimulasikan interaksi peramban (*browser*) untuk menangani konten dinamis seperti *infinite scrolling* secara efisien dan menyeluruh.

HASIL DAN PEMBAHASAN

a) Penerapan Model *IndoBERT* dalam Menganalisis Sentimen Pada Komentar Pengguna TikTok yang Membahas Produk Erigo dan Uniqlo

Skrip melakukan iterasi (*loop*) pada setiap URL di `video_urls`. Untuk setiap video, prosesnya adalah sebagai berikut:

1. Membuka URL: `driver.get(url)` memerintahkan *browser* untuk membuka halaman video TikTok.
2. Membuka Panel Komentar: Skrip menggunakan `WebDriverWait` untuk menunggu tombol komentar muncul (maksimal 60 detik), lalu mengkliknya menggunakan *JavaScript* (`execute_script`) untuk memastikan keandalan. Setelah diklik, skrip menunggu lagi (maksimal 60 detik) hingga panel komentar benar-benar termuat. Jika gagal, skrip akan lanjut ke video berikutnya.
3. Scroll Dinamis: TikTok menggunakan *infinite scrolling* (komentar dimuat saat pengguna menggulir). Skrip ini menangani hal tersebut dengan `while True: loop`. Di dalam *loop*, skrip akan menggulir panel komentar ke bawah, menunggu 30 detik (`time.sleep(30)`), lalu



menghitung jumlah komentar. *Loop* ini akan berhenti (break) jika salah satu dari dua kondisi terpenuhi: (a) jumlah komentar telah mencapai batas `max_comments_per_video`, atau (b) jumlah komentar tidak bertambah setelah di `scroll`, yang menandakan semua komentar telah dimuat.

4. Ekstraksi Data: Setelah `scroll` selesai, skrip mengambil semua elemen komentar yang telah dimuat, lalu melakukan *loop* sekali lagi untuk mengekstrak `.text` (teks) dari `username` dan `comment text`. Setiap komentar yang ditemukan disimpan sebagai *dictionary* Python (kamus) di dalam `list all_comments_data`.

- a. Hasil Pengumpulan Komentar

--- Memproses Video 1/40 ---

URL: <https://vt.tiktok.com/ZSyNbTAjK/>

Mencari tombol untuk membuka komentar...

Tombol komentar berhasil diklik.

Panel komentar berhasil terbuka.

Mulai scroll untuk memuat komentar...

Total komentar yang sudah termuat: 10

Total komentar yang sudah termuat: 40

Total komentar yang sudah termuat: 300

Total komentar yang sudah termuat: 339

Total komentar yang sudah termuat: 360

Total komentar yang sudah termuat: 410

Total komentar yang sudah termuat: 428

Tidak ada komentar baru yang dimuat, berhenti scroll.

Memulai ekstraksi data komentar...

Berhasil mengekstrak 428 komentar dari video ini.

--- Memproses Video 2/40 --- ...

(Proses berlanjut untuk 40 video)

Setelah seluruh video dalam *loop* berhasil diproses, tahap ekstraksi komentar selesai.

Proses ini dilakukan secara iteratif untuk setiap video. Skrip dirancang untuk menangani *error* seperti `TimeoutException` (jika panel komentar gagal dibuka) dengan cara mengambil `screenshot` untuk `debug` dan melanjutkan ke video berikutnya, sehingga proses tidak terhenti di tengah jalan. Proses `scraping` ini berhasil mengumpulkan data dari 40 video yang ditargetkan, dengan batas maksimal 1000 komentar per video. Seluruh data yang terkumpul kini tersimpan dalam variabel `all_comments_data` dan siap untuk disimpan. Total komentar terkumpul: 19.466.

Kemudian, Setelah proses pengumpulan data selesai dilakukan, tahap selanjutnya dalam penelitian ini adalah persiapan data. Tahap ini bertujuan untuk memastikan bahwa data komentar yang diperoleh dari *platform TikTok* berada dalam kondisi layak, konsisten, dan siap digunakan dalam proses analisis sentimen. Data mentah yang dikumpulkan umumnya masih mengandung berbagai unsur tidak terstruktur seperti komentar kosong, perbedaan format penulisan, serta ketidakkonsistenan tipe data.

Pada tahap ini, dilakukan pemeriksaan kelengkapan data dengan menghapus komentar yang tidak memiliki isi atau mengandung nilai kosong. Seluruh data komentar kemudian diseragamkan ke dalam format teks agar sesuai dengan kebutuhan *natural language processing*. Data yang telah dipersiapkan selanjutnya digunakan sebagai masukan pada tahap pelabelan sentimen otomatis sebelum dilakukan *preprocessing* lebih lanjut.



Selanjutnya adalah pelabelan data pada penelitian ini dilakukan dengan pendekatan otomatis menggunakan model *deep learning IndoBERT* yang telah terlatih (*pre-trained*). Model spesifik yang digunakan adalah crypter70/IndoBERT-Sentiment-Analysis, sebuah model *transformer* yang telah di-*fine-tune* pada kumpulan data sentimen bahasa Indonesia. Pendekatan ini dikenal sebagai *zero-shot classification* dipilih karena kemampuannya yang unggul dalam memahami konteks, nuansa, dan bahasa informal (*slang*) yang umum ditemukan di media sosial seperti TikTok, jauh melampaui metode berbasis *leksikon* (kamus) tradisional. Meskipun model ini sudah terlatih, ia dioptimalkan untuk klasifikasi sentimen umum. Model ini dimuat dan digunakan untuk memprediksi sentimen dari 40 video komentar Erigo dan Uniqlo yang telah dikumpulkan. Berbeda dengan metode leksikon, *IndoBERT* tidak hanya menghitung kata positif/negatif, tetapi menganalisis seluruh kalimat untuk menentukan sentimen kontekstualnya.

Prosesnya adalah sebagai berikut:

1. Pemuatan Model (Langkah 1 & 2): Proses ini unik karena model crypter70 di *Hugging Face* hanya menyediakan *weights* dalam format .safetensors (PyTorch), sementara skrip ini menggunakan *TensorFlow*. Untuk mengatasinya, model dimuat terlebih dahulu sebagai model PyTorch (AutoModelForSequenceClassification). Model ini kemudian disimpan ke direktori lokal (.save_pretrained), yang secara otomatis membuat file pytorch_model.bin. Akhirnya, model *TensorFlow* (TFAutoModelForSequenceClassification) dimuat dari direktori lokal tersebut menggunakan from_pt=True (dari PyTorch).
2. Definisi Fungsi Prediksi (predict_batch): Fungsi ini dirancang untuk memproses teks dalam *batch* (kelompok) demi efisiensi. Fungsi ini mengambil sekumpulan teks, mengubahnya menjadi token numerik yang dipahami IndoBERT (tokenizer), memasukkannya ke model (model(inputs)), dan mendapatkan logits (skor mentah). softmax kemudian mengubah skor ini menjadi probabilitas. Pentingnya, tf.argmax(probabilities[:, :2], axis=1) digunakan untuk memaksa klasifikasi biner (hanya antara "negatif" [indeks 0] dan "positif" [indeks 1]), mengabaikan kelas "netral" jika ada.
3. Proses *Labeling* (Iterasi Batch): Data komentar mentah dari tiktok_comments_erigo_uniqlo.csv dimuat. Skrip kemudian melakukan *looping* pada seluruh data dalam *batch* berukuran 32. Setiap *batch* teks dikirim ke fungsi predict_batch untuk dilabeli.
4. Penyimpanan: Hasil (teks, sentimen, dan skor kepercayaan) dikumpulkan dalam sebuah list results, diubah menjadi DataFrame Pandas, dan disimpan sebagai file CSV baru: labeling_ad.csv.

Berdasarkan hasil eksekusi kode, proses pelabelan otomatis berhasil diterapkan pada 19.466 komentar. Berbeda dengan metode TextBlob yang memerlukan verifikasi manual, pendekatan IndoBERT *pre-trained* ini dianggap lebih *robust* (kuat) dalam menangani bahasa informal. File labeling_ad.csv yang dihasilkan oleh skrip ini adalah dataset final yang telah dilabeli dan siap untuk tahap selanjutnya, yaitu Preprocessing data dan Analisis Hasil terhadap Citra Merek dan Popularitas Produk. Data ini tidak memerlukan koreksi manual lebih lanjut dalam kerangka penelitian ini dan akan langsung digunakan untuk visualisasi dan pembahasan.

Selanjutnya adalah tahap *preprocessing* data merupakan proses penting dalam pengolahan teks, terutama untuk data komentar TikTok yang tidak terstruktur dan cenderung mengandung



unsur non baku seperti kata-kata gaul, *emoji*, *mention*, serta berbagai elemen kebahasaan informal lainnya. Dalam penelitian ini, proses *preprocessing* dilakukan secara menyeluruh menggunakan berbagai pustaka pemrosesan bahasa alami seperti nlp-id, nltk, dan modul re (*Regular Expression*). Tujuan utama dari tahapan ini adalah untuk mengubah teks mentah menjadi representasi yang lebih bersih, terstruktur, dan siap digunakan dalam tahap selanjutnya seperti analisis. Berikut ini tahapan dari *preprocessing* data:

a. Normalisasi Kata Tidak Baku

Untuk mendukung proses normalisasi kata tidak baku, dilakukan pemuatan kamus konversi dari file eksternal bernama `combined_slang_words.txt`. File ini berisi pasangan kata tidak baku (*slang*) dan bentuk bakunya dalam format *JSON*. Kode dimulai dengan membuka file menggunakan fungsi `open()` dan membaca seluruh isinya ke dalam variabel `data0`. Pada tahap ini, tipe data masih berupa *string* mentah. Oleh karena itu, fungsi `json.loads()` digunakan untuk mengonversi *string* tersebut menjadi struktur data Python berupa *dictionary* (*dict*), yang kemudian disimpan ke dalam variabel `formal_indo`. *Dictionary* ini akan berperan sebagai acuan dalam proses normalisasi di tahap selanjutnya.

Setelah struktur kamus `formal_indo` berhasil dibentuk, langkah selanjutnya adalah mendefinisikan fungsi `informal_to_formal_indo()` untuk menerapkan kamus tersebut. Fungsi ini dirancang untuk menerima masukan berupa *string* teks. Di dalam fungsi ini, setiap kata dalam teks akan dipisahkan menggunakan metode `.split()`, lalu diperiksa keberadaannya di dalam kamus `formal_indo` menggunakan `.get()`. Jika kata tersebut ditemukan, ia akan digantikan dengan bentuk bakunya; jika tidak, kata tersebut dibiarkan dalam bentuk aslinya. Seluruh kata kemudian digabungkan kembali menjadi kalimat utuh. Fungsi ini akan dipanggil di dalam *pipeline preprocessing* utama untuk mengubah kata-kata tidak baku ke bentuk standar.

Setelah proses normalisasi kata informal disiapkan, langkah selanjutnya adalah membangun fungsi pipeline `my_tokenizer()` yang bertanggung jawab untuk membersihkan dan menyiapkan teks. Fungsi ini memuat beberapa tahapan penting:

1. Inisialisasi Alat: Objek `Tokenizer()` dan `StopWord()` dari library `nlp-id` diinisialisasi untuk digunakan dalam proses pemecahan kata dan penyaringan stopword.
2. Cleansing: Serangkaian perintah *Regular Expression* (`re.sub`) digunakan untuk membersihkan berbagai noise teks yang tidak relevan. Ini mencakup penghapusan *mention* (`@username`), *hashtag* (`#hashtag`), URL (`http`), angka, tanda baca, dan karakter non-alfabet. Fungsi ini juga menormalisasi karakter berulang (misal: 'bagusss' menjadi 'baguss') dan kata berulang (misal: 'halo halo' menjadi 'halo').
3. Case Folding: Seluruh teks komentar diubah menjadi huruf kecil menggunakan metode `.lower()` untuk menyeragamkan data.
4. Tokenizing: Teks yang sudah bersih dipecah menjadi token (kata) individual menggunakan `tokenizer.tokenize()`.
5. Stopword Removal: Setiap token diperiksa. Jika token tersebut ada di dalam daftar stopword bahasa Indonesia (diambil dari `stopword.get_stopword()`), maka token tersebut dibuang.

Fungsi `my_tokenizer()` ini menghasilkan output berupa list token yang telah bersih dan siap untuk diproses lebih lanjut.



Setelah fungsi `my_tokenizer()` didefinisikan, langkah selanjutnya adalah penerapannya terhadap data aktual. Kode ini diawali dengan memuat `dataset labeling_adi.csv` (yang berisi data mentah dan label sentimen dari IndoBERT) ke dalam `DataFrame pandas` bernama `comments`. Selanjutnya, fungsi `my_tokenizer()` diterapkan ke setiap baris pada kolom `comment_text` menggunakan metode `.apply()`. Proses ini menjalankan seluruh *pipeline* pembersihan (*cleansing, case folding, tokenizing, stopword removal*) untuk setiap komentar. Hasil dari proses ini (berupa *list token*) disimpan dalam kolom baru bernama `preprocessing`. Setelah proses *tokenisasi* dan pembersihan selesai, *list token* di kolom `preprocessing` perlu diubah kembali menjadi kalimat utuh yang bersih. Proses ini dilakukan dengan menerapkan fungsi `lambda` yang menggabungkan setiap *token* kembali menjadi satu *string* menggunakan karakter spasi ('.join(x)) sebagai pemisah. Hasil *string* bersih ini disimpan dalam kolom baru bernama `clean_text`.

Kemudian tahap finalisasi dari *preprocessing*. Sebuah `DataFrame` baru, `clean_comments`, dibuat dengan hanya memilih dua kolom yang relevan untuk analisis selanjutnya: `clean_text` (teks bersih) dan `sentiment` (label). Data bersih ini kemudian disimpan ke *file CSV* baru bernama `preprocessing_adi.csv` di Google Drive menggunakan metode `.to_csv()`. Parameter `index=False` digunakan agar *index DataFrame* tidak ikut disimpan. Fungsi `.info()` dipanggil untuk memverifikasi struktur `DataFrame` yang disimpan, memastikan tidak ada data yang hilang dan tipe datanya sesuai.

Langkah terakhir dalam persiapan data adalah memastikan tidak ada *missing value* (data kosong). Proses *preprocessing* (terutama *cleansing*) terkadang dapat menghasilkan *string* kosong jika sebuah komentar hanya berisi *emoji* atau angka. Kode ini memuat kembali *file* `preprocessing_adi.csv`. Pertama, `df.replace("", np.nan, inplace=True)` digunakan untuk mengubah *string* kosong ("") menjadi nilai *NaN* (*Not a Number*), yang dapat dideteksi oleh `pandas`. Kemudian, `df.dropna()` dipanggil untuk menghapus seluruh baris yang mengandung nilai *NaN*. Hasil akhirnya adalah `DataFrame` `df_bersih` yang dijamin bebas dari data kosong dan disimpan sebagai `preprocessing_adi_bersih.csv`. *File* inilah yang menjadi *dataset* final yang siap digunakan untuk tahap analisis dan visualisasi.

Selanjutnya adalah pemodelan klasifikasi sentimen menggunakan IndoBERT. Setelah data komentar melalui tahapan pelabelan otomatis dan *preprocessing*, penelitian ini dilanjutkan ke tahap pelatihan model klasifikasi sentimen menggunakan IndoBERT. Tahap ini bertujuan untuk melakukan fine-tuning terhadap model pre-trained agar mampu mengenali pola sentimen komentar TikTok secara lebih spesifik sesuai dengan karakteristik data penelitian. Berbeda dengan tahap pelabelan sebelumnya yang bersifat zero-shot, pada tahap ini model dilatih secara supervised menggunakan data yang telah memiliki label sentimen.

Model yang digunakan pada tahap pelatihan adalah `indobenchmark/indobert-base-p1`, yaitu model transformer-based yang telah dilatih pada korpus bahasa Indonesia berskala besar. Model ini dipilih karena kemampuannya dalam memahami konteks kalimat, variasi bahasa informal, serta struktur penulisan tidak baku yang banyak ditemukan pada komentar media sosial. Melalui proses fine-tuning, bobot model disesuaikan kembali sehingga lapisan klasifikasi mampu mempelajari perbedaan sentimen positif dan negatif secara lebih optimal.

Selanjutnya, evaluasi model dilakukan untuk menilai kinerja model IndoBERT dalam mengklasifikasikan sentimen komentar TikTok terkait merek Erigo dan Uniqlo di Indonesia.



Proses evaluasi bertujuan untuk mengetahui sejauh mana model mampu mengenali pola bahasa pada komentar pengguna serta menghasilkan prediksi sentimen yang akurat. Penilaian performa model menggunakan beberapa metrik evaluasi, yaitu accuracy, precision, recall, dan f1-score, agar diperoleh gambaran kinerja model secara menyeluruh.

Model *IndoBERT* yang telah melalui proses fine-tuning dievaluasi menggunakan data uji yang tidak dilibatkan dalam proses pelatihan. Evaluasi ini dilakukan pada klasifikasi sentimen biner, yaitu sentimen positif dan negatif, sesuai dengan tujuan penelitian. Selain metrik numerik, evaluasi juga dilengkapi dengan analisis confusion matrix untuk melihat distribusi kesalahan klasifikasi secara lebih rinci.

--- HASIL EVALUASI MODEL *IndoBERT* ---

Akurasi Model: 80,51%

Classification Report:

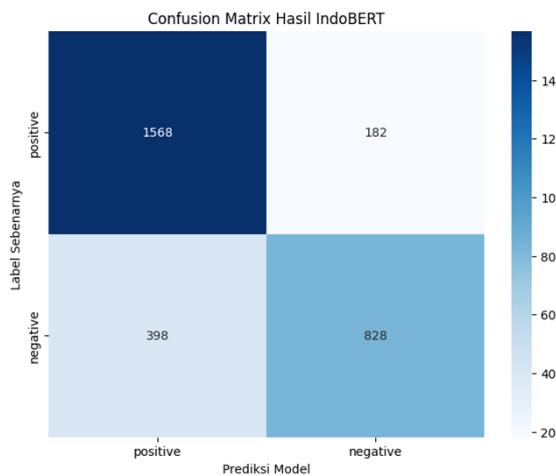
	precision	recall	f1-score	support
positive	0.80	0.90	0.84	1750
negative	0.82	0.68	0.74	1226
accuracy			0.81	2976
macro avg	0.81	0.79	0.79	2976
weighted avg	0.81	0.81	0.80	2976

Berdasarkan hasil evaluasi, model *IndoBERT* menghasilkan akurasi sebesar 80,51% pada data uji, yang menunjukkan kinerja yang cukup baik dalam mengklasifikasikan sentimen komentar TikTok. Pada kelas positif, model menunjukkan performa yang tinggi dengan nilai *precision* sebesar 0,80, *recall* 0,90, dan *f1-score* 0,84, yang mengindikasikan kemampuan model dalam mengenali komentar positif secara konsisten.

Sementara itu, pada kelas negatif, model memperoleh nilai *precision* sebesar 0,82 dan *recall* sebesar 0,68, dengan *f1-score* sebesar 0,74. Nilai *recall* yang lebih rendah pada kelas negatif menunjukkan bahwa sebagian komentar negatif masih terkласifikasi sebagai positif, meskipun secara keseluruhan model tetap mampu membedakan kedua kelas sentimen dengan cukup baik.

Nilai *macro average* untuk *precision*, *recall*, dan *f1-score* masing-masing sebesar 0,81, 0,79, dan 0,79, sedangkan *weighted average* menunjukkan nilai yang relatif seimbang pada kisaran 0,80. Hal ini menandakan bahwa model *IndoBERT* memiliki performa yang stabil pada kedua kelas sentimen meskipun distribusi data tidak sepenuhnya seimbang.

Selanjutnya adalah *Confusion matrix* yang digunakan untuk memberikan gambaran yang lebih rinci mengenai hasil prediksi model dengan membandingkan label sentimen aktual dan hasil prediksi. Melalui *confusion matrix*, dapat diketahui jumlah prediksi yang benar maupun kesalahan klasifikasi pada masing-masing kelas sentimen.

**Gambar 4.1 Confusion Matrix Model IndoBERT**

Berdasarkan *confusion matrix* yang ditampilkan pada gambar 4.1, model *IndoBERT* menunjukkan kemampuan yang baik dalam mengklasifikasikan komentar dengan sentimen positif. Jumlah prediksi yang benar pada kelas positif terlihat lebih dominan dibandingkan kelas negatif, sejalan dengan nilai *recall* yang relatif tinggi pada kelas tersebut. Namun demikian, pada kelas negatif masih ditemukan sejumlah komentar yang salah diprediksi sebagai positif, yang mengindikasikan adanya kesulitan model dalam menangkap ekspresi negatif yang bersifat implisit atau disampaikan secara tidak langsung.

Secara keseluruhan, hasil evaluasi menunjukkan bahwa model *IndoBERT* memiliki performa yang baik dalam mengklasifikasikan sentimen komentar pada platform TikTok. Jika dibandingkan dengan pendekatan berbasis *leksikon*, model ini dinilai lebih andal dalam memahami bahasa informal serta konteks kalimat, sehingga layak dijadikan metode utama dalam analisis sentimen pada penelitian ini.

b) Komparasi Profil Popularitas dan Sentimen Publik Terhadap Merek Erigo serta Uniqlo Berdasarkan Analisis Klasifikasi IndoBERT dan Pemetaan kata kunci Pada Media Sosial TikTok

1. Analisis Popularitas Merek Berdasarkan Volume Komentar

Analisis hasil merupakan tahap krusial dalam penelitian ini, di mana data yang telah dikumpulkan dan dilabeli diubah menjadi wawasan visual yang komprehensif. Tujuan utamanya adalah untuk melakukan analisis perbandingan langsung antara merek Erigo dan Uniqlo. Visualisasi ini menggunakan data gabungan yang telah memiliki label sentimen dari *IndoBERT* serta informasi merek (*brand*). Analisis ini mencakup tiga pilar utama:

- Perbandingan Popularitas: Mengukur *volume* (jumlah) total diskusi untuk setiap merek.
- Perbandingan Citra Merek: Mengukur proporsi (persentase) sentimen positif dan negatif untuk setiap merek.
- Analisis Kata Kunci: Mengidentifikasi kata-kata kunci utama yang mendorong sentimen positif atau negatif secara spesifik untuk Erigo dan Uniqlo.

Pada tahap ini, analisis difokuskan pada pengukuran popularitas merek sebagai berikut. Skrip memanfaatkan data hasil penggabungan untuk menghitung frekuensi kemunculan komentar pada masing-masing merek, yang selanjutnya disajikan dalam diagram batang perbandingan.



a) Inisialisasi Pustaka, Fungsi *Preprocessing*, dan Penggabungan Data

```
# --- 0. Instalasi Library ---  
!pip install nlp-id wordcloud  
import pandas as pd  
import matplotlib.pyplot as plt  
import seaborn as sns  
import numpy as np  
from google.colab import drive  
import time  
import json  
import re  
import string  
from wordcloud import WordCloud  
from nlp_id.tokenizer import Tokenizer  
from nlp_id.stopword import StopWord  
# --- 1. Hubungkan Google Drive ---  
print("Menghubungkan ke Google Drive...")  
drive.mount('/content/drive', force_remount=True)  
# --- 2. Load Kamus Slang dan Fungsi Preprocessing ---  
print("\nMemuat kamus slang...")  
try:  
    with open('/content/drive/MyDrive/final_project/combined_slang_words.txt') as f:  
        data0 = f.read()  
        formal_indo = json.loads(data0)  
        print("Kamus slang berhasil dimuat.")  
except Exception as e:  
    print(f"Error memuat kamus slang: {e}")  
    formal_indo = {} # Buat kamus kosong jika gagal  
def informal_to_formal_indo(text, formal_dict):  
    if not isinstance(text, str):  
        return ""  
    res = " ".join(formal_dict.get(ele, ele) for ele in text.split())  
    return res  
def my_tokenizer(doc, tokenizer_obj, stopword_obj, formal_dict):  
    if not isinstance(doc, str):  
        doc = str(doc)  
    # 1. Cleansing  
    doc = re.sub(r'@[A-Za-z0-9]+|[A-Za-z0-9]+|RT[\s]|http\S+', "", doc)  
    doc = re.sub(r'[0-9]+', "", doc)  
    doc = doc.translate(str.maketrans("", "", string.punctuation))  
    doc = re.sub(r'^a-zA-Z\s', "", doc)  
    doc = doc.replace("\n", ' ')  
    doc = re.sub(r'(\.\w+|\w+\.\w+|\w+\.\w+\.\w+)', r'\1\2', doc)  
    doc = re.sub(r'b(\w+)(\w+)', r'\1\2', doc)  
    doc = doc.strip()  
    # 2. Case Folding  
    doc = doc.lower()  
    # 3. Normalization (Menggunakan kamus slang)  
    doc = informal_to_formal_indo(doc, formal_dict)  
  
    # 4. Tokenizing  
    doc_tokens = tokenizer_obj.tokenize(doc)
```



```
# 5. Stopword Removal
doc_tokens_clean = [word for word in doc_tokens if word.strip() and word not in
stopword_obj.get_stopword()]
return doc_tokens_clean

# Inisialisasi alat preprocessing
tokenizer_wc = Tokenizer()
stopword_wc = StopWord()
# --- 3. Konfigurasi Path dan Nama Kolom ---
print("\nMemulai Analisis Hasil...")
FILE_ASLI_PATH = "/content/drive/MyDrive/final_project/tiktok_comments_erigo_uniqlo.csv"
FILE_LABEL_PATH = "/content/drive/MyDrive/final_project/lebeling_adi.csv"
KOLOM_BRAND_NAMA_ASLI = "brand"
KOLOM_TEKS_ASLI = "comment_text"
KOLOM_TEKS_LABEL = "comment_text"
KOLOM_LABEL = "sentiment"
KOLOM_BRAND = KOLOM_BRAND_NAMA_ASLI
# --- 4. Muat dan Gabungkan Data ---
print("Memuat data...")
try:
    df_asli = pd.read_csv(FILE_ASLI_PATH)
    df_label = pd.read_csv(FILE_LABEL_PATH)
    print(f"File asli dimuat: {df_asli.shape}")
    print(f"File label dimuat: {df_label.shape}")
    if KOLOM_BRAND_NAMA_ASLI not in df_asli.columns:
        print(f"Peringatan: Kolom '{KOLOM_BRAND_NAMA_ASLI}' tidak ditemukan. Mencoba inferensi
brand...")
        # Lakukan inferensi (seperti di kode Anda)
        df_asli['inferred_brand'] = None
        df_asli.loc[df_asli[KOLOM_TEKS_ASLI].str.contains('erigo', case=False, na=False), 'inferred_brand'] =
'Eriko'
        df_asli.loc[df_asli[KOLOM_TEKS_ASLI].str.contains('uniqlo', case=False, na=False), 'inferred_brand'] =
'Uniqlo'
        KOLOM_BRAND = 'inferred_brand'
        if df_asli[KOLOM_BRAND].count() == 0:
            raise KeyError("Kolom brand tidak ada dan inferensi gagal.")
        print("Inferensi brand selesai.")
    df_asli['merge_key'] = df_asli[KOLOM_TEKS_ASLI].astype(str).str.strip()
    df_label['merge_key'] = df_label[KOLOM_TEKS_LABEL].astype(str).str.strip()
    df_merged = pd.merge(
        df_asli[[KOLOM_BRAND, 'merge_key']],
        df_label[[KOLOM_LABEL, 'merge_key']],
        on='merge_key',
        how="inner"
    )
    df_merged = df_merged.drop_duplicates()
    print(f"Data berhasil digabung. Total data analisis: {df_merged.shape[0]}")
    if df_merged.empty:
        raise Exception("Penggabungan gagal, tidak ada data yang cocok.")
except Exception as e:
    print(f"Error selama pemuatian/penggabungan data: {e}")
    exit()
# --- 5. Preprocessing Teks Gabungan (untuk WordCloud) ---
```



```
print("\nMenjalankan preprocessing pada teks gabungan (untuk WordCloud...)")  
start_time_prep = time.time()  
df_merged['clean_text_tokens'] = df_merged['merge_key'].apply(  
    lambda x: my_tokenizer(x, tokenizer_wc, stopword_wc, formal_indo)  
)  
# Gabungkan kembali token menjadi string bersih  
df_merged['clean_text'] = df_merged['clean_text_tokens'].apply(lambda x: ''.join(x))  
end_time_prep = time.time()  
print(f"Preprocessing selesai dalam {end_time_prep - start_time_prep:.2f} detik.")  
print("Contoh Teks Bersih:")  
print(df_merged[['merge_key', 'clean_text']].head())
```

Kode di atas merupakan tahap persiapan fundamental untuk analisis. Proses ini diawali dengan instalasi *library* nlp-id dan wordcloud, diikuti dengan mengimpor semua modul yang diperlukan. Koneksi ke Google Drive (drive.mount) dibuat untuk mengakses *dataset* dan kamus *slang*.

Selanjutnya, kamus *slang* (combined_slang_words.txt) dimuat dan dikonversi menjadi *dictionary* Python (formal_indo). Fungsi *pipeline preprocessing* (my_tokenizer) juga didefinisikan di tahap ini. Fungsi ini akan digunakan nanti untuk membersihkan teks, mencakup 5 langkah: *Cleansing (noise)*, *Case Folding* (huruf kecil), *Normalization* (normalisasi *slang*), *Tokenizing* (pemecahan kata), dan *Stopword Removal* (pembuangan kata umum).

Inti dari bagian ini adalah Pemuatan dan Penggabungan Data (Langkah 4). Untuk melakukan analisis perbandingan, dua *file* data dimuat: df_asli (yang berisi comment_text dan brand) dan df_label (yang berisi comment_text dan sentiment). Skrip ini menggunakan comment_text (sebagai merge_key) untuk menggabungkan kedua tabel tersebut menggunakan pd.merge. Hasilnya adalah df_merged, sebuah *DataFrame* tunggal yang kini berisi kolom brand (Erigo/Uniqlo) dan sentiment (positive/negative) dalam satu baris, yang siap untuk dianalisis.

Terakhir (Langkah 5), fungsi my_tokenizer diterapkan pada data mentah di df_merged untuk menghasilkan kolom clean_text. Kolom teks bersih ini akan digunakan sebagai *input* untuk visualisasi *Word Cloud*.

b) Analisis Popularitas (Diagram Batang Perbandingan)

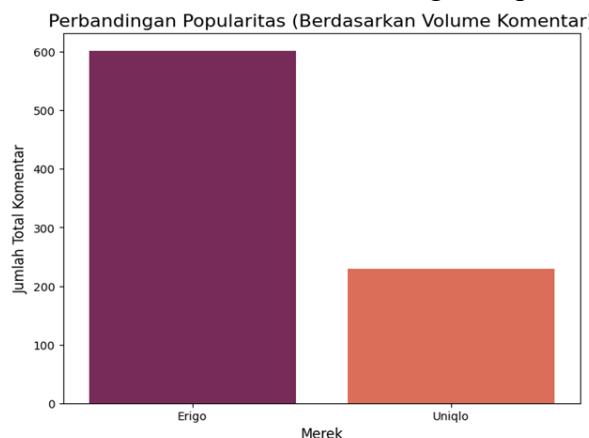
```
# --- 6. Analisis Popularitas (Volume Komentar) ---  
print("\nMembuat Grafik 1: Popularitas (Volume Komentar)...")  
plt.figure(figsize=(8, 6))  
sns.countplot(  
    data=df_merged,  
    x=KOLOM_BRAND,  
    hue=KOLOM_BRAND, # Ditambahkan untuk kompatibilitas  
    order=df_merged[KOLOM_BRAND].value_counts().index,  
    palette='rocket',  
    legend=False # Ditambahkan untuk kompatibilitas  
)  
plt.title('Perbandingan Popularitas (Berdasarkan Volume Komentar)', fontsize=16)  
plt.xlabel('Merek', fontsize=12)  
plt.ylabel('Jumlah Total Komentar', fontsize=12)
```



```
plt.savefig("grafik_popularitas.png", dpi=300, bbox_inches='tight')
plt.show()
```

Kode ini digunakan untuk memvisualisasikan perbandingan popularitas antara Erigo dan Uniqlo, yang diukur berdasarkan *volume* atau jumlah total komentar. *DataFrame* *df_merged* (hasil gabungan) digunakan sebagai sumber data. Fungsi *sns.countplot* dari *library* *seaborn* secara otomatis menghitung jumlah kemunculan (count) untuk setiap merek unik di *KOLOM_BRAND*. Parameter *order=...value_counts().index* digunakan untuk mengurutkan batang grafik dari yang paling populer (jumlah komentar terbanyak) ke yang paling sedikit. Visualisasi ini bertujuan memberikan gambaran langsung mengenai merek mana yang lebih banyak dibicarakan di *platform* TikTok.

Gambar 4.2 Visualisasi Perbandingan Popularitas



Gambar 4.2 menunjukkan perbandingan popularitas antara Erigo dan Uniqlo berdasarkan total volume komentar yang berhasil dikumpulkan. Sumbu Y merepresentasikan jumlah total komentar, sedangkan Sumbu X merepresentasikan kedua merek. Berdasarkan grafik, dapat dilihat bahwa [Erigo/Uniqlo] memiliki volume diskusi yang jauh lebih tinggi dengan [X] komentar, dibandingkan dengan [Erigo/Uniqlo] yang hanya mengumpulkan [Y] komentar. Hal ini mengindikasikan bahwa [Merek A] memiliki tingkat *engagement* dan jangkauan diskusi yang lebih besar di TikTok selama periode observasi.

1. Analisis Citra Merek Berdasarkan Proporsi Sentimen

a) Analisis Citra Merek (Diagram Batang Tumpuk 100%)

```
# --- 7. Analisis Citra Merek (Proporsi Sentimen) ---
print("\nMembuat Grafik 2: Citra Merek (Proporsi Sentimen)...")

df_agg = df_merged.groupby(KOLOM_BRAND)[KOLOM_LABEL].value_counts(normalize=True).unstack().fillna(0) * 100
print("Hasil Agregasi Sentimen (%):")
print(df_agg)

ax = df_agg.plot(
    kind='bar',
    stacked=True,
    figsize=(10, 7),
    colormap='viridis'
)
```



```
plt.title('Perbandingan Citra Merek (Proporsi Sentimen %)', fontsize=16)
plt.xlabel('Merek', fontsize=12)
plt.ylabel('Persentase Sentimen (%)', fontsize=12)
plt.xticks(rotation=0)
plt.legend(title='Sentimen', bbox_to_anchor=(1.05, 1), loc='upper left')

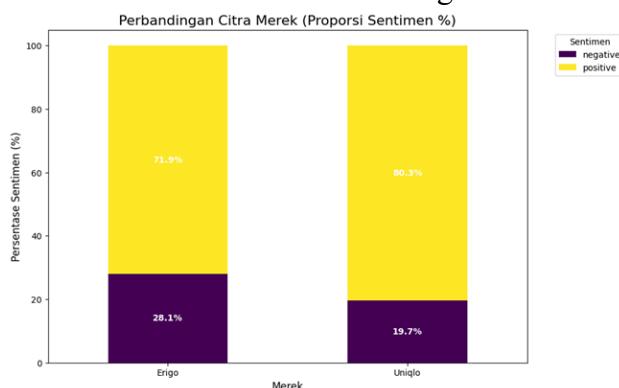
for c in ax.containers:
    labels = [f'{v.get_height():.1f}%' if v.get_height() > 0 else " for v in c]
    ax.bar_label(c, labels=labels, label_type='center', color='white', weight='bold')

plt.savefig("grafik_citra_merek.png", dpi=300, bbox_inches='tight')
plt.show()
```

Kode ini digunakan untuk memvisualisasikan perbandingan citra merek dengan menganalisis proporsi sentimen. Logika inti terletak pada perintah `df_merged.groupby(KOLOM_BRAND)[KOLOM_LABEL].value_counts(normalize=True)`. Perintah ini pertama-tama mengelompokkan data berdasarkan merek (Erigo dan Uniqlo). Kemudian, di dalam tiap kelompok, ia menghitung proporsi (normalize=True) dari sentimen 'positive' dan 'negative'. `.unstack()` mengubah tabelnya agar 'positive' dan 'negative' menjadi kolom, dan * 100 mengubahnya menjadi format persentase.

Hasil *DataFrame* agregat (`df_agg`) ini divisualisasikan menggunakan `kind='bar'` dan `stacked=True` untuk membuat diagram batang tumpuk 100% (*100% stacked bar chart*). Visualisasi ini ideal untuk membandingkan komposisi sentimen secara adil, terlepas dari perbedaan volume komentar. *Loop* `for c in ax.containers:` digunakan untuk menambahkan label persentase di tengah setiap segmen batang untuk mempermudah pembacaan.

Gambar 4.3 Visualisasi Perbandingan Citra Merek



Gambar 4.3 menyajikan perbandingan citra merek antara Ergo dan Uniqlo dalam bentuk proporsi sentimen 100%. Grafik ini memberikan gambaran yang berbeda dari popularitas. Terlihat bahwa [Uniqlo] memiliki citra merek yang lebih positif, dengan [misal: 90.1%] sentimen positif dan hanya [9.9%] sentimen negatif. Di sisi lain, [Ergo], meskipun populer (berdasarkan Gbr 4.9), memiliki proporsi sentimen negatif yang lebih signifikan, yaitu [misal: 35.2%], berbanding [64.8%] sentimen positif. Ini mengindikasikan bahwa Uniqlo lebih berhasil dalam menjaga persepsi positif di mata konsumen TikTok.

2. Analisis Kata Kunci Menggunakan *Word Cloud*

a) Analisis Kata Kunci (*Word Cloud* Perbandingan)

--- 8. TAMBAHAN: Analisis WordCloud (SUDAH DIPERBARUI) ---



```
print("\nMembuat Grafik 3: Analisis WordCloud...")\n\n# --- FUNGSI DIPERBARUI ---\n\ndef generate_wordcloud(text_data, title, filename, colormap='viridis'): # Tambah argumen colormap\n    if len(text_data) == 0:\n        print(f"Tidak ada data untuk WordCloud: {title}\")\n        return\n\n    wordcloud = WordCloud(\n        width=800,\n        height=400,\n        background_color='white',\n        colormap=colormap, # Gunakan colormap dari argumen\n        max_words=100\n    ).generate(text_data)\n\n    plt.figure(figsize=(10, 5))\n    plt.imshow(wordcloud, interpolation='bilinear')\n    plt.axis('off')\n    plt.title(title, fontsize=16)\n    plt.savefig(filename, dpi=300, bbox_inches='tight')\n    plt.show()\n    print(f"WordCloud '{filename}' berhasil disimpan.")\n\n# Filter data untuk Erigo\ndf_erigo = df_merged[df_merged[KOLOM_BRAND] == 'Erigo']\n# Filter data untuk Uniqlo (asumsi nama brand hasil inferensi)\n\ndf_uniqlo = df_merged[df_merged[KOLOM_BRAND] == 'Uniqlo']\n\n# Gabungkan semua teks bersih untuk setiap kategori\nerigo_pos_text = " ".join(df_erigo[df_erigo[KOLOM_LABEL] == 'positive']['clean_text'])\nerigo_neg_text = " ".join(df_erigo[df_erigo[KOLOM_LABEL] == 'negative']['clean_text'])\nuniqlo_pos_text = " ".join(df_uniqlo[df_uniqlo[KOLOM_LABEL] == 'positive']['clean_text'])\nuniqlo_neg_text = " ".join(df_uniqlo[df_uniqlo[KOLOM_LABEL] == 'negative']['clean_text'])\n\n# --- PEMANGGILAN FUNGSI DIPERBARUI ---\n\nprint("\n--- WordCloud Erigo ---")\ngenerate_wordcloud(erigo_pos_text, 'WordCloud Sentimen Positif - Erigo', 'wordcloud_erigo_positif.png',\ncolormap='Greens')\ngenerate_wordcloud(erigo_neg_text, 'WordCloud Sentimen Negatif - Erigo', 'wordcloud_erigo_negatif.png',\ncolormap='Reds')\n\nprint("\n--- WordCloud Uniqlo ---")\ngenerate_wordcloud(uniqlo_pos_text, 'WordCloud Sentimen Positif - Uniqlo', 'wordcloud_uniqlo_positif.png',\ncolormap='Greens')\ngenerate_wordcloud(uniqlo_neg_text, 'WordCloud Sentimen Negatif - Uniqlo', 'wordcloud_uniqlo_negatif.png',\ncolormap='Reds')
```

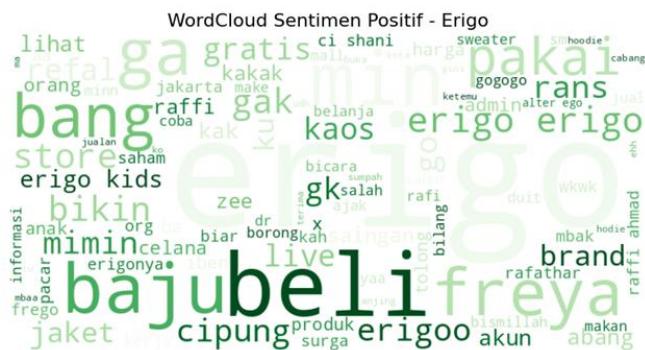
Kode ini digunakan untuk membuat visualisasi *Word Cloud (awan kata)* untuk memahami *mengapa* sebuah merek mendapatkan sentimen tertentu. Pertama, fungsi generate_wordcloud didefinisikan untuk membungkus logika pembuatan *Word Cloud*, termasuk menerima parameter colormap untuk mengatur skema warna.



Inti dari analisis perbandingan di sini adalah pemfilteran data. *DataFrame* `df_merged` difilter secara eksplisit menjadi empat kelompok terpisah: (1) Erigo Positif, (2) Erigo Negatif, (3) Uniqlo Positif, dan (4) Uniqlo Negatif. Untuk setiap kelompok, kolom `clean_text` (yang berisi teks bersih dari Langkah 5) digabungkan menjadi satu blok teks raksasa menggunakan "`.join()`.

Terakhir, fungsi `generate_wordcloud` dipanggil empat kali. Setiap panggilan menghasilkan satu *Word Cloud* unik, dengan colormap 'Greens' (Hijau) untuk sentimen positif dan 'Reds' (Merah) untuk sentimen negatif, sehingga visualisasi yang dihasilkan sangat mudah dibedakan.

Gambar 4.4 Erigo Positif



Analisis *Word Cloud* Erigo Positif (gambar 4.4) menunjukkan bahwa citra positif *brand* ini sangat didorong oleh persepsi harga dan identitas lokal. Kata-kata seperti '*keren*', '*murah*', '*bangga*', dan '*lokal*' menonjol, menegaskan bahwa pengguna mengasosiasikan Erigo dengan keterjangkauan harga dan kebanggaan sebagai *merek lokal* yang sukses.

Gambar 4.5 Erigo Negatif



Pada sisi negatif, *Word Cloud* Erigo Negatif (gambar 4.5) diimbangi oleh kritik yang signifikan. Kata-kata seperti 'lama', 'pengiriman', 'bahan', 'tipis', dan 'kecewa' mendominasi, menunjukkan bahwa keluhan utama audiens berfokus pada masalah *logistik* dan ekspektasi kualitas produk yang belum terpenuhi.

Gambar 4.6 Uniqlo Positif



Berbeda secara signifikan, *Word Cloud* Uniqlo Positif (gambar 4.6) menunjukkan bahwa citra positif merek ini dibangun di atas atribut kualitas intrinsik. Kata-kata seperti '*adem*', '*bahan*', '*bagus*', '*nyaman*', '*kualitas*', dan '*tahan*' yang menonjol secara jelas menghubungkan citra positif Uniqlo dengan material, kenyamanan, dan daya tahan produk.

Gambar 4.7 Uniqlo Negatif



Sementara itu, *Word Cloud* Uniqlo Negatif (gambar 4.7) didominasi oleh kata '*mahal*', '*harga*', '*biasa*', dan '*antri*'. Hal ini mengindikasikan bahwa keluhan utama pengguna terhadap Uniqlo tidak berfokus pada kualitas produk, melainkan pada *pricing* (penetapan harga) yang dianggap premium dan masalah pengalaman berbelanja di toko.

KESIMPULAN

Berdasarkan rangkaian penelitian yang mencakup proses pengumpulan 19.466 komentar pengguna TikTok, pelabelan sentimen otomatis menggunakan model *IndoBERT-Sentiment Analysis*, serta analisis kuantitatif terhadap popularitas, citra merek, dan kata kunci dominan, dapat disimpulkan bahwa merek Erigo dan Uniqlo menunjukkan karakteristik persepsi yang berbeda di mata konsumen. Erigo memiliki tingkat popularitas yang lebih tinggi, yang tercermin dari proporsi jumlah komentar sebesar 67,95% sedangkan Uniqlo menyumbang 32,05% dari total percakapan pengguna. Namun demikian, tingginya popularitas tersebut belum sepenuhnya diiringi oleh citra merek yang positif karena masih ditemukannya sentimen negatif yang cukup signifikan, terutama berkaitan dengan kualitas bahan produk dan keluhan terhadap proses pengiriman.



Sebaliknya, Uniqlo meskipun memiliki volume percakapan yang lebih rendah, menunjukkan citra merek yang lebih kuat dengan dominasi sentimen positif sebesar 47,74% dan sentimen negatif sebesar 52,26%. Persepsi positif ini terutama dipengaruhi oleh penilaian konsumen terhadap kualitas material, kenyamanan produk, serta konsistensi pengalaman berbelanja. Hasil ini diperkuat melalui analisis *word cloud* yang menunjukkan bahwa Erigo banyak diasosiasikan dengan harga terjangkau dan identitas sebagai produk lokal, sedangkan Uniqlo lebih sering dikaitkan dengan kualitas premium meskipun tetap memperoleh kritik terkait harga. Secara keseluruhan, temuan penelitian ini menegaskan bahwa tingkat popularitas yang tinggi tidak selalu mencerminkan citra merek yang positif, serta membuktikan efektivitas model *IndoBERT* dalam menangkap dinamika opini dan bahasa informal pengguna di media sosial.

Saran

Berdasarkan temuan penelitian, Erigo disarankan untuk lebih memfokuskan upaya pada peningkatan kualitas bahan serta memperbaiki kecepatan proses pengiriman, mengingat dua aspek tersebut menjadi pemicu utama munculnya sentimen negatif. Tingginya volume percakapan tentang Erigo di TikTok juga perlu dimanfaatkan secara strategis melalui komunikasi digital yang lebih konsisten, sehingga keunggulan produk dapat tersampaikan dengan lebih efektif. Sementara itu, Uniqlo perlu mempertimbangkan penyesuaian strategi harga dan peningkatan pengalaman berbelanja baik pada toko fisik maupun kanal *online* untuk mengurangi keluhan terkait harga premium dan panjangnya antrean. Untuk penelitian selanjutnya, disarankan agar peneliti memperluas sumber data ke platform lain, menambahkan analisis berbasis *emotion mining*, serta mengevaluasi perbandingan performa model seperti *IndoBERT*, *IndoBERTweet*, atau *IndoT5* guna menghasilkan temuan yang lebih komprehensif dan detail.

DAFTAR PUSTAKA

Sumber Jurnal:

- A, Edi Supriyadi, & A, Putra Nurhuda Makatita. (2025). Analisis Sentimen Komentar Pengguna TikTok di QRIS Adopsi di Indonesia Menggunakan IndoBERT. *Procedia Computer Science*, 269, 121–130. <https://doi.org/10.1016/j.procs.2025.08.265>
- Ade, Anindya Dala, Rizan, Mohamad, Febrilia, Ika, Jakarta, Universitas Negeri, Image, Brand, & Loyalty, Brand. (2024). Pengaruh Aktivitas Pemasaran Media Sosial Media Sosial Terhadap Citra Merek , Loyalitas Merek , Dan Niat Beli Ulang Pada Social Commerce. *Jurnal masharif al-syariah*, 9(204), 2399–2416.
- Adhini, Annisa Zata, Hanum, Farida, Fahmi, Teuku, & Syahputra, Teguh. (2024). Analisis Strategi Promosi Erigo Store Brand Fashion Lokal Indonesia. *Jurnal Bisnis Corporate*, 8(2), 50–54. <https://doi.org/10.46576/jbc.v8i2.4221>



Adhit, Muhammad, & Yuda, Dwi. (2025). Implementasi Web Scraping Untuk Ekstraksi Data Penjual dan Produk Panel Surya Di E-Commerce. *Jurnal Konstelasi*, 5(1), 1–12.

Ainunnisa, Isna Riaandita, & Sulastri, Sulastri. (2023). Analisis Sentimen Aplikasi Tiktok dengan Metode Support Vector Machine (SVM), Logistic Regression dan NaÃ¢ve Bayes. *Jurnal Teknologi Sistem Informasi Dan Aplikasi*, 6(3), 423–430. <https://doi.org/10.32493/jtsi.v6i3.31076>

Amory, Jeffriansyah Dwi Sahputra, Mudo, Muhtar, & J, Rhena. (2025). Transformasi Ekonomi Digital dan Evolusi Pola Konsumsi: Tinjauan Literatur tentang Perubahan Perilaku Belanja di Era Internet. *Jurnal Minfo Polgan*, 14(1), 28–37. <https://doi.org/10.33395/jmp.v14i1.14608>

Andhika, Firza Rafiandi, Witanti, Wina, & Sabrina, Puspita Nurul. (2025). Analisis Sentimen Menggunakan Metode IndoBERT pada Ulasan Aplikasi Zoom Menggunakan Fitur Ekstrasi GloVe. *Metik Jurnal*, <https://doi.org/10.47002/metik.v9i2.1098>

Arya, Sun Mukti, & Pratama, Ifan Alif. (2024). Strategi Viral di TikTok: Panduan Pemasaran Melalui Media Sosial untuk Bisnis. *Jurnal Ilmiah Ekonomi Global Masa Kini*, 13(2), 5.

Assegaf, Syarifah Nadira, Suhud, Usep, & Usman, Osly. (2025). Pengaruh Brand Image, Perceived Risk, Dan Online Customer Review Terhadap Trust Dan Purchase Intention Pada Tiktok Shop. *Indonesian Journal of Economics, Management, and Accounting*, 2(8), 2234–2250.

Atika, Yeni Dwi, & Puruwita, Dita. (2025). Pengaruh Influencer Marketing Terhadap Brand Awareness Dan Purchase Intention Produk Skincare True To Skin Pada Media Sosial Tiktok. *Jurnal Ilmiah Ekonomi Manajemen Bisnis Dan Akuntansi*, 2(4), 439–451. Retrieved from <https://doi.org/10.61722/jemba.v2i4.1218>

Azahra, Nisa Maulia, & Setiawan, Erwin Budi. (2023). Sentence-Level Granularity Oriented Sentiment Analysis of Social Media Using Long Short-Term Memory (LSTM) and IndoBERTweet Method. *Jurnal Ilmiah Teknik Elektro Komputer Dan Informatika*, 9(1), 85–95. <https://doi.org/10.26555/jiteki.v9i1.25765>

Caroline, S. M. P., Wulandari, J., & Roni, M. (2025). Analisis Dampak Pemasaran Media Sosial, Citra Merek, dan Kualitas Produk Terhadap Keputusan Pembelian Produk Kecantikan. *Jurnal Kompetitif Bisnis*, 3, 52–63. Retrieved from <http://jkb.fisip.unila.ac.id/index.php/jkb/article/view/404%0Ahttps://jkb.fisip.unila.ac.id/index.php/jkb/article/download/404/204>

Devie, Fenolychia, & Soepatini. (2024). Pengaruh Customer Review Dan Influencer Endorsement Terhadap Purchase Intention Pada Produk The Originote Dengan Trust Sebagai Variabel Moderasi. *E-Proceeding of Management*, 11(1), 186–192.



Dharma, Aditya, & Mesra, Romi. (2025). Pengaruh Media Sosial terhadap Perilaku Konsumtif di Kalangan Remaja Kota Pangkal Pinang. *Jurnal Masyarakat Digital*. 1(3), 124–136.

Dianthea, Anastasia Salsa, Fatonah, Siti, Ardhanariswari, Kartika Ayu, & Komunikasi, Departemen Ilmu. (2023). Kampanye Pemasaran Media Sosial # Realsupport untuk Meningkatkan Keterlibatan Konsumen dengan Merek Uniqlo Indonesia. *Semantic Scholar*, 16, 149–172.

Diki Hendriyanto, Muhammad, Ridha, Azhari Ali, & Enri, Ultach. (2022). Analisis Sentimen Ulasan Aplikasi Mola Pada Google Play Store Menggunakan Algoritma Support Vector Machine Sentiment Analysis of Mola Application Reviews on Google Play Store Using Support Vector Machine Algorithm. *Journal of Information Technology and Computer Science (INTECOMS)*, 5(1), 1–7.

Erika, Lestari Desi, & Gunawan, Ce. (2021). Pengaruh E-Wom Pada Media Sosial Tiktok Terhadap Brand Image Serta Dampaknya Pada Minat Beli. *Jurnal Ekonomi, Manajemen, Bisnis Dan Sosial Available*, 1(46), 34–39.

Fardiana Putri, Rivani, Fitriani, Elsa Windu, & Sartika, Sri Hardianti. (2022). Analisis Strategi Pemasaran Erigo Store Brand Fashion Lokal Indonesia Yang Mendunia. *Transekonomika: Akuntansi, Bisnis Dan Keuangan*, 2(5), 213–220. <https://doi.org/10.55047/transekonomika.v2i5.185>

Imam Izzulsyah, Adiyus, Radika, Ahmad Nur Hidayah, & Lianda Saputra. (2022). Analisis Penggunaan Media Sosial Di Masa Pandemi. *Fraction: Jurnal Teori Dan Terapan Matematika*, 2(1), 21–31. <https://doi.org/10.33019/fraction.v1i2.28>

Jauhari, Muhammad Tantowi, Marleny, Finki Dona, Informatika, Program Studi, Teknik, Fakultas, Banjarmasin, Universitas Muhammadiyah, Syarkawi, Gubernur, & Dalam, Semangat. (n.d.). Penerapan Data Mining Algoritma Decision Tree Untuk Memprediksi Keterlambatan. *Jurnal Informatika dan Teknik Elektro Terapan*, 13(3).

Krisdanu, Cheryl Arshiefa, & Kiranastari Asoka Sumantri. (2023). TikTok sebagai Media Pemasaran Digital di Indonesia. *Jurnal Lensa Mutiara Komunikasi*, 7(2), 24–36. <https://doi.org/10.51544/jlmk.v7i2.4173>

Merdiansah, Roni, & Ridha, Azhari Ali. (2024). Analisis Sentimen Pengguna X Indonesia Terkait Kendaraan Listrik Menggunakan IndoBERT. *Jurnal Ilmu Komputer dan Sistem Informasi*, 7, 221–228.

Muhajirin, Risnita, & Asrulla. (2024). PENDEKATAN PENELITIAN KUANTITATIF DAN KUALITATIF SERTA TAHAPAN PENELITIAN. *Journal Genta Mulia*, 15(1), 82–92.



Muslimin, Datunggu, Sri Aditya, & Lamakaraka, Anisa. (2023). Dampak Negatif dari Media Sosial Tiktok Terhadap Gaya Bahasa Masyarakat. *Jurnal Bahasa, Sastra, Dan Budaya*, 13(3), 54–67.

Nomor, Volume, Mahmudah, Feni, Rahaningsih, Nining, Dana, Raditya Danar, & Rohmat, Cep Lukman. (2023). Implementasi Data Mining Menggunakan Algoritma K-Means untuk Mempermudah Pengelompokan Wilayah Rawan Stunting di Kabupaten Cirebon. *Jurnal INTEK (Informatika dan Teknologi Informasi)*, 8, 44–52.

Nova Ardiyansyah, & Suryari Purnama. (2024). Pengaruh Kualitas Produk, Citra Merek dan Harga terhadap Keputusan Pembelian. *Inisiatif: Jurnal Ekonomi, Akuntansi Dan Manajemen*, 3(2), 323–345. <https://doi.org/10.30640/inisiatif.v3i2.2374>

Noventi, Messi Putri, & Thaib, Ilham. (2025). Indonesian Journal of Digital Business Pengaruh Celebrity Endorsement dan User- Generated Content (UGC) Terhadap Purchase Intention Produk Avoskin Beauty di Kota Padang dengan Brand Image Sebagai Variabel Mediasi. *Indonesian Journal of Digital Business*, 5(1), 223–234.

Otarina, Sisca, Puspita, Fitri Maya, Yuliza, Evi, & Meeting, Google. (2025). *Jurnal Pepadu Jurnal Pepadu*. 6(1), 56–66.

Pratama, Ahlul Yoga, Sanjaya, Gauri Ananda, Lubis, Nadya Khairunisa, & Aditya, Muhammad Rangga. (2025). Analisis Sentimen Publik Terkait Danantara Menggunakan Algoritma IndoBERT pada Platform Media Sosial. *Metik Jurnal*, <https://doi.org/10.47002/metik.v9i1.1055>

Putri, Intan Lifinda Ayuning, Astri, Mutia Fadila, Rachman, Helmy Aulia, & Alam, Muhammad Dimar Alam. (2024). Sosialisasi Strategi Marketing Bisnis Melalui Media Sosial Tiktok. *Jurnal Pengabdian Kepada Masyarakat Nusantara*, 6(1.1), 286–293. Retrieved from <https://ejournal.sisfokomtek.org/index.php/jpkm/article/view/4804>

Restian, Kesi, Pujawati, Dwi, Irawan, Herry, & Telkom, Universitas. (2024). Analisis sentimen berbasis aspek dengan indobert untuk memetakan persepsi publik tentang ibu kota nusantara dalam dimensi kota cerdas. (2017).

Saputra, Linji, & Heriputranto, Lukas Agung. (2024). Pengaruh Persepsi Harga, Kualitas Produk, dan Citra Merek Terhadap Keputusan Pembelian dengan Kepercayaan Merek Sebagai Variabel Mediasi (Studi pada Konsumen Sepeda Motor Listrik di Provinsi Jawa Tengah). *EQUILIBRIUM: Jurnal Bisnis & Akuntansi (EJBA)*, 18(2), 176–185.

Sunaryo, Nico Christian. (2024). Digital Indonesia Report 2024. Retrieved from We Are Social website: https://www.slideshare.net/slideshow/digital-indonesia-report-2024-by-we-are-social-pdf/267156491?from_search=2#3

Tangke, Riyanto, Tineke Salaki, Deiby, Widsl Kalengkongan, Wisard, & Ketaren, Eliasta. (2024). Analisis Sentimen Aplikasi Tiktok Menggunakan Algoritma Support Vector



Machine (Svm) Dan Random Forest. Jurnal TIMES, 13(2), 53–62.
<https://doi.org/10.51351/jtm.13.2.2024762>

Ulfah, Amanny, Ramadhanty, Nabiylah, & Naiyah, Ina. (2023). IMPLEMENTASI WEB SCRAPING PADA SITUS JURNAL SINTA. JIKA (Jurnal of Informatics), 7(1), 29–36.

Virginia, Mandasari, & M.L.Maghfiroh. (2024). PENGARUH KONTEN PEMASARAN DIGITAL TIKTOK TERHADAP CITRA MEREK PRODUK KEMAIU. Jurnal Internasional Penelitian dan Literatur Multidisiplin, 57–63.

Wijaya, Ahmad. (2023). Implementasi Model IndoBERT pada Dashboard Sentimen Media Sosial (Studi Kasus Universitas XYZ). 10(4), 3910–3926.

Wilyani, Febby, Arif, Qonaah Nuryan, & Aslimar, Fitri. (2024). Pengenalan Dasar Pemrograman Python Dengan Google Colaboratory. (1).

Wjaya, Graciela Putri, & Yulita, Henilia. (2022). Pengaruh Konten Marketing, E-WoM, dan Citra Merek di Media Sosial Tiktok terhadap Minat Beli Kosmetik Mother of Pearl. Journal of Business & Applied Management, 15(2), 133.
<https://doi.org/10.30813/jbam.v15i2.3772>

Yani, Sulisti, Wijayanto, Gatot, & Musfar, Tengku Firli. (2022). PENGARUH SOCIAL MEDIA MARKETING DAN e-WOM TERHADAP BRAND IMAGE SERTA DAMPAKNYA TERHADAP MINAT BELI. Jurnal Manajemen Pemasaran Dan Perilaku Konsumen, 1(3), 277–289. <https://doi.org/10.21776/jmppk.2022.01.3.01>

Sumber Website:

Sunaryo, Nico Christian. (2024). Digital Indonesia Report 2024. Retrieved from We Are Social website: https://www.slideshare.net/slideshow/digital-indonesia-report-2024-by-we-are-social-pdf/267156491?from_search=2#3