

ANALISIS MULTI-MODEL UNTUK SENTIMEN DAN PREDIKSI RATING PELANGGAN E-COMMERCE

**The Hack 2025 – CCI Data Research
Kelompok 1**



**CENTRAL COMPUTER
IMPROVEMENT**

Disusun Oleh:

Muhammad Ammar Ridho	:	103052400009
Putri Khairamulya Ramadhini	:	1301220176
Ovi Orlanda Br Ginting	:	

**Telkom University
Bandung
2025**

DAFTAR ISI

BAB I	3
1.1. <i>Relevansi dan Latar Belakang</i>	3
1.2. <i>Rumusan Masalah</i>	4
1.3. <i>Tujuan</i>	4
1.4. <i>Batasan Masalah</i>	5
BAB II	6
2.1. <i>Sumber dan Kualitas Dataset</i>	6
2.2. <i>Exploratory Data Analysis (EDA)</i>	7
2.3. <i>Preprocessing Data</i>	7
BAB III	8
BAB IV	9
BAB V	10
DAFTAR PUSTAKA	11

BAB I

DEFINISI MASALAH

1.1. Relevansi dan Latar Belakang

Industri *e-commerce* di Indonesia mengalami pertumbuhan yang sangat pesat dalam dekade terakhir, namun di sisi lain, berbagai keluhan konsumen terus muncul dan memengaruhi tingkat kepuasan pelanggan terhadap platform. Data dari Badan Penyelesaian Sengketa Konsumen (BPKN) menunjukkan tren pengaduan konsumen *e-commerce* yang fluktuatif namun tetap signifikan: dari 280 kasus pada 2017 meningkat drastis menjadi 3.256 kasus pada 2021, kemudian menurun menjadi 929 kasus pada 2023, dan 162 kasus hingga November 2024 (Jurnal BEST, 2024).

Meskipun terjadi penurunan sejak puncak pada 2021, jumlah pengaduan masih menunjukkan masalah sistematis yang perlu diatasi. Masalah utama yang konsisten dilaporkan meliputi produk tidak sesuai deskripsi, keterlambatan pengiriman, dan layanan pelanggan yang kurang responsif. Yang lebih mengkhawatirkan, tingkat penyelesaian kasus pengaduan di sektor *e-commerce* masih kurang dari 50%, menunjukkan kompleksitas penanganan dan keterbatasan sumber daya dalam layanan pelanggan (Jurnal BEST, 2024).

a. Masalah Pengiriman

Keterlambatan pengiriman dan pembatalan layanan secara sepihak menjadi keluhan dominan. Platform besar seperti *Shopee* dan *Tokopedia* mendapat sorotan khusus karena lonjakan permintaan pada musim tertentu membuat logistik mengalami hambatan, terutama untuk daerah di luar Jawa dan wilayah Indonesia Timur (IDL Cargo, 2025).

b. Produk Tidak Sesuai Deskripsi

Ketidaksesuaian produk termasuk barang cacat dan palsu merupakan keluhan utama di platform seperti *Shopee*, *Tokopedia*, dan *Bukalapak* (Jurnal BEST, 2024).

c. Ulasan dan Rating Tidak Akurat

Ulasan palsu dan rating menyesatkan mempengaruhi keputusan pembelian konsumen, terutama pada *platform marketplace* terbesar seperti *Shopee* dan *Tokopedia* (Jurnal BEST, 2024).

d. Transparansi Harga

Konsumen mengeluhkan harga yang tidak transparan, termasuk biaya antar tersembunyi dan promo yang membingungkan seperti harga yang dinaikkan sebelum diskon *flashsale* (Jurnal BEST, 2024).

1.2. Rumusan Masalah

Berdasarkan latar belakang tersebut, rumusan masalah dalam penelitian ini adalah:

- a. Bagaimana kerja sistem analisis sentimen untuk mengklasifikasikan ulasan pelanggan e-commerce?
- b. Faktor-faktor apa saja yang paling berpengaruh terhadap rating pelanggan dan bagaimana memprediksinya secara akurat?
- c. Bagaimana mengidentifikasi ketidaksesuaian antara sentimen ulasan dan rating untuk mendeteksi potensi manipulasi atau bias ulasan?

1.3. Tujuan

Berdasarkan masalah di atas, proyek ini bertujuan untuk mengembangkan sistem analisis multi-model yang mampu menganalisis sentimen ulasan pelanggan e-commerce, memprediksi rating berdasarkan faktor operasional, dan mengidentifikasi ketidaksesuaian antara sentimen dan rating untuk deteksi potensi manipulasi ulasan.

Tujuan spesifik sebagai berikut;

- a. Mengembangkan model *NLP* (*Natural Language Processing*) untuk klasifikasi sentimen ulasan dengan akurasi $\geq 85\%$
- b. Membangun model prediksi rating dengan *MAE* (*Mean Absolute Error*) ≤ 0.5 dan $R^2 \geq 0.75$ berdasarkan parameter seperti *delivery_delay_hours*, *price*, *freight_value*, dan *seller_grade*
- c. Mengidentifikasi inkonsistensi sentimen-rating dengan *precision* $\geq 80\%$ untuk deteksi anomali ulasan
- d. Menyediakan *dashboard* interaktif untuk visualisasi insights dan faktor-faktor yang mempengaruhi kepuasan pelanggan

1.4. Batasan Masalah

Untuk menjaga fokus dan realisme proyek, batasan ruang lingkup ditetapkan sebagai berikut:

- a. Data terbatas pada dataset ulasan *e-commerce*, fokus pada kategori produk umum.
- b. Metodologi menggunakan *supervised learning* untuk klasifikasi sentimen tiga kategori (positif, negatif, netral) dan *machine learning konvensional* untuk prediksi rating.
- c. Implementasi sistem prototype dengan antarmuka sederhana, tidak dioptimalkan untuk *deployment* produksi skala besar.
- d. Evaluasi terbatas pada metrik standar (*accuracy*, *precision*, *recall*, *F1-score*, *MAE*, *RMSE*, R^2) tanpa analisis psikologi konsumen mendalam.

BAB II

DATASET

2.1. Sumber dan Kualitas Dataset

Proyek ini menggunakan dua dataset utama yang saling melengkapi untuk analisis sentimen dan prediksi rating.

a. Dataset Review Tokopedia dan Shopee (*Kaggle*)

review_id	review	rating
1	slow delivery	1
21384	Good quality Accommodating seller Well-packaged ????	3
40895	Baik..sesuai price .. thanks.	4
57264	I love it ?????	5

Tabel 1 *Dataset Review Tokopedia dan Shopee (kaggle)*

Dataset Review Sentimen ini berisi 68.855 baris data berupa teks ulasan pelanggan dan rating kepuasan dalam skala 1 hingga 5. Ulasan mencerminkan pengalaman dan opini pelanggan setelah menggunakan produk atau layanan, sedangkan rating memberikan penilaian numerik atas tingkat kepuasan mereka. Kombinasi kolom review dan rating ini menjadi dasar untuk melatih model analisis sentimen, sekaligus membantu memahami pola kepuasan pelanggan di platform e-commerce.

a. *E-Commerce Analytics for Delivery and Review Prediction (Hugging Face)*

order_id	price	freight_value	seller_grade	...	delivery_delay_hours	review_score
e481f51c...	29.99	8.72	Mediocre	...	-170.58	4
53cdb2fc...	118.7	22.76	Mediocre	...	-128.54	4
47770eb9...	159.9	19.22	Mediocre	...	-413.89	5
949d5b44...	45.0	27.2	Mediocre	...	-311.52	5

Tabel 2. *E-Commerce Analytics for Delivery and Review Prediction (huggingface)*

Dataset E-Commerce Analytics for Delivery and Review Prediction berisi data transaksi publik dari Olist Store di Brasil, dengan total sekitar 100.000 pesanan yang

tercatat pada periode 2016 hingga 2018. Dataset ini memuat informasi detail terkait pengiriman, harga, reputasi penjual, karakteristik produk, lokasi penjual dan pelanggan, serta riwayat pembelian. Data ini digunakan untuk memprediksi skor ulasan pelanggan secara numerik dan menganalisis faktor-faktor operasional yang memengaruhi tingkat kepuasan di ekosistem e-commerce.

2.2. Exploratory Data Analysis (EDA)

2.3. Preprocessing Data

BAB III

METODOLOGI

BAB IV

EVALUASI DAN ANALISIS

BAB V

KESIMPULAN

DAFTAR PUSTAKA

- IDL Cargo. (2025). *Logistik Lebaran 2025: Siap-siap dengan Permintaan Meledak*. Diakses dari <https://idlcargo.co.id/blog/logistik-lebaran-2025-siap-siap-dengan-permintaan-meledak>
- Jurnal BEST. (2024). Analisis pengaduan konsumen e-commerce di Indonesia. *Jurnal BEST*, 1(1), 1-15. Diakses dari <https://jurnalbest.com/index.php/mrbest/article/download/303/127>
- OJS STIAMI. (2024). Pengaruh harga dan kualitas layanan terhadap kepuasan pelanggan e-commerce. *BIJAK: Business and Economics Journal*, 15(2), 123-135. Diakses dari <https://ojs.stiami.ac.id/index.php/bijak/article/view/3427>
- Prosiding ICOSTEC. (2023). Faktor-faktor yang mempengaruhi rating pelanggan dalam e-commerce. *Prosiding International Conference on Science and Technology*, 2(1), 45-52. Diakses dari <https://prosiding-icostec.respati.ac.id/index.php/icostec/article/view/50>
- Repository Jurnal ISI. (2024). Analisis pengaruh informasi produk terhadap kepuasan konsumen online. *Information Systems International*, 8(1), 67-78. Diakses dari <https://journal-isi.org/index.php/isi/article/view/437>