



# CAPSTONE PROJECT

Analisis Sentimen Ulasan Produk Tokopedia dengan IBM Granite



Presented By:

Syalva Audina Pratiwi





# Agenda



- 03 Raw Dataset Link
- 04 Project Overview
- 05 Analysis Process
- 06 Insight & Findings (Visualisasi)
- 07 Conclusion & Recommendations
- 08 AI Support Explanation



# RAW DATASET LINK

Raw Dataset ini digunakan sebagai bahan analisis sentimen (positif, netral, negatif) dan menjadi dasar untuk menghasilkan insight & rekomendasi.

Dataset: Tokopedia Reviews (40.607 ulasan) – Diambil dari Kaggle

# 1

[https://drive.google.com/file/d/1jXsB5uh\\_ug0jxl\\_5MCtJWB\\_J3y\\_NU042/view?usp=sharing](https://drive.google.com/file/d/1jXsB5uh_ug0jxl_5MCtJWB_J3y_NU042/view?usp=sharing)

## Kolom Penting:

- text - isi ulasan pelanggan
- rating - skor bintang 1-5
- category - kategori produk
- product\_name - nama produk



# PROJECT OVERVIEW

## Latar Belakang



- E-commerce seperti Tokopedia memiliki ribuan ulasan produk setiap hari.
- Ulasan ini penting untuk memahami kepuasan pelanggan dan perbaikan layanan.
- Tantangan: ulasan terlalu banyak untuk dibaca manual, sering berisi bahasa campuran & tidak terstruktur.

## Permasalahan

- Penjual kesulitan mengetahui tren keluhan pelanggan dan hal positif dari produk/jasa.
- Tidak ada sistem otomatis yang merangkum insight utama dari ulasan.

## Tujuan Proyek

- Mengklasifikasikan ulasan menjadi positif, netral, negatif.
- Mengidentifikasi pola utama dalam ulasan pelanggan.
- Memberikan rekomendasi berbasis AI untuk peningkatan layanan

## Pendekatan

- Menggunakan dataset Tokopedia (40.607 ulasan).
- Preprocessing teks → pembuatan label sentimen dari rating.
- Model AI IBM Granite 3.0-2B Instruct dijalankan di Google Colab.
- Evaluasi performa (classification report, confusion matrix).
- AI summarization untuk insight & rekomendasi.

# Analysis Process

...

## Dataset

- Dataset ulasan Tokopedia (40.607 review, CSV format).
- Dipilih karena representatif & mencakup berbagai kategori produk.



## Cleaning

- **Bersihkan teks:** hapus emoji, tanda baca berlebih, dan URL.
- Konversi rating bintang menjadi label sentimen:
  - 1-2 ★ → Negatif
  - 3 ★ → Netral
  - 4-5 ★ → Positif
- Alasan: agar model AI punya "label emas" sebagai pembanding hasil prediksi.

## Modeling

- Gunakan IBM Granite 3.0-2B Instruct (via Hugging Face Transformers).
- Teknik: text classification (untuk sentimen) + summarization (untuk insight).
- Alasan: model ini ringan, open-source, dan cocok untuk teks berbahasa campuran (Indonesia + Inggris).

## Evaluation

- Gunakan classification report (precision, recall, F1-score, accuracy).
- Visualisasi dengan confusion matrix.
- Alasan: metrik ini memberi gambaran detail tentang performa klasifikasi di tiap kelas (positif, netral, negatif).

## Insight Generation

- Gunakan AI untuk merangkum top positives, top issues, dan rekomendasi.
- Alasan: membantu penjual fokus pada kekuatan & kelemahan utama tanpa harus membaca ribuan ulasan.



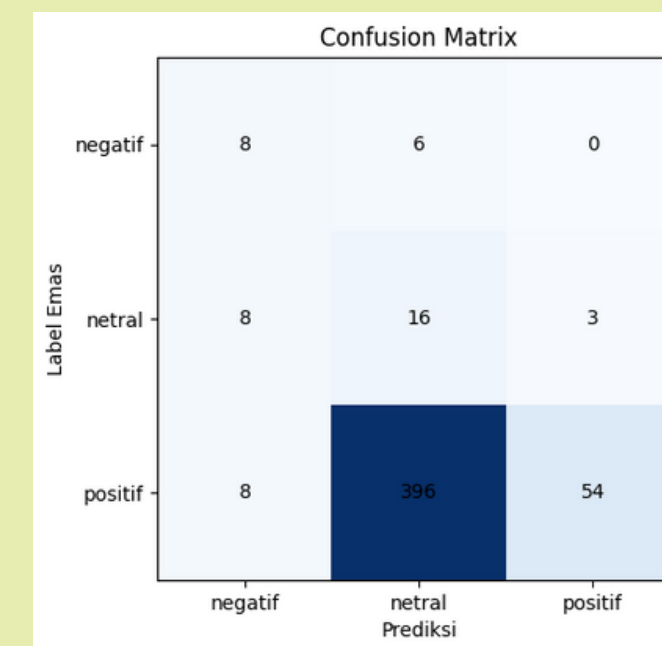
# Insight & Findings

## Hasil Analisis Sentimen (500 sampel uji):

- Akurasi keseluruhan: 16% (rendah akibat imbalanced data).
- Model lebih banyak memprediksi positif, meskipun kenyataannya ada ulasan netral/negatif.
- Menunjukkan perlunya perbaikan dataset & model agar seimbang

## Temuan Utama dari Ulasan:

- **Top Positives:**
  - Fast response penjual
  - Harga terjangkau
  - Kualitas & packaging memuaskan
- **Top Issues:**
  - Stok barang sering habis
  - Deskripsi produk kurang jelas
  - Keterlambatan pengiriman



**Confusion matrix** menunjukkan bahwa model masih memiliki kelemahan dalam mengklasifikasikan ulasan. Sebagian besar ulasan **positif** justru diprediksi sebagai **netral**, sehingga akurasi keseluruhan menjadi rendah. Meskipun beberapa ulasan negatif dan netral berhasil diprediksi dengan benar, ketidakseimbangan data membuat model cenderung bias terhadap kelas tertentu.

Hal ini menandakan perlunya penyeimbangan dataset serta peningkatan model agar hasil klasifikasi lebih akurat dan representatif.

## Interpretasi Logis:

- Mayoritas pelanggan puas (positif), tetapi ada masalah operasional (stok & pengiriman) dan informasi produk.
- Insight ini relevan bagi penjual, **kekuatan** ada di pelayanan & harga, **kelemahan** ada di manajemen stok dan deskripsi produk.

# CONCLUSION & RECOMMENDATIONS



## Conclusion



- AI dapat membantu menganalisis ribuan ulasan Tokopedia dengan cepat.
- Hasil klasifikasi masih terbatas (akurasi rendah) karena dataset imbalanced dan ulasan banyak mengandung bahasa campuran.
- Meski begitu, AI berhasil memberikan insight kualitatif tentang kepuasan pelanggan dan masalah utama.

## Recommendations



- Manajemen Stok → perbaiki sistem inventori untuk mengurangi keluhan “barang habis”.
- Deskripsi Produk → tambahkan detail produk yang lebih lengkap & jelas untuk mengurangi salah paham.
- Pengiriman → perbaiki proses logistik agar keterlambatan dapat diminimalisir.
- Layanan Pelanggan → pertahankan kekuatan fast response & kualitas packaging sebagai nilai jual utama.
- Pengembangan AI Lanjutan → seimbangkan dataset dan coba model lebih besar untuk meningkatkan akurasi klasifikasi.



# AI SUPPORT EXPLANATION



Dalam proyek ini, AI berperan penting untuk menganalisis ulasan pelanggan secara otomatis. Model Large Language Model (LLM) Granite 3.0-2B Instruct digunakan di Google Colab untuk melakukan sentiment classification, yaitu mengelompokkan ulasan ke dalam kategori positif, netral, atau negatif. Selain itu, AI juga dimanfaatkan untuk summarization, merangkum ribuan ulasan menjadi insight utama seperti kekuatan produk, permasalahan yang sering muncul, serta rekomendasi perbaikan. Dengan cara ini, AI mendukung proses analisis agar lebih cepat, efisien, dan mampu memberikan gambaran menyeluruh tanpa harus membaca seluruh dataset secara manual.





# THANK YOU

Analisis Sentimen Ulasan Produk Tokopedia dengan IBM Granite