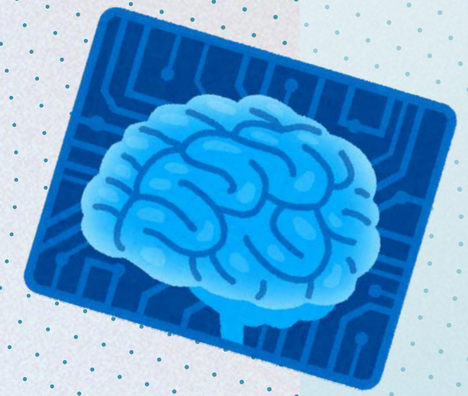
Customer Review Analysis Using IBM Granite via Replicate API

Raw Dataset Link

This dataset was obtained from the Mendeley Data website. The dataset used PREDICT-ID, is a collection of Indonesian product reviews sourced from the e-commerce platform Tokopedia.

📎 https://data.mendeley.com/datasets/574v66hf2v/1

🎯Analysis Focus: Computers and Technology Category



Project Overview

Project Objective:

Automatically extract insights from customer reviews of technology products.

Problem Solved:

- Difficulty understanding large amounts of feedback.
- No automated system for analyzing customer opinions.

Background:

- Many customer reviews are not fully utilized.
- Manual analysis is time-consuming and inefficient.

Solution Approach:

- Category filtering.
- Text preprocessing.
- Analysis using IBM Granite via API.

Analysis Process

Data Collection

The PREDICT-ID dataset is derived from Mendeley Data and contains Indonesian-language product reviews from Tokopedia. This dataset includes several columns, such as Customer Review, Product Category, and others.

Data Cleaning & Filtering

Only data in the Computers & Technology category was selected. The text was cleaned by converting capital letters to lowercase, removing URLs and mentions, removing unnecessary non-alphanumeric characters, and removing excess whitespace.

Large Language Model Processing via API

The ibm-granite/granite-3.1-8b-instruct model is used through the Replicate API.

Each cleaned review is analyzed by LLM to generate:

- Summarization: Summarizing the review text
- Sentiment Analysis: Positive, Negative, Neutral
- Emotion Detection: Emotions such as Happy, Disappointed, Satisfied, Angry, Neutral
- Complaint Detection: Yes/No
- Seller Suggestion: A brief recommendation

Visual Analysis

Visualization is done to display:

- Distribution of sentiment
- Dominant emotions
- Dominant complaints

Processing and Storage of Results

All LLM results are stored in a new DataFrame, then exported to CSV format for further analysis.

Insight & Findings

💡 Insight:

1. While most reviews are positive, there are hidden complaints.
2. Dominant emotions: Satisfaction & Disappointment.
3. Main complaints: Delivery and technical support.
4. Customers provide concrete suggestions, such as improving service and communication.

Insight & Findings

📈 Tokopedia Review Data Insight Visualization:

- Sentiment Distribution

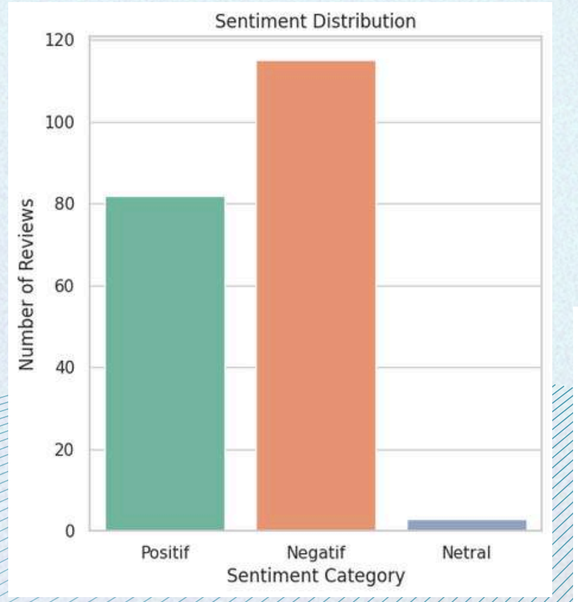
The graph shows that the majority of reviews are negative, followed by positive, and a few are neutral.

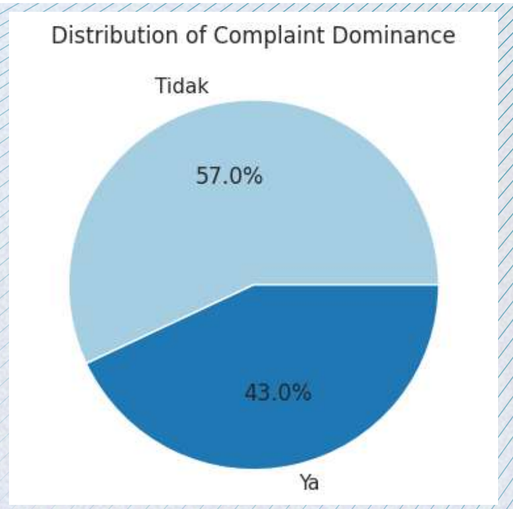
- Dominant Emotions

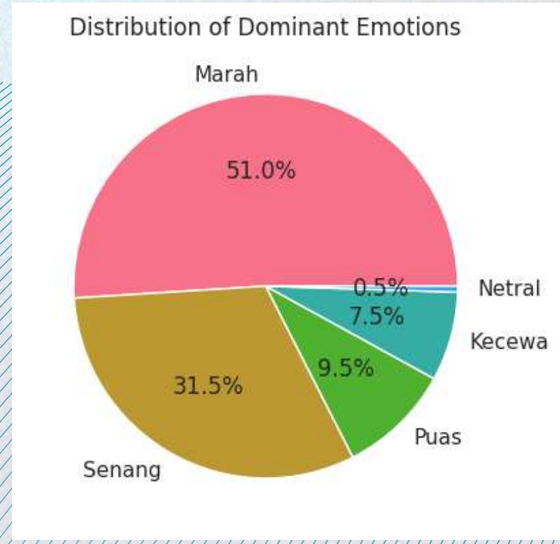
The most common emotion was anger (51%), followed by happiness (31.5%), satisfaction (9.5%), disappointment (7.5%), and neutral (0.5%).

- User Complaints

43% of reviews contained complaints, while 57% did not.







Conclusion & Recommendations

Conclusion:

- AI helps extract insights from unstructured text.
- Automated analysis saves time and provides accurate insights.

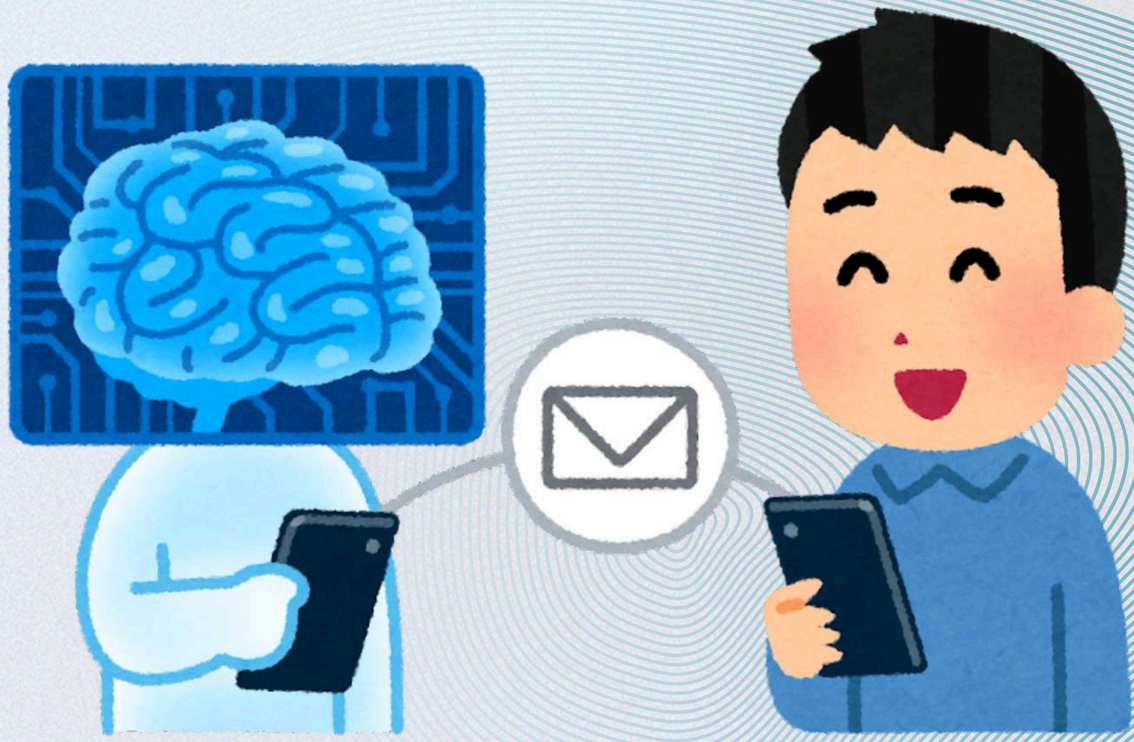
Recommendations:

- Speed up the delivery process.
- Improve post-purchase service.
- Implement an LLM-based automated complaint detection system.

AI Support Explanation

Model: IBM Granite LLM via Replicate API
AI Functions Used:
📌 Summarization: Summarizes reviews into one sentence.
📌Sentiment Analysis: Classifies positive, negative, and neutral.
📌Emotion Detection: Angry, Happy, Disappointed, Neutral, etc.
📌 Complaint Detection: Yes/No.
📌Recommendation: Suggestions for sellers based on review content

Platform: Google Colab + Python + Prompt Engineering.



Thank You