

# Product Review Analysis Report

This report outlines the analysis of a dataset containing 1,000 customer reviews for electronic products across categories such as Smartphones, Laptops, Smart Home, Wearables, and Audio. The dataset includes features like review text, ratings, and labeled sentiments. The project addresses three tasks: developing a document processing pipeline for semantic search, applying a Large Language Model (LLM) for summarization and Q&A, and conducting sentiment analysis with visualizations.

## 1. Document Processing Pipeline

- **Approach:** Reviews were loaded from a text file using LangChain's TextLoader and split into documents with CharacterTextSplitter (chunk size: 500, overlap: 50). Embeddings were generated using OpenAIEmbeddings and stored in a Chroma vector database.
- **Implementation:** The `retrieve_semantic_reviews` function retrieves the top\_k semantically similar reviews for a given query, enabling concept-based searches. This leverages the vector database for efficient similarity matching.

## 2. LLM Application

- **Category Summaries:** The `generate_category_summary` function samples up to 10 reviews per category and uses OpenAI's GPT-3.5-turbo to generate summaries, identifying overall performance, praised features, common issues, and improvement suggestions.
- **Q&A System:** The `product_qa` function retrieves relevant reviews via semantic search and uses the LLM to answer user questions based on those reviews. The Q&A system does the following:
  1. Uses semantic retrieval to fetch top-k relevant reviews for the question.
  2. Constructs a grounded prompt with system instruction and review snippets.
  3. Returns the LLM answer strictly based on provided text.

## 3. Sentiment Analysis & Classification

- **Model:** A pre-trained model (j-hartmann/sentiment-roberta-large-english-3-classes) was used to classify review sentiments as positive, neutral, or negative.
- **Performance:** Predictions were compared to the dataset's sentiment labels, which had an Accuracy of 0.805

Built a Streamlit app with interactive tabs:

1. By Category: Stacked bar charts of sentiment proportions.
2. Over Time: Line charts tracking monthly counts of positive, neutral, and negative reviews.
3. By Rating: Heatmap correlating star ratings with sentiment labels.
4. By Feature: Horizontal bar charts showing sentiment distribution for top 10 features.

## **Conclusion**

The project successfully implemented a semantic search pipeline, LLM-driven insights, and sentiment analysis. This framework offers valuable tools for understanding customer feedback and guiding product enhancements.