Robo-Reviews

Javi Ansoleaga

Project Overview

RoboReviews is a project designed to analyze customer reviews using AI to generate valuable insights for product improvement and consumer guidance.

Main Goals:

1. **Sentiment Analysis**: Classify reviews as positive, negative, or neutral to help understand customer sentiment.
2. **Product Clustering**: Cluster products into 4-6 meaningful categories for better market understanding.
3. **Review Summarization**: Use generative AI to summarize reviews and recommend the top products for each category.

RoboReviews:

Turning customer reviews into... Something, eventually...

Why This Is Still Awesome:

- Despite the challenges, the groundwork is in place. Every good project starts somewhere
- If nothing else, I've learned a whole lot during this project. Specially how not to build a model or two.
- Plus, the preprocessing is almost (sadly) a project on its own

What is your story?

Model 1: Sentiment Analysis
This one's almost functional! It can tell a happy customer from an angry one... most of the time. Let's call it a 'work-in-progress with potential.

Model 2: Product Category Clustering
Technically, there's code. Realistically, it's just a very confused collection of products that refuses to behave. Consider it a philosophical experiment in chaos.

Model 3: Generative AI for Summarization
Barely started, but the ideas are solid. If ideas could be considered finished models, then I'd have a masterpiece!

Demo

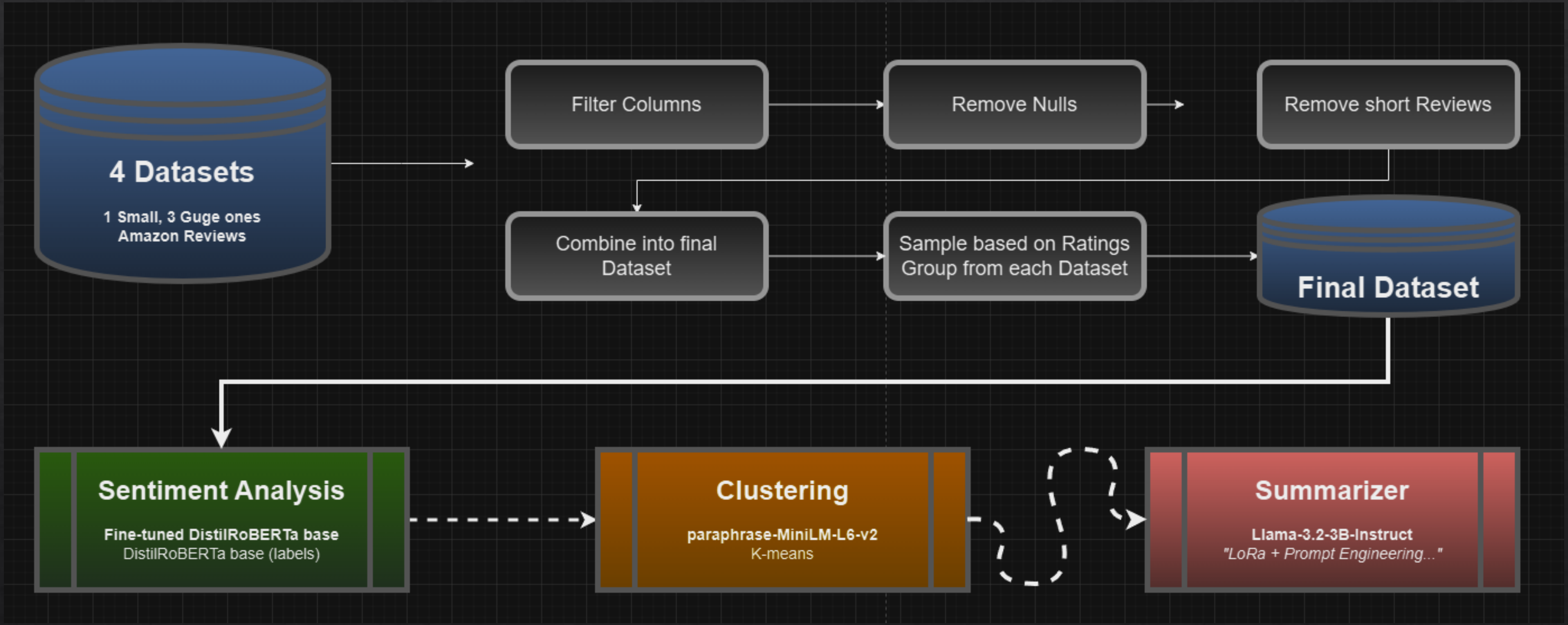
- Add a short demo of how your product can be used
- And/Or add an example of a chat review based on your product

Introduction

- Explain the real world problem this solves
- Leave a point of connection to your methodology

Now that you have seen our super cool product and know why this matters, let me show how we did it!

Methods



4 Datasets

1 Small, 3 Guge ones
Amazon Reviews

Filter Columns

Remove Nulls

Remove short Reviews

Combine into final Dataset

Sample based on Ratings Group from each Dataset

Final Dataset

Sentiment Analysis

Fine-tuned DistilRoBERTa base
DistilRoBERTa base (labels)

Clustering

paraphrase-MiniLM-L6-v2
K-means

Summarizer

Llama-3.2-3B-Instruct
"LoRa + Prompt Engineering..."

Model 1: Sentiment Analysis

The goal was to classify customer reviews into positive, negative, or neutral categories.

Problems:

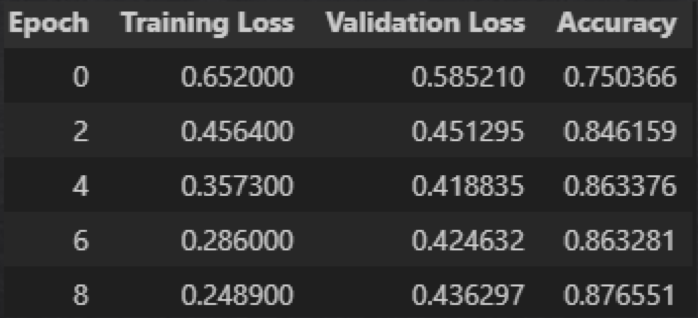
- The review ratings are not always consistent with the sentiment.
- Lack of True sentiment labels for training the model.

Approach:

- Generate labels with a simpler approach to then train and evaluate the model.
- **VADER**: Initially used, but it struggled with nuanced sentiment.
- **DistilRoBERTa**: Switched to using the base model before training to generate the labels.

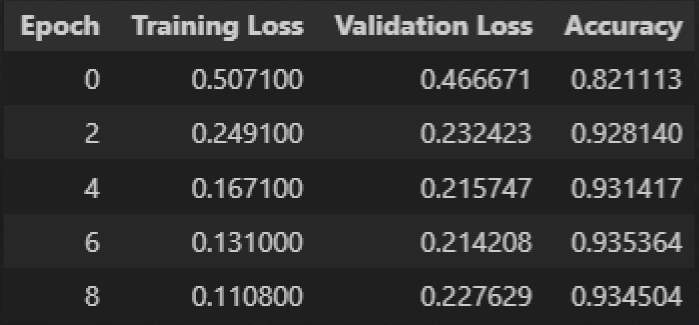
Model 1 - Evaluation

- VADER

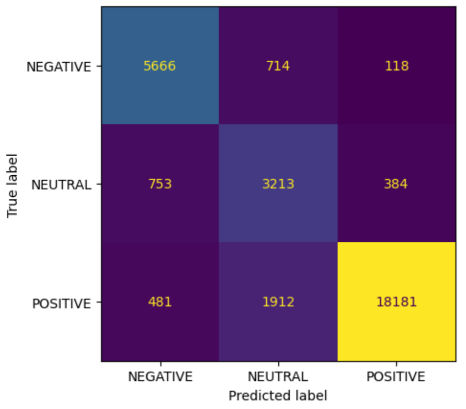


Epoch	Training Loss	Validation Loss	Accuracy
0	0.652000	0.585210	0.750366
2	0.456400	0.451295	0.846159
4	0.357300	0.418835	0.863376
6	0.286000	0.424632	0.863281
8	0.248900	0.436297	0.876551

- DistilRoBERTa



Epoch	Training Loss	Validation Loss	Accuracy
0	0.507100	0.466671	0.821113
2	0.249100	0.232423	0.928140
4	0.167100	0.215747	0.931417
6	0.131000	0.214208	0.935364
8	0.110800	0.227629	0.934504



True label \ Predicted label	NEGATIVE	NEUTRAL	POSITIVE
NEGATIVE	5666	714	118
NEUTRAL	753	3213	384
POSITIVE	481	1912	18181

Takeaway

Project Recap:

- Model 1:
  Functional but not Perfect! Managed to classify reviews into sentiment categories.
- **Model 2:**
  Technically built, practically a bit of a mes (clustering didn't quite work out as planned).
- **Model 3:**
  Barely started, but I laid some foundations…

Takeaway

Lessons Learned:

- Spent way too much time on data preprocessing, which the Project was definetly not about.
- Got distracted by investigation, learning better ways of doing things and side experiments instead of focusing on the project goals.
- **Main Lesson:** Finish first, then optimize.

Final Note:

- Despite running out of time, I learned a lot about both what to do and what *not* to do in projects like this.

Repository: Grimngor/RoboReviews (github.com)