Robot Reviews

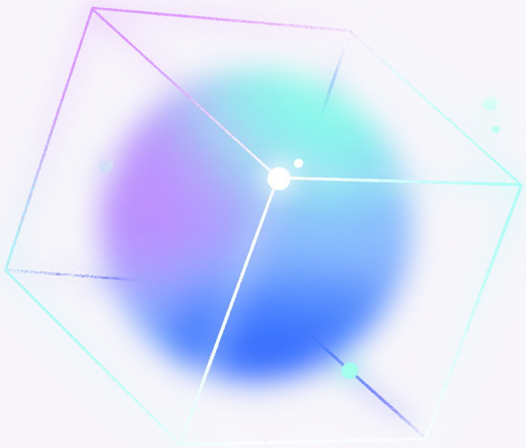
Robot Reviews

Freddy Rivero - Diego Inclan - Saiqa Mehdi

Overview

- Models used: DistilBERT, XNet and Roberta
- We worked as a trio to develop three powerful models, with each member specializing on one model.
- This focused approach allowed for fine-tuning and optimization, leading to superior performance.
- Our collaboration maximized efficiency and ensured every model excelled!

What is your story?



- Worked as a group in each model
- Challenging dataset
- Deep exploration on pre-trained models and their characteristics
- Develop a model that can efficiently scrape and analyze a large database of product reviews, understand the sentiment of each review, automatically categorize the reviews, and extract relevant data to train a generative language model capable of outputting human-like product reviews on demand.
  This would allow for the creation of high-quality, personalized reviews at scale, tailored to specific products or use cases.

Demo

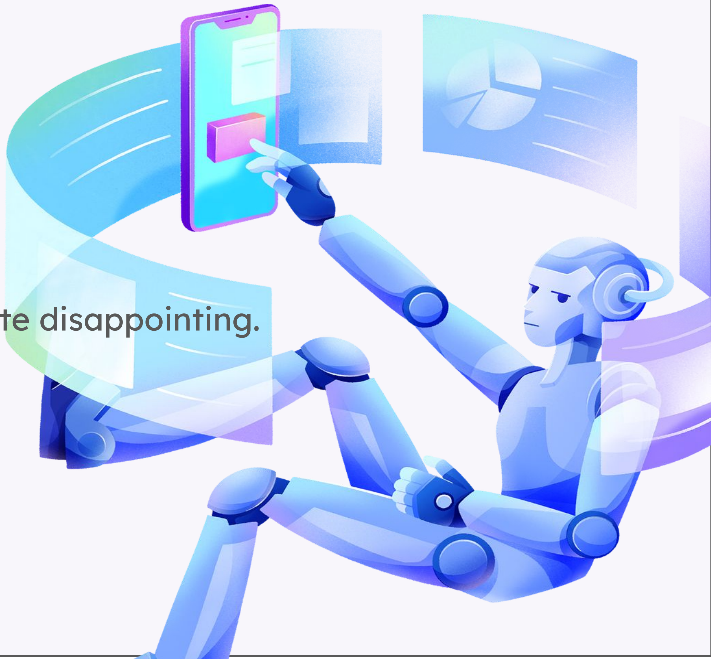
User input for generating a review based on a Negative sentiment for a product called "Samsung Phone" in the "Smartphone" category using N-shot prompting:

Example Demo Input:

- Product Name: Samsung Phone
- Category: Smartphone
- Sentiment: Negative

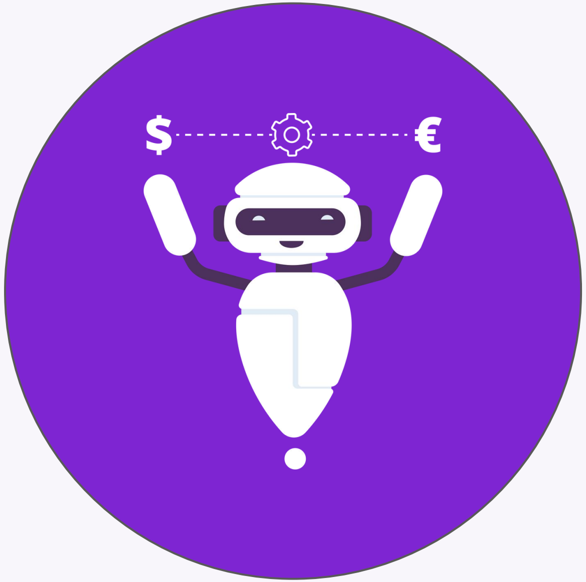
Generated Review:

I bought the Samsung Phone in the Smartphone category and it was quite disappointing. The phone camera was not as as good as they said on the description.



Introduction

- An intelligent review aggregator that provides real-time, actionable insights for businesses looking to improve customer satisfaction and product quality
- Overview of current trends and analysis of product performance.
- Trend analysis to identify products with consistently high or low ratings.



Classificator
DistilBERT with LoRA Configuration

Methods

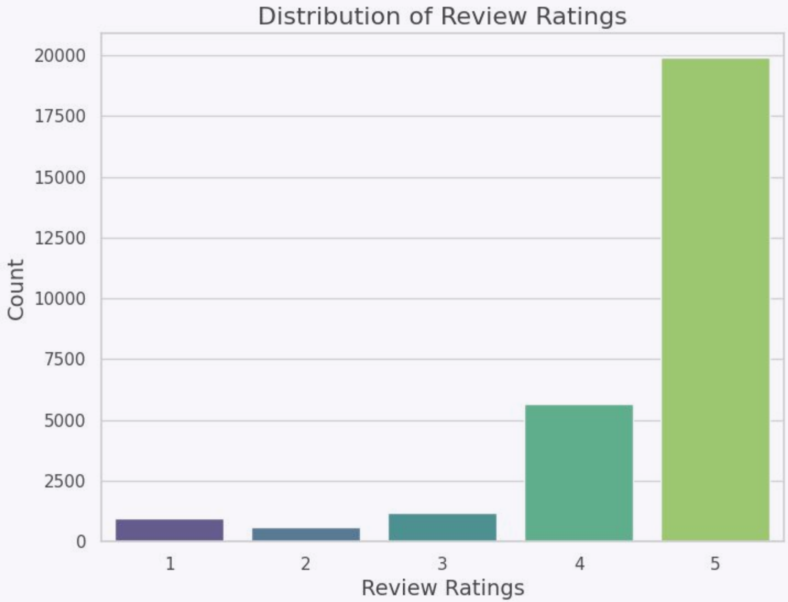
- Little text preprocessing due to transformers vs. choosing relevant columns:
  - 'Name' - 'brand' - 'primary Categories' - 'reviews.text' - 'reviews.rating')
- Dataset imbalanced
- Neutral class being under under represented or being labeled as negative due to deleting missing values so instead we balanced the DS

Classificator - DistilBERT with LoRA adaptation

Classify customer reviews into positive, negative, or neutral categories to help the company improve its products and services.

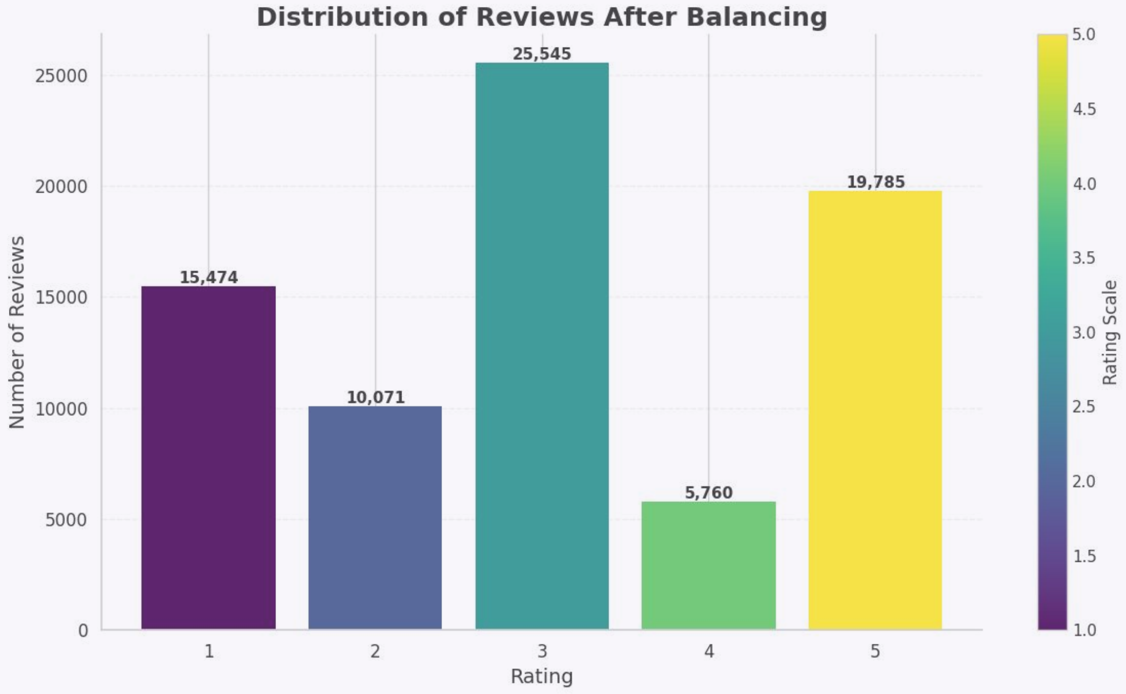
- Imbalanced dataset
- Removing Missing Values challenge

5	19897
4	5648
3	1206
2	616
1	965



Classificator - DistilBERT with LoRA adaptation

Ensured that the model won't be biased towards the majority class (which could skew results in sentiment classification.



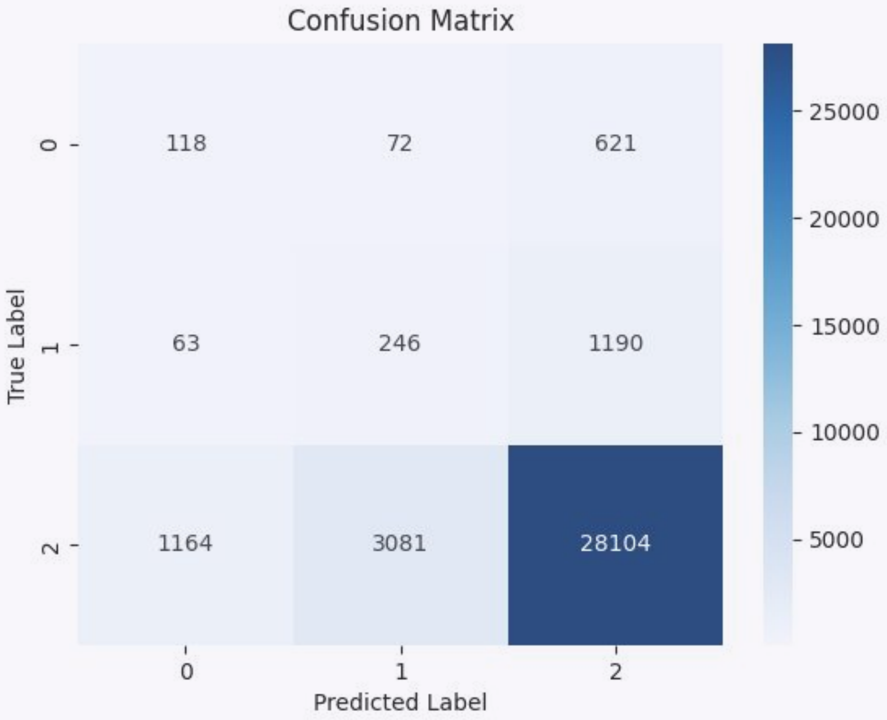
Classificator - DistilBERT with LoRA configuration

- Fine-Tuning via LoRA
- 3-fold Cross-Validation

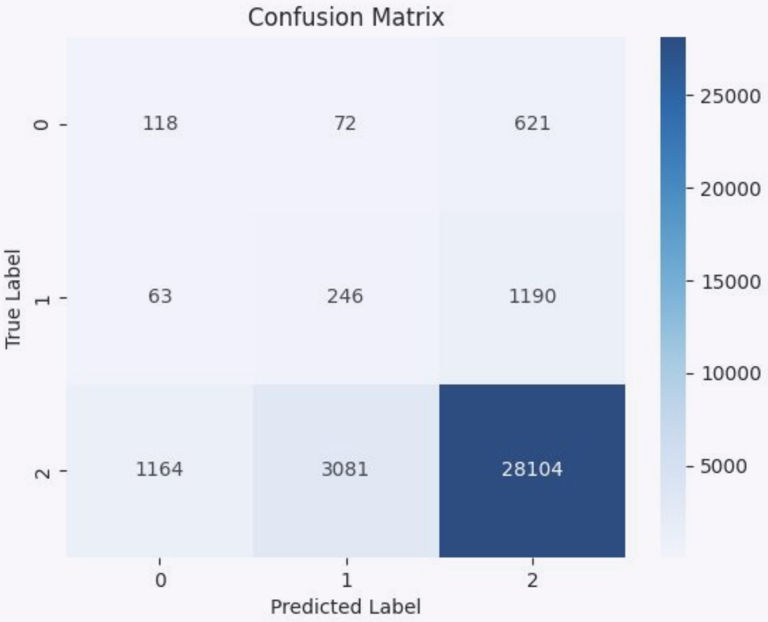
Classificator - DistilBERT with LoRA adaptation

vg Accuracy: 0.9198147060742481
vg Precision: 0.920401176453891
vg Recall: 0.9198147060742481
vg F1 score: 0.9197869333589356

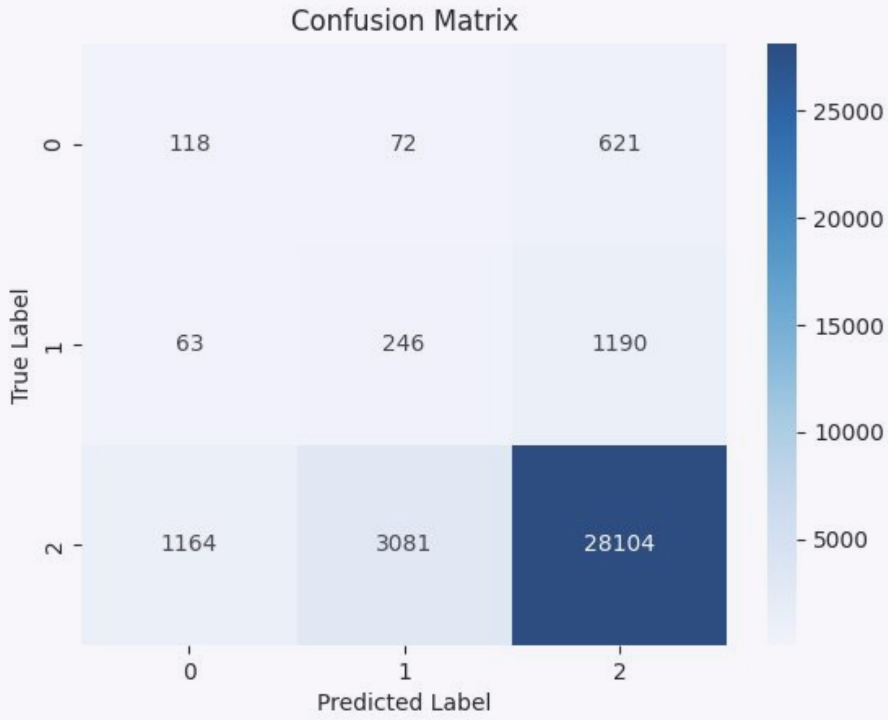
Classificator - DistilBERT with LoRA adaptation



Classificator - DistilBERT with LoRA adaptation
Evaluation



Classificator - DistilBERT with LoRA adaptation



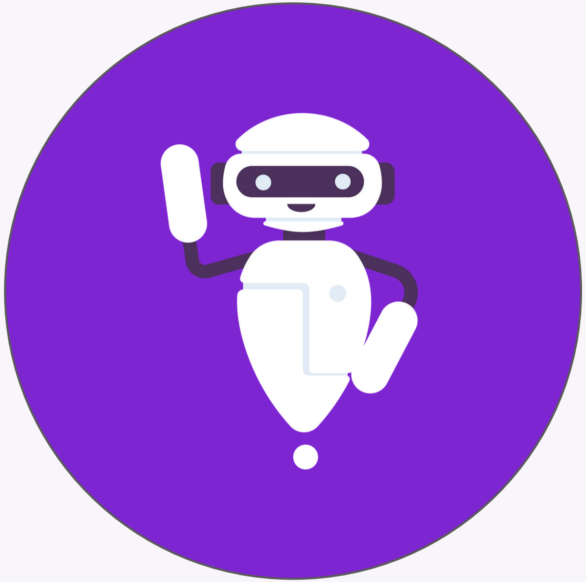
- Ambiguous reviews

"You get what you pay for **Was not to happy** but for a 3 year old it's ok"

Rating: 3

"did not power up...at all. **I am so happy** that my grand-niece wasn't here and doesn't know about this yet, she would have been crying."

Rating: 1



Clustering Model
XLNet

Clustering Model - XLNet

Models Experimentation: XLNet / Albert / Electra

- Cluster product categories into just 4-6 of them

Meta-Category	Keywords
Ebook Readers	kindle, ereader
Batteries	battery, charge, AAA, AA, alkaline
Accessories	keyboard, mouse, laptop stand, case, headphones, adapter, speakers, charger, cables, remote controls, docker, TV fire sticks, docker
Tablets	Ipad, Kids Tablets, Fire Tablets, Amazon Tablets
Non-Electronics	nespresso, pod, pet carrier, coffee

XLNet Fine-Tuning Hyperparameters

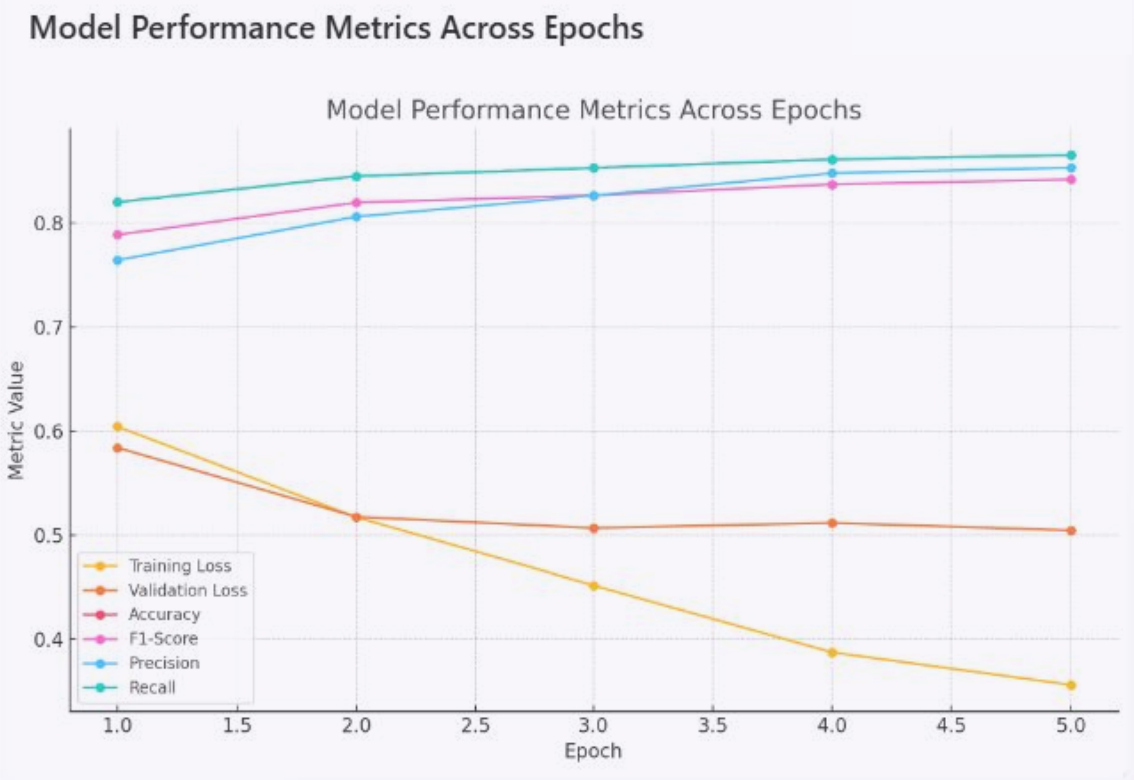
Main Challenge

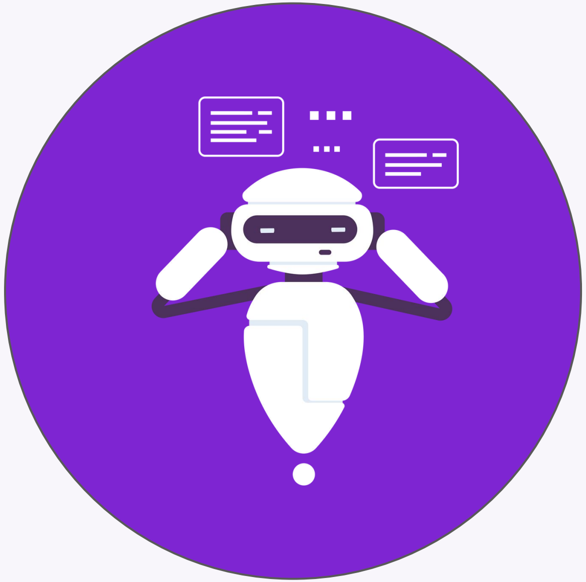
Finding a balance between managing GPU memory constraints and model performance

Change	Original Setting	Modified Setting	Reason for Change
Precision	FP32	FP16 (Mixed Precision)	Reduced memory usage by using 16-bit precision, which allows training on larger models and datasets (FP16 training).
Gradient Accumulation	No gradient accumulation	`gradient_accumulation_steps=2`	Simulated larger batch size while reducing memory footprint by accumulating gradients over multiple steps.
Batch Size	`train_batch_size=16`	`train_batch_size=8`, `eval_batch_size=8`	Reduced batch size to prevent OutOfMemory errors on the GPU.
Number of Epochs	3	5	Increased number of epochs to allow more training iterations for improved learning and model stability.

Clustering Model - XLNet

Evaluation

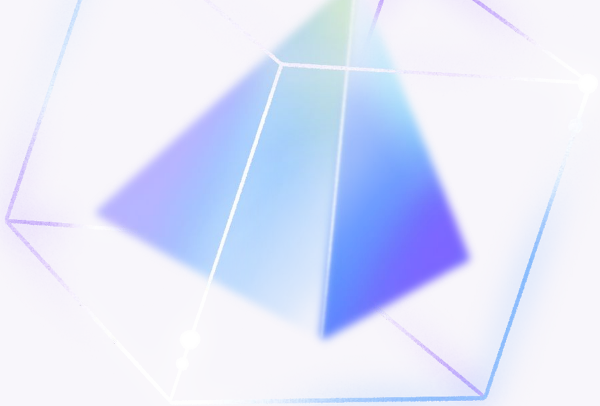




Summarizer Model
Roberta with GPT2

Summarizer Model - Roberta with GPT2

- Use Generative AI to summarize reviews into an article which recommends the top products for each category.
- There's a necessity to use the N-Shot approach to fine tune the model.
- As well the use of different reviews to avoid falling into hallucination or misinterpretation of the user input.



Summarizer Model - Roberta with GPT2

- Model used to identify the sentiment: *Cardiffnlp/twitter-roberta-base-sentiment-latest*
- Roberta was pre-trained on 10K Twits to be able to detect

Output example:

- Full Model Output for Review: Bought as gift for my granddaughter who is 10. She loves it. Plays games, watches videos and listens to music.

Label: Positive, Score: 0.96% Accuracy

- Full Model Output for Review: Worst batteries I have ever purchased. I use in the mouse to my laptop. Doesn't hold a charge for more than three days with limited usage (maybe 4 hours/day). Not worth the price.

Label: Negative, Score: 0.93% Accuracy

Summarizer Model - Roberta with GPT2

Evaluation

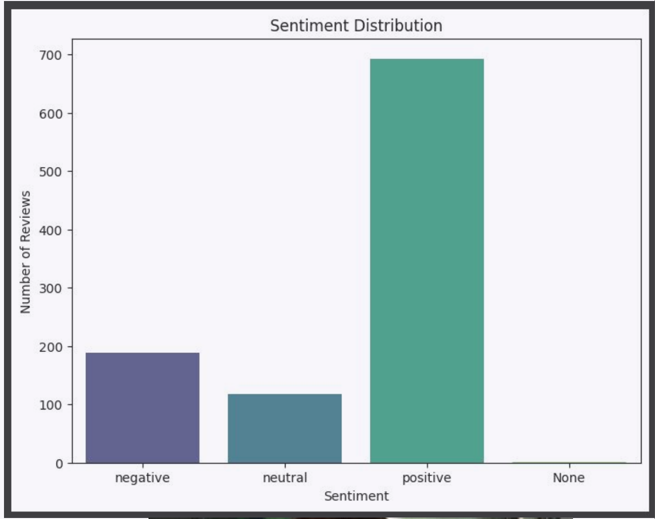
Summarizing Reviews into an Article: Use GPT-2 to generate an article recommending top products by summarizing positive reviews, categorized by product type.

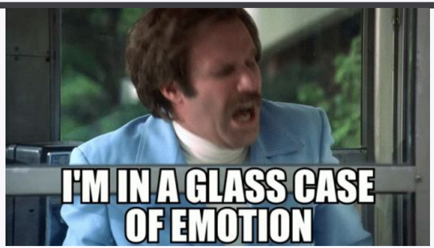
Generating the Article: Filter for positive, high-confidence reviews and generate an article for each product category using GPT-2 with N-shot prompting.

Evaluation: Compare the generated article’s content with the precision and confidence scores from the sentiment analysis model.

Why This Works: N-shot prompting guides GPT-2 to generate structured articles, and confidence scores ensure only reliable reviews are summarized.

Explain Complex Evaluation: Precision checks the accuracy of sentiment classification, while confidence scores verify the reliability of the selected reviews.





Takeaway

- Recap your project and show us again the final version you came up with

- If we had more time we would explore

  Classification Model: Combining text sentiment with rating

  Clustering Model: Roberta, LDA, Zero-shot Learning

  Summarization Model: Apply model to real case scenario

Github Repo's:

Diego Inclán - _KonKon28/Robo-Review-Diego (github.com)_

Saiqa Mehdi - _SaiqaMehdi/RoboReviews-Project (github.com)_

Freddy Rivero - _fredsmeds/RobotReviews: Group project on automated reviews from Amazon products (github.com)_

Thank you!

Any questions?