Amazon Sentiment Analysis

Ahmad Khan, Julieanna Li, Justin N. , Mehrnaz Bastani , Neha Shome, Siddhartha Mangamuri

Table of Contents

Dataset Overview

- Dataset: **Amazon product reviews** on Kaggle
  - 568,000 customer reviews
- Focused on specific characteristics of each review
  - Text
  - Summary
  - Timestamp
- Dataset itself is large and has **174,779 duplicate rows**
- No null values present in dataset
- Reviews span multiple years and have multiple categories

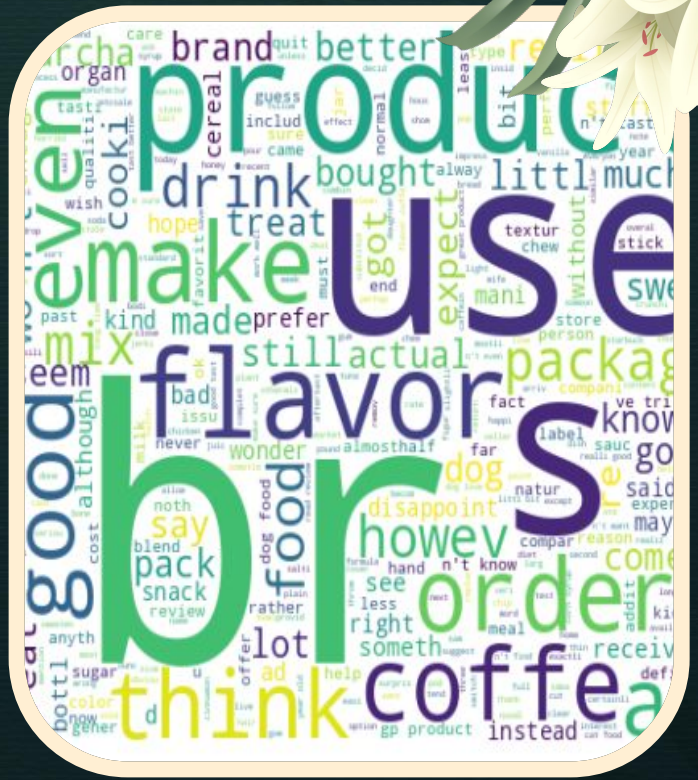
Model Training

- Used Pretrained Hugging Face Models
  - Twitter-roBERTa-base for Sentiment Analysis
    - Specialized version of roBERTa (Robustly Optimized BERT Pre Training Approach)
    - Trained on ~124M tweets (Jan 2018 - Dec 2021)
    - Self-supervised natural language processing (NLP)
    - Max ‘score’ of each label identifies each review’s ‘emotional tone’
    - Negative, Neutral, and Positive Sentiments

Exploratory Data Analysis

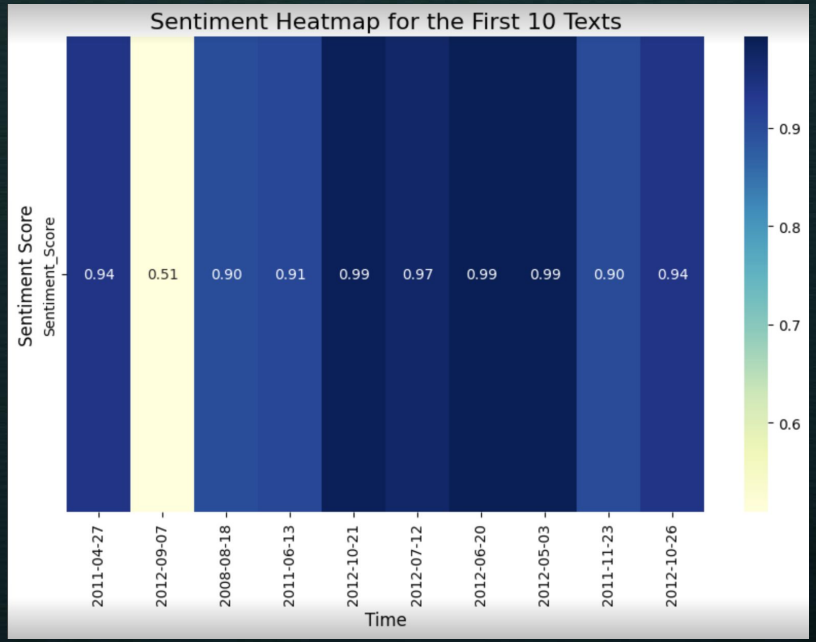
- Words like good, love, and recommend indicate favorable reviews, with a focus on product quality, taste, and flavor.
- Terms like disappoint, problem, and price suggest some concerns related to product issues and cost.

WordCloud



Sentiment Trends Over Time

- From the first 10 texts analyzed, there is no distinct pattern observed over time.
- Though, the sentiment scores are generally between 0.90 and 0.99, with a single anomaly.
- More texts have to be analyzed for more accurate analysis (a limitation on our end).



Real-Time Sentiment Analysis

- The cakesters melt in your mouth. I love these so much, that I am writing this review at 4:30am. Couldn't sleep, so needed a little cakester and milk.
- Score: 0.984004
- Sentiment: Positive

```
from transformers import pipeline

sentiment_pipeline = pipeline("sentiment-analysis")
```

This initializes a pre-trained Hugging Face sentiment analysis model.

```
review_text = "..."
```

Contains the review text to be analyzed by the model.

```
result = sentiment_pipeline(review_text)[0]
```

Feeds the review_text into the sentiment analysis pipeline, and extracts the first result (in case the pipeline is applied to multiple inputs).

```
print(f"Review: {review_text}")
print(f"Predicted Sentiment: {result['label']}")
print(f"Confidence: {result['score']:.2f}")
```

Displays the result (the predicted sentiment and confidence score)

Next Steps and Conclusion

Expanding Data for Visualizations

Due to limitations in loading the full dataset, our next steps will be to adjust our visualizations to incorporate a more comprehensive dataset, rather than focusing on the smaller sections we were initially able to load.

This project provides better insights into customer feedback, helping predict trends and improve business strategies and customer experience!

Thank You!

CREDITS: This presentation template was created by Slidesgo, including icons by Flaticon, and infographics & images by Freepik