Predicting Positivity and Negativity Trends on Amazon Food Reviews using Sentiment Analysis

Introduction:

In today's digital world, online reviews hold immense power. They not only shape customer decisions but also directly impact a brand's reputation. To achieve 100% customer satisfaction, understanding the sentiment behind these reviews is crucial. This project uses a rich dataset of Amazon food reviews, which encompasses a wide variety of products, user profiles, and most importantly, emotions. Our goal is to extract valuable insights from this data, identifying patterns and trends that show what customers truly feel about the food they're buying. Not only explore what people say, but also go deeper to understand why some reviews contradict the given star rating. By using sentiment analysis tools like VADER and Roberta, we'll uncover the emotional undercurrents of these reviews. This analysis will look at reviews with seemingly illogical sentiment, such as positive reviews with a single star rating. Ultimately, this project aims to use the power of customer reviews to provide businesses with actionable insights that can be used to improve both products and services.

Dataset:

This dataset consists of reviews of fine foods from amazon. Reviews include product and user information, ratings, and plain text reviews. Data includes reviews of more than 500,000. This allows us to analyze what people say about the food, how they rate it.

Description of Technologies used:

- **VADER:** Valence Aware Dictionary and Sentiment Reasoner is sentiment analysis tool used for to analyze the sentiment for text, which can be informal and full of slang. VADER assigns a sentiment score based on the positive, negative and neutral categories to text based on predefined word associations.
- **ROBERTA:** Robustly Optimized BERT Pre-training Approach is a NLP model that is used to perform sentiment analysis on the review texts, thereby improving the sentiment classification accuracy. This model is used to understand the relationships between words and their context. It works highly effectively for sentiment analysis.
- **Negative sentiment 5-star review:** It is a type of rating-text mismatch review when someone rates something with 5 stars, which usually indicates the highest satisfaction, but their comments or review express dissatisfaction or criticism. It's like saying "everything seems great, but there are some problems I want to point out."
- **Positive sentiment 1-star review:** It is a type rating-text mismatch review when someone gives the lowest rating (1 star) but expresses positive feelings or satisfaction in their comments. It might seem contradictory, but it could happen if the person appreciated certain aspects of the product or service while still finding significant flaws or issues that led them to give a low rating.

Recent Developments:

- I got to know the discrepancies between the numerical rating and the textual review provided by user for a productor service.
- There is some inconsistency between text reviews and score ratings for a product or service posted review.

1. Text-rating review discrepancy (TRRD): an integrative review and implications for research, this paper shows the necessity of considering both text reviews and score ratings to have reliable survey results.
2. A Study on Text-Score Disagreement in Online Reviews, this paper explains that the reviews ranked with middle scores include a mixture of positive and negative aspects.

Data Preprocessing:

This dataset has already undergone preprocessing, so we proceed directly to analysis.

Data Visualization:

Data visualization is a crucial part of this project as it helps to understand the patterns, trends and correlations in the data.

The distribution of scores given by user visualized using a histogram. This helps us to understand the overall sentiment of reviews.

The ratio of 'HelpfulnessNumerator' column to 'HelpfulnessDenominator' column is plotted to understand the distribution of helpful reviews .

This provides insights into how many reviews users found helpful.
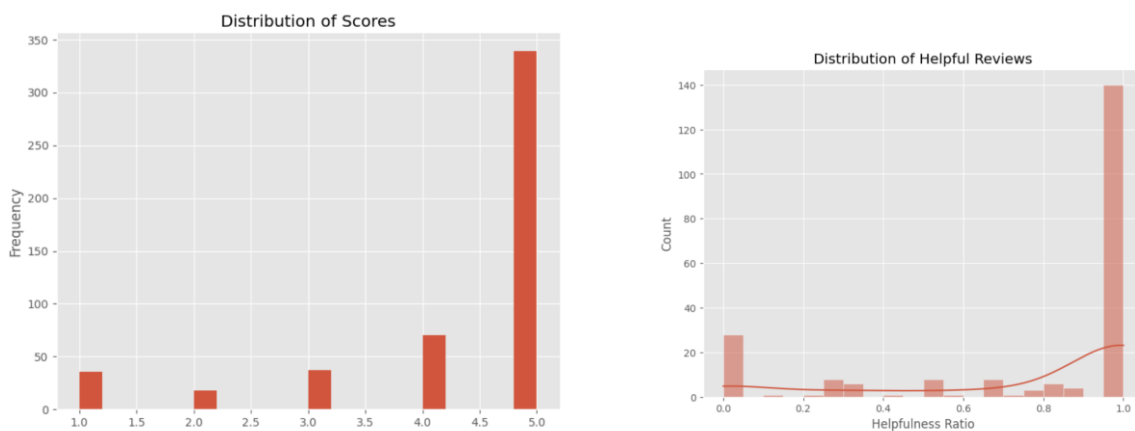
Ialso created word cloud to visually represent the positive words that set a positive tone of the review.

Methods:

1. **VADER:**

Valence Aware Dictionary and Sentiment Reasoner is tool that is specifically used to tune the sentiments expressed in social media. It is used in this project to analyze the sentiment of the text reviews.

Here VADER used a 'polarity_scores' method to calculate the sentiment scores of the text and this method has returned a dictionary with 4 items 'neg', 'neu', 'pos' and 'compound'. These represent proportions of negative, neutral, positive scores.

This is applied to the text column which contains reviews to calculate the sentiment scores for each review.

These sentiment scores are visualized to understand the sentiment of the reviews and how they connected with scores given the users.


VADER Compound Score by Amazon Star Review
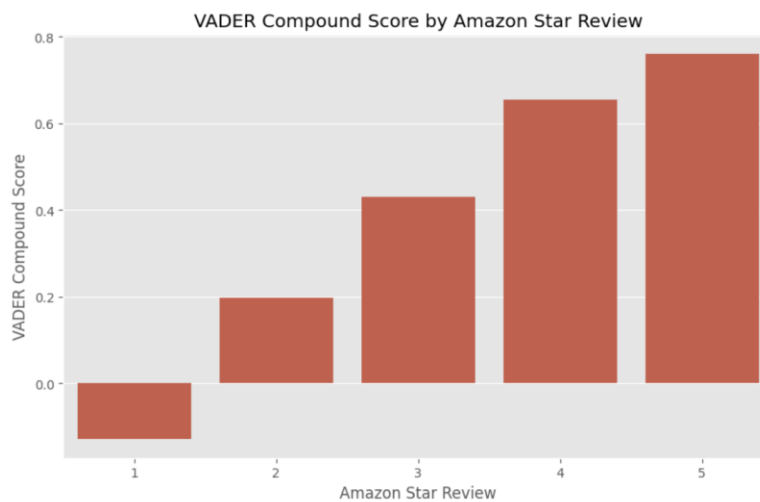
2. **ROBERTA**

Robustly Optimized BERT Pretraining Approach is a variant of Bidirectional Encoder Representations from Transformers which is used to provide more robust and optimized performance.

ROBERTA understands the context of words in the sentence.

Here it reads the reviews written by the users and returns a prediction for the sentiment of the text.

It determines if the review is positive, negative and neutral. The sentiment scores are visualized using pair plots to understand the sentiment of the revies and how they correlate with the scores given by users.

ROBERTA Positive Score — ROBERTA Neutral Score — ROBERTA Negative Score
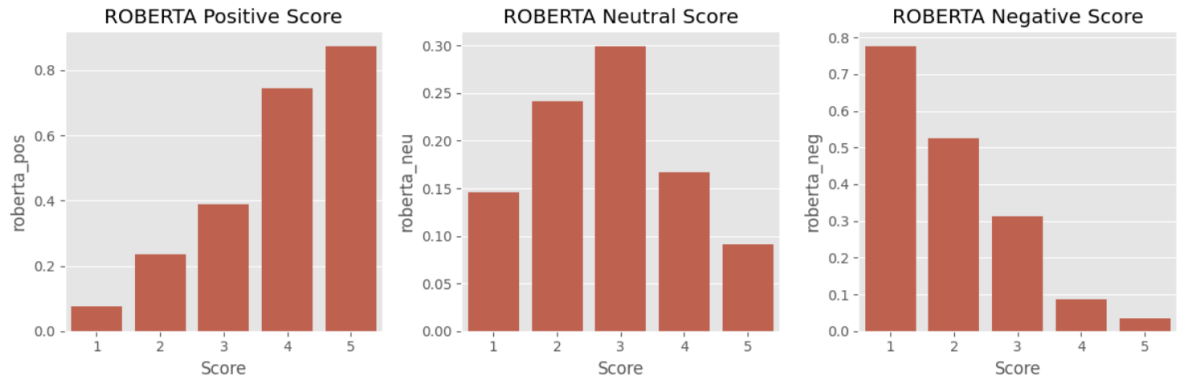
3. **Rating-Text Mismatch Reviews:**

   This is a situation where the sentiment expressed in the review does not match the number stars given by the user.

   Here both VADER and ROBERTA scores are used to analyze the review text and calculate sentiment scores.

   These scores compared with the user-given star ratings to identify potential rating-text mismatches.

```
VADER sentiment star-1 review:
Serveice delivery with the seller was excellent. The product was not. Will not order again. Bad taste. Iam not sure if it was o
ld or that how it tastes. I throw it right away, i coudl not drink it.
```

```
VADER sentiment star-5 review:
I drive OTR...Over the Road truck and this helps to keep me alert. It has no sugar high & no crash. I'm on a autoship which i l
ike. The price is very good. I've seen it over 2 bucks in the store.
```

Model Evaluation:

Here we have compared the sentiment analysis results from VADER and ROBERTA models using metrics like accuracy, precision, recall and F1 score.

```
VADER Model Metrics:
Accuracy: 0.85
Precision: 0.88
Recall: 0.94
F1-score: 0.91

ROBERTA Model Metrics:
Accuracy: 0.91
Precision: 0.93
Recall: 0.95
F1-score: 0.94
```

Based on these scores we can conclude that:

- ROBERTA model outperforms VADER in terms of accuracy, precision, recall and F1 score, this tells that it is more effective in sentiment analysis for this dataset.
- The high precision and recall scores for both models show that they have good performance in classifying positive and negative sentiments.

Conclusion:

This project successfully implemented sentiment analysis on the dataset by using VADER and ROBERTA models. While ROBERTA model overperformed VADER, both models were able to calculate sentiment scores for reviews. This project shows the potential use of sentiment analysis and data visualization to gain insights from the user reviews.

References:

- [Amazon Fine Food Reviews (kaggle.com)](#)
- [2004.13851.pdf (arxiv.org)](#)
- [Topic Modeling using Gensim-LDA in Python | by Aravind CR | Analytics Vidhya | Medium](#)
- [Local Consumer Review Survey 2024: Trends, Behaviors, and Platforms Explored - BrightLocal](#)
- [Learning Distributed Representations from Reviews for Collaborative Filtering | Proceedings of the 9th ACM Conference on Recommender Systems](#)
- [Text-rating review discrepancy (TRRD): an integrative review and implications for research | Future Business Journal | Full Text (springeropen.com)](#)