Sentiment Analysis Report

Dataset Description

The dataset used in this analysis consists of Amazon product reviews. It contains text reviews provided by customers for various products available on Amazon. Each review includes the text of the review itself, along with other metadata such as ID, product name, reviewer username, ratings, and timestamps.

Preprocessing Steps

Several preprocessing steps were applied to the review text before sentiment analysis:

- Text Cleaning:
    - Removing stop words (common words like "the", "a", "an") that don't contribute to sentiment.
    - Removing non-alphabetic tokens (punctuation, emojis).
    - Lemmatization: Converting words to their base form (e.g., "loved" becomes "love").
- Handling Missing Values: Rows with missing review text are dropped

Sentiment Analysis Methods

TextBlob is a rule-based approach utilizes a sentiment lexicon to assign polarity scores to the text, ranging from -1 (negative) to 1 (positive). Based on the score, the sentiment is classified as positive, neutral, or negative.

The commented out section at the end of the code which explains an alternative sentiment analysis method uses spaCy Transformers. This uses pre-trained transformer models to perform sentiment analysis. It is potentially more accurate compared to TextBlob. However, it is also more computationally expensive and slower to process large datasets, which is why it was not included in the main code.

Evaluation of Results

The distribution of sentiment labels (positive, neutral, negative) provides insights into the overall sentiment of the reviews. This distribution helps understand customer sentiment towards the products. Analysis using the TextBlob model of 100 of the reviews shows significantly more positive reviews compared to negative or neutral.

Models' Strengths and Limitations

Strengths:

- Efficiency: The TextBlob sentiment analysis model is efficient and easy to use. It provides a quick way to analyse sentiment in text data without extensive configuration or training.
- Ease of Implementation: The TextBlob sentiment analysis model does not require extensive preprocessing or parameter tuning, making it suitable for rapid prototyping and experimentation.

Limitations:

- Limited Context Understanding: TextBlob sentiment analysis relies on simple heuristics and does not capture nuanced contextual information present in reviews. It may struggle to accurately classify sentiment in complex or ambiguous reviews.
- Dependency on Preprocessing: The performance of TextBlob sentiment analysis heavily depends on the quality of text preprocessing. Inaccurate or incomplete preprocessing can lead to biased or incorrect sentiment analysis results.
- Limited Customization: TextBlob sentiment analysis does not offer extensive customization options or support for domain-specific sentiment lexicons. It may not be suitable for specialized or domain-specific sentiment analysis tasks.

Conclusion

Overall, the TextBlob sentiment analysis model provides a quick and straightforward approach to analyse sentiment in Amazon product reviews. While it offers simplicity and ease of implementation, it has limitations in capturing contextual nuances and may not provide accurate results in all scenarios. The spaCy transformer-based model may provide an alternative higher accuracy method but is also more computationally expensive and slower to process large datasets.