Sentiment Analysis Report: Amazon Consumer Reviews of Products

Description of the Dataset:

The data used was from Kaggle, that was a sample dataset regarding consumer reviews of Amazon products like the Kindle, Fire TV Stick, batteries and more provided by Datafiniti. The data lists reviews of over 34,000 entries that cover various sentiments regarding the products. The dataset includes basic information regarding the product along with ratings, review text, dates, and more that can be used for analysis.

Preprocessing Steps:

The dataset from Kaggle has been cleaned preprocessed to ensure that analysis is more accurate. The dataset was loaded using pandas and had any missing values within the 'reviews_text' column dropped using dropna() to filter incomplete data. Each review within the dataset was converted to a lowercase string using str() with lower() and stripped of any extra whitespace with strip(). Now that the dataset was more standardized, spacy was used to tokenize the text along with using .is_stop attribute to filter out common words that don't affect sentiment. The last step to reduce noise was to join the remaining tokens into a string allowing the model to focus on key word association.

Evaluation of Results:

I reviewed the model on the first 5 reviews of the dataset and was able to confirm that it is accurately evaluating the reviews. Currently each review was accurately placed based off the polarity score and the tone of the overall review matching. Review 2 is one to deep dive into based on the context of the review, though deemed negative, it is regarding buying items in bulk and may not accurately describe the product review. The model appears to be able to correctly identify sentiment but may need further refinement.

Testing sentiment on sample reviews:

- Review 1: I order 3 of them and one of the item is bad quality. Is missing backup spring so I have to put a pcs of aluminum to make the battery work.
    - Sentiment: Negative - Polarity Score: -0.4500
- Review 2: Bulk is always the less expensive way to go for products like these
    - Sentiment: Negative - Polarity Score: -0.3333
- Review 3: Well they are not Duracell but for the price i am happy.
    - Sentiment: Positive - Polarity Score: 0.8000

- Review 4: Seem to work as well as name brand batteries at a much better price
    - Sentiment: Positive - Polarity Score: 0.5000
- Review 5: These batteries are very long lasting the price is great.
    - Sentiment: Positive - Polarity Score: 0.2450

Models Strengths and Limitations:

Strengths:

- Model is simple and will run quickly on datasets without heavy computing
- Uses spacy and TextBlob for robust preprocessing and easy polarity scoring
- Handles basic Sentiment Analysis well

Limitations:

- The removal of stop words may impact the models ability to apply context on short reviews
- The model does not handle slang or special characters
- TextBlob is a rules-based and may not interpret things such as tone or sarcasm within reviews