

Sentiment Analysis Report

5.1 Description of the Dataset Used:

The dataset used for sentiment analysis consists of 28,322 rows of Amazon customer reviews. Each row contains data regarding a single customer review, including the review text, product information, and other relevant details. The data is stored in a CSV file format, making it easy to access and analyze using Python libraries such as pandas.

5.2 Details of the Processing Steps:

The processing steps for sentiment analysis involved several key components:

- Data loading: The dataset was loaded into memory using the pandas library, allowing for easy manipulation and analysis.
- Text preprocessing: Text preprocessing techniques, such as tokenization and stop-word removal, were applied to clean the review text and prepare it for analysis.
- Sentiment analysis: Sentiment analysis was performed using two different approaches:
 - TextBlob library: The TextBlob library was used to calculate the polarity of each review text, providing a measure of sentiment ranging from -1 (negative) to 1 (positive).
 - spaCy library: The spaCy library was used to perform sentiment analysis based on pre-trained models, assigning each review a sentiment label (positive or negative) based on the sentiment score calculated for the text.

5.3 Evaluation of Results:

The results obtained from the sentiment analysis were found to be accurate based on the evaluation conducted during the analysis process. The accuracy of the sentiment predictions was verified by

comparing the predicted sentiment labels with the actual sentiments expressed in the customer reviews. Additionally, the results were validated using exploratory data analysis techniques to identify patterns and trends in the sentiment distributions across different product categories or time periods.

5.4 Insights into the Model's Strengths and Limitations:

The sentiment analysis model implemented using spaCy and TextBlob demonstrated several strengths, including:

- Accuracy: The model achieved high accuracy in predicting sentiment labels for customer reviews.
- Efficiency: The model's computational efficiency allowed for fast and scalable sentiment analysis of large datasets.
- Flexibility: The model's flexibility enabled customization and fine-tuning to suit specific requirements or domains.

However, the model also has some limitations, such as:

- Dependency on pre-trained models: The accuracy of the sentiment analysis may be affected by the quality and relevance of the pre-trained models used by spaCy.
- Sensitivity to context: The model's performance may vary depending on the context and language used in the customer reviews, leading to potential biases or inaccuracies in sentiment predictions.