**T21 - Capstone Project – NLP Applications – Sentiment Analysis Report**

**5.1. Description of the Dataset Used**

The dataset used for this sentiment analysis includes a wide range of information related to Amazon product reviews, such as identifiers (ID, name, ASIN), product attributes (brand, categories, keys, manufacturer) and detailed review information (date, purchase status, recommendation status, review ID, number of helpful votes, rating, source URLs, reviews text, review title, and reviewer's location and username).

**5.2. Details of the Preprocessing Steps**

The preprocessing focuses on the reviews.text column, cleaning the text for sentiment analysis. The primary steps include removing non-alphabetic characters to focus solely on words, lowercasing to ensure uniformity, and eliminating stopwords and punctuation to reduce noise, so as to focus on the meaningful content of the reviews. This approach streamlines the text, preparing it for accurate sentiment analysis.

**5.3. Evaluation of Results**

The sentiment analysis shows an incorrect classification of the first review as negative, although the actual review is quite positive. This shows that the model might not always understand when people are saying good things in a less direct way. The person that submitted this review uses words that would also imply a negative sentiment and thus, the model fails to interpret the review in a nuanced way of expression of satisfaction. However, the model accurately identified the positive sentiments of the second and third reviews.

**5.4. Insights into the Model's Strengths and Limitations**

Strengths: The model performs well at understanding clearly positive or clearly negative text.

Limitations: The model may not understand fully the context or the complexity of human sentiment. The mistake with the first review shows it can miss the mark with longer or less straightforward positive comments. The reliance on cleaned text might also oversimplify the analysis, potentially overlooking subtleties conveyed in the original reviews.