T26 - Natural Language Processing with SpaCy

**Practical Task 1**

To plan for this script, I have broken down the required steps necessary below.

Task plan breakdown:

1. **Download the dataset of product reviews.**
   1. The dataset is essentially a csv file and we are mainly interested in the review strings in the review.txt column.
2. **Implementing a sentiment analysis model using spaCytextblob.**
   1. We will use spacytextblob to check each review to determine its polarity. Polarity of 1 indicates a very positive sentiment, polarity of -1 indicates a very negative sentiment whereas a polarity of 0 indicates a neutral sentiment.
3. **Preprocess the text data**
   1. Dropping the NA values.
   2. Extracting the review text per row.
   3. Removing the stop words leaving only the more important words for sentiment analysis.
4. **Creating a dataframe for sentiment analysis.**
   1. Creating a data frame with 3 columns (reviews, sentiment, sentiment score)
   2. Running the scikit packages to make sentiment predictions on the product reviews.

Model output based after training and testing the data:

A screenshot of a computer screen

Description automatically generated

Precision is generally quite high for negative, neutral and positive reviews.

The model has a perfect recall for positive reviews but scores quite low for neutral reviews and also quite poorly on the negative reviews.