# Sentiment Analysis Report

## Task 21: Capstone Project - NLP Applications

E. Thompson
ET23090010130

## Description of the dataset

The dataset used for this sentiment analysis was the 'Consumer Reviews of Amazon Products' dataset from Kaggle, '1429_1.csv'. The dataset has 34660 rows and 21 columns. It contains basic product and review information, including the rating and text of the reviews, for Amazon products such as the Fire Tablet, Echo, Kindle and Fire TV Stick. The data originates from Datafiniti's Product Database, and is a sample from a large dataset.

## Preprocessing steps

To preprocess this data, firstly only the required columns were selected: the text of the reviews, found in reviews.text, and the ratings found in the reviews.ratings column. These columns contained a small number of null values. The 34 rows with null values were dropped, resulting in 34626 rows.

100 rows from the dataset were sampled and preprocessed for analysis. This number was chosen due to the high ratio of positive reviews to negative and neutral. A sample size of 100 was able to capture a small number of negative and neutral reviews in order to better evaluate the performance of the model.

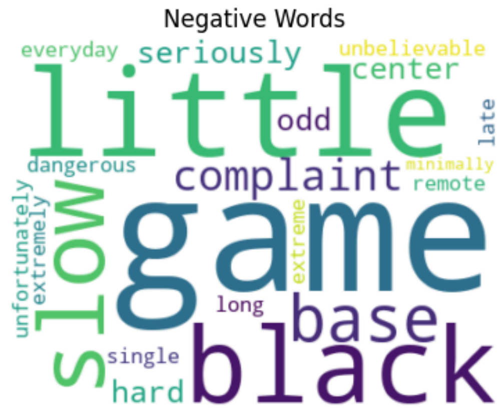
The review text in the sample was preprocessed using the 'preprocess' function. This function first converts each review into a SpaCy Doc object. Each Token in the Doc object which isn't a stop word or punctuation is lemmatized and made lower case. The if statement strips out both stop words and punctuation. The remaining words are rejoined into a single string and returned by the function.

## Results

From the sample, two reviews that had been identified as positive, two as negative and two as neutral were examined:

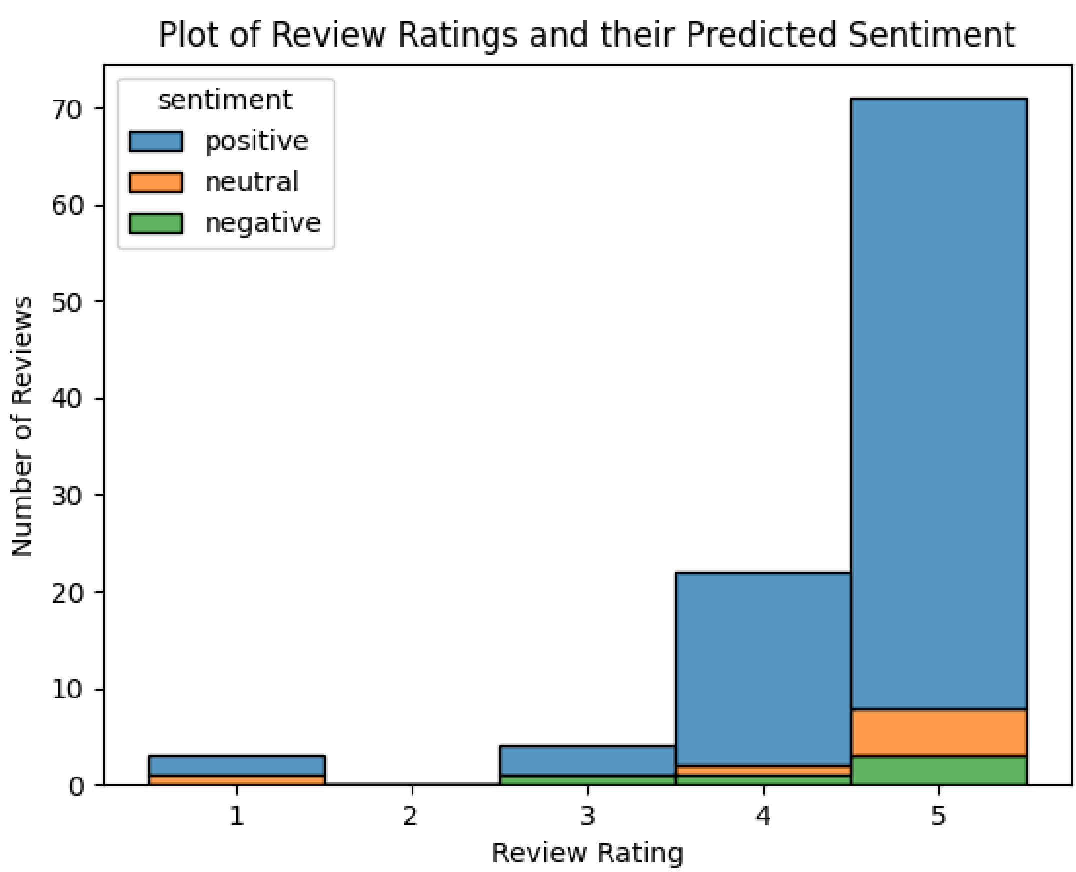
- The two 'positive' reviews were correctly identified. The sentiment strength was fairly low, at 0.32 and 0.45, considering how enthusiastic both reviews read to a human.
- The two reviews identified as being neutral were incorrect, as they both read positively and were given a rating of 5.0 by the reviewers.
- The two 'negative' reviews were also incorrect. They had reviewer ratings of 5.0 and 4.0 respectively, the first was slightly positive sounding, the second neutral in tone. The first only had a very slight negative polarity, of -0.03 whilst the second was given a stronger negative sentiment, of -0.35. The difference between the two was as expected.

A word cloud was generated from the main sample (100 reviews) to show the sentiment predicted for words in the reviews:


The greater number of positive to negative reviews is reflected in both the quantity and quality of words shown. The largest positive words are clearly positive in nature, whereas the largest negative words are not obvious. The word ‘game’ being perceived as negative could skew many reviews as the products allow games to be played and may be mentioned in this context, rather than ‘to game’ a system, for example.

A stacked bar chart was generated to show the distribution of the reviews and how well the model performed, using the review rating as a guide to the actual sentiment behind the review. This is only a rough guide, as reviewers will not always use the rating system correctly, or write a review which matches the rating they have given.



The reviews with a rating of 1.0 were examined to see how well the model had identified their sentiment, as these would be expected to be negative. Of the three, the review identified as ‘neutral’ was actually negative. One of the ‘positive’ reviews, with a polarity of 0.6, was negative. The other was correctly identified as being positive, with a polarity of 0.45. For the latter, it would appear that the reviewer probably had not intended to give a low rating.

The three reviews with a 5.0 rating which the model had identified as negative were also examined. All of these read positively, although not very strongly so. The polarity shows that the sentiment strength was only very slightly negative; -0.03, -0.08 and -0.04.

The similarity found between “This thing looks nice, runs nice and feels nice and for the price I don't think you can find a better tablet.” and “Great tablet to teach young children . Grandchildren love it to watch favorite you tube shows”, was 0.8. This high similarity score makes sense, as these reviews are both positive reviews about tablets.

## Evaluation of the model

The model correctly identifies the majority of positive reviews. It appears to have more trouble in identifying negative and neutral reviews.

This is reflected in the words shown in the word cloud - the positive words are very obviously positive, whereas the negative words identified are a mixture of very apparent words such as ‘slow’, ‘dangerous’ and ‘hard’, and neutral words such as ‘little’, ‘game’ and ‘black’. Some of these words can be negative in some contexts but are unlikely to be used in a negative manner in these reviews.

One of the two incorrect neutral reviews returned uses the phrase “I would have been a fool to buy it” of another product. This could potentially be the cause of the confusion, as the model may not have been able to identify that this referred to a different product.

Another source of confusion for the model may be in the removal of stop words during preprocessing. If words such as ‘can’t’, ‘not’ and ‘didn’t’ are removed, it changes the meaning of a phrase to the opposite. This could explain “Bought this during the black friday sale for $35 and can't complain at all. Keeps my kids busy, and runs hearthstone, that's all I need.” being incorrectly identified as negative, for example. An improvement to the model could be to retain these words and enable it to understand how these words switch the sentiment.

A particularly long negative review that the model identified as positive mentioned a retailer called ‘Best Buy’ several times, which could potentially have caused the polarity to have come back as very positive (0.6) rather than negative. A possible improvement could be to accurately identify and remove brand names from the reviews before conducting sentiment analysis.

Overall, it works well in identifying most of the positive reviews correctly. There seems to be a bit of bias towards negativity, as the positive samples did not score as highly as might be expected, and some that read as neutral/slightly positive came back as slightly negative. However, it also misidentified a negative review as neutral. It would be interesting to see how well it would perform given only low rated reviews, as the sample contained few for the model to be tested on. A larger SpaCy model could be tried to see if this improves the performance.