Summary

Description of the dataset -

This dataset contained a list of over 28000 consumer reviews for Amazon products like the Kindle, Fire TV Stick, Batteries and more provided by Datafiniti's Product Database. The dataset included the following information for every review - 'id' 'dateAdded' 'dateUpdated' 'name' 'asins' 'brand' 'categories', 'primaryCategories', 'imageURLs' 'keys' 'manufacturer','manufacturerNumber', 'reviews.date' ,'reviews.dateSeen', 'reviews.didPurchase', 'reviews.doRecommend', 'reviews.id', 'reviews.numHelpful', 'reviews.rating', 'reviews.sourceURLs', 'reviews.text', 'reviews.title' ,'reviews.username' ,'sourceURLs'.

Preprocessing Steps –

The Preprocessing steps included - loading the data, dropping null values if any, and lastly using the .strip() and .lower() methods to remove any leading or trailing whitespaces and normalizing everything into small caps.

Evaluation of Results –

The results showed that the model seems to be working decently well. As outlined inside the code file, the negative, neutral and positive reviews were assigned the right polarity scores.

Strengths and Limitations –

As mentioned above, the model seems to be able to identify neutral, negative and positive sentiments decently well. So it can be said that the model can be used for general classification.

There is lots of room for improvement though. Especially in terms of the model's accuracy. The positive sentiment sample in our code file can be said to very positive or extremely positive so we would expect a score quite close to 1. But the model's output was only 0.45, positive but still not ideal.