

Sentiment analysis report

This report is made to support my task from document 'Natural Language Processing with SpaCy' found in Dropbox: [here](#)

A description of the dataset used.

The data set I have used for this project is the 'amazon_product_reviews' from Kaggle: <https://www.kaggle.com/datasets/datafiniti/consumer-reviews-of-amazon-products>

This data contains a list of over 32,000 consumer reviews for Amazon products provided by Datafiniti's Product Database. This dataset is just a sample of the original dataset

Details of the preprocessing steps.

The data from this dataset was processed by removing the null values from the 'review.text' column which was used for this task. We also used a sample text to test the functions created for the preprocessing.

Once the null values were removed, I removed the stop words and anything else that's not an alphabet and then converting that into string as saving it into another column for comparison and understanding if the code have done its job correctly.

Evaluation of results.

After processing we were able to see that the model is working well where it's able to display the polarity of the customer review if it was Positive, Negative or neither of these. Giving us a sense of understanding of users who are pleased or displeased of the product.

We have also compared the values of 2 different comments from the data frame to get an understanding of the similarities of the customer's review

Insights into the model's strengths and limitations.

I have done a very basic preprocessing of the data, and it will really benefit by usage of Regex to remove very specific signs, and punctuations/errors from the reviews and using lemmatisation to get the root form of the words.

This will help in collating similar reviews and more complex data handling that could help in further processing the data for understanding trends, and customer requirements with respect to the current model of data.

Though I am sure it can be used similarly on other data sets as well to come to different conclusions.