**Technical Report on Sentimental Analysis of Amazon Reviews**

OVERVIEW

1. Project Background and Description

Sentiment analysis refers to the use of natural language processing, text analysis, computational linguistics, and biometrics to systematically identify, extract, quantify, and study affective states and subjective information.

[Sentiment analysis](https://www.paralleldots.com/sentiment-analysis" \t "https://towardsdatascience.com/_blank)is contextual mining of text which identifies and extracts subjective information in source material, and helping a business to understand the social sentiment of their brand, product or service while monitoring online conversations, like Codepth has prepared a project for Amazon and customer reviews on it, to basically understand the market reaction and the ‘sentiments’ of the customers and buyers. The project can very well be used in areas which have a direct relation with customer reviews and for the companies that considers them to be of some significance.

1. Project Scope

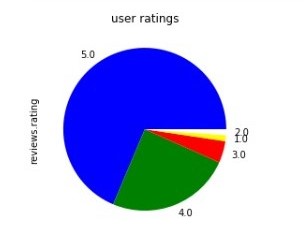
Sentiment analysis of product reviews, an application problem, has recently become very popular in text mining and computational linguistics research. Here, we want to study the correlation between the Amazon product reviews and the rating of the products given by the customers. We use both traditional machine learning algorithms including Naive Bayes analysis, Support Vector Machines, Knearest neighbor method and deep neural networks such as Recurrent Neural Network(RNN), Recurrent Neural Network(RNN). By comparing these results, we could get a better understanding of the these algorithms. They could also act as a supplement to other fraud scoring detection methods.

In this project, we investigated if the sentiment analysis techniques are also feasible for application on product reviews form Amazon.com.

Within the study, different machine learning algorithms are compared, trained and tested on a dataset containing product reviews from Amazon.com which are randomly selected from dataset available from Kaggle containing multiple reviews.

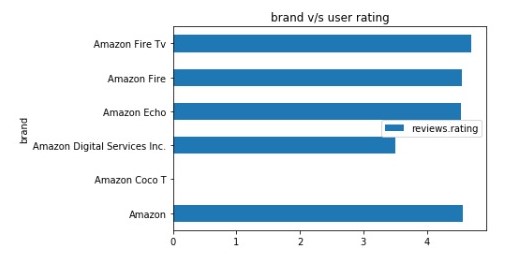
This resulted in a very accurate classification, with the best results for reviews on furniture products (Accuracy = 0.92). In conclusion, LSTM networks are very suitable for classification of the sentiment on product reviews and the results do not change significantly for different categories.

1. Graphical Representations used in the analysis and their significance



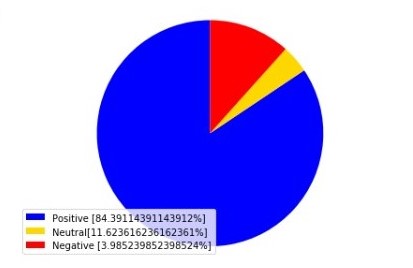
The first graph in the analysis project is the “*user ratings”* **pie chart**, which basically represents the reviews given by the different users/customers/buyers to the particular products of Amazon (the brand) from different categories, also giving their reviews (in title form and in the text form) about the same.

The reviews can be anything that the customer chooses from a range of 1 to 5. The pie chart divides the ratings given by the different customers in proper proportions, carefully focussing on products which the customers feel are better and high-rated.



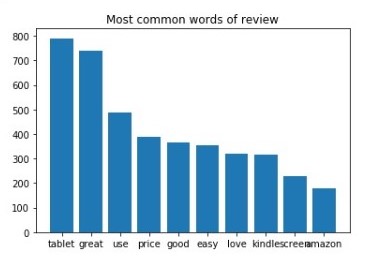
The second graph is the *“brand v/s user rating”* **horizontal bar graph**. Amazon has divided itself into different ‘sub-brands’ like Amazon Fire TV, Fire, Echo, Digital Services Inc. and Coco TV and the products which are covered under them are rated by the customers/users/buyers and then comapred in the bar graph as to see which sub-brand’s products have got the highest or lowest review ratings.

This shows how the different services provided by a particular brand (in this case, Amazon) are compared to each other to find out which one of those is being most liked by the customers (judging by their review ratings) and which one still needs to upscale their sales.



The third graphical representation is the *“review analysis”* **pie-chart** that compares the percentage of the three different types of reviews, that are the- positive, neutral and negative review ratings given by the customers/users/buyers, that are found out after a proper review analyis, using stop words. In the legend in the chart, the percentages of the number of those particularly characterised review ratings are written and are divided into 3 parts very clearly with the use of three different colours; blue (positive reviews), gold (neutral reviews) and red (negative reviews).

This clearly defines how are the current sales and ratings of these particular products of the brand and is the direction that the brand is progressing in is upto the mark or not.



The fourth graph is the **vertical bar graph** for *“Most common words of reviews”.* This graph shows us which ones are the most commonly used words in the textual reviews given by the customers/users/buyers as a whole so that the company can focus more on those words and understand the overall picture of what the customers like as a whole about the brand and it’s multiple products.

The graph also shows the frequency of the used words; the number of times these particular words are used in all of the reviews, which will help the company to focus on the ‘key’ points of their brand.