

**Sentiment Analysis of Amazon Cellphone reviews**

**TEAM RSS**

*Rahul Sharma rks5737@psu.edu*

*Shreyesh Doppalapudi* [*szd5630@psu.edu*](mailto:szd5630@psu.edu)

*Shivang Suchak svs6946@psu.edu*

**School of Graduate Professional Studies**

DAAN 862– Analytical Programming with Python

December, 2019

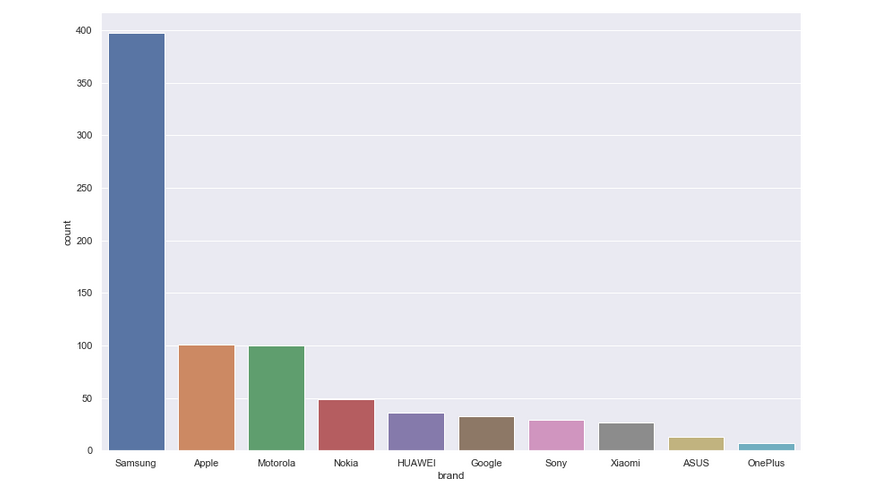
**Abstract**

Predicting the type of sentiment associated with the reviews provided by the users on a cell phone through Amazon website. Multinomial Naive Bayes model was trained for the prediction as the scenario involved text classification. This prediction would be useful for a new user who is trying to look for specific things before buying a new cell phone online by automatically identifying the prominent emotion of the review.

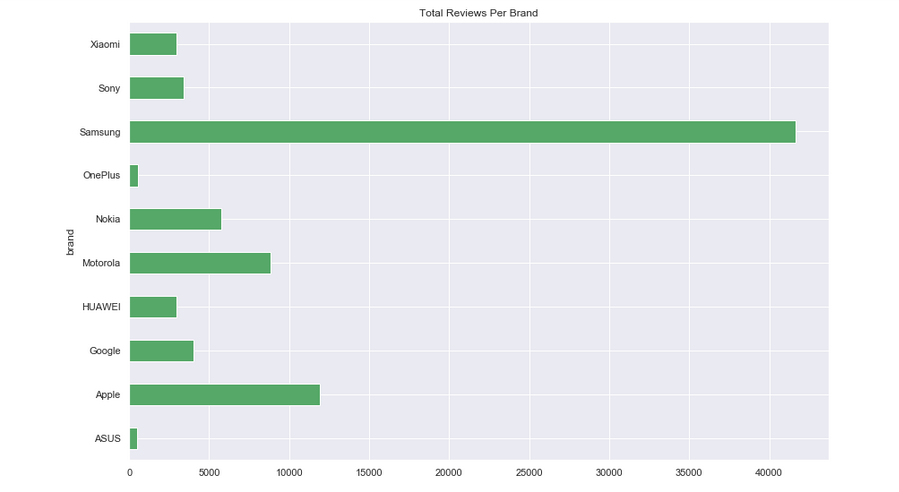
**Introduction**Whenever a user tries to find a reliable review before buying a cellphone online, it becomes very confusing as the number of reviews of all phones by famous brands would be very large in number. It becomes cumbersome to go through each review; this model becomes very useful as people can filter out which type and brand of phone they want/do not want to use.

**Source of Data, Characteristics of Data**

* The data was inspired from Kaggle.com which describes the reviews of user who bought cellphone via Amazon. Link: [*https://www.kaggle.com/grikomsn/amazon-cell-phones-reviews*](https://www.kaggle.com/grikomsn/amazon-cell-phones-reviews)
* The data included labels like Brand, Reviews, Rating, Image, URL, etc. with sample size of around 82,000. There were two datasets namely items and reviews
* There was a total of 10 brands, with a total of 792 items.

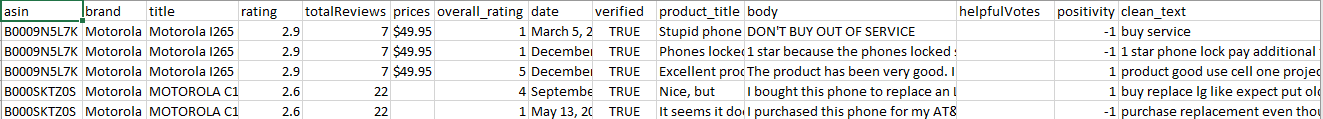


* Next, we considered the average rating per brand as well as total number of reviews brand wise.

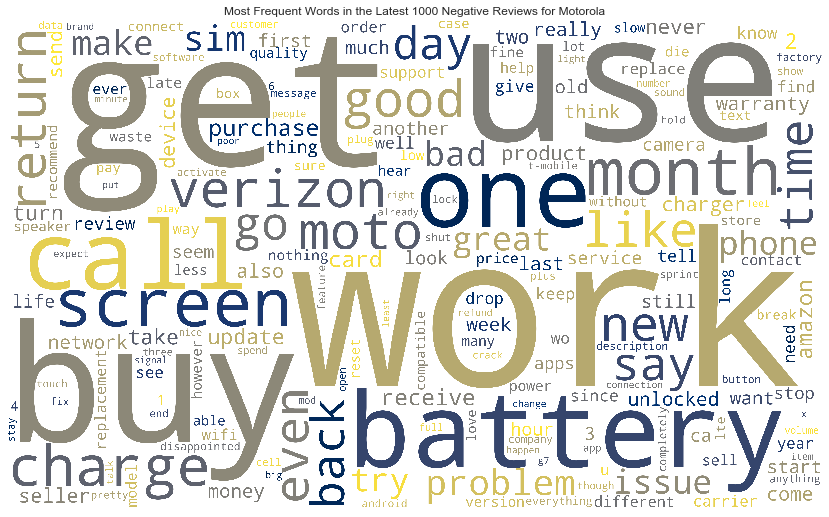
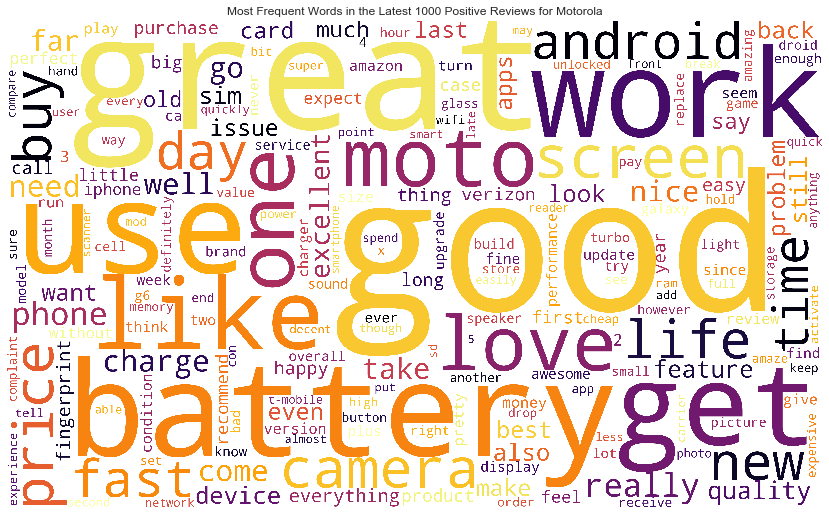


**Data cleaning and Preprocessing**

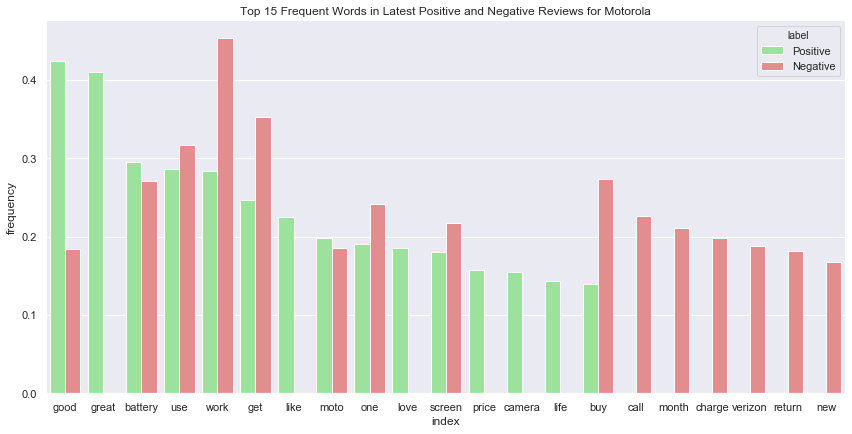
* First, the items were grouped **Brand** wise. Next, the items dataset and reviews dataset were merged.
* Dropped unnecessary columns which were assumed to not influence the sentiments of the reviews. Those columns were: url, image, reviewUrl, name
* Only **verified** users were considered in order to make the data more reliable.
* Product reviews text were labelled in the form of positive, negative, and neutral using the rating as a basis (Negative- Rating < 3, Neutral- Rating = 3, Positive- Rating > 3)
* The body of the review was taken, vectorized at a word-level and then, lemmatized to extract the root word from the actual word based on the Part-of-Speech (POS) tagging
* A sample of the dataset after it has been pre-processed is a s below:



* Word cloud was created for 3 brands namely Motorola, Samsung, Nokia to better understand the structure underlying sentiment of the reviews
* Positive words, negative words as well the most frequently used words of these brands were taken into consideration.
* Top words in Positive and Negative reviews of Brand Motorola

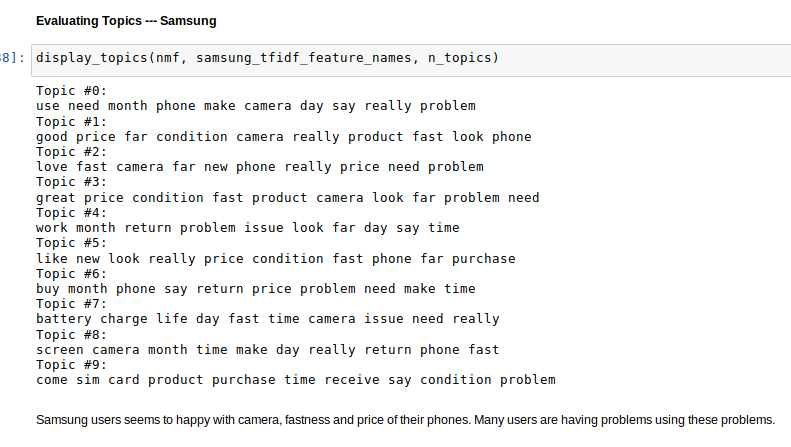


* + Similarly, the Top words in positive and negative reviews as well as Top 15 frequent words



**Topic Modelling**

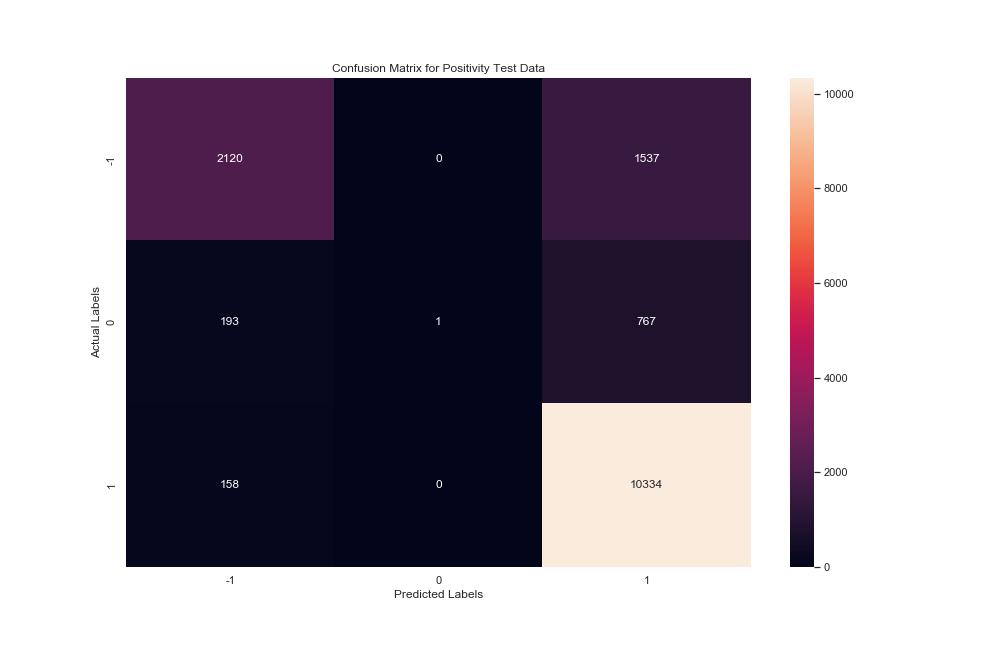
* In order to reflect how important a word is to a document in a collection or corpus, term frequency–inverse document frequency, TF- IDF vectorization was used. The (Non-negative Matrix Factorization) NFM was used.



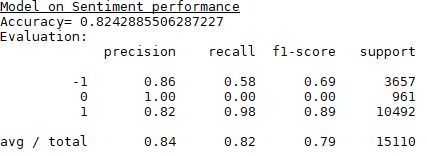
**Multinomial Naïve Bayes**

Similarly, the evaluation topics were carried out for other brands to get the insight on which set of words were eminent in the review for that specific Brand.

* As part of the next step, two Multinomial Naïve Bayes models were built to predict the sentiment(positivity) and the rating of the review using the body of the review as a predictor.
* The dataset was spilt into training and test sets in the ration of 80:20.
* TF-IDF matrix was created for both training and test sets as described in the steps before and the training set was fed as an input to both the models.
* The performance of both the models on the test set was compared using the classification metrics of accuracy, precision, recall and F-1 score.
* **Results and Discussion:**
  + - The results of the model on sentiment prediction are given as below:

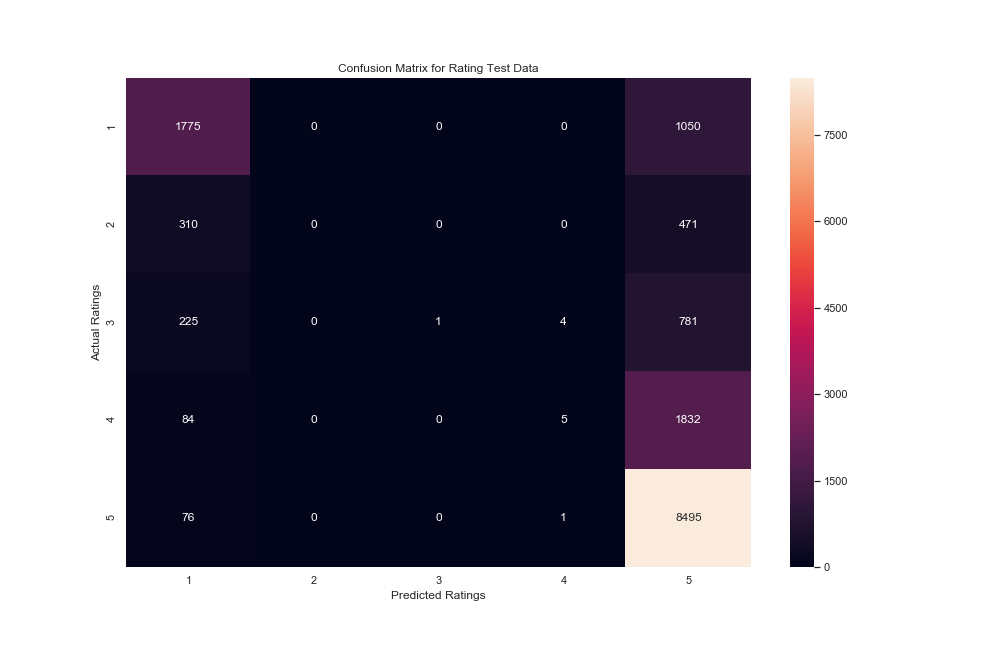


* + - The classification matrix of sentiment analysis is shown below:

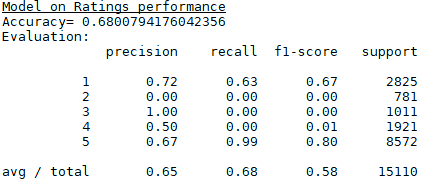
****

**\* *1, 0, -1 represent Positive, Neutral and Negative Sentiments respectively***

* As can be seen from the confusion matrix and classification above, the model has a good performance on positive and negative sentiment text, but the performance is low on the neutral sentiment text. The model is finding it difficult to differentiate the neutral sentiment from the positive and negative sentiment, which affects the precision of both the positive and negative sentiments classification.
* In addition, the recall on negative sentiment is on the lower side, which means the model could not differentiate the negative sentiment from the positive sentiment in a few cases. The inherent negation in sentences, which is difficult to identify, might be a reason for the low performance in these cases.
  + - The results of the model on rating prediction are given as below:



* + - The classification matrix of rating performance is shown below:



* As with the Sentiment model, the performance of the Ratings is low on the ratings in the middle (2, 3, 4) while it is better on the extreme ratings (1, 5). Again, the model is not able to differentiate between what text constitutes an extreme rating and that for an intermediate rating.
* The recall on the negative extreme rating is also low. The model is not able to leverage the inherent negation in text related to a negative review which is like the sentiment model.

**Conclusion**

* There are two questions that need to be answered in terms of the motivation behind the project
* **Question 1:** Why did you choose this project?
  + **Answer:** One of our teammates recently purchased a Cell phone via Amazon and he had to go through a lot of trouble while scrolling reviews as he didn’t know how reliable and how many reviews would be enough to validate the product, this experience was one of the main motives behind building this model.
* **Question 2:** What is the social/commercial/scientific benefit of your solution?
  + **Answer:** The commercial aspect would be that the product would become more reliable as the reviews would be from verified people and the emotions would already be known by the user. The social aspect would be to help the user understand the product as well as other people’s opinions. The scientific aspect is to understand the process of automatic text classification and to apply it to a real-world problem.
* Both the models performed well in the identification of positive sentiments and ratings, while it could not identify intermediate sentiment or ratings. The performance on negative sentiment and ratings was affected by the inherent negation on the text.
* The identification of negation should be a priority for the improving the model performance. Once identified, the root word could be updated to reflect the negation of the word to improve the identification of negation for the model, thereby improving its performance.
* The usage of more advanced machine learning models (Deep Learning models such as ANN, RNN or CNN) can improve the performance of the model on intermediate sentiments and ratings.
* If the performance of the model can be improved, the model can be used as a check to identify the reviews that corroborate their rating. This validation can help the user to make an informed choice while selecting the product while also informing him of the terms that the users use to describe their sentiment on the product, which can be found through identification of the best-performing predictors from model.