**Project End Goal**

* To create a recommender system based on the review text at Make level.

***Balaram Remala:***

* Assigned task: To work with the reviews associated with ‘Laptops’.

**Achieved Goals:**

* Data Munging:(2 weeks)

The data is in the form of group of reviews in json files where each json file have some list of review for each product along with their product specification. That way we have around 2000 products. Out of that 2000 products, few of them do not have proper product description, so I have removed those null values.

* Data Filtering: (3 Days)

Initially, the data is filtered from null values and a new column called “Make” is added to the dataframe for analysis based on brand name.

* Data Cleaning: (1 Week)

The data review content text is cleaned from short words and slang words using slang word dictionaries and word replacements.

* Applying Classifier: (2 Weeks)
  + 1. A Naïve Bayes Classifier is fed with a list of positive, negative and neutral words given from the txt file and trained to classify the content of our review and return if it is positive/negative/neutral as “Tag”.
    2. Once the Tag is obtained, it is added to the main dataframe and a review on counts of this column is done.
* Data Visualization: (2 Days)

Data is visualized based on makes with most positive review, most negative reviews and ratio of positive and negative counts are taken to get which brands are relatively better.

source: <https://textblob.readthedocs.io>

**Target Goals:**

* Check with the best classifier other than Naive Bayes based on the accuracies and apply to the data.
* Review of products based on time and date range.
* Data Visualization based on “Tag” and other features of the product.

All the above activities are done in a single ipython notebook linked below: <https://github.com/UNCG-CSE/CORPUS/blob/master/Project_Balaram/Project_Balaram.ipynb>