Sentimental analysis on Yelp reviews

By: Rohit Bhoopalam
(1001100534)
&
Pawan Puttaswamy
(1001094272)

Objective:

Perform Natural Language Processing on Yelp business data and predict attributes that benefits business management.

Overview:

Performing sentimental analysis based on reviews to predict whether a review is positive or negative. Location analysis for any business, based on category. Extracting most liked food items of a restaurant based on the reviews.

Dataset:

Yelp Dataset (https://www.yelp.com/dataset_challenge/dataset) and we want to get restaurant menu using the Yelp API.

End Product:

1. On searching a restaurant, a chart will be displayed showing the positive and negative review trend for that restaurant over a period of time.
2. On selecting a city and category of the business, a set of charts will be displayed showing relative statistics for that city and that business category for that particular location, which can be helpful to decide the business location for an entrepreneur.
3. Display most liked and disliked food items of a restaurant based on the reviews. Only the food items that are listed in the menu of a particular restaurant are used to extract the food sentiments. This can be further extended to suggest the best restaurant around a user's location and the food item one is willing to eat (using the menu of the near-by restaurants).

Challenges:

1. To come up with an approach to categorize a review into a positive or a negative review.
2. Gathering relevant features to decide whether the location is suitable or not. We may have to cluster the location of the businesses to gather the statistics.

3. Extracting the food items from the reviews and predicting the reviewer's sentiment towards those food items into a positive or a negative food review. Then aggregating the results to show most liked and disliked food of a restaurant.

Design:

1. a) Preprocessing of the text from the reviews.
   b) Performing lexical analysis on the preprocessed reviews.
   c) Performing sentimental analysis on the tokenized reviews.
   d) Classifying the assessed review into a positive or a negative review.
2. a) Selecting features to decide on the location's suitableness.
   b) To summarize and display the findings for that location.
3. a) Getting food items from the restaurant's menu using Yelp API based on the restaurant's name.
   b) Performing sentimental analysis on the reviews for those food items.
   c) Aggregating the results to show most liked and disliked food items.

Evaluation:

1. Using the reviewer's rating, which is provided in the Yelp dataset we can evaluate whether the review is positive or negative.
2. We can manually evaluate the number of positive and negative responses for food items for few restaurants.

Partitioning the tasks:

We want to select different tasks and work parallelly. We want to discuss the approach and progress and help each other from time to time.