# Recommender System for Yelp Users

**Team Members:** Hardik Manoj Katariya 1001030751

**Objective:** To design a recommender system that would help predict what rating a user would give to a given item (business in our case) or recommend a particular item (business) to the user.

**Overview:** Yelp data set provides us with reviews given by people for various businesses. We can perform analysis on these reviews and can perform matching between different users. We can also compute relationships between different businesses and can develop recommendations based on them.

**Data Mining Tasks:**

1. Will need to perform sentiment analysis for a business based on the reviews given by users. This would help in finding out the popularity of the particular business as well as help in knowing what things the user is interested in. This would require classification.
2. Will also need to find out similarity between different users based on the reviews.
3. Also will need to find out the relation between different businesses.

**Plan:**

I plan to build a good classifier that would help accurately identify if the review is positive, negative or neutral. I would then try to build a user-user as well as item-item collaborative filtering recommender system using the above sentiment analysis. The user-user collaborative filtering would help in predicting the rating the user would give to a particular business as well as in recommending a particular item to the user. The item-item collaborative filtering would mainly help in recommending the item to the user.

**Challenges:**

Many challenges would arise from dataset. I would need to perform data cleaning. Building a classifier that would accurately differentiate between positive, negative and neutral reviews would be challenging as it is difficult to identify sarcasm. The accuracy of the classifier would be the biggest challenge. Also being new to the field of recommender system, it would be time consuming to develop it.

**Address Challenges:**

In order to build a good classifier, I would take reference from Amazon Sentiment Analysis Tutorial, Stanford NLP Software and other good algorithms which provide good accuracy. Also I would try to learn to build a recommender system as early as possible from various online resources.

**Evaluate Efficacy:**

For evaluating the accuracy of the classifier to classify the reviews, I could first manually classify a particular amount of tweets into positive, negative or neutral and then run the classifier on those reviews and find out the accuracy of the classifier. The overall efficacy of the recommender system could be determined by using various metrics like rank metrics, basic decision support metrics etc.