**SI 670 Final Project Proposal**

**Yelp Restaurants Prediction, Evaluation and Recommendation**

Team: The Sleep Deprived

Mumber: Meixin Yuan, Yuan Cheng, Xinyi Zhao

**Motivation:**

In our project, we are interested in what kinds of factors or attributes are important for the success of businesses, including restaurants, hotels, schools or medical agencies, etc. Fortunately, Yelp Dataset provides us good resource to implement this kind of analysis and prediction. Yelp is a public company of U.S. which publish crowed-sourced reviews about different kinds of businesses by developing, hosting, and marketing the Yelp.com. In addition, we also want to build a good recommendation engine to provide corresponding recommendations to users.

**Datasets:**

The datasets come from Yelp Dataset (<https://www.yelp.com/dataset>) which include huge information about the metadata of all kinds of businesses and users, as well as the user reviews of those businesses. The complete datasets have five json files, and three of them would be used in our project.

1. yelp\_academic\_dataset\_business.json, the attribute and related information about each business
2. yelp\_academic\_dataset\_tip.json, the users history about the users and the businesses which they obtained services from.
3. yelp\_academic\_dataset\_user.json, the detailed information about the users including the connections of other users.

**Methods and Evaluations:**

1. Set up proper prediction models, based on the information of the businesses and the rates (“stars”), predict the rates of the test data by these models. Find the most accurate prediction algorithm.

Candidate models: Linear Regression, SVR, MLP, Random Forest Regression, XGboost.

Evaluation: MAE, MSE.

1. Once find the best prediction model, evaluate the importance of each features and find the most important features that influence the success of the business. Figure out the association of these features with the rates of the business.
2. Based on the users’ history, recommend businesses to related users.

Candidate model:

a) Collaborative filtering: SVD, SVDpp, KNNBasic etc. Recommend restaurants by other users who have similar preference of this user. Evaluation: MAD, RMSE (yelp\_academic\_dataset\_tip.json).

b) Content-based filtering. Recommend similar type of restaurants based on the users history. (yelp\_academic\_dataset\_tip.json).

**Computing:**

We are going to use Jupiter Notebook to implement the most part of computing work.

**Existing Work:**

**Duty of each group member:**