

## **Project Objective**

The objective of the project being proposed is to improve the yelp user suggestions. Suggestions will be more personalized. We plan to achieve this by identifying the similarities of reviews, ratings, and tips of the users with other users and categorize the users into categories and then suggest the future options which other users of same category visited or rated high in the recent time. It is a better way to provide suggestions to users.

## **Project Tasks**

Implementing categorization on the users (using user-defined categories, will talk more about it further down the document).

Using link analysis to create a pool of users in addition to each of the category to find better suggestions.

Classification to predict a restaurant option for the particular user, based on the above categories.

## **Deliverables**

At this moment I am unable to upload anything other than report, will update the repository with the working code soon.

## **Methods and Algorithms**

I am writing an algorithm that will categorize the users of yelp data set. I have selected a small data set initially to check if the data structure is working as expected.

I have planned to use Naïve Bayes Classifier for the Classification of restaurants in a particular category for a particular user

## **Evaluation Plan**

I am working on a data structure that will handle all the tasks of this project. I am changing the program constantly to make it efficient enough to have good performance. I tried to run the yelp data set to categorize the users.

Furthermore, I have planned to build classifier for the above categories. The classifier here will be the restaurants. The Restaurants will be cut down to a limited number since, the restaurant options come from the category, the user belongs to.

Link analysis is what I haven't figured out completely, but will use it to create the pool of users or to strengthen the suggestion performance by building a network.

## **Change of Plan since Proposal**

I have not decided to change the major parts of the project proposal but on how to deal with the intermediate tasks. Initially I decided to classify so that I would be able to categorize the users.

## **Difficulties Encountered**

I am working around with the implementation of link analysis to create the pool among users with respect to their behaviour.

## **Addressing the difficulties**

### Categories of users

- Foodies (Users with more reviews, check-ins, votes)
- Regular (Users with average number of reviews, check-ins, votes)
- Rare Visitors (Users with very less number of reviews, check-ins, votes)

Once the users are categorized over the above categories, I will work on link analysis. The link analysis results are expected to either give more understanding of the categories with the network built or adds additional aspects to the category with the existing category.

Link analysis to link the current user with the other users of the same category with respective to the behaviour of the users.

### Tasks to be accomplished

#### **Pending Tasks**

Implementation of link analysis.

Implementing classification on the categories of the users.

Testing the overall data set on the code so far written.

#### **Expected Challenges**

Efficient Data Structure on the Whole:

The evaluation plan seems to working theoretically, but the algorithms that I plan to choose and the data structure built around it along with the data mining techniques all-together will be inefficient in the beginning but will be improved over the time.

Pooling users with link analysis implementation:

I am expecting a learning curve here, that will help me understand the categories and to create a dataset that will help me to classify the restaurant for a particular user.s