Title:

Understanding the similarities of users by performing classics data mining techniques on data-set to provide users with better future suggestions.

Team Information: UPENDRA MADAM - 1001050282

Objective and overview of the project. Why is it interesting and significant?

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 and classify the users into categories and 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.

What are the data mining tasks you will perform on the Yelp dataset?

Link Analysis, Classification are some of the techniques I will be using during the course of this project.

What do you plan to deliver at the end of the semester? How would you present the outcome of your study? How would you demo your work? What will you place on your website?

I will visualize the results in a website, and a report written through my understanding of the entire project. Future suggestions for a particular user would be output for the project, these will be shown as part of the demo.

What are the challenges in this project?

Identifying the hidden patterns between the users and representing the associated information. The hidden patterns here are the similarities, identifying them and then writing an optimized algorithm to accomplish this is another challenge in the project, as we are dealing with big data-set.

How do you plan to address the challenges? How would you design and implement the solution?

By implementing the classification technique, I will categorize the data of users and their behaviour from the available yelp data-set and by implementing link analysis, I wish to create a network through which I will be connecting the different users.

The design and the implementation of the proposed project would rely on how effectively a network is constructed between the different users. Considering the multiple factors like check-ins, tips, reviews and positive or negative feedbacks given by users I categorize them and form the network.

Using optimized data structure through-out the project is important in elevating the performance, to achieve this, I plan to learn new ways of writing efficient algorithms to deal with such a data-set.

How would you evaluate the efficacy of your solution?

A layman can consider the same previous check-ins or the type of reviews to that of the people within the network. A better evaluation would be finding out the similarities of the details from the information and by linking all the factors that bring them along. The more the factors are similar the same network they belong to.

How would you partition the tasks and coordinate among team members?

Project purpose is to find un-identified similarities between the users, the following could be considered as tasks,

- Understanding the reviews, ratings to categorize the users.
- Create a network of users considering different patterns.
- Linking their experiences to give better future suggestions.

Since I wish to do the project alone, work is not distributed with others.