

Yelp Based Recommendation System

GA Data Science - Final Project Outline

Setu Upadhyaya
February 18, 2014

1. Problem to be solved

Build a recommendation system based on Yelp data which is provided as a dump for academic purchases.

2. Description of dataset

This project is based on the Yelp dataset which is provided for the Berkeley area. Data includes information about businesses, reviews and users.

3. Hypothesis

My hypothesis is that most people's tastes for restaurants can be clustered. By identifying which cluster a user falls into, I believe I can easily predict other restaurants they will like.

4. Statistical methods I plan to use and why

This is a classification problem. I will use supervised approaches to classify a restaurant based upon various features. I believe KNN may be a good algorithm here. I will try others as well (logistic regression, random forest, etc.).

5. Applications the finding may have

This project will provide more insight into what factors determine which restaurants you will like. I will personally apply these to finding good places to eat.

Initially, I will deliver a python script which will output restaurant recommendations. I also hope to place this on the internet as an interactive tool. However, this may be beyond the scope of the course.

REFERENCES:

Yelp dataset: https://www.yelp.com/academic_dataset