Stephen Paff

Yelp Project Proposal Description

Written on April 11th, 2017

**Overview:** This project is my submission to [Yelp’s 2017 Dataset Challenge](https://www.yelp.com/dataset_challenge). In this Project Proposal Description, I will describe my initial approach. What do I intend to analyze, and how do I intend to conduct my analysis.

**Project Description:**

1. **What is the problem I want to solve?**

Yelp’s challenge is to use the dataset of local business reviews, business attributes, check-ins, and tips from various cities around the world to understand cultural tends in consumer preferences. I am interested in conducting clustering analysis to determine patterns in consumer preferences. In particular, I will address the following questions:

1. What patterns are there connecting favorable restaurants for a particular consumer, and how can one group restaurants based on consumer preferences?
2. For consumers who like a particular restaurant, what other restaurants or types of restaurants do they prefer as well?
3. How could one cluster restaurants based on similarities consumer preferences?
4. Can we provide informed recommendations for new restaurants based on a consumer’s ratings?
5. **Who is my hypothetical client and how would they use this data?**

My hypothetical client is Yelp. They would use this data to categorize restaurants based on common consumer preferences and generate a recommendation model based on previous restaurant preferences.

1. **What data am I going to use for this? How will I acquire this data?**

I am going to use the datasets provided on Yelp’s dataset [webpage](https://www.yelp.com/dataset_challenge).

1. **What is my basic approach to solving this problem?**

I will conduct network and clustering analysis of restaurants based on consumer preferences. For consumer preferences, I will use ratings primarily and secondarily reviews. I see the project as having the following features:

1. Determining which restaurants have common consumers rating and/or reviewing their business
2. Conduct a cluster analysis of consumers to determine patterns for which consumers like and dislike similar restaurants
3. Analyze common features of these restaurants to generate groupings
4. Provide an algorithm to provide restaurant recommendations for a consumer based on his/her ratings of other restaurants
5. **What are my deliverables?**

I intend to write and post a report outlining my algorithm, analysis, and findings, as well as develop a PowerPoint presentation.