

Project Proposal

Understanding the factors that improve Business ratings on Yelp.

Sravan Kumar Reddy Mummadi

Objective

Model a system that provides suggestions for improving ratings and positive reviews of businesses that consistently under-perform on yelp.

My system helps in answering some interesting queries like “how much existing yelp ratings and reviews affect the future ratings and reviews”. It helps in business making decisions similar to "Is it feasible to open a Chinese/Indian restaurant at some specific location", "how far on average can a user travel just to enjoy their favourite food items", "how the distance from home/work place affects the ratings"? This is very significant to business owners as it plays very important role in improving their sales and revenues.

Data Mining Models

Classification

Classification tasks to be performed on users based on the reviews and ratings they provided. Different types of classifiers and classification algorithms are to be used and implemented on the data before we derive any intermediate conclusions.

Association Rules

Association rules are to be used to link similar business scenarios. This would help us out to extract factors and reasons that helped in improving Yelp ratings and reviews.

Other Models

Other models are to be implemented as and when needed while deriving some intermediate results.

Outcomes

Online Data Analytics

We plan to deliver data analytical results which focus on factors that lead to poor ratings or reviews of any business. We would like to present the results as pictorial diagrams (histograms, pie charts) and if possible a research paper. All the intermediate results and models used will be clearly explained in the paper. I would like to present all the analysed

data and diagrams on the website. Graphs will be used to depict how 'good ratings' and reviews are being followed every month.

Project Challenges

Finding Subject of the review:

Finding the subject of the review is very important to identify the real cause of the poor rating. It is also needed to reason out many intermediate conclusions. Text parsing is to be used to find out the reasons for poor rating and suggesting the locations for expanding the business.

Finding the hidden factors that improve the business rating and reviews:

Applying correct models are very challenging to find out the real factors that affecting the business reviews and ratings.

Proposed solutions:

- We may have limit to some major factors that affect yelp ratings.
- Taking help of mathematical models and nltk parser to parse text and finding the accurate subject of the review.

Design and Implementation

We plan to implement the project in python, using natural language parsers, and other data analytic libraries to parse all the reviews and derive the factors that affect ratings and reviews.

Following are some of the factors that affect reviews and ratings:

- Location of the restaurant
- Amenities of the restaurant
- Specific item of the menu
- Cost, Quality of the product
- Customer service etc.

Testing

We plan to model uncertainty in our models and conclusions through testing methodologies like hypothesis, null, alternative testing.

Project Roles

Sravan Kumar Reddy Mummadi

Planning to implement all the front end (web design), data analytics and backend tasks.