

Proposed project for CMPS 242 Machine Learning Fall 2016

Team members:

Greeshma Swaminathan (gswamina@ucsc.edu)

Neha Ojha (nojha@ucsc.edu)

Jianshen Liu (jliu120@ucsc.edu)

Alex Bardales (abardale@ucsc.edu)

PROPOSAL:

We will use the Yelp dataset to analyze what are the factors that make a business successful. This learning will help in predicting the chances of a new business becoming successful and can also provide insights for improving existing businesses. We will perform exploratory analysis on the available data to focus our efforts on those business categories that have the most reviews. We hope to observe interesting aspects about different businesses based on their location.

Our project will be useful for new businesses. As a business is starting out, it has no Yelp ratings. We will identify the attributes that give the highest likelihood of success. In parallel, our results will also tell us which attributes are not important for Yelp ratings, allowing business owners to focus efforts where the return on investment is high. Following the recommendations obtained from our results, a business should also be able to obtain a higher Yelp rating.

Before using machine learning algorithms on the data, we plan to preprocess the data in order to eliminate outliers and remove redundant information. This would ensure that we have clean data before using it to make predictions.