Yelp Project Data Memo

Erasmo Rivas

2022-10-20

Overview of Data Set:

The dataset I will be working with is the Yelp data set, specifically the subset of Resaurants. The data is available here: https://www.kaggle.com/datasets/yelp-dataset/yelp-dataset. I will be working with a mixture of continuous and discrete predictors, for example, one of the continuous predictors is numbers of reviews, and a discrete or categorical predictor would be whether the restaurant provides Wifi. There is about 40,889 data points and there are 47 predictors.

Research Questions:

I would like to see if the stars (the rating) can be predicted from the predictors. I will use a classification approach. I think number of reviews and potentially pricing of restaurants might be important. The goal is primarily predictive, but a secondary goal might be inference on predictors that lead to a high rating.

Project Timeline:

Data Preparation – finished have cleaned and converted to working structure Exploratory Data Analysis – 10/30 Fit Model 1 – 11/5 Fit Model 2 – 11/10 Fit Model 3 – 11/15 Fit Model 4 – 11/20

Interpret results – 11/25 Finish Write up/Narrative – 12/03