**Introduction**

Yelp reviews provide insights about the place based on which people can take the decision accordingly. Textual reviews of people are subjective in nature. It is possible that two people have written reviews with the same number of positive or negative words but they might have given different ratings in terms of stars. So to get help from the review one must go through plenty of textual reviews which is very tedious and time consuming. Predicting star rating based on sentiments of all the textual reviews will help users to cut all subjectivity of the reviews and take decisions swiftly.

**Technical Approach**

1. Dataset selection and preprocessing
   According to the technical approach discussed in the paper for predicting the stars starts with selecting the dataset. Dataset contains the textual reviews as well as the star given by the customers. According to the paper, the dataset is divided into two parts where 90% data is used as training data and the rest is used as testing data.
2. Feature selection
   After pre-processing the dataset the next step is the feature selection. Three methods have been discussed in the paper to select the features.
   a. Baseline : Uses top **words** present as main feature
   b. Feature Engineering I : Does **part of speech** analysis
   c. Feature Engineering II : Uses top **adjective** after part of speech analysis
3. Models
   The given paper has classified this problem as a regression problem. So suggested models for this problem are Linear regression, support vector regression , decision tree regression. These methods are useful only for predicting the rating of the business whereas for sentiment analysis naive bayes and support vector machine techniques are suggested.
4. Evaluation and results
   In the given approach in paper prediction of the stars for the rating is done by the regression technique so for evaluating the approach root mean square error(RMSE) is being used instead of accuracy parameter.

   Results :
   It has been observed in the paper that after experimenting with each feature selection technique and each model linear regression with baseline feature selection technique gives the least root mean square error which is 0.6014.

**Strength and weaknesses of the papers**

| Paper | Strength | Weakness | How to address |
|---|---|---|---|
| Predicting a Business' Star in Yelp from Its Reviews' Text Alone | • This paper suggests very simple approach for prediction | • Does not consider many important feature like friends influence | • Rather than considering features from only raw data this approach should add other features also. |
| Feature Extraction based on Semantic Sentiment Analysis | • This approach considers semantic analysis also instead of just calculating word frequencies. | • This paper has an approach for feature extraction only so rating prediction for the businesses is not there.<br>• No influence from friends is considered in rating prediction | • We will add Method to predict the rating from sentiment analysis |
| Predicting Yelp Ratings Using User Friendship Network Information | • Considers influence from the friends while predicting the rating | • Visualization and testing is difficult | • This method can be implemented for only one business at first so it doesn't become too complex. |

Table 1.

**Connection between the papers**

Predicting user ratings in a Yelp network is the main heading for the problem. We have studied three research papers related to this topic. One of the research papers provides the overall technical approach to predict the star rating for any business on the yelp.This paper provides the feature selection methods as well as suggest some of the regression models for learning methodology.

First paper was more focused on the overall approach for predicting the rating stars whereas the second paper is more focused on doing sentiment analysis on the textual data. This paper suggests a Naive Bayes based approach to classify the text.Other paper that we have studied is Predicting Yelp Ratings Using User Friendship Network Information. Which considers the influence of the friend on user ratings. This paper also suggests the baseline model to predict the rating.

**Promising further research questions emerging from these papers**

The important study that can be done after reading these three research papers is that in what way and how much the sentiment analysis and friendship network on yelp influences the rating of the business.

**An idea of a better model for something**

The model which gives appropriate weights to both sentiment analysis and friendship network influence for the star rating prediction.

**Predicting a Business' Star in Yelp from Its Reviews' Text Alone on Yelp challenge dataset**

| Model | Baseline (Top Frequent Words from Raw Data) | Feature Engineering I (Top Frequent Words after POS) | Feature Engineering II (Top Frequent Adjectives after POS) |
|---|---|---|---|
| Linear Regression | **0.6014** | 0.6488 | 0.6052 |
| Support Vector Regression | 0.7278 | 0.7298 | 0.7135 |
| Support Vector Regression-n | 0.6296 | 0.6791 | 0.6766 |
| Decision Tree | 0.6689 | 0.6698 | 0.6766 |

Table 2. Representing **RMSE** for each model and feature selection technique pair

**Reference**

1. Predicting a Business' Star in Yelp from Its Reviews' Text Alone Mingming Fan,Maryam Khademi
2. Feature Extraction based on Semantic Sentiment Analysis Mohammed Almashraee and Adrian Paschke
3. Predicting Yelp Ratings Using User Friendship Network Information Wenqing Yang (wenqing), Yuan Yuan (yuan125), Nan Zhang (nanz)