

Recommending Restaurants to investors by Predicting User Ratings and Overall trend of Business growth

-Harshal Gurjar

Critical Thinking:



Critical thinking continued

- ▶ Recognize problems, to find workable means for meeting those problems
- ▶ Gather pertinent (relevant) information
- ▶ Recognize unstated assumptions and values
- ▶ Interpret data, to appraise evidence and evaluate arguments
- ▶ Recognize the existence (or non-existence) of logical relationships between propositions
- ▶ Draw warranted conclusions and generalizations
- ▶ Put to test the conclusions and generalizations at which one arrives

How well is the Business?
How does the future of this business
looks like?



Factors considered

- ▶ Investor will like to invest in the restaurants which are popular among the customers
- ▶ Such restaurants are likely to succeed in the future
- ▶ So to identifying which restaurants to invest in, we will consider
 - ▶ which restaurants have good feedback from user
 - ▶ which restaurants have good ratings
 - ▶ which restaurants are likely to receive good feedback from user
 - ▶ which restaurants are likely to receive good ratings
 - ▶ Trend of business in future

Why do I need to understand the customer's feedback?

- ▶ Customer satisfaction is most important factor in any business.
- ▶ Major reason for business downfall is dissatisfaction among customers
- ▶ Happy customers = growing business

If we want to improve,
we must first
understand the current
trend of the business.
Simple, isn't it?

Trend Analysis

- ▶ User ratings
- ▶ User Review
- ▶ Review counts
- ▶ Trend in user ratings
- ▶ Trend in user Review
- ▶ Important factors in understanding the customer satisfaction

Data Collection

- ▶ Data sets
 - ▶ yelp_academic_dataset_review.json
 - ▶ stars
 - ▶ yelp_academic_dataset_user.json
 - ▶ User_id
 - ▶ Text
 - ▶ Review_count
 - ▶ Average_stars

Data Transformation

- ▶ Convert Json data to csv file for easy access
 - ▶ yelp_academic_dataset_review.csv
 - ▶ yelp_academic_dataset_user.csv
 - ▶ integratedUserAndReviewData.csv
 - ▶ Review length

Assumptions and Scope

- ▶ All data provided is trustworthy
- ▶ Project calculates overall business strength
- ▶ Data is consistent throughout
- ▶ Considering only top 100000 entries due to hardware constraints

Solution

- ▶ **Exploratory Data Analysis** - to understand the pattern in data
- ▶ **Generalized Linear Regression**
- ▶ **Sentiment analysis** to get length of useful words
- ▶ **Model Building** to predict overall trend of business
 - ▶ average stars, word count in review and number of reviews will be used to build the model.
 - ▶ **Annova Test** to test the result
 - ▶ nested likelihood ratio test

Initial Observations

- ▶ Result shows that users average stars, review length or review counts are terrible predictors of users' ratings
- ▶ However when all this factors are considered together, it is good predictor of user ratings and overall business trend
- ▶ They will therefore be the basis for the model.

Model Validation

- ▶ The given dataset is divided into training and testing dataset
- ▶ Training = 60%
- ▶ Testing = 40 %
- ▶ Model is trained using training set
- ▶ Model is fitted and tested for testing set
- ▶ Confusion matrix is used to determine the results

Results

- ▶ Model achieves 62% of accuracy in predicting user ratings for future as well as 59% for predicting the trend in the business
- ▶ Considering the factors that customers ratings can also depend upon the external factors such as there emotional state, Financial state, Scenario before giving ratings, we can surely say that this prediction level is a good achievement

Future Work

- ▶ We can identify growth of particular category of restaurant such as food, ambience, service etc
- ▶ For this we just need to tweak the algorithm to that granularity
- ▶ Extensive sentiment analysis and NLP for predicting the tone of reviewer
- ▶ Considering the revenue of restaurant for identifying customer satisfaction

References

- ▶ https://en.wikipedia.org/wiki/Linear_regression
- ▶ <https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=4&cad=rja&uact=8&ved=0ahUKEwiynlC8x8rTAhXBOyYKHe7VAM0QFgg7MAM&url=http%3A%2F%2Fcs229.stanford.edu%2Fproj2014%2FYun%2520Xu%2C%2520Xinhui%2520Wu%2C%2520Qinxia%2520Wang%2C%2520Sentiment%2520Analysis%2520of%2520Yelp%27s%2520Ratings%2520Based%2520on%2520Text%2520Reviews.pdf&usg=AFQjCNHCJ83fnobr9CpBPBtziGuFRDAVSsw>
- ▶ <http://minimaxir.com/2014/06/reviewing-reviews/>
- ▶ https://en.wikipedia.org/wiki/Generalized_linear_model
- ▶ <https://www.slideshare.net/ohassta/critical-thinking>
- ▶ <http://leonidzhukov.net/hse/2011/seminar/papers/chi09-tie-gilbert.pdf>
- ▶ <http://www.utdallas.edu/~ryoung/phdseminar/DataMiningComparison-Melody.pdf>
- ▶ <http://www.criticalthinking.org/pages/defining-critical-thinking/766>

Thank You