Predicting Restaurant Review Sentiment from Yelp Reviews

Soorya Paturi

Rating Yelp Reviews

-Leaving a review on Yelp is the most common way to rate a restaurant.

-A review consists of a text review as well as a star rating in the range of (1-5).

-Sentiment prediction can add to or replace star rating.

Why is it important to predict review sentiment?

-Sentiment prediction on a text review adds clarity of meaning.

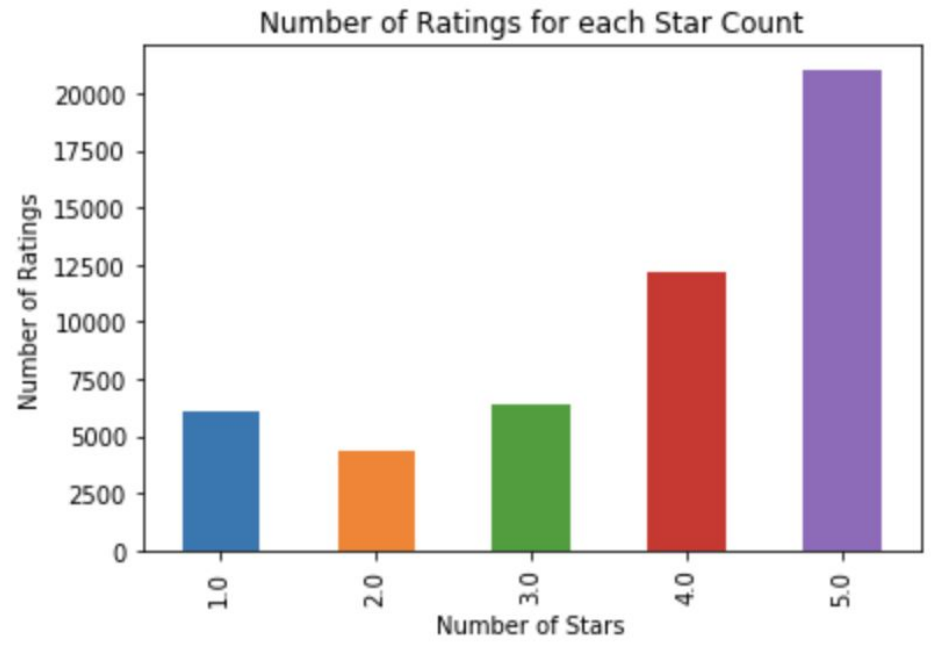
-Another metric can be added to the review.

-More conversational style of reviews with sentiment analysis.

Background on Data Set

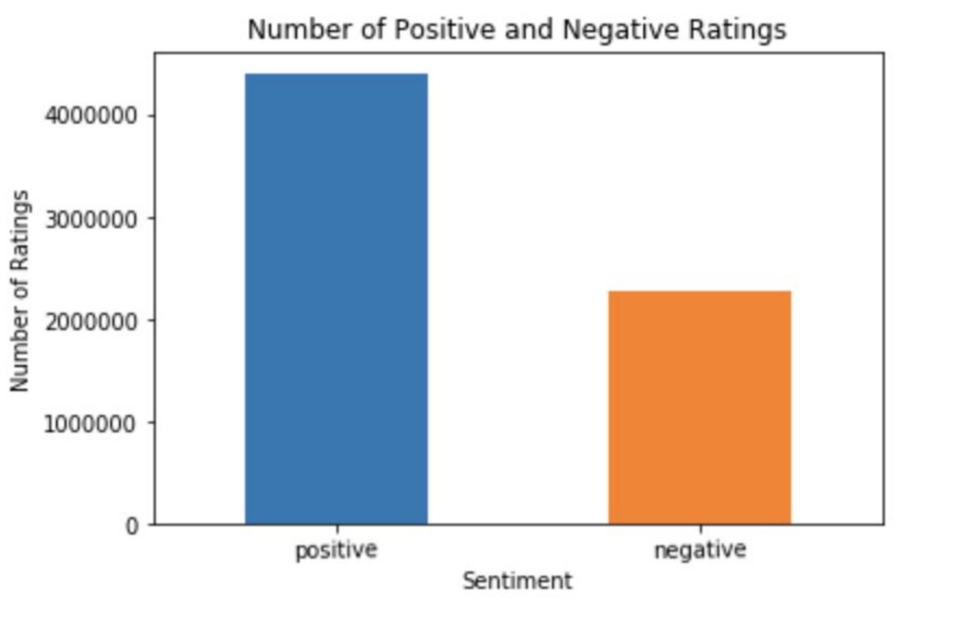
-Contains 5.2 million reviews from 174,000 businesses in 11 metropolitan areas.

-Subsetted to a random sample of 50k reviews of restaurants in Las Vegas.



Data Wrangling

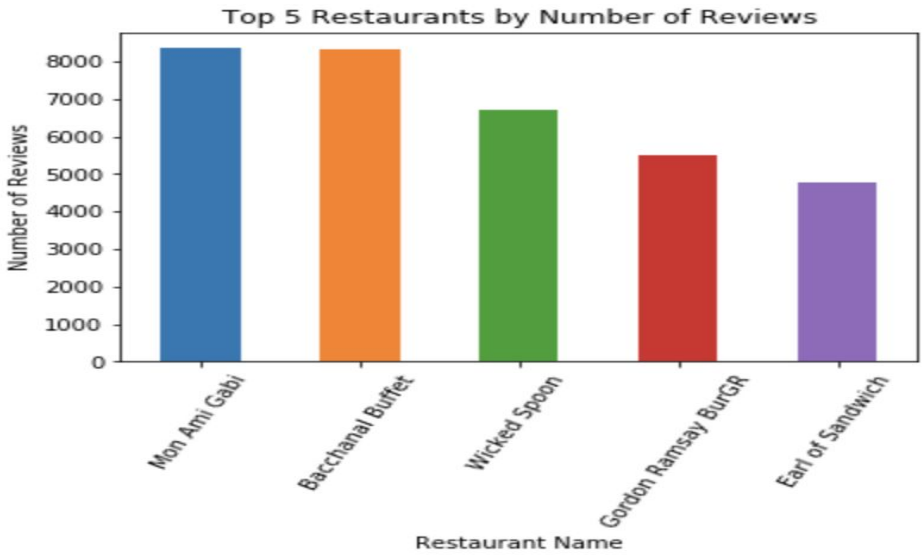
-Imbalanced data
-Imbalance reduced with relabeling
-Missing values accounted for
-Outliers explored

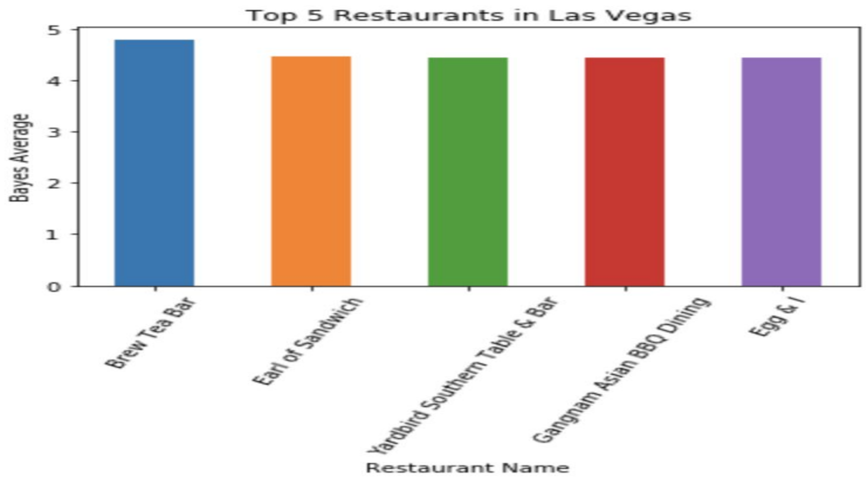


Ranking Restaurants

-Number of Reviews

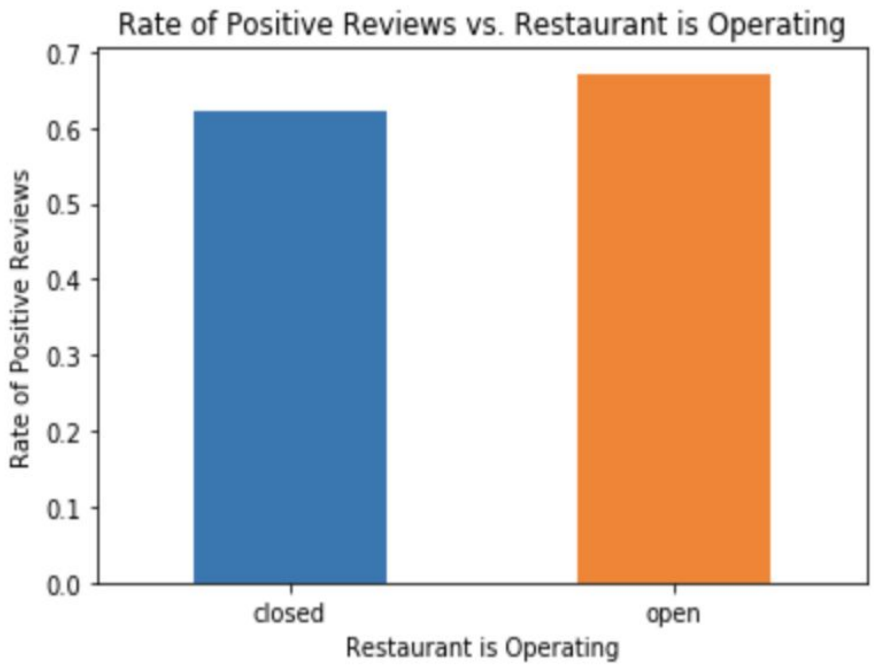
-Bayesian Average





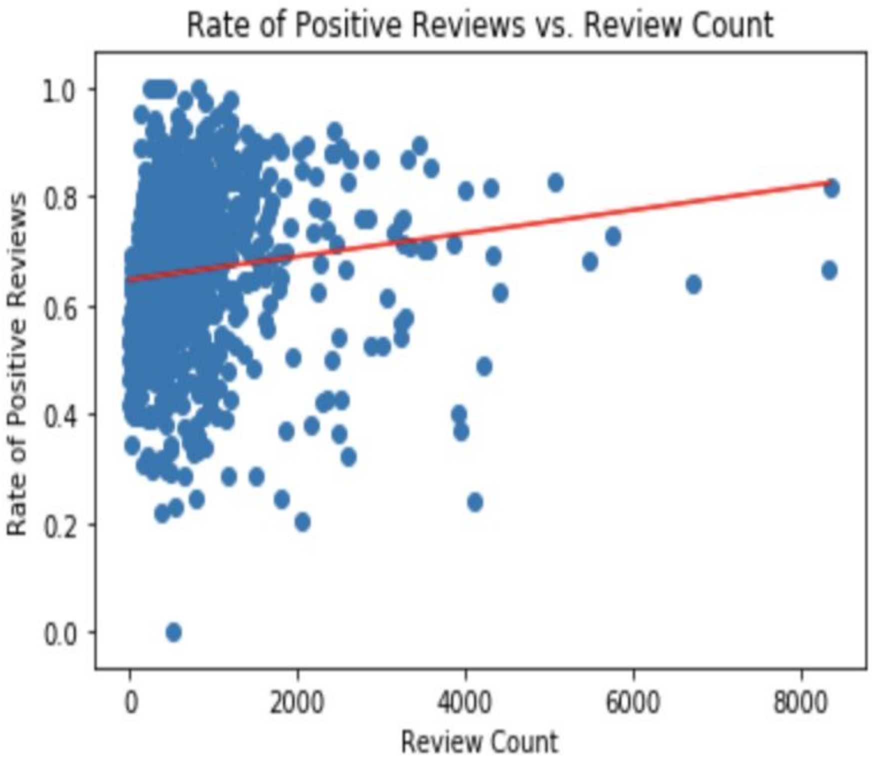
Restaurant Operation

-Open vs Closed
-Negative Reviews affect closure
-Might be a useful predictor



Restaurant Review Count

-Review count is a measure of popularity
-Weak correlation to rate of positive sentiment



Price Range

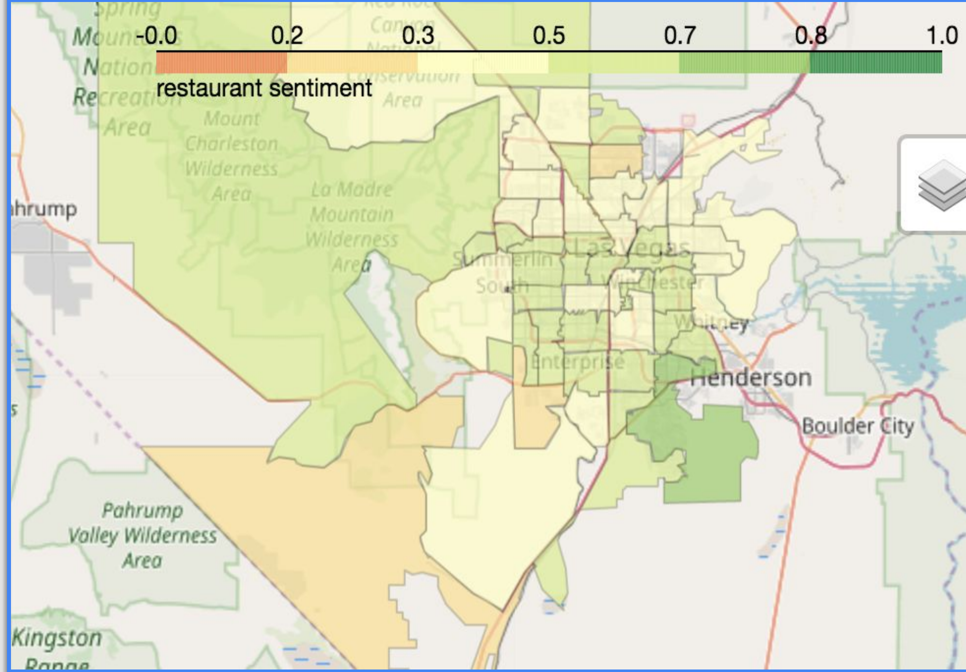
-Price Range has four options: (1-4).

-Highest rate of positive reviews is in price range (4).

Geographic Representation

-Heat map of positive review rate by zip code in Las Vegas

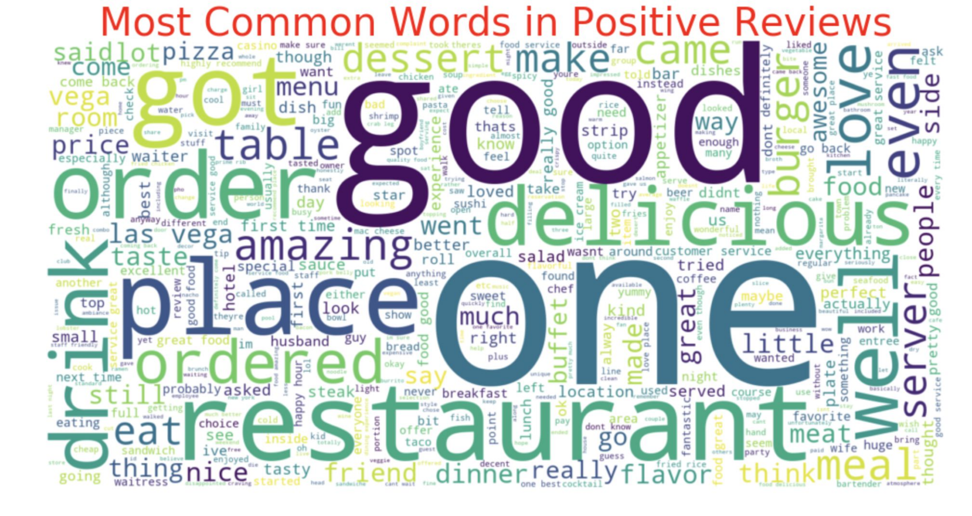
--Too many zip codes to hot one encode



TextBlob

-Finds sentiment polarity for each review

--Wordcloud visualizes most common words in positive reviews



Word2Vec

-Neural network that trains with the words in the data set to create vector representations of each word

-Size 300 was used for word vectors and stop words were not removed

Great:

```
[('fantastic', 0.8009390234947205),
 ('wonderful', 0.7734380364418030),
 ('fabulous', 0.7601549625396729),
 ('terrific', 0.7351025342941284),
 ('excellent', 0.7282377481460571),
 ('awesome', 0.7204433679580688),
 ('phenomenal', 0.6875117421150208),
 ('amazing', 0.6724933385848999),
 ('exceptional', 0.6698110103607178),
 ('outstanding', 0.6648397445678711)]
```

Awful:

```
[('terrible', 0.8226549625396729),
 ('horrible', 0.8048520684242249),
 ('alright', 0.7285017371177673),
 ('disgusting', 0.7106912136077881),
 ('subpar', 0.7092010974884033),
 ('sucked', 0.6660960912704468),
 ('gross', 0.6531112790107727),
 ('lousy', 0.6510778665542603),
 ('stellar', 0.6461423039436340),
 ('lacking', 0.6347830295562744)]
```

Feature Selection

-Only text review was used.

-Allows for generalization.

-Text was preprocessed and the converted to Word2Vec representation.

-Word vectors averaged in ensemble methods but not in neural network.

Models

Benchmark-Label all applications as the majority class

Neural Network-multiple dense layers with activation functions as well as pooling and bidirectional short term memory.

Random Forest-250 estimators, minimum sample leaf=1

Gradient Boosted Tree-250 estimators, minimum sample leaf=1, learning rate =1

Voting Classifier-soft voting with both random forest and gradient boosted tree

Results

-AUC and accuracy are looked at because of class imbalance.

-Classifier with highest AUC: Deep Neural Network

-Neural network is higher is all metrics except FNR.

-Ensemble methods did surprisingly well even though word vectors were averaged.

TABLE II

BENCHMARK CLASSIFICATION METRICS

Model	Accuracy	FPR	FNR	BER	F1-Score	AUC
ZeroR Baseline	0.6335	1.0	0.0	0.5	0.7977	0.5

TABLE III

MACHINE LEARNING CLASSIFICATION METRICS

Model	Accuracy	FPR	FNR	BER	F1-Score	AUC
Random Forest	0.8354	0.3733	0.0588	0.2160	0.8836	0.9118
Gradient Boosted Tree	0.8602	0.2547	0.0815	0.1681	0.8971	0.9261
Voting Classifier	0.8563	0.2871	0.0710	0.1790	0.8956	0.9236
Deep Neural Network	0.8774	0.1978	0.0836	0.1407	0.9078	0.9409

Recommendations and Conclusion

-Replace star rating with percentage of reviews that are positive

-Removes difficulty of sentiment interpretation from text.

-Should be placed next to reviews and model can be changed as performance metrics increase.

-Mean star rating does not give any information about variance of ratings.

Future Work

-More samples in training set.

-Majority class can be downsampled

-More hyperparameters can be tuned

-Additional fields from the restaurant can be used.