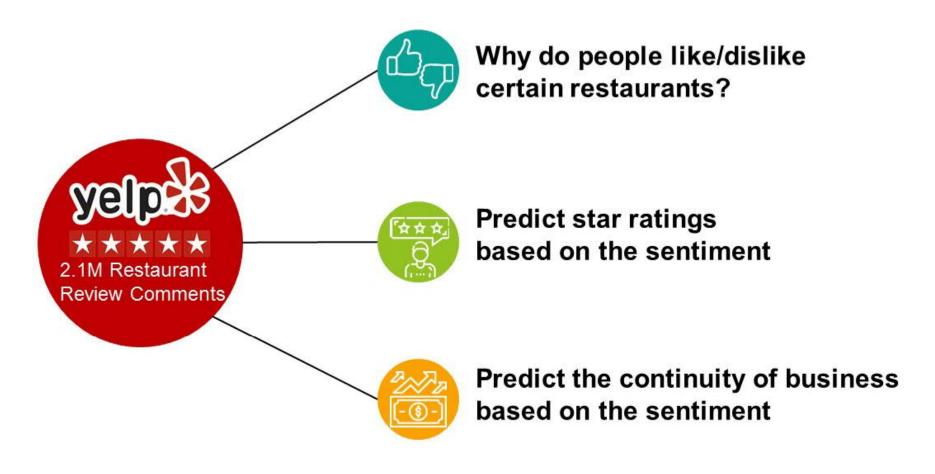


Sentiment analysis on Yelp Reviews

Terrance Xia

Business Problems & Data



Modeling Process



- Remove stop words, symbols, punctuation, non-English language
- Oversampling
- Train / test split

- Sentiment ratings
- Number of reviews
- Number of check ins
- Review length

- Logistic regression
- Random forest
- XGBoost
- Model optimization
- Star rating (Accuracy: 51%)
- Business continuity (Accuracy: 89%)

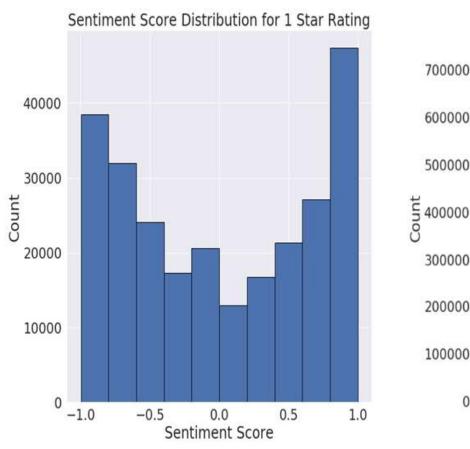
Common Words Star 5 vs Star 1

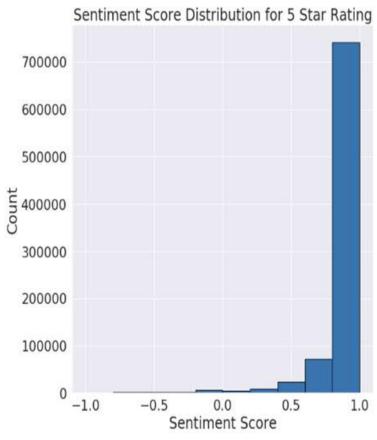


Common Words Star 5 Rating Restaurant

Common Words Star 1 Rating Restaurant

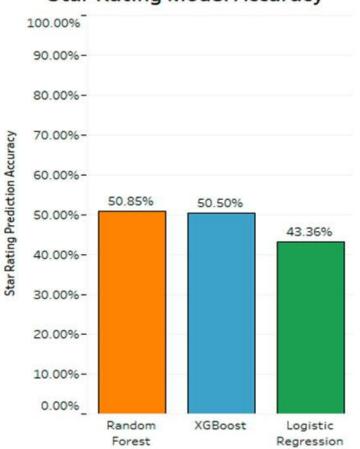
Review Comments Sentiment Distribution Star 1 vs Star 5



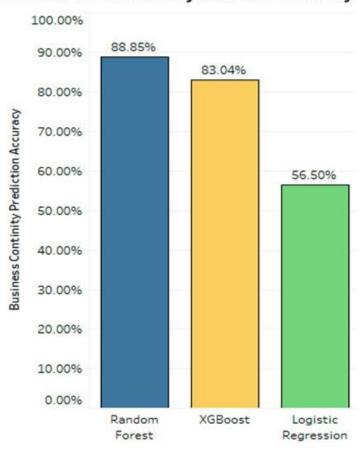


Model Accuracy Comparison





Business Continuity Model Accuracy



Model Application





- Improve star rating on Yelp for restaurant owners
- Change reputation for restaurants on Yelp
- Attract more customers

- Insights to potential investors
- Additional evaluation criteria for goodwill