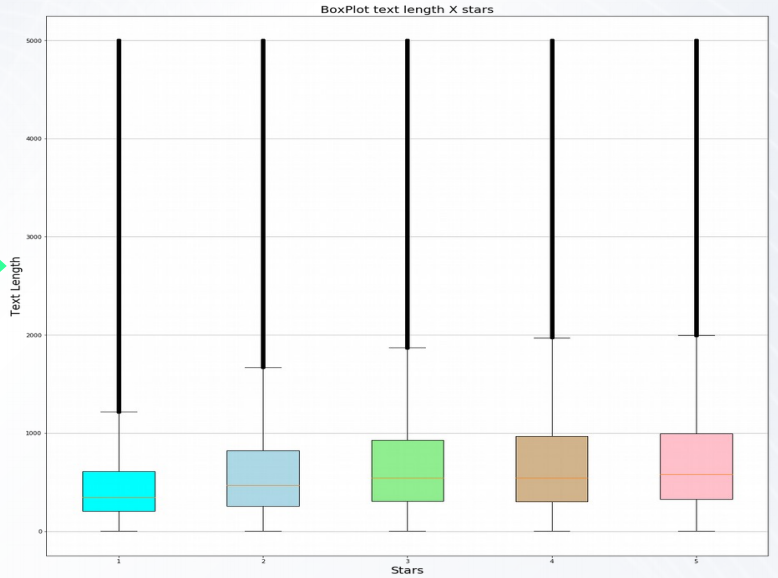
Recommendation system for Yelp based on predicted review rate

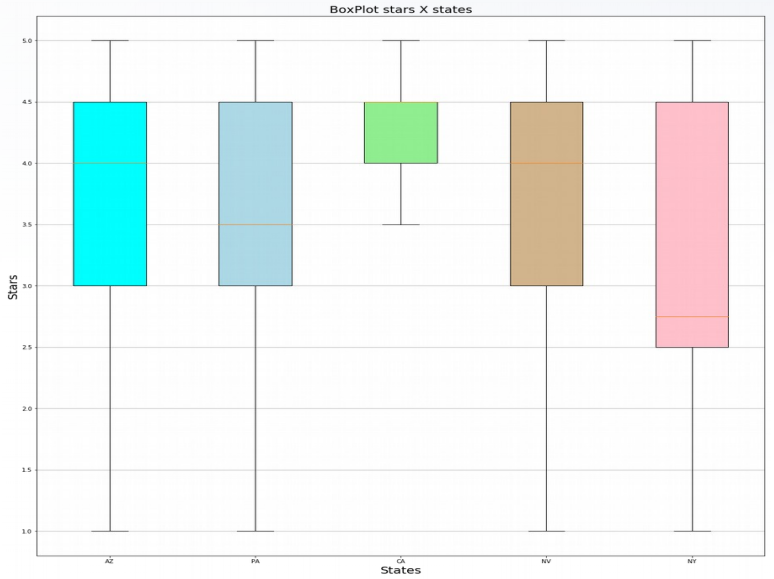
Hua Wei
2018-Feb-20

Motivation and Yelp Dataset

- Target: Improve the performance of recommender system with predicted review rate

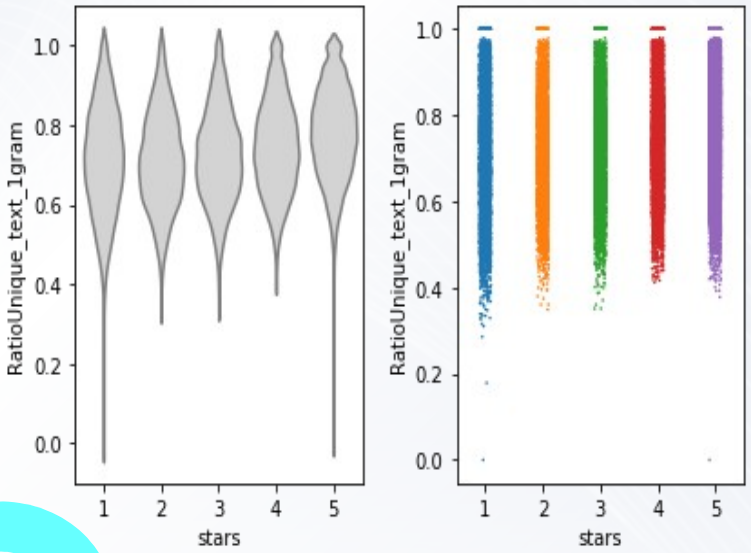
- Dataset:
  - review.json: review content, like text, stars(rate), user_id, business_id, etc.
  - user.json: user information, like number of friends, followers, etc
  - business.json: restaurant details, like average rate
  - tips.json: tip text from a user for a restaurant

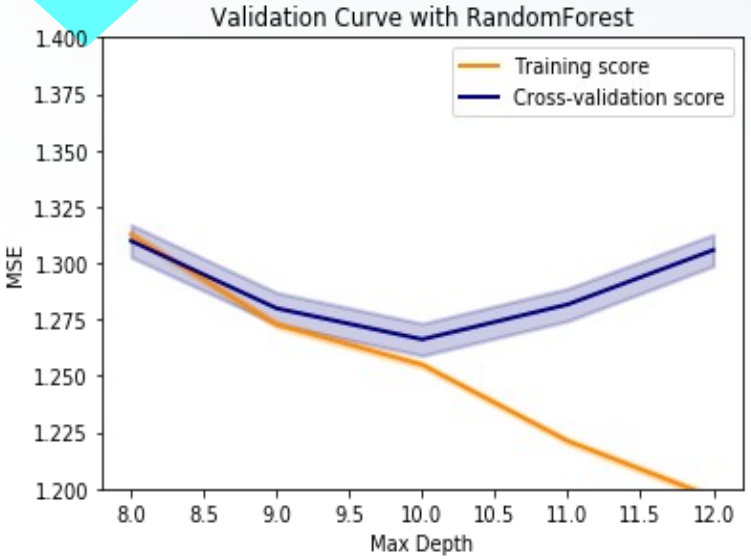




Feature engineering and Machine learning system

- Feature engineering:
  - Text: review text(tip text). Count, distance, TF-IDF based similarity
  - User and business info
- Machine learning system:
  - Predicted rate: PredictRateModels.ipynb
  - Recommendation with predicted rate: RecommendationSysTest.ipynb





Conclusion, Plan and Challenge

- Build a vanilla data analysis framework to predict review rate with 10% data
- Build a baseline recommendation system with surprise package

- Try to build rate predict models for every metropolitan area
- More feature engineer on tips.json, etc
- Tune model with Bayesian optimization

- Parallel processing on dataset (necessary for text feature engineer)
- Need to think a better way to use predicted rate in recommendation system