

Restaurants recommendation system for Yelp user

Hua Wei
2018-Mar-28

Motivation

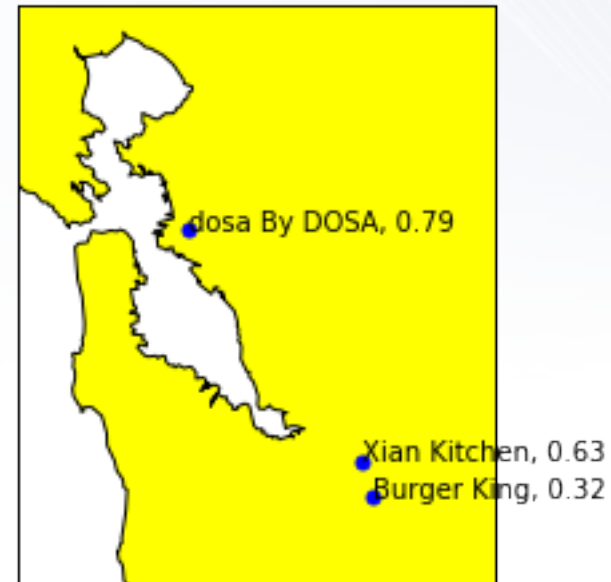
- Target: Get recommendeds restaurant within a certain range for a specific user
- Example: Heroku APP under construction...

Input

User ID (unique ID to obtain user history, like rate, review for all restaurants)

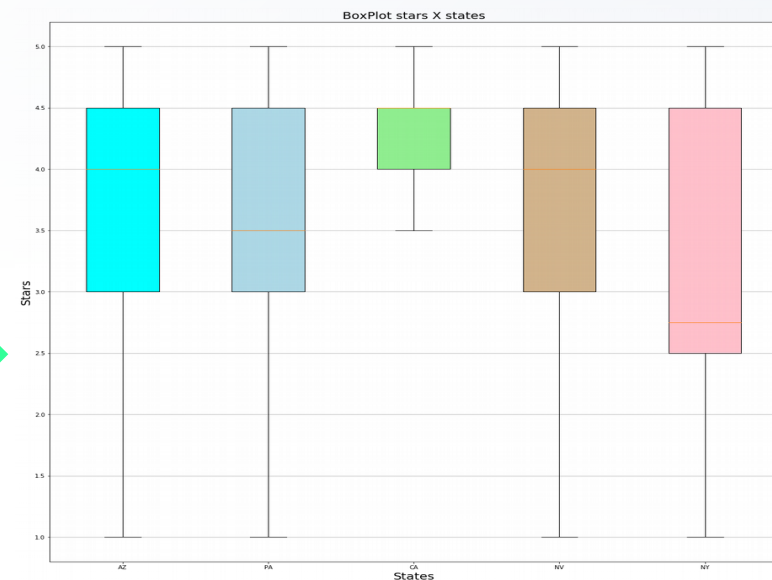
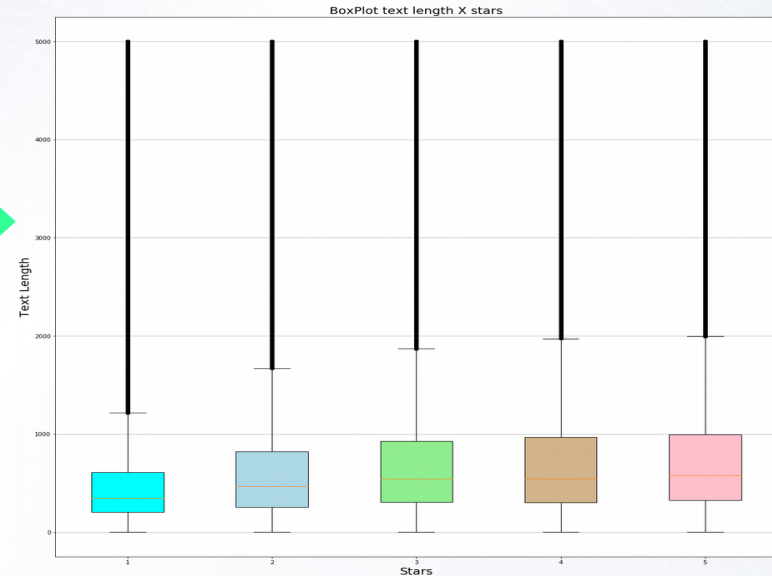
Range (In miles)

Output



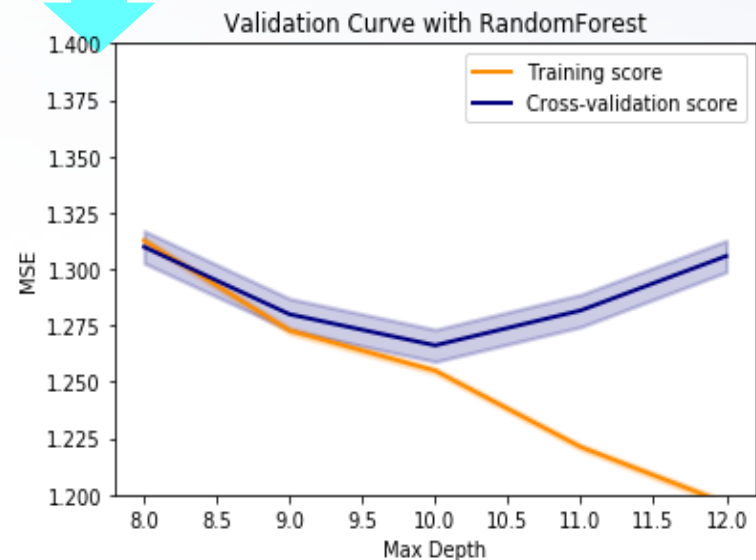
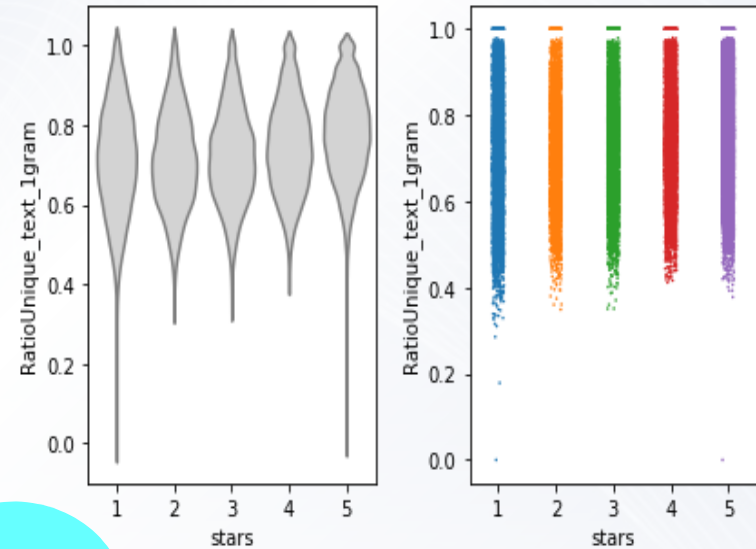
Yelp Dataset

- 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:
[RecommendationSystemTest.ipynb](#)



Conclusion, *Plan* and *Challenge*

- Build a vanilla data analysis framework to predict review rate with 20% 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 (distributed system) on dataset