Yelp Recommendation System

Yelp Help : Aayush Birla, Sai Chaitanya Sanivada, Samyak R Jain, Himanshu Kandpal

Motivation and About the Project

Since they were first used by Netflix, Amazon; **Recommendation Systems** have become an essential part of our lives. These systems have become so common that we tend to forget the apparent ease with which these systems are designed to optimize our consumer choices.

In this project, we aim to use the **Yelp Dataset [1]** to build a restaurant recommendation system which would help users to find restaurants suited to their liking. We are trying to answer the question : "**give me more restaurants like this one**" for the user.

Data and Labels

Yelp Dataset

Consists of separate files wherein each file is composed of a single object type, one JSON-object per-line.

- **business.json -** Contains business data including location data, attributes, and categories.
- **review.json -** Contains full review text data including the user_id and the business_id of the written review.
- **user.json -** User data including the user's friend mapping and all the metadata associated with the user.
- **checkin.json -** Check ins on a business.
- **tip.json -** Tips written by a user on a business. Tips are shorter than reviews and tend to convey quick suggestions.

References

[1]https://www.yelp.com/dataset/documentation/main
[2]https://www.kaggle.com/rohanrao/tutorial-on-reading-large-datasets

Model

1] Exploratory Data Analysis

We tried to clean up the data by checking for missing values and incorrect data-types.
Explored the dataset looking for any apparent relationships between the variables. In the process, we generated some plots.
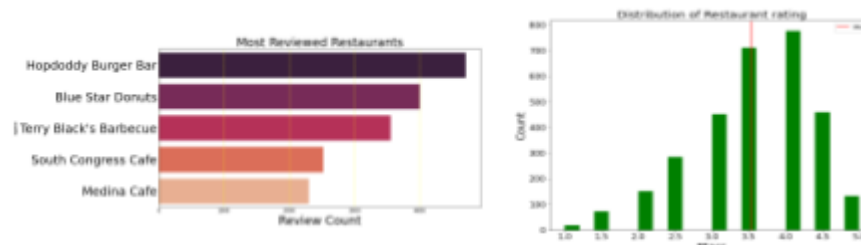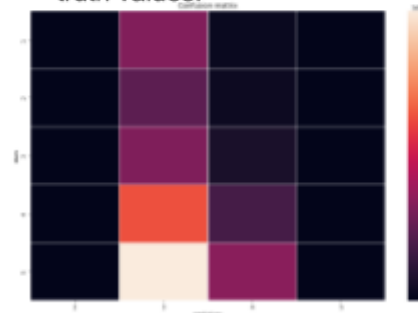

2] Baseline Model

Created a baseline model than can predict a user's rating for a given restaurant and recommend similar restaurants based on the item-item similarities matrix.
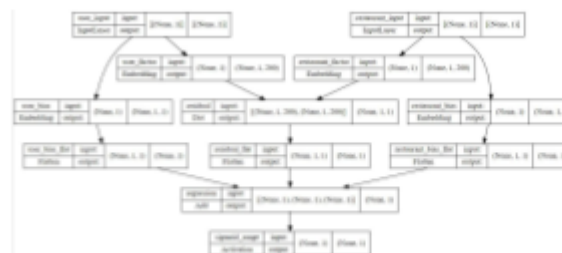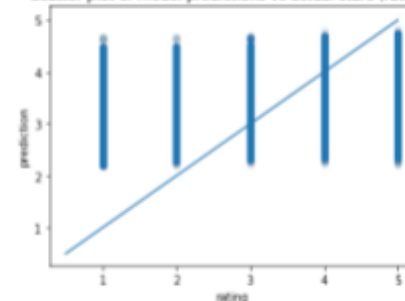

Results

Created a confusion matrix for the predictions and the truth values.

Conclusion and Future Work

- In conclusion, we have used matrix factorization techniques with regression to predict the user - item rating.
- Recurrent Neural Networks can be used in the future as incorporating time in the model is important.
- A few restaurants serving seasonal foods can have difference in their ratings corresponding to the current season.
- We can include other businesses such as saloons, Hotels along with restaurants in our model.
- We can use the business reviews text data available in the dataset to perform sentiment analysis and predict business ratings
- We can also use the photos available in the dataset to find similarities between different restaurants.