Restaurant Recommendation System

By Team - **Ensemblers**

● ● ●

Ishani Bari

Preetpal Singh
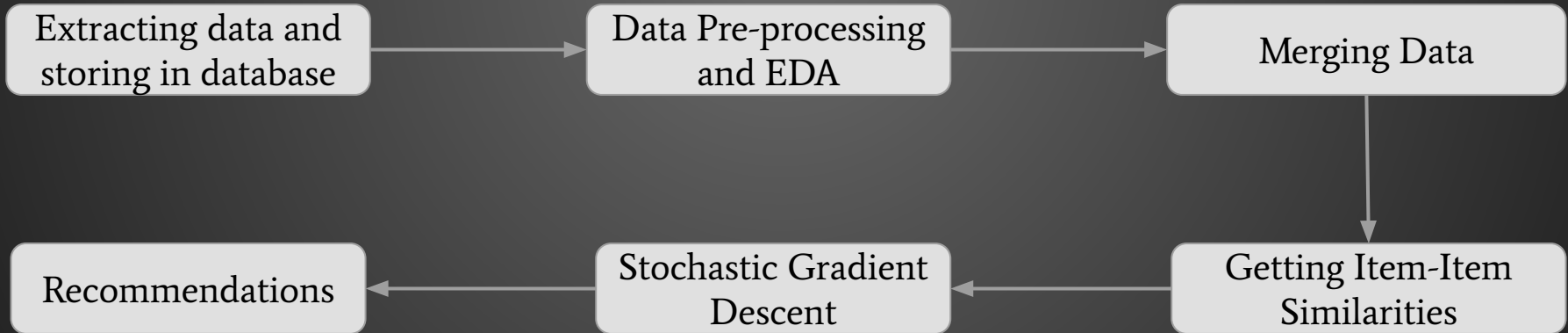
Rampunit Kumar

Umang Agarwal

Outline

- Introduction

- Methodology

- Results

- Conclusion

- Future Scope
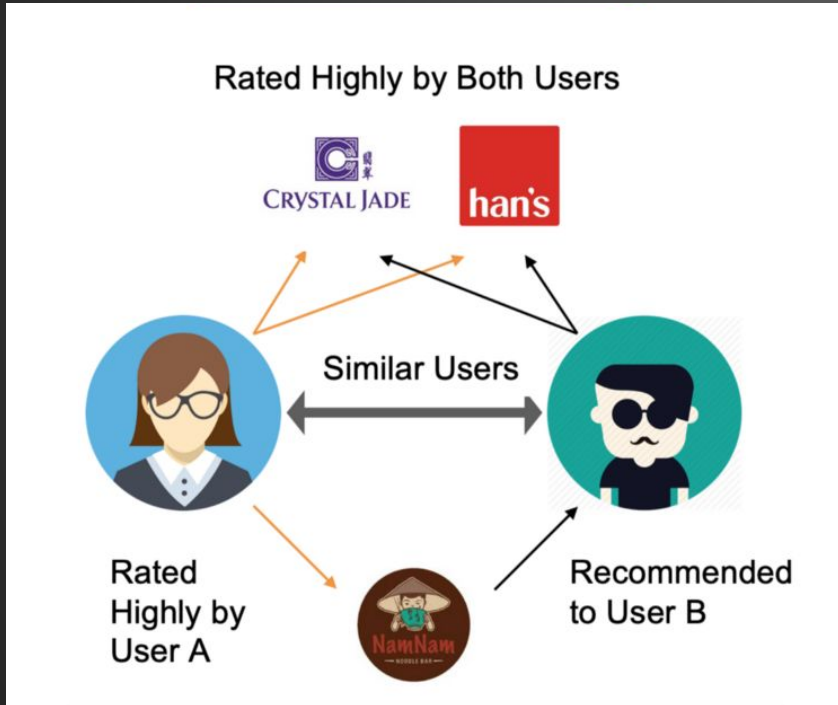
Introduction -

Problem Statement :

We need to create a recommendation system for restaurants using collaborative filtering (CF).

Approach towards the problem statement :

Extracting data and storing in database	→	Data Pre-processing and EDA	→	Merging Data

Recommendations	←	Stochastic Gradient Descent	←	Getting Item-Item Similarities

Introduction

Method used for recommendation -

Source - https://medium.com/@kencheah/vegas-baby-1213890614eb

Collaborative Filtering Using

Stochastic Gradient Descent

Methodology

Dataset Description -

The Yelp dataset has 5 JSON files containing about 3.7 million reviews from 687,000 users for 86,000 businesses.

Data Preprocessing done :

- Unpacking the json into columns.
- Keeping businesses having category as restaurant/food.
- Unpacking nested json columns.
- Handling null values.
- Merging the nested dataframes.
- One hot encoding the categorical columns which we get from nested dataframes
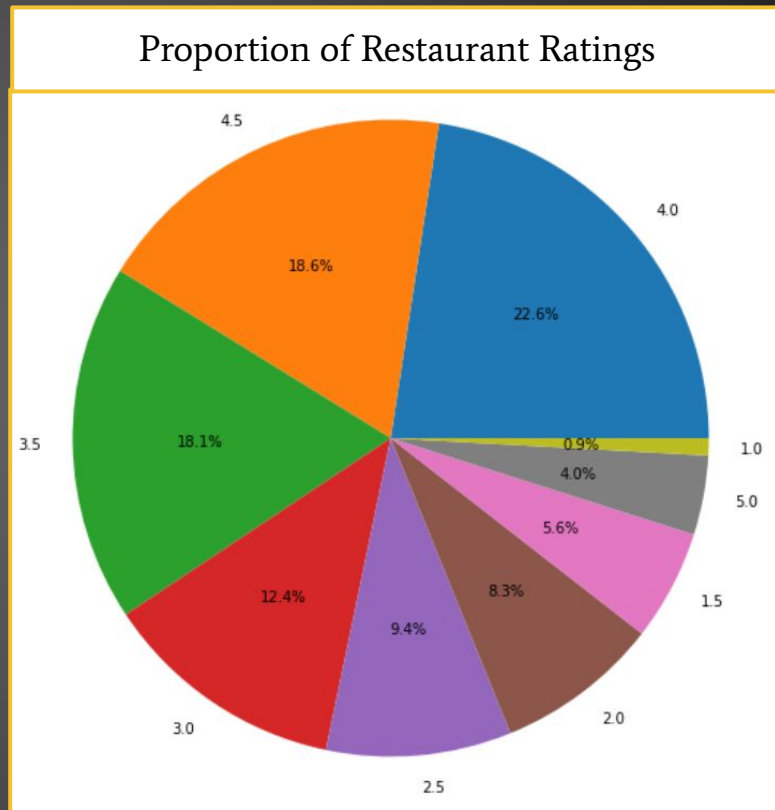

Fig1 - Proportion of Restaurant Ratings

Methodology

Packages used - SQLite, Pandas, Numpy, Matplotlib, Seaborn, Sklearn, Tensorflow, Scipy
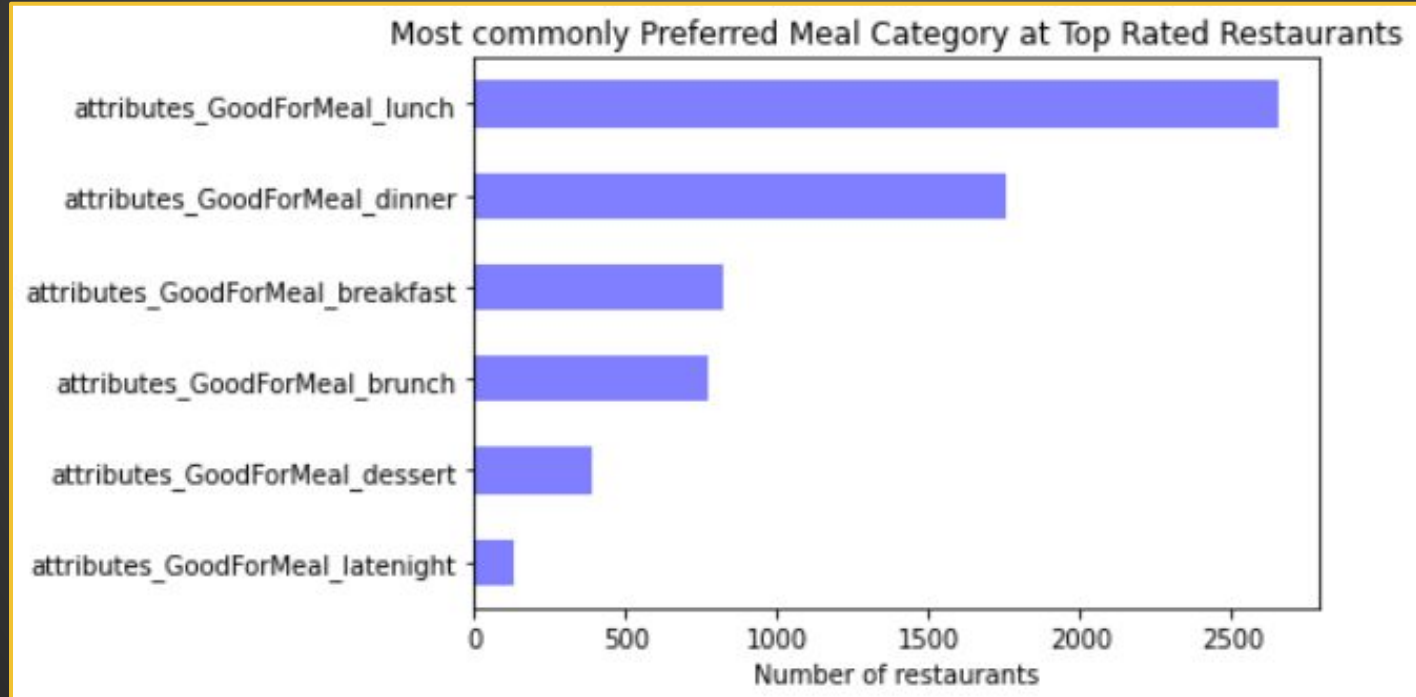
Insights Gained after pre-processing and EDA-

Fig2 - Most commonly preferred meal category at top rated restaurants

Methodology

Insights Gained after pre-processing and EDA -
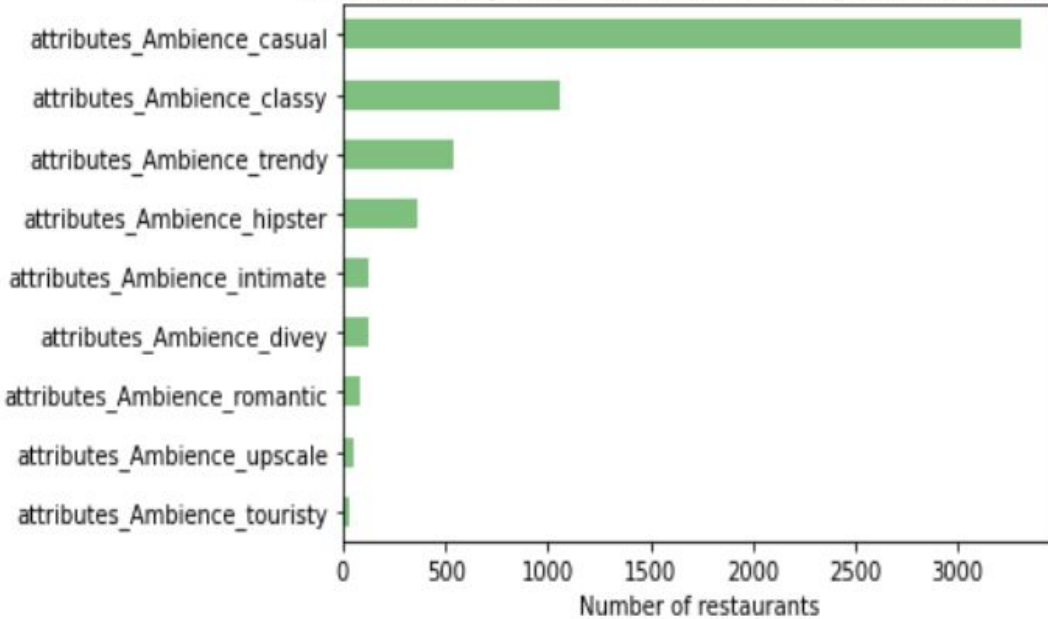

Fig.3 - Most commonly preferred ambience at top rated restaurants
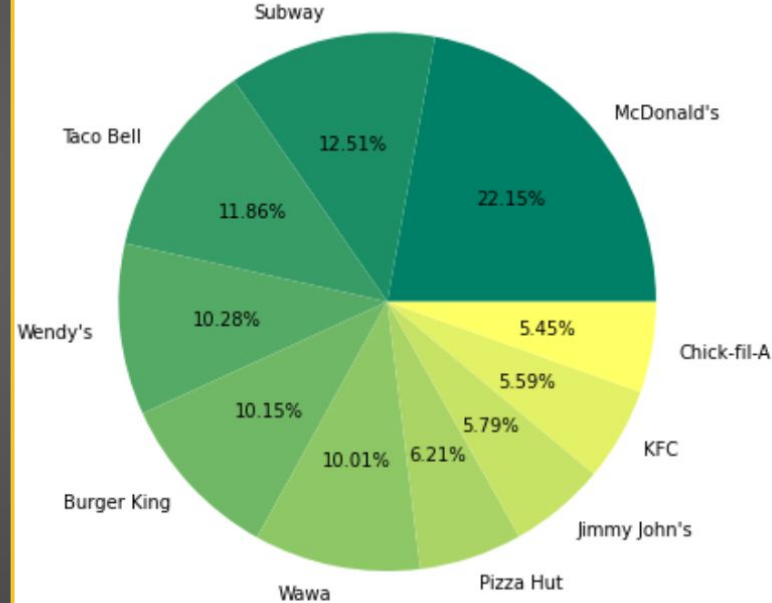

Fig.4 - Mostly Reviewed Restaurants

Results

Collaborative Filtering Model based on Latent Factors Architecture
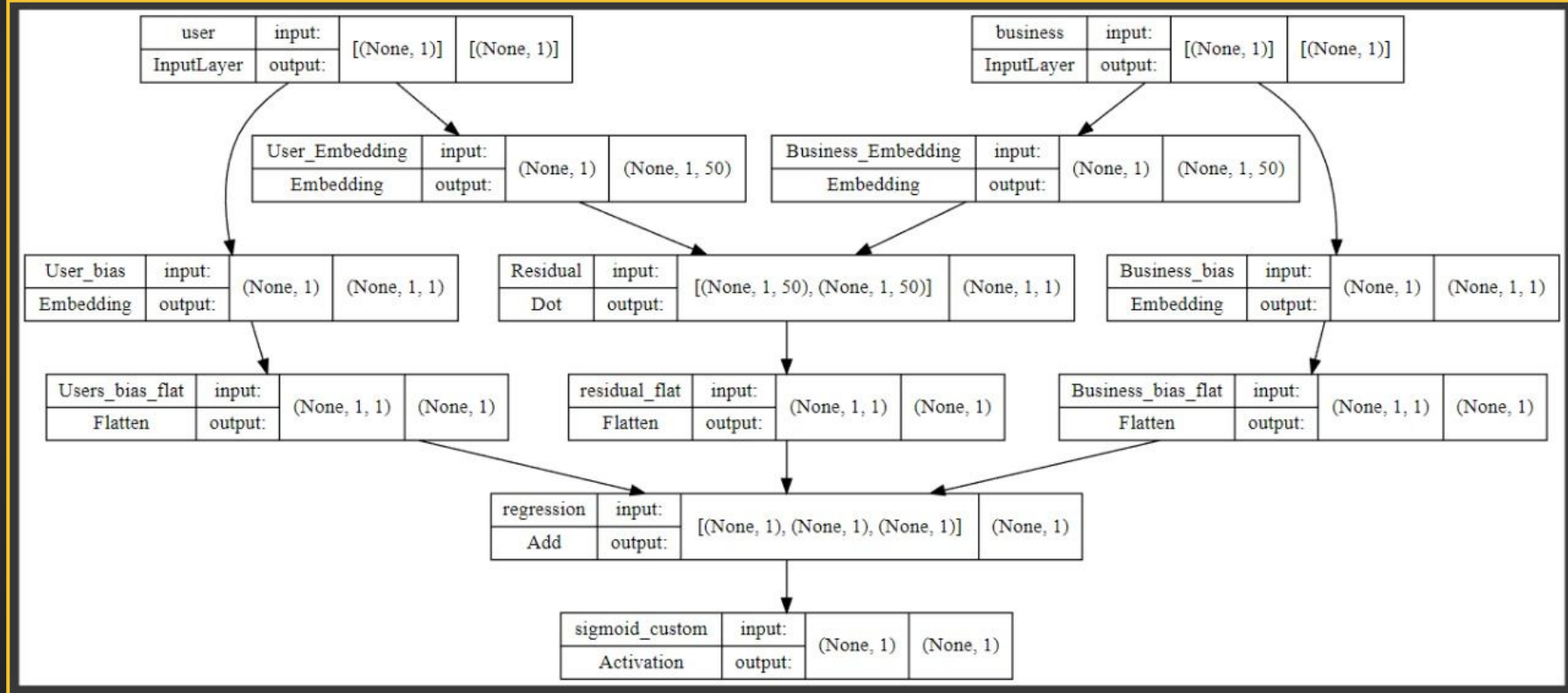

Fig.5 - Collaborative Filtering Model based on Latent Factors Architecture

Results

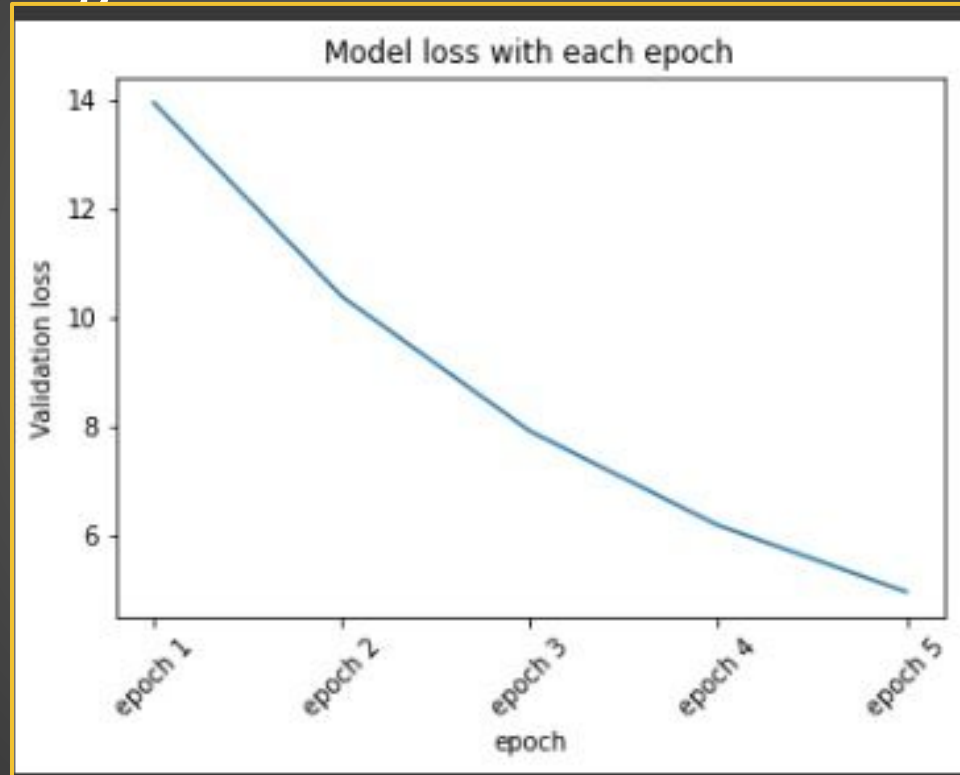## Collaborative Filtering Model Evaluation

Fig.6 - Model loss with each epoch

Conclusion

We did item based collaborative filtering to recommend restaurants to similar users. To optimize the collaborative filtering we used Stochastic Gradient Descent (SGD) method that has three hyperparameters, the learning rate η, regularization coefficient λ, and L the number of latent features to include.

Future Work

Hybrid model of content based and collaborative filtering to improve our recommendation system.