Recommendation System

Collins Kanyiri

AIMS

- create a unique experience for each of our customers while subtly increasing company ROI
- building a tailored, unique recommendation system that can effectively suggest movies to our users
- attract new users to our platform
- expand the platform by investing in newer and relevant content for users to enjoy.

Objectives

1. Use Collaborative Filtering to predict what users would rate movies that have not been seen before.

2. Find the optimum model to use

3. apply model to the data

4. Draw conclusions based on our data and model results.

observations

- Created genre labels in order to determine the most popular genres in the dataset; top 3 were Comedy, Drama, and Action
- Analyzed the distribution of ratings: a rating of 4 was the most frequent
- Analyzed the top 10 movies overall, as well as by genre, to use as a future comparison to post-model EDA to determine presence of popularity bias

Recommendation Modeling

- Used the base model ,KNN-baseline model. Without tuning any hyperparameters
- Used a rating scale of 0-5
- Explored various recommendation algorithms within the Surprise library but settled for a tuned GridSearchCV-SVD model.
- To help validate our models, we incorporated GridSearchCV while also running K-folds Cross Validation (3-folds). This helped us validate our model, but more importantly, efficiently tune the hyperparameters we fed in the model.

Conclusions

Overall, our model does a fairly decent job of estimating users' ratings, with an approximate error of 0.8563

Our model is a purely collaborative filtering model, and therefore does not address the "cold start problem"

future work

- look to incorporate aspects of a content based filtering model
- Incorporation of features, such as genres and year, into our matrix factorization models