

Hi, Omar, sorry for interrupt, our team have consulted you on the forum about topic, our team are permitted to choose the dataset on Kaggle as topic 0, the detail of our project as below:

For the dataset: <https://www.kaggle.com/rounakbanik/the-movies-dataset>

Detailed of my team(team number:45) shows as below:

Yuke Li z5176261

Zanning Wang z5224151

Yuankai Zhao z5233940

Zhenming Wang z5140192

At the first stage, we plan to build three different types of recommender systems:

(1) Item-based collaborative filtering

Collaborative filtering based systems use the actions of users to recommend other items. In general, they can either be user based or item based. Which means the users like me, also looked at these other items. Item based collaborative filtering uses the patterns of users who browsed the same item as me to recommend me a product (users who looked at my item also looked at these other items). User-based approach is often harder to scale because of the dynamic nature of users, whereas items usually don't change much, so item-based approach often can be computed offline .We will train a KNN models to cluster similar movies based on user's ratings and make movie recommendation based on similarity score of previous rated movies.

(2) Demographic Filtering

Some of recommender system recommend the movie to the most audience based on the popularity or the genre of the movie. In other word, the system recommends the same movies to the users with similar demographic features. Due to the different habit of the audience, this filtering is considered as a less accuracy system. We treat It as the worst case to compare the accuracy of the first system.

(3) Content Based filtering

The system recommended the similar items based on a particular item. For example, this system uses item metadata, such as genre, director, description, actors, etc. For movies, the general idea behind these recommender systems is that if a person liked a particular item, he or she will also like an item that is similar to it. (user who like particular director or actors will likely be recommended the similar items)

At the second stage, by comparing three model of accuracy and the capacity of the prediction, we may work on the following issues:

- (1) Add cross validation
- (2) Find the best K of KNN algorithm
- (3) Evaluating model accuracy for KNN
- (4) cold start problem,
- (5) popular bias,

(6) data sparsity problem

Detailed of my team shows as below:

Yuke Li z5176261 45

Zanning Wang z5224151 45

Yuankai Zhao z5233940 45

Zhenming Wang z5140192 45