Recommendation Systems - Group Assignment
Technical report

Transformations followed on the section 1:
To create the base matrix for the content based recommendation systems we followed the following steps:

- We used the function dcast from the package reshape2.
- We transformed the classes of the variable TagValue into new columns.
- We counted the number of users for the values of the matrix.
- Each row is being uniquely identified by ArtistID.
- We applied some transformations to the variables month and year.

To create the base matrix for the collaborative filtering recommendation systems we followed the following steps:

- We transformed the values of the variable ArtistID into new columns.
- We counted the number of users for the values of the matrix.
- Each row is being uniquely identified by a user.

Steps followed on the section 2:
For the CF functions in step2, we created 3 functions: 1. CF function on the user basis for one user. 2. CF function on the user basis for all of the users. 3. CF function on the item basis.

The steps we are like the following for the CF function on the user basis for one user:

1. As required by the instruction, we calculated similarity using the Pearson correlation.
2. We then filtered the similarity matrix for the specific user.
3. Using the specific similarity matrix, we selected similar users based on the threshold.
4. Based on the similarities of those users, we then make predictions.

The steps are like the following for the CF function on the user basis for all users:

1. As required by the instruction, we calculated similarity using the Pearson correlation.
2. Using the similarity matrix, we selected similar users based on the threshold.
3. Based on the similarities of those users, we then make predictions.

The steps are like the following for the CF function on the item basis for all users:

1. As required by the instruction, we calculated similarity using the Pearson correlation.
2. Using the similarity matrix, we selected similar items based on the threshold.
3. 3. Based on the similarities of those items, we then make predictions

Argumentations on the use of specific recommendation techniques, evaluation metrics and hybridization techniques.

Content based recommendation systems help to give general recommendations of artists based on their similarity with other artists. Collaborative filtering recommendation systems (CFRS) help to give recommendations based on the preferences of other users and are mostly used by big commercial websites. Our user based CFRS tries to estimate the preference of users for particular artists by analyzing the preferences of similar users. On the other hand, the item based CFRS tries to estimate the preference by analyzing the preferences that users have to similar artists.

There are different ways to evaluate our recommendations but the ones we consider in this project are MAE (mean absolute error) and the F1. MAE measures the average magnitude of the errors in a set of predictions. While F1 is the weighted harmonic mean of the test's precision and recall and ranges from 0 to 1. MAE is frequently used as a common measure of forecast error in time series analysis. On the other hand F1 score is mostly used when the False Negatives and False positives are crucial, also F1 score is a better metric when there are imbalanced classes.

Advantages, disadvantages and improvements of the best recommendation system.

We see that we achieve a better RMSE for hybrid model using linear regression but Cluster based filtering achieves betters results for other evaluation parameters. Henceforth, we chose Cluster based collaborative filtering output as the best recommendation