Optimization of the Hybrid Movie Recommendation System Based on Weighted Classification and User Collaborative Filtering Algorithm

Now-a-days due to increasing demand and access of online information technology leads to major data accumulation which is called as BigData and its become mandatory to process such BigData to provide recommendation to known or unknown users on which items to purchase or which movie to watch. In past many recommendation algorithms was introduced such as Content Based Recommendation (CBR (which matches similarity between user items and then perform recommendation and Collaborative Filtering (this will matches similarity between similar behaviour user and then perform recommendation).

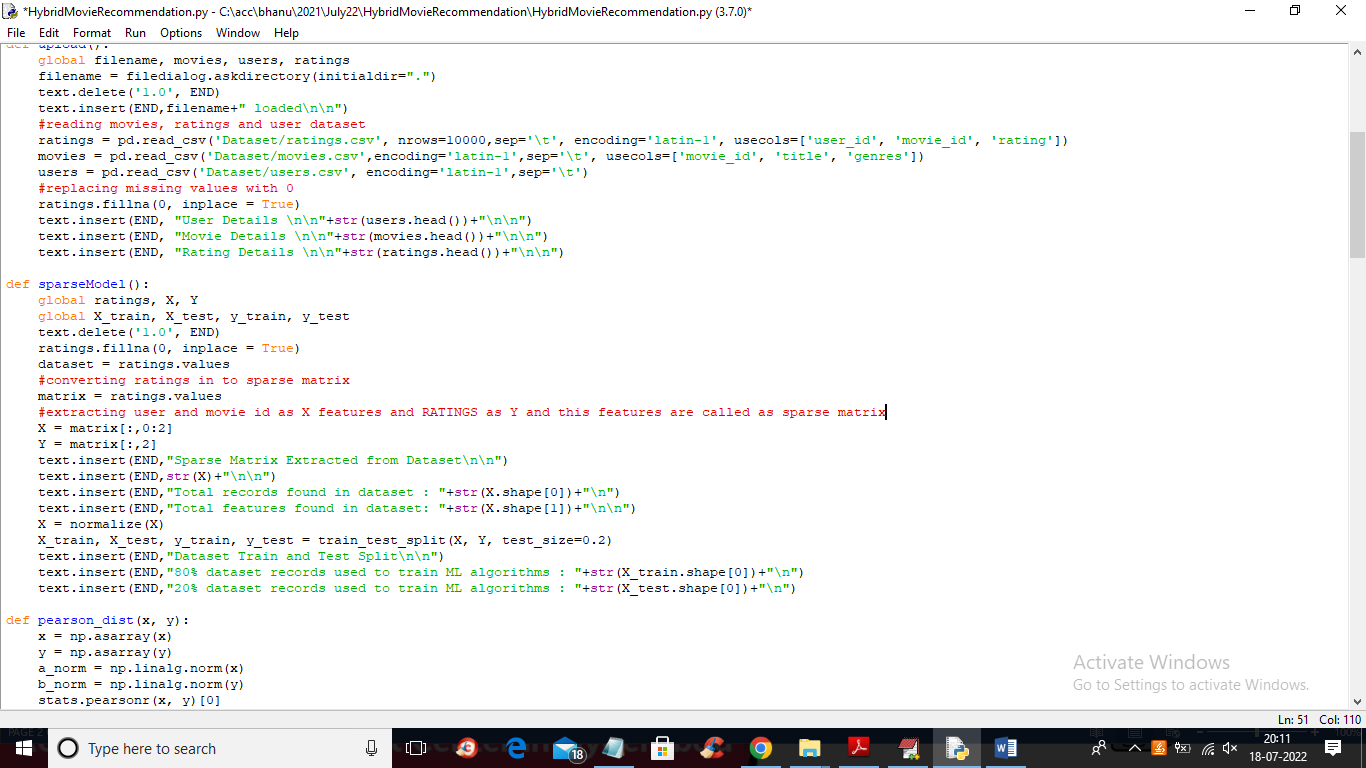
All existing algorithms will generate high dimensional array from dataset by taking matrix with users and their ratings and if dataset contains huge record then matrix contains more ratings columns which lead to high dimensional array and this will consume more time and recommendation will be inaccurate.

To overcome from above problem author of this paper introduced Hybrid Recommendation with weighted classification. Propose algorithm consists of following modules

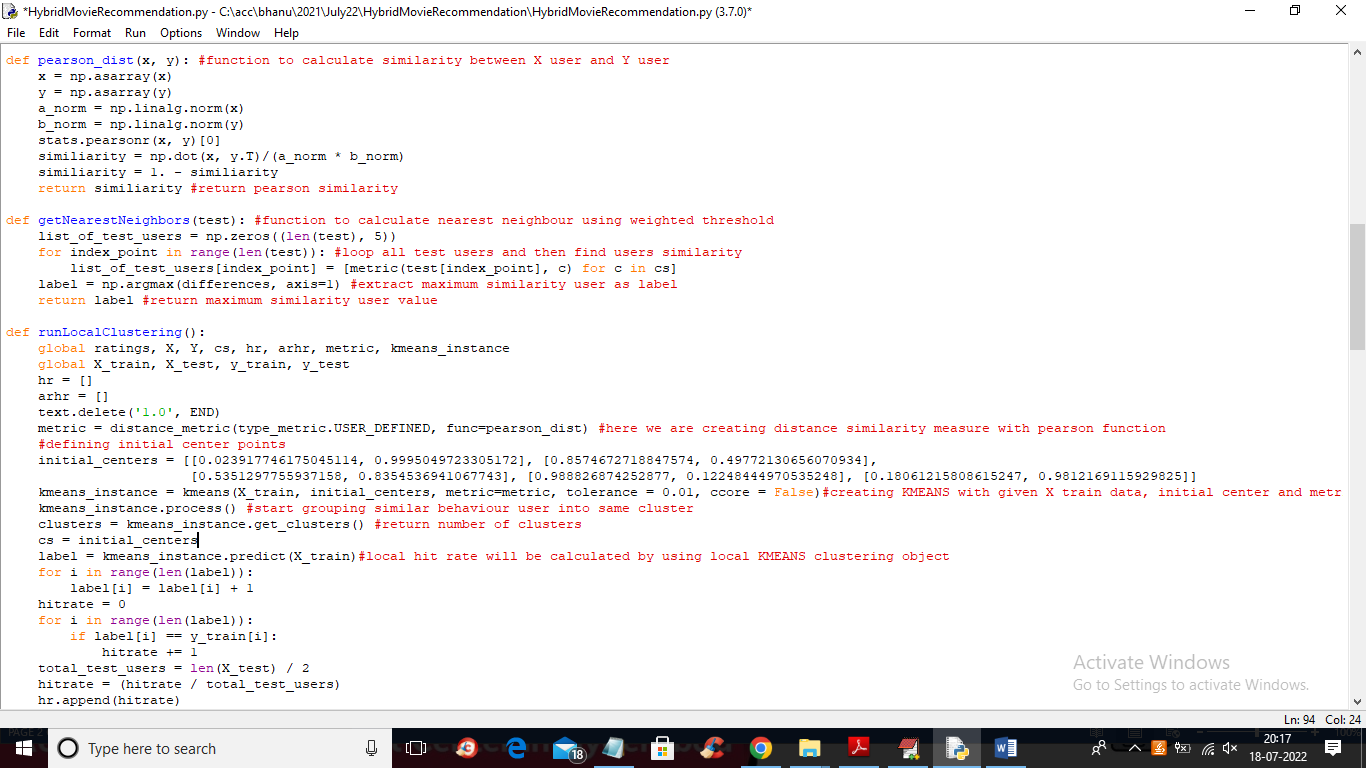
1. Sparse Linear Model: using this module we will extract required features from dataset such as User\_ID, Movie\_ID and ratings to build basic recommendation model. Extracted features will be split into TRAIN and TEST data where application will use 80% dataset size for training and 20% dataset size for testing.
2. Local Recommendation Model Training: extracted sparse features will get trained with clustering algorithm called KMEANS. KMEANS will group similar behaviour users in to same cluster. Similar behaviour user’s similarity will be calculated using Pearson Spearman function. KMEANS trained model will be applied on 20% test data to predict LOCAL HIT RATE (refers to correct number of ratings prediction) and Average hit rate.
3. Top-N personalized recommendation using Weighted Classification: In this module we will take USER ID & MOVIE ID as input and then extract Cluster Centres from KMEANS and then NEAREST NEIGHBORS will be calculated between INPUT and CLUSTER CENTRES by using PEARSON formula and this formula will return MAX Threshold similarity users and their movies as RECOMMENDATION LIST. Calculating correct rating prediction using NEAREST NEIGHBORS is known as Global HIT RATE and its average is called as Average Hit Rate.

In above 3 modules just we have use 3 features (user id, movie id and ratings) for clustering and recommendation so it will be called as LOW dimension array and solved the problem of High dimensional array.

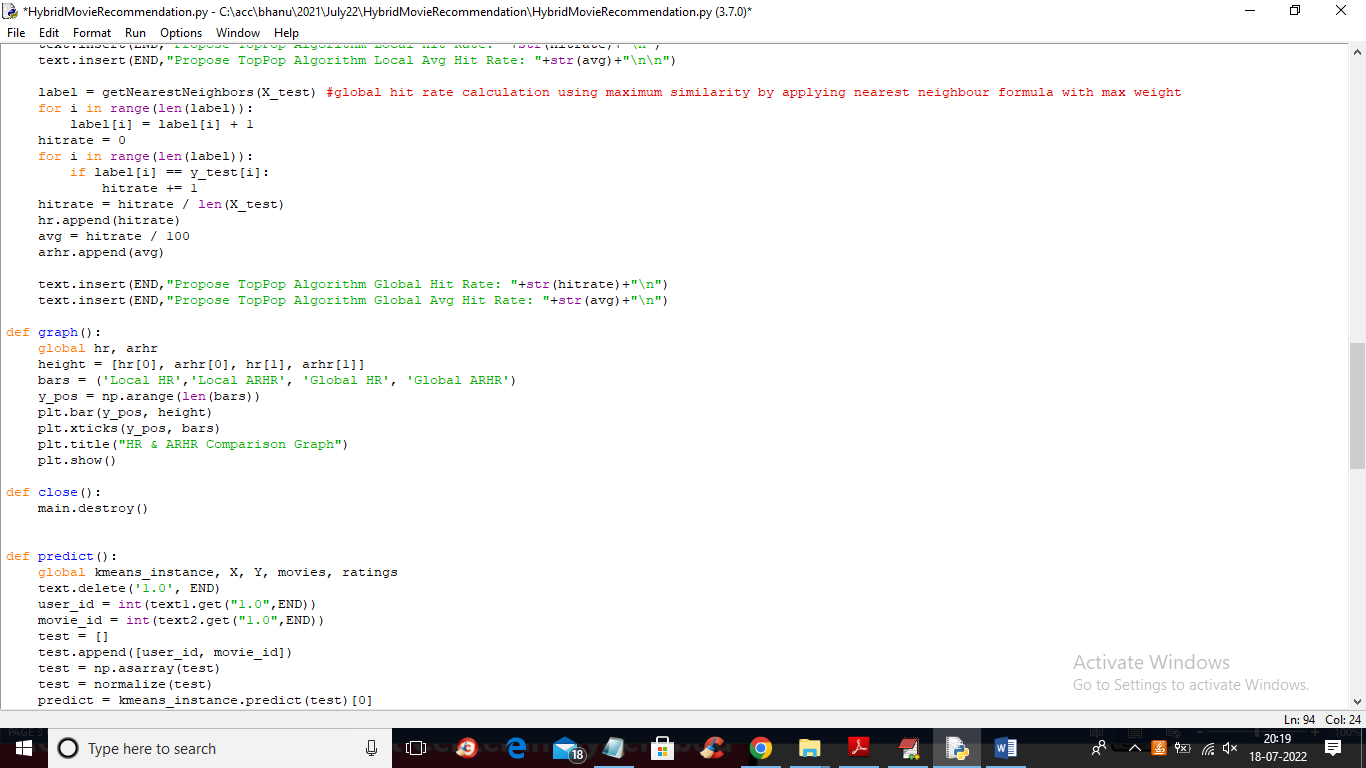
In below screen we are showing code which implemented above 3 modules



In above screen read red colour comments to know sparse features model extraction from dataset. In below screen showing code for PEARSON formula and KMEANS algorithm



In above screen read red colour comments to know about PEARSON, KMEANS and local hit rate calculating using KMEANS prediction. In below screen we are calculating GLOBAL HIT RATE using nearest NEIGHBOR



In above screen at red colour line you can see we are finding similar behaviour user using Nearest Neighbour function.

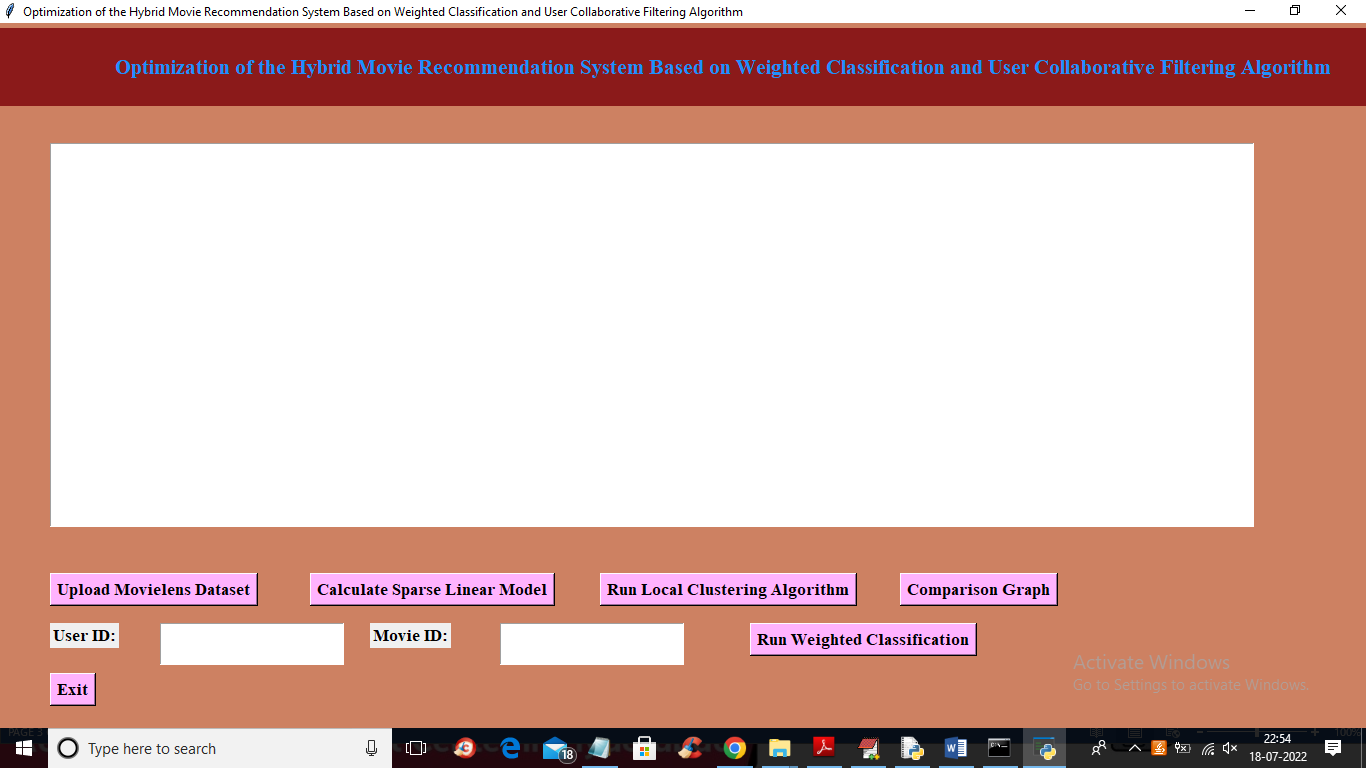
MODULES DETAILS

To implement this project we have designed following modules

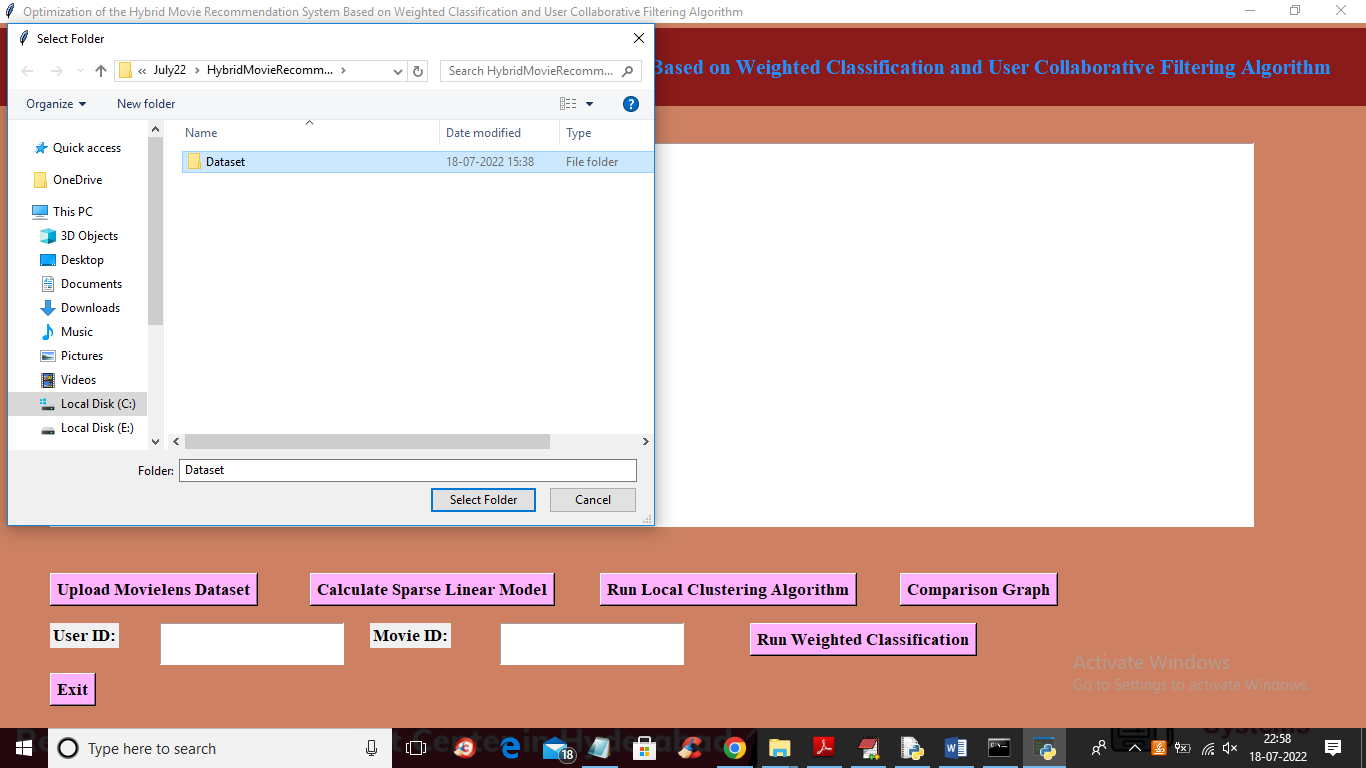
1. Upload Movielens Dataset: using this module we will upload dataset to application
2. Calculate Sparse Linear Model: using this module we will read dataset and then build SPARSE LINEAR features model
3. Run Local Clustering Algorithm: sparse features will be input to KMEANS to group similar user into same cluster by using PEARSON function
4. Comparison Graph: using this module we will plot local and global HR, ARHR graph
5. Run Weighted Classification: using this module we will accept User and Movie ID and then using weighted classification will predict movie recommendation

SCREEN SHOTS

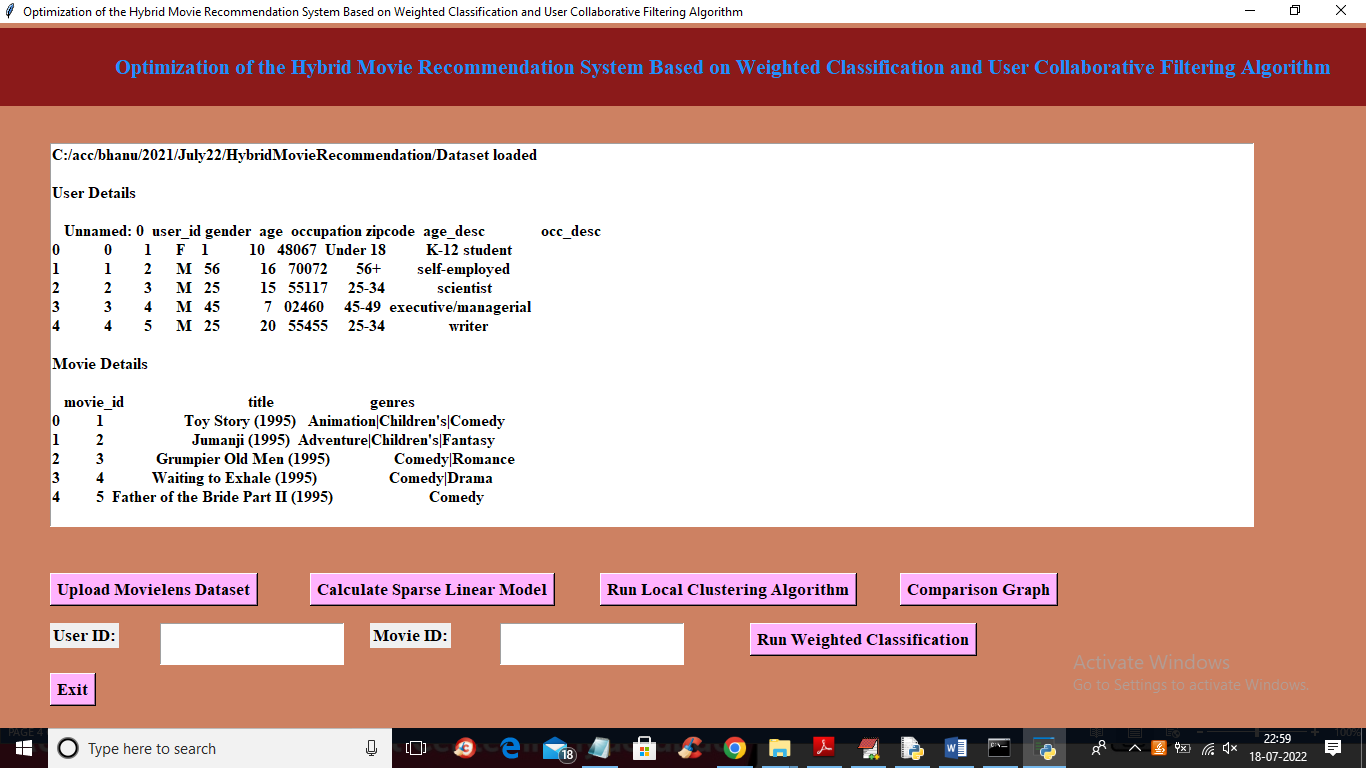
To run project double click on ‘run.bat’ file to get below screen



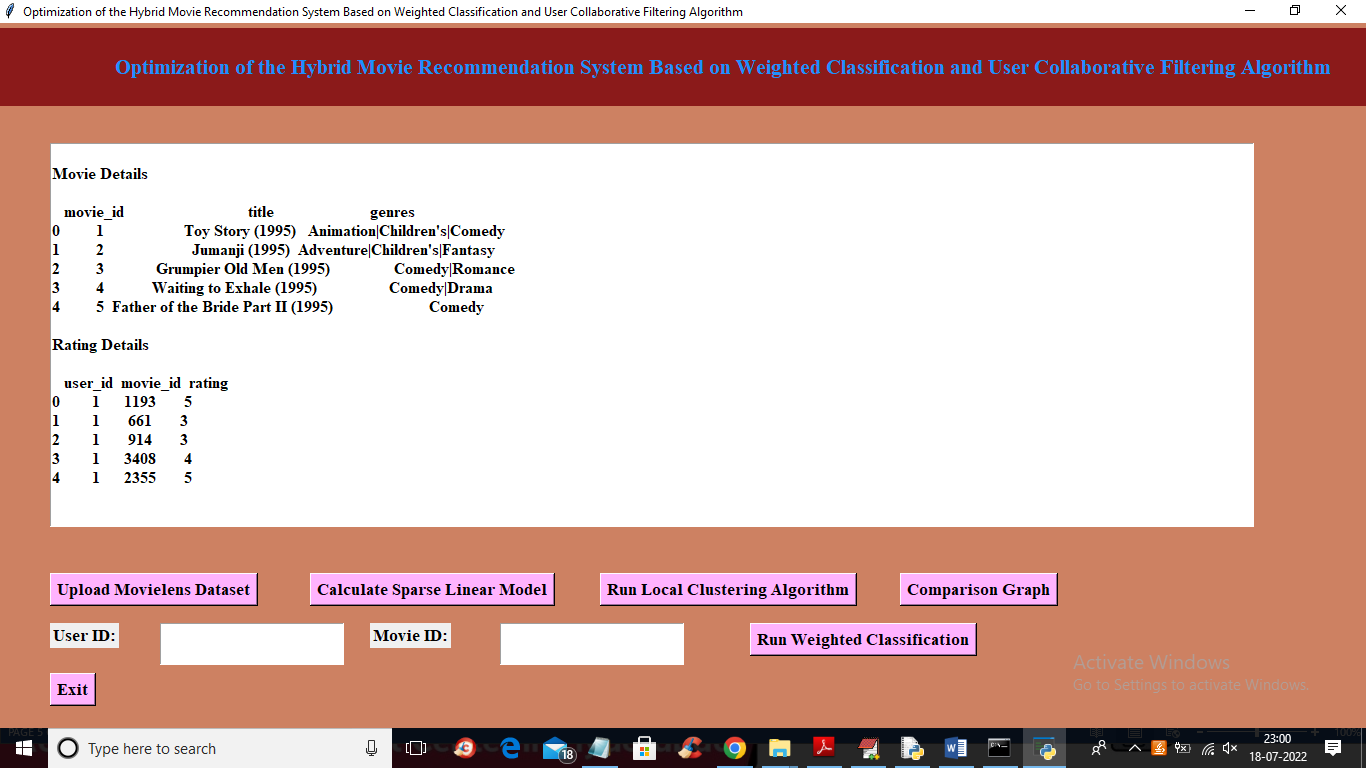
In above screen click on ‘Upload Movie Lens Dataset’ button to upload dataset and get below output



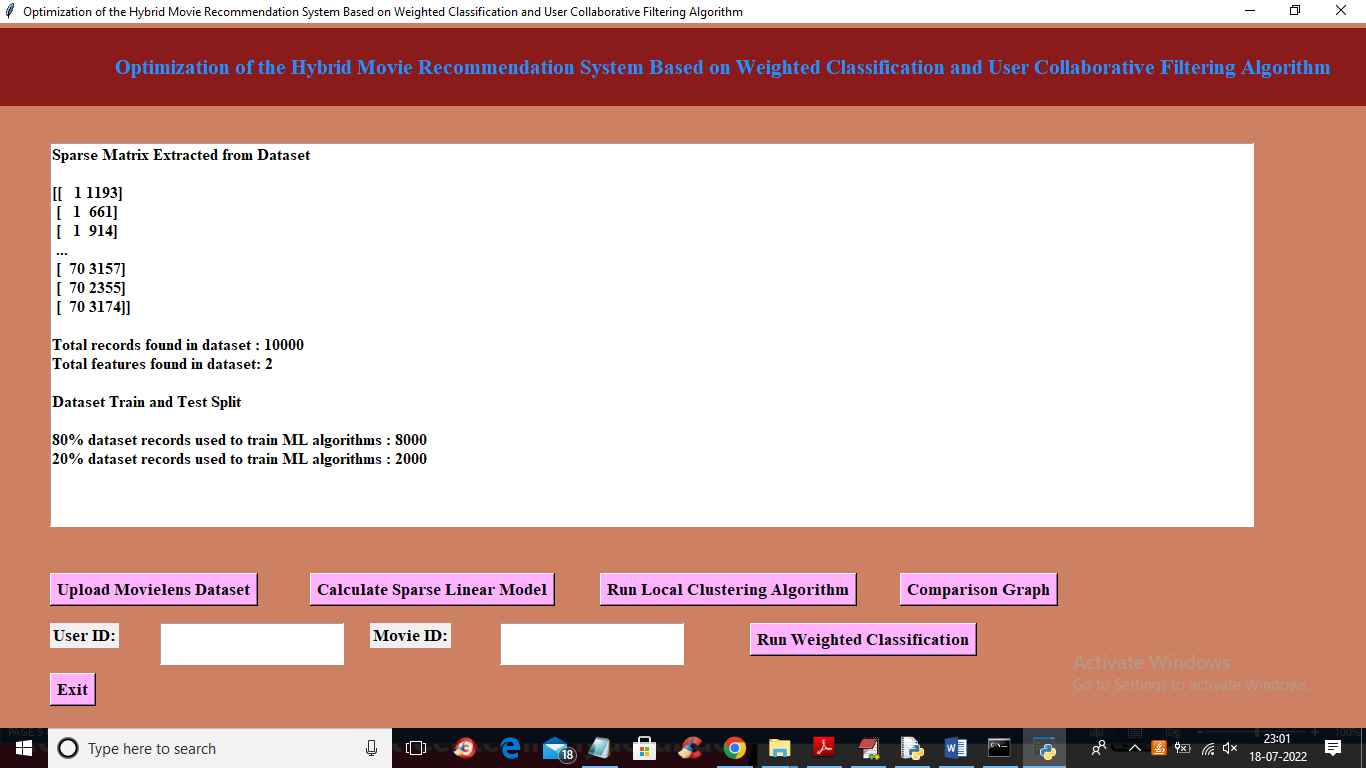
In above screen selecting and uploading ‘Dataset’ folder and then click on ‘Select Folder’ button to upload dataset and get below output



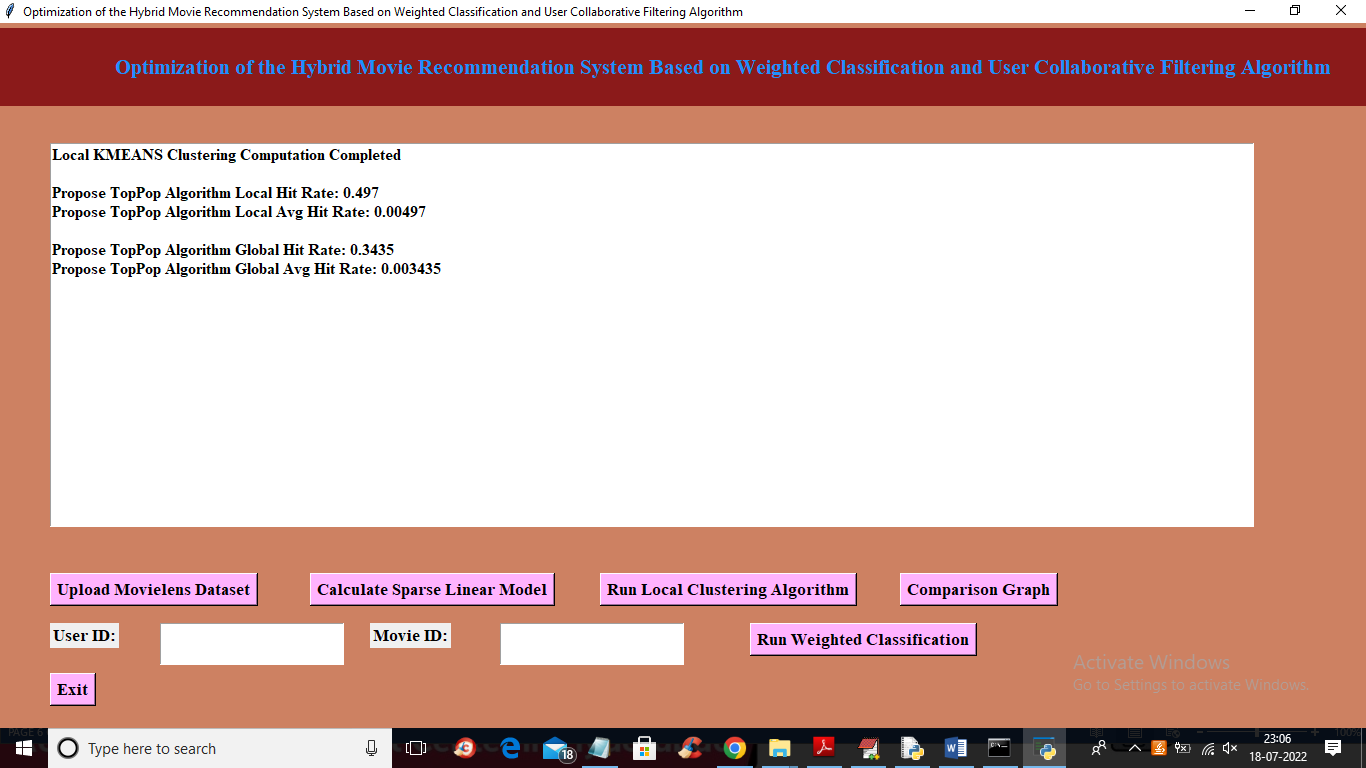
In above screen user details and movie details are loaded and scroll down above screen to view RATING details



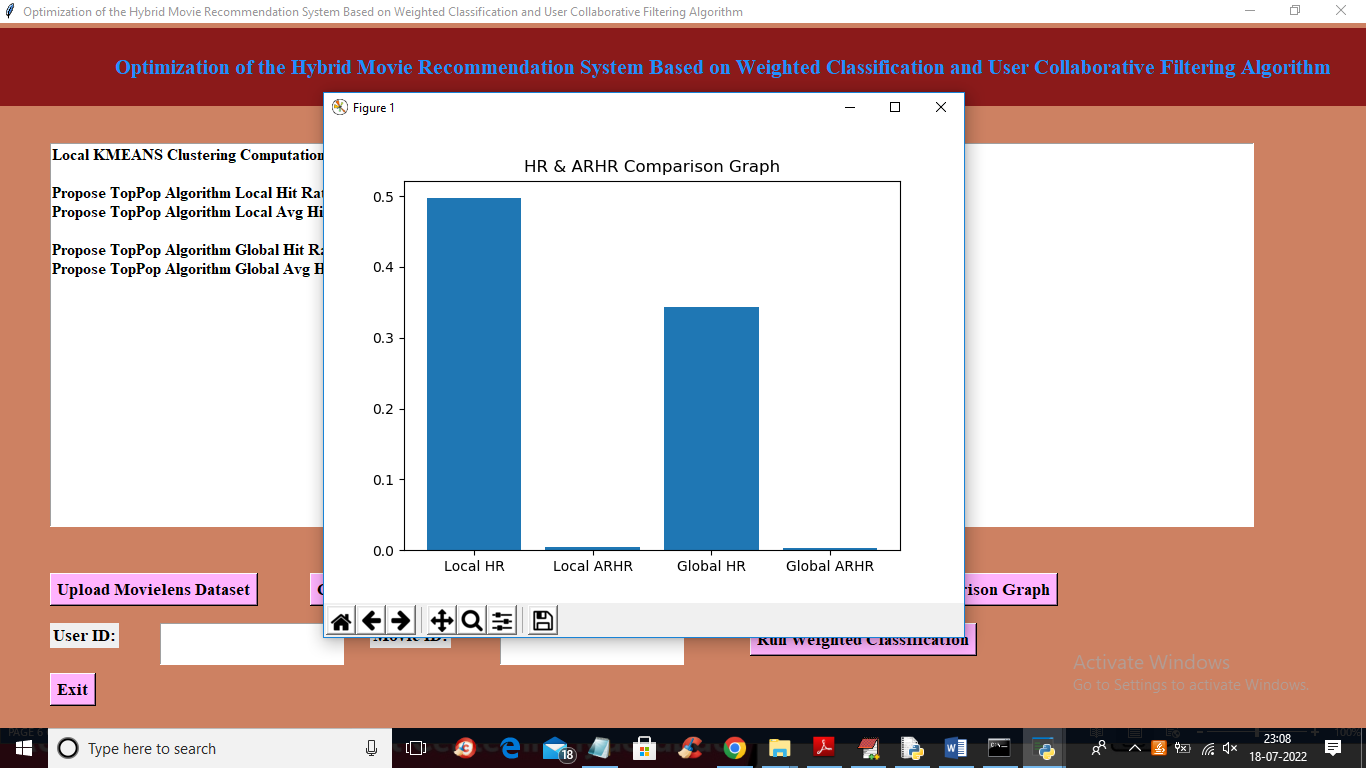
In above screen Rating details also loaded and now click on ‘Calculate Sparse Linear Model’ button to extract sparse features and get below output



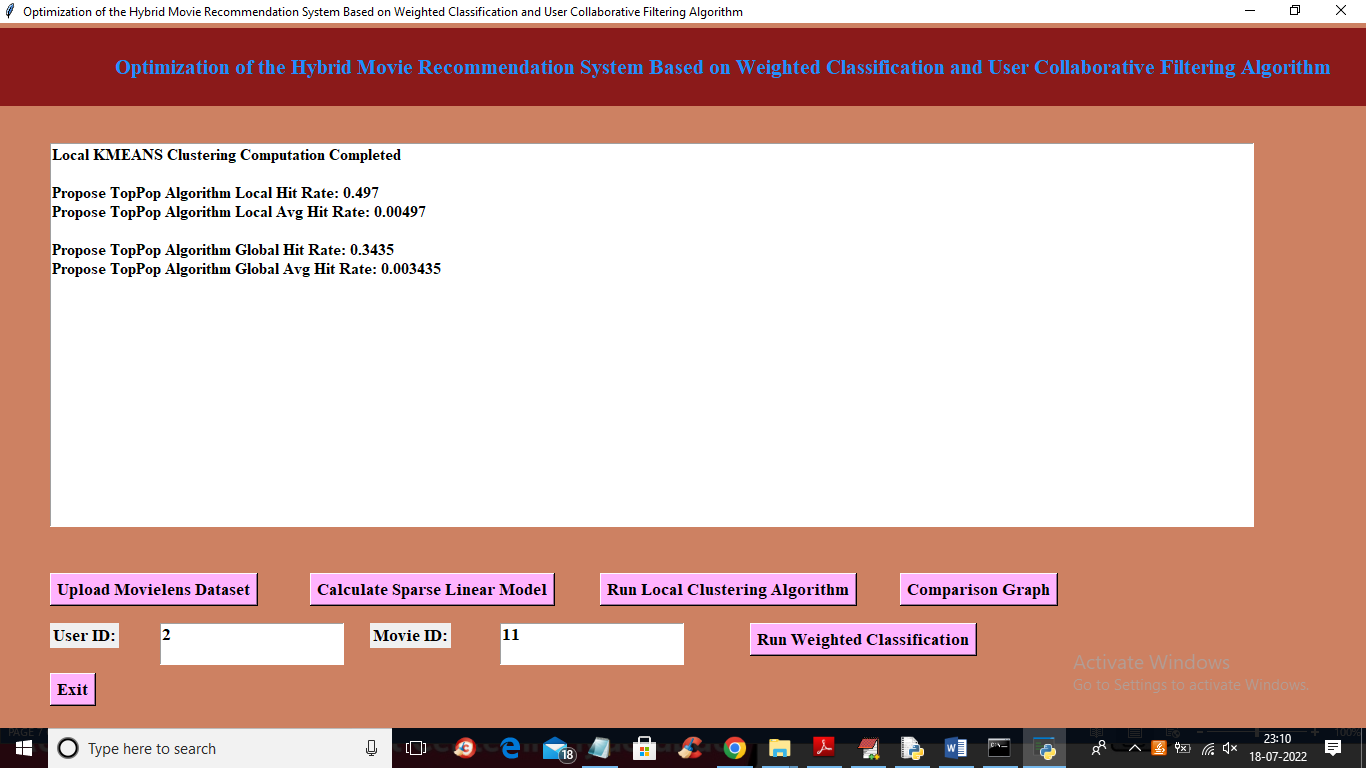
In above screen sparse in square bracket we can see sparse matrix features and then we can see dataset size and then train and test data split details and now dataset is ready and now click on ‘Run Local Clustering Algorithm’ to build clusters and get below output



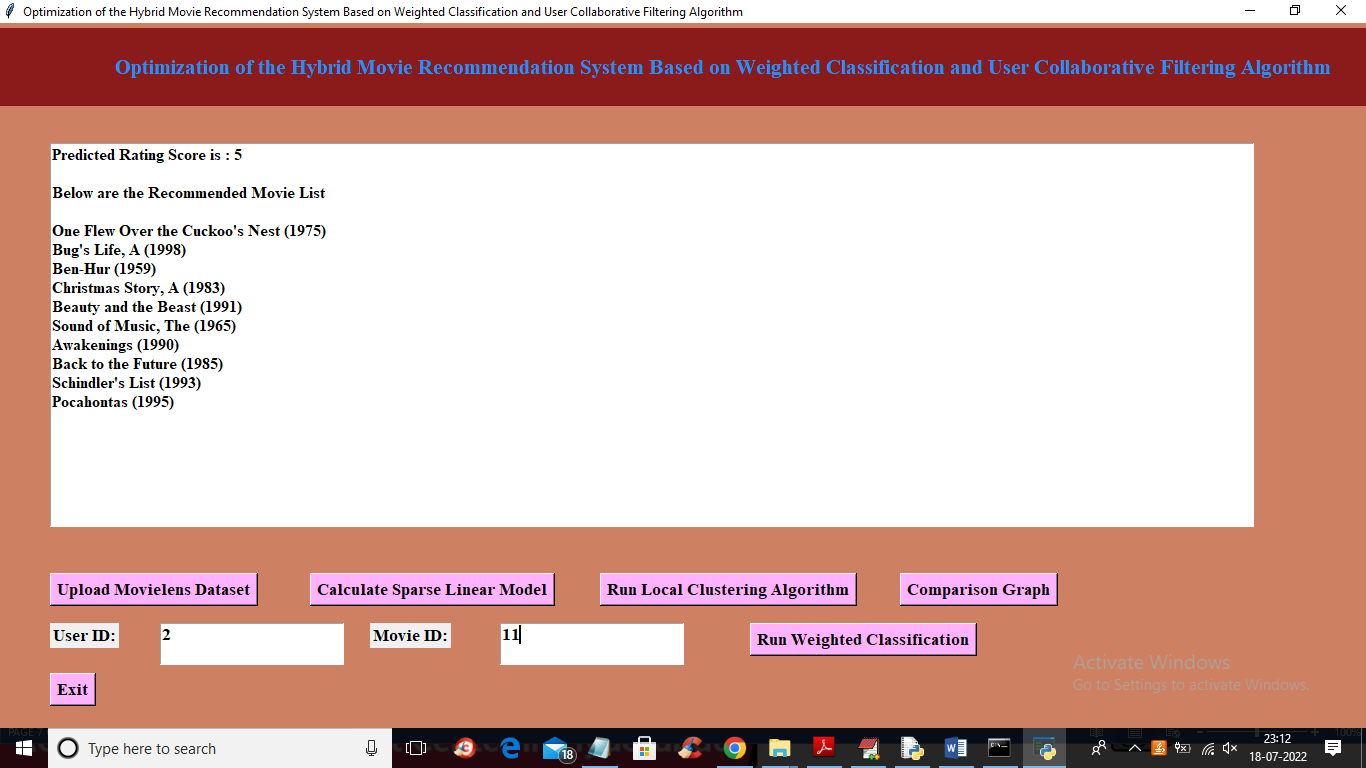
In above screen clustering task completed and we can see local and global HIT Rate and now click on ‘Comparison Graph’ button to plot local and global comparison graph



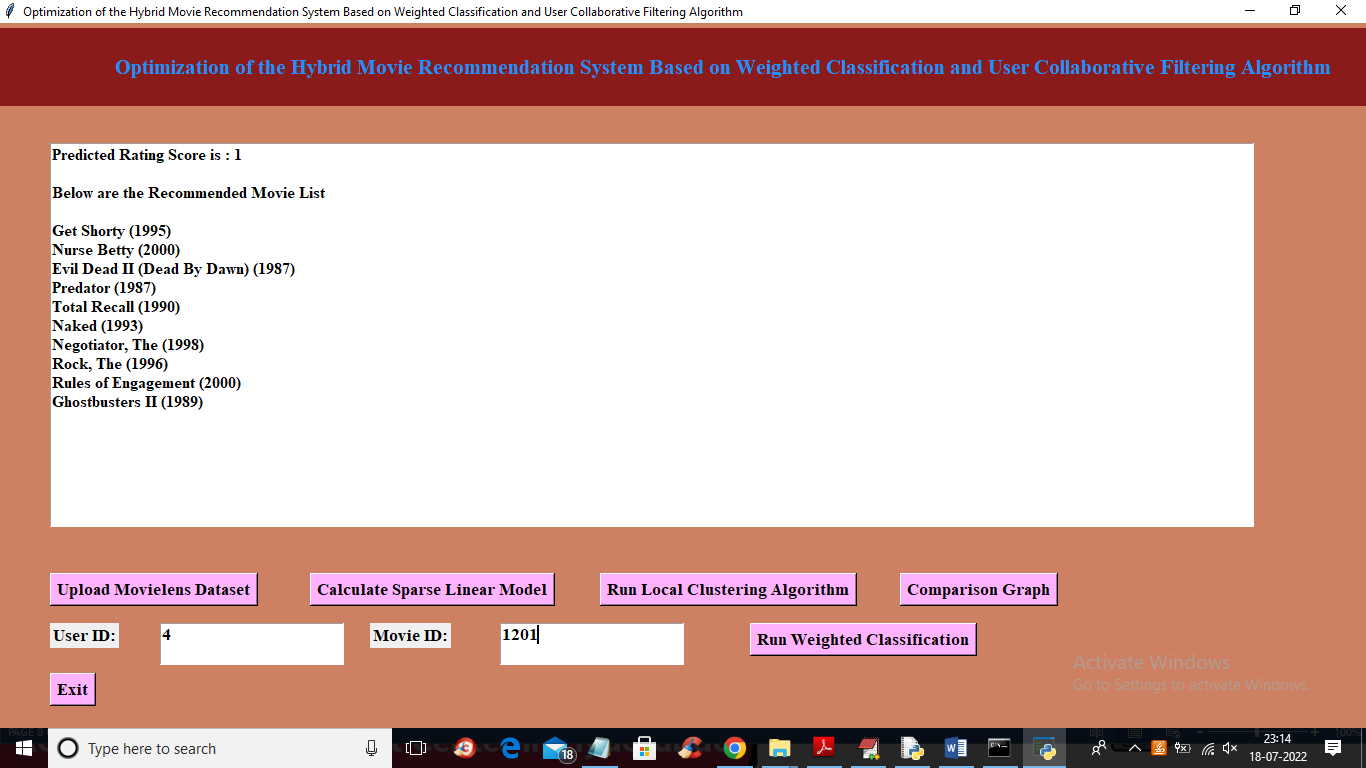
In above screen x-axis contains local and global Hit Rate and y-axis contains values and now enter user id and movie and then click on ‘Run Weighted Classification’ button to predict move recommendation



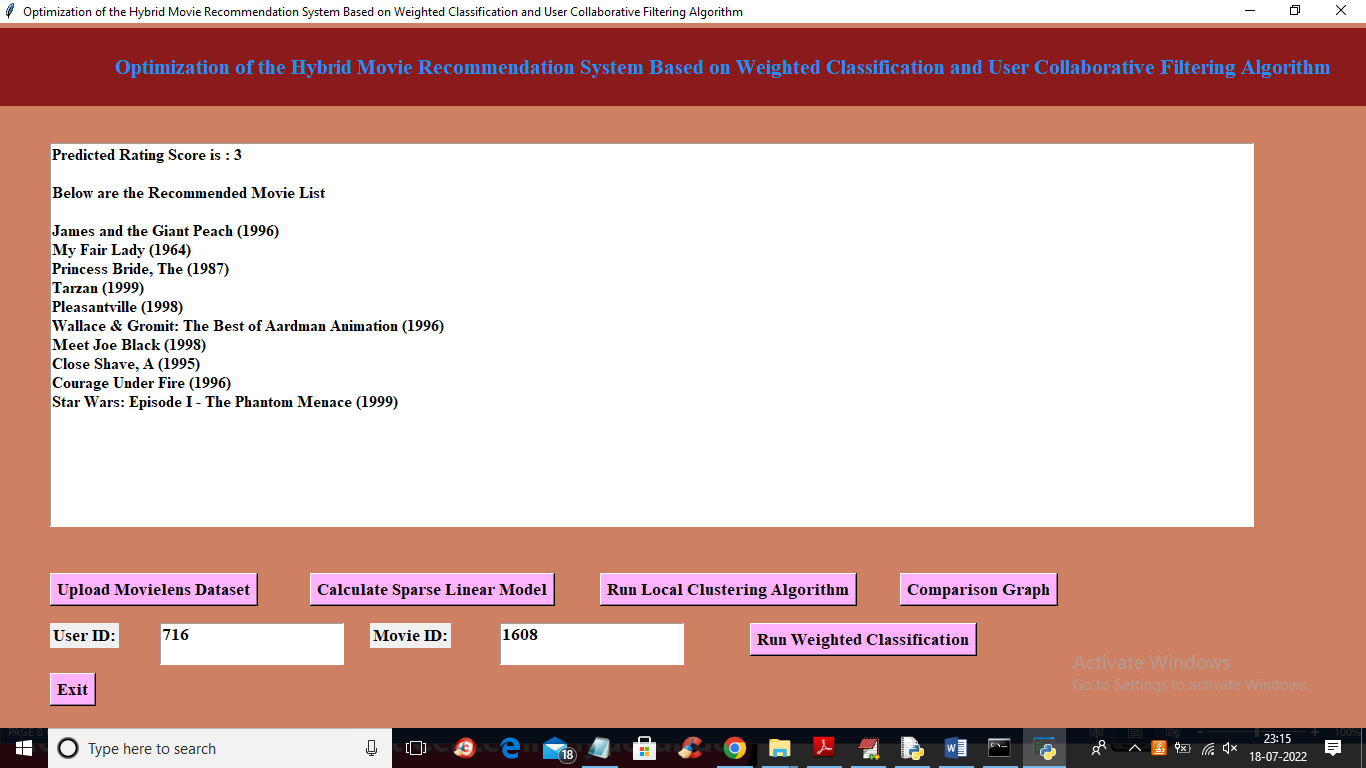
In above screen I entered user ID as 2 and Movie ID as 11 and then click on ‘Run Weighted Classification’ button to get below recommendation list

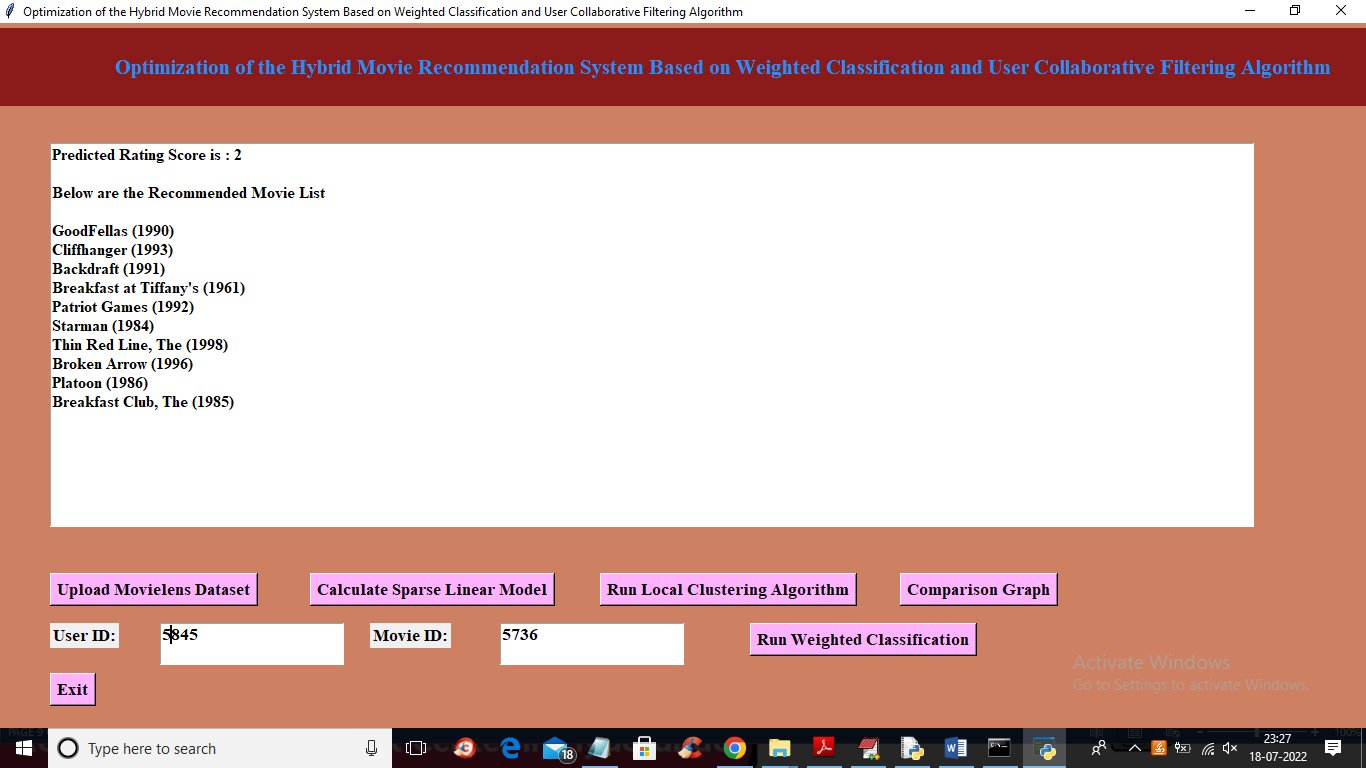


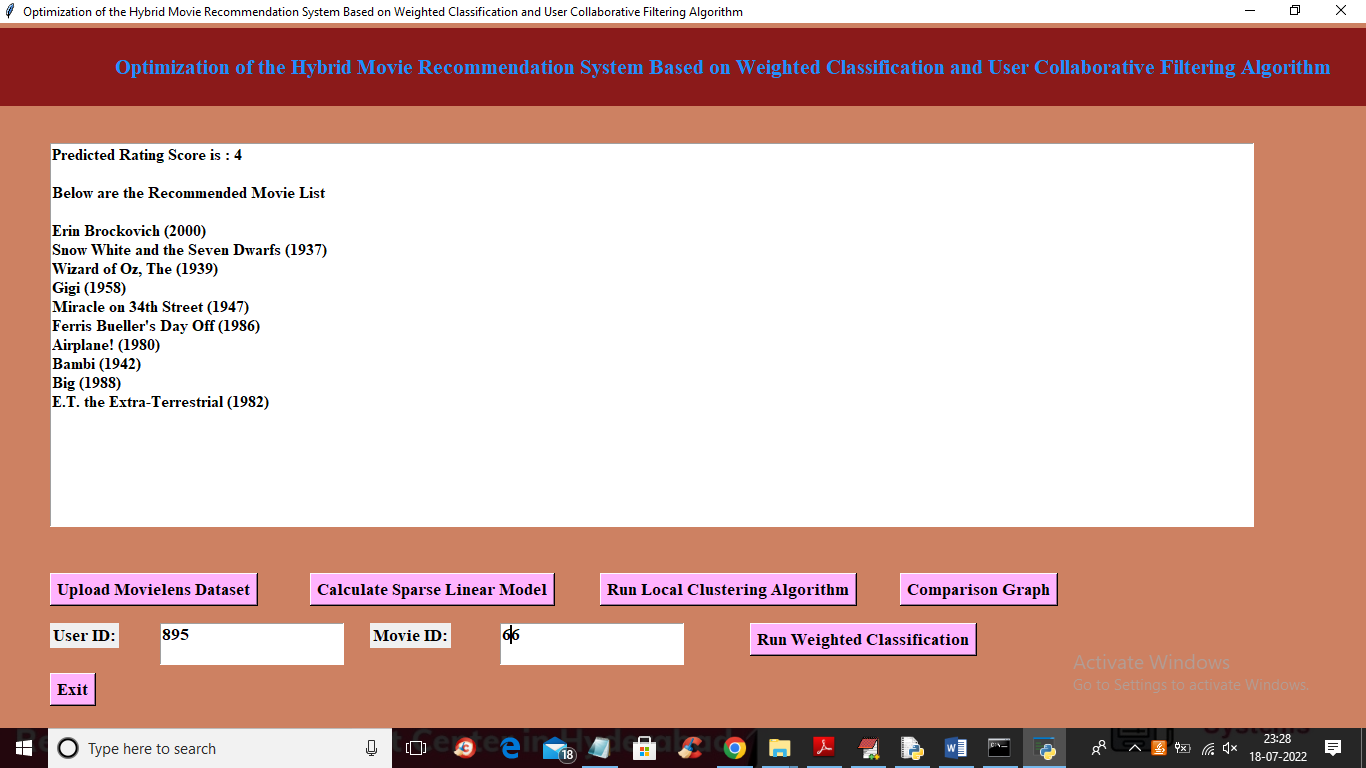
In above screen for given user id and movie ID we got predicted rating as 5 and then based on predicted rating score and user behaviour displaying top 10 recommended movies. Similarly enter any user id and movie id to get recommendation list



In above screen for 4 and 1201 we got above output







Similarly enter user and movie id and get recommendation list