* Function description

|  |  |
| --- | --- |
| Function Name | Description |
| print.time | It prints current system time. It is just debugging function which is used to measure performance of the code. |
| find.avg.rating.by muv | It is used to calculate average rating of each movie by movie.id. It is used to build naïve model |
| find.corr.user | It is used to find correlation between users for different algorithms like Euclidean, manhattan and Lmax for different training set. |
| predict.rating | This function is used to predict the rating for a particular movie watched by a user. Here, I have taken weighted average of user correlation and rating. ∑ (Corr value \* rating by similar user) / ∑ (Corr. value) |
| predict.over.testdata | This function is used to iterate all user-movie pairs from test data and pass one by one to predict.rating function to predict rating for each pairs. It returns testdata with predicated value by all algorithms |
| predict.over.testdata.with.age | This function is used to iterate all user-movie pairs from test data and pass one by one to predict.rating function. Here, we will only predict rating for best k value for 3 algorithms. |
| find.avg.rating.by.muv.10m | It is used to find average rating for each movie present in 10M dataset. |
| predict.over.testdata.10m | It is used to predict rating over different testsets in 10M data. |
| find.corr.user.10m | This function is used to calculate correlation factor for 10m data. Here, if there are less than 4 movies common among user pairs then |
| find.corr.over.train.data.10m | Here, we cannot follow the same method to find correlation for 100k because we cannot create matrix of 71k x 71k in any of the language on average machine. So, I am splitting data by user and finding correlation for 10m data. |

* Guidelines
* **Flag IsDebugMode:**
  + True <- In this case, program will not do any kind of computation but read existing data saved in csv files. **(Precomputed CSV files will be provided on request.(800 MB files))**
  + False <- In this case, complete system will run and it will compute everything from scratch
* **Flag Is10MSupported:**
  + True <- If you want to run this recommendation system on supercomputer or fast server, choose this option
  + False <- For normal computation, In this option, it will take a part of data and compute rating
* For 100k data set, put all data in **ml-100k** directory
* For 10M data set, put all data in **ml-10M100K\Clean** directory.
* For 10M data set, please use **Formal10MData.sh** to format all data as “::” separator is not supported by R. so this script will replace “::” to “,”
* There is a discrepancy between training and test data set for 10M data. Please provide correct files in proper directory and current system will able to process everything.