a) First let's look at the ratings.csv file. The first few columns of ratings.csv is as shown in the picture below:
##

	userId	movieId	rating	timestamp
0	1	31	2.5	1260759144
1	1	1029	3.0	1260759179
2	1	1061	3.0	1260759182
3	1	1129	2.0	1260759185
4	1	1172	4.0	1260759205

##

b) MAE, Mean Absolute Error measures the average magnitude of the errors in a set of predictions RMSE, Root Mean Squared Error is a quadratic scoring that measures the average magnitude of the error.

RMSE will always be larger or equal to the MAE. The greater the difference between them, the greater the variance in the individual errors in the sample. Both measures can range from 0 to infinity. Lower the values of MAE and RMSE, the better.

c) RMSE and MSE for Probabilistic Matrix Factorization (PMF) is as shown below:
##
Output:

```
RMSE and MAE for five-fold PMF:
{'test_mae': array([0.76888102, 0.77694119, 0.78394215, 0.78330379, 0.78947437]), 'test_
rmse': array([0.99355912, 1.00720332, 1.01273677, 1.01490639, 1.02196274]), 'fit_time':
(3.1727535724639893, 3.1059951782226562, 3.1052229404449463, 3.1319563388824463, 3.12854
2900085449), 'test_time': (0.144212007522583, 0.1401071548461914, 0.12485003471374512, 0
.1422138214111328, 0.1229918003082754)}
```

RMSE and MAE for five-fold PMF:
{'test_mae': array([0.76888102, 0.77694119, 0.78394215, 0.78330379,
0.78947437]),
'test_rmse': array([0.99355912, 1.00720332, 1.01273677, 1.01490639,
1.02196274]),
'fit_time': (3.1727535724639893, 3.1059951782226562,
3.1052229404449463, 3.1319563388824463, 3.128542900085449),
'test_time': (0.144212007522583, 0.1401071548461914,
0.12485003471374512, 0.1422138214111328, 0.1229918003082754)}

##
RMSE and MSE for User Based Collaborative Filtering is as shown below:

##
Output:

```
RMSE and MAE for five-fold User Based Collaborative Filtering:
{'test_mae': array([0.7141557 , 0.7064661 , 0.70278552, 0.70853913, 0.70401333]), 'test_
rmse': array([0.93101601, 0.91979979, 0.92004312, 0.92286234, 0.92019265]), 'fit_time':
(0.2722916603088379, 0.276334285736084, 0.28095579147338867, 0.2749972343444824, 0.27584
52892303467), 'test_time': (1.305346965789795, 1.2526752948760986, 1.2655532360076904, 1
.2794365882873535, 1.250403881072998)}
```

RMSE and MAE for five-fold User Based Collaborative Filtering:
{'test_mae': array([0.7141557 , 0.7064661 , 0.70278552, 0.70853913,
0.70401333]),
'test_rmse': array([0.93101601, 0.91979979, 0.92004312, 0.92286234,
0.92019265]),
'fit_time': (0.2722916603088379, 0.276334285736084,
0.28095579147338867, 0.2749972343444824, 0.2758452892303467),
'test_time': (1.305346965789795, 1.2526752948760986,
1.2655532360076904, 1.2794365882873535, 1.250403881072998)}

##
RMSE and MSE for Item Based Collaborative Filtering is as shown below:
##
Output:

```
RMSE and MAE for five-fold Item Based Collaborative Filtering:
{'test_mae': array([0.69668777, 0.6834048 , 0.69497176, 0.68137891, 0.68707677]), 'test_
rmse': array([0.91582721, 0.89923181, 0.91691483, 0.89652777, 0.90592884]), 'fit_time':
(7.029903888702393, 6.840514659881592, 6.929088354110718, 6.81726861000061, 6.9107584953
308105), 'test_time': (4.7717673778533936, 4.765731334686279, 4.694500923156738, 4.67408
3471298218, 4.624952793121338)}
```

{'test_mae': array([0.69668777, 0.6834048 , 0.69497176, 0.68137891,
0.68707677]),
'test_rmse': array([0.91582721, 0.89923181, 0.91691483, 0.89652777,
0.90592884]),
'fit_time': (7.029903888702393, 6.840514659881592, 6.929088354110718,
6.81726861000061, 6.9107584953308105), 'test_time':
(4.7717673778533936, 4.765731334686279, 4.694500923156738,
4.674083471298218, 4.624952793121338)}

##
d) The average mean value of PMF, User based and Item Based collaborative filtering:
##
Output:

```
Average MAE of Probabilistic Matrix Factorization (PMF):  0.7805085024399361
Average RMSE of Probabilistic Matrix Factorization (PMF):  1.0100736692057644
```

```
Average MAE of User based Collaborative Filtering:  0.7071919558018237
Average RMSE of User based Collaborative Filtering:  0.9227827816757037
```

```
Average MAE of Item based Collaborative Filtering:  0.6887040013969324
Average RMSE of Item based Collaborative Filtering:  0.9068860902612874
```

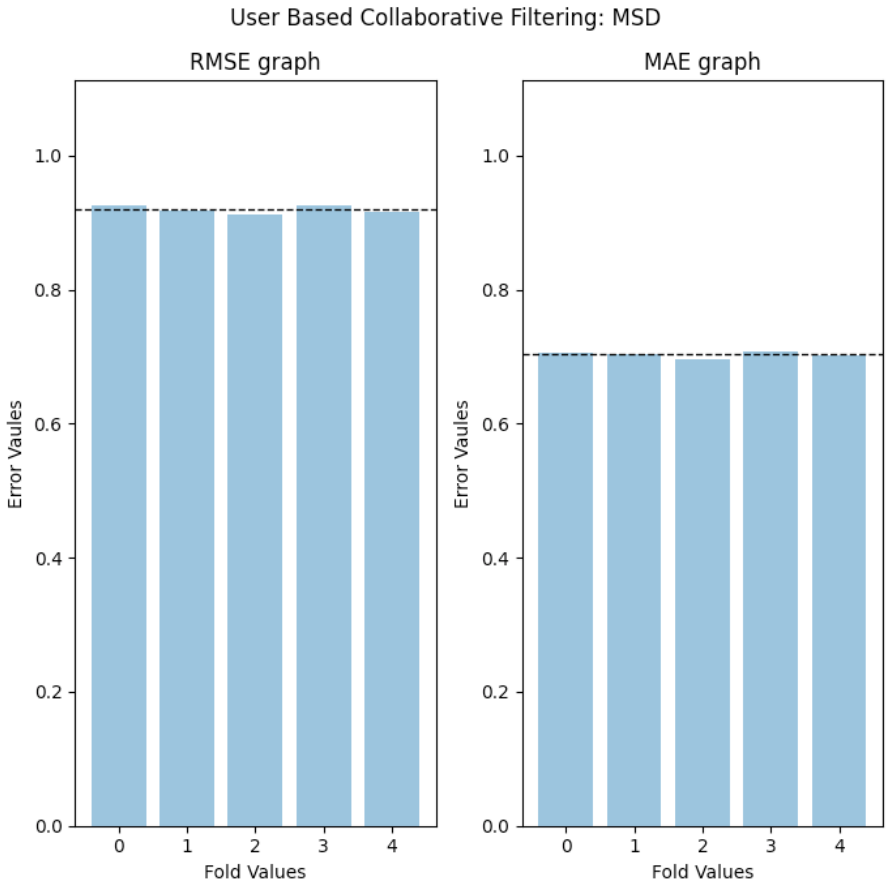
##

Lower MAE and RMSE indicates better fit. Hence, based on the average MAE and RMSE values, Item based Collaborative Filtering has a better fit.

e) The following are the graphs for MSE and RMSE over five-fold cross validation for Item and User based collaborative filtering with MSD, Pearson and Cosine distance. The dashed line represents the mean value for the corresponding value.

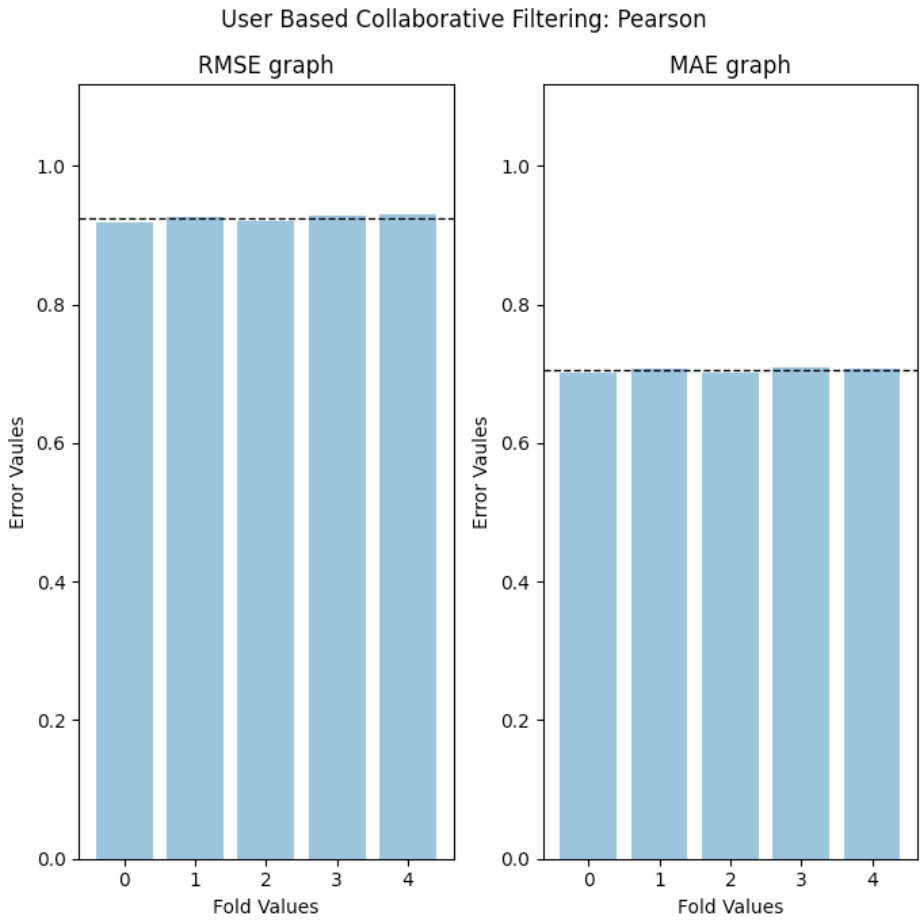
1) User-based Collaborative filtering using MSD:
##
Output:

```
RMSE:  [0.91995992 0.91516118 0.91864677 0.92968895 0.91343338]
RMSE mean:  0.919378040396331
MAE:  [0.70246464 0.69703964 0.70412455 0.71342336 0.69828168]
MAE mean:  0.70306677371561O9
```



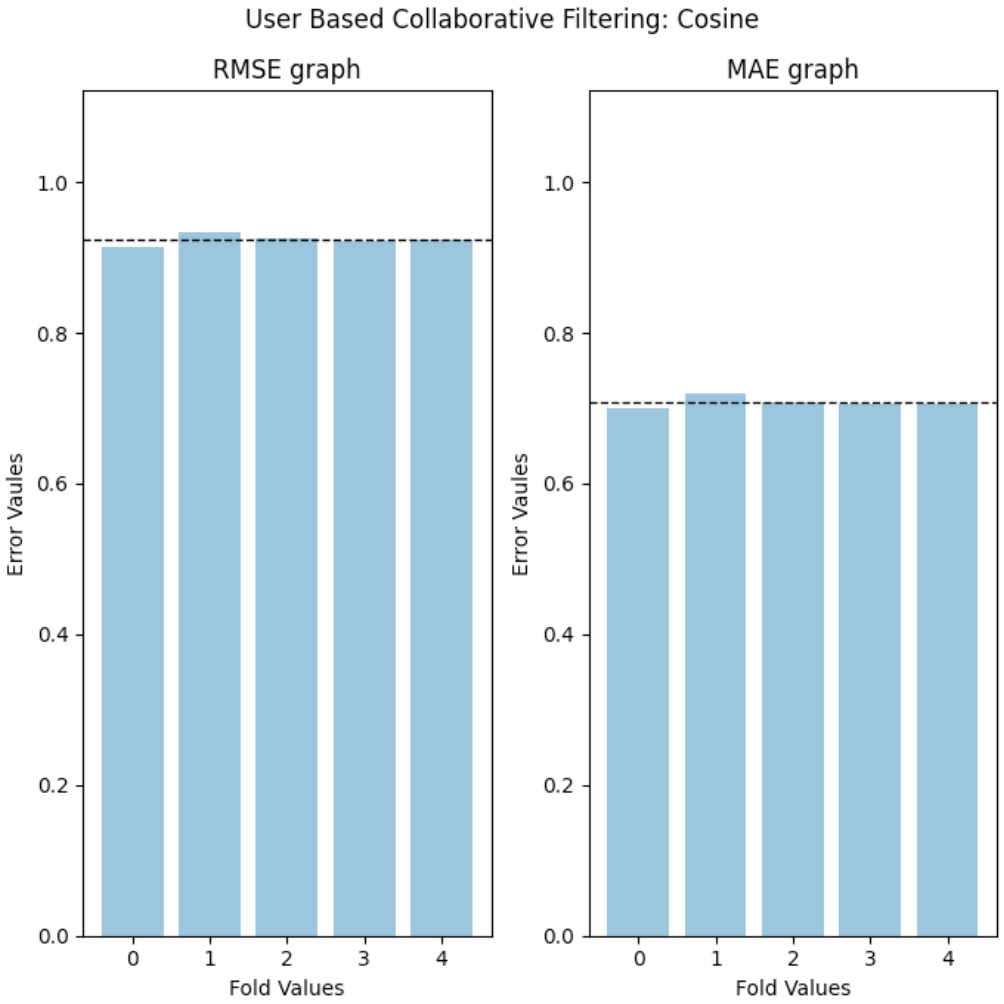
##
2) User-based Collaborative filtering using Pearson similarity:
##
Output:

```
Done computing similarity matrix.
RMSE:  [0.91859457 0.92663286 0.92087915 0.92785224 0.93102311]
RMSE mean:  0.9249963866311152
MAE:  [0.70049484 0.70725768 0.70079237 0.71013973 0.70793123]
MAE mean:  0.7053231697587835
```



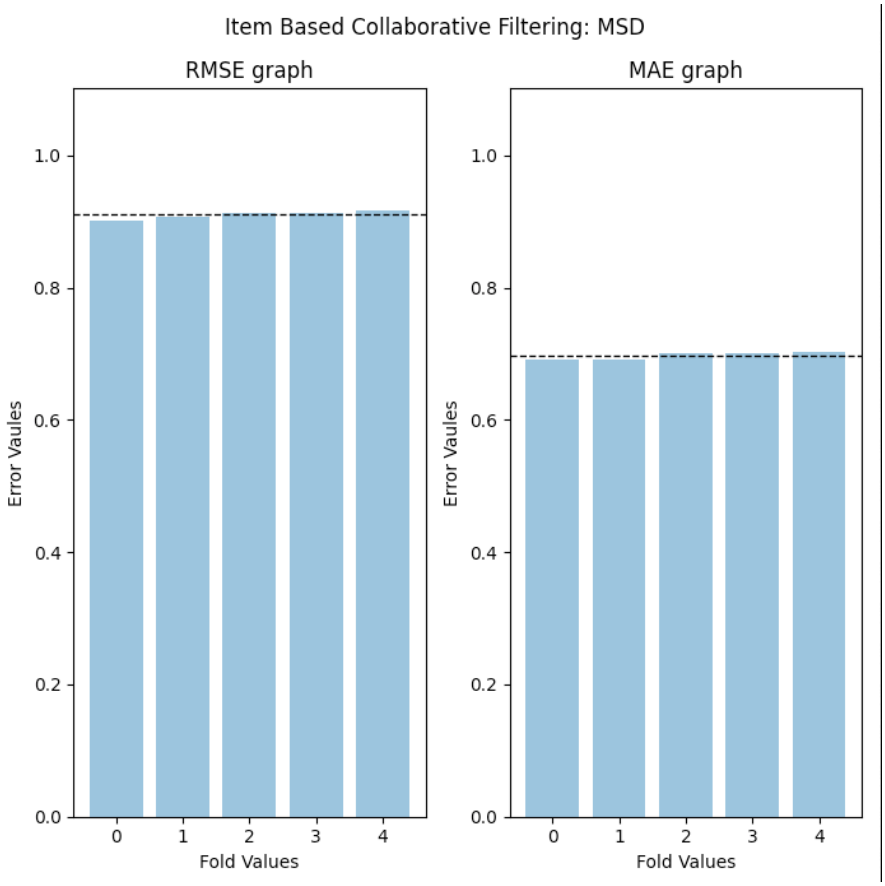
##
3) User-based Collaborative filtering with Cosine similarity:
##
Output:

```
RMSE:  [0.91382918 0.93446099 0.92477793 0.92152215 0.92343715]
RMSE mean:  0.9236054800674249
MAE:  [0.70051696 0.7188593  0.70698608 0.70598475 0.70561073]
MAE mean:  0.707591565830958
```



##
4) Item based Collaborative Filtering with MSD:
##
Output:

```
RMSE:  [0.9017048  0.90802927 0.91362797 0.91276238 0.91781909]
RMSE mean:  0.9107887016237737
MAE:  [0.69143847 0.69217989 0.70029861 0.70047108 0.70323328]
MAE mean:  0.6975242643933586
```



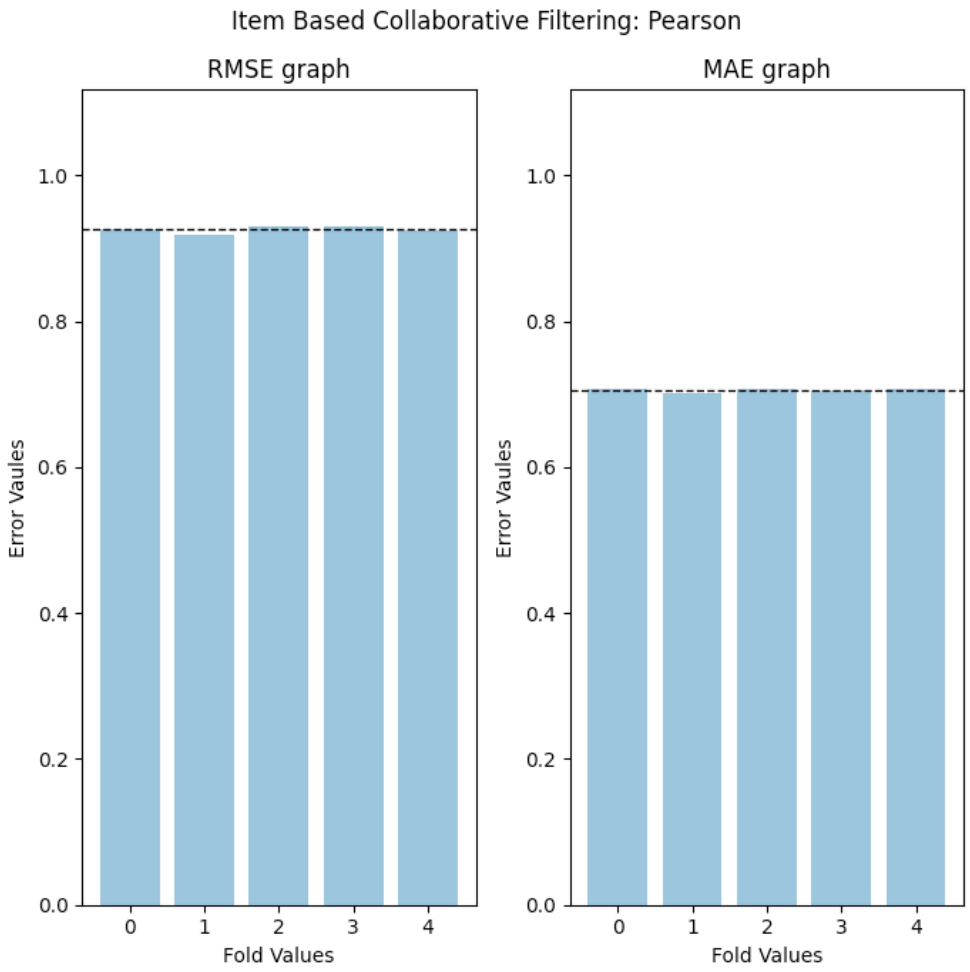
##

5) Item based Collaborative Filtering with Pearson similarity:

##

Output:

```
RMSE:  [0.92610148 0.91765773 0.92963856 0.93115523 0.924585  ]
RMSE mean:  0.9258276007593409
MAE:  [0.70717239 0.70138695 0.70681952 0.70491089 0.70699121]
MAE mean:  0.7054561927885127
```



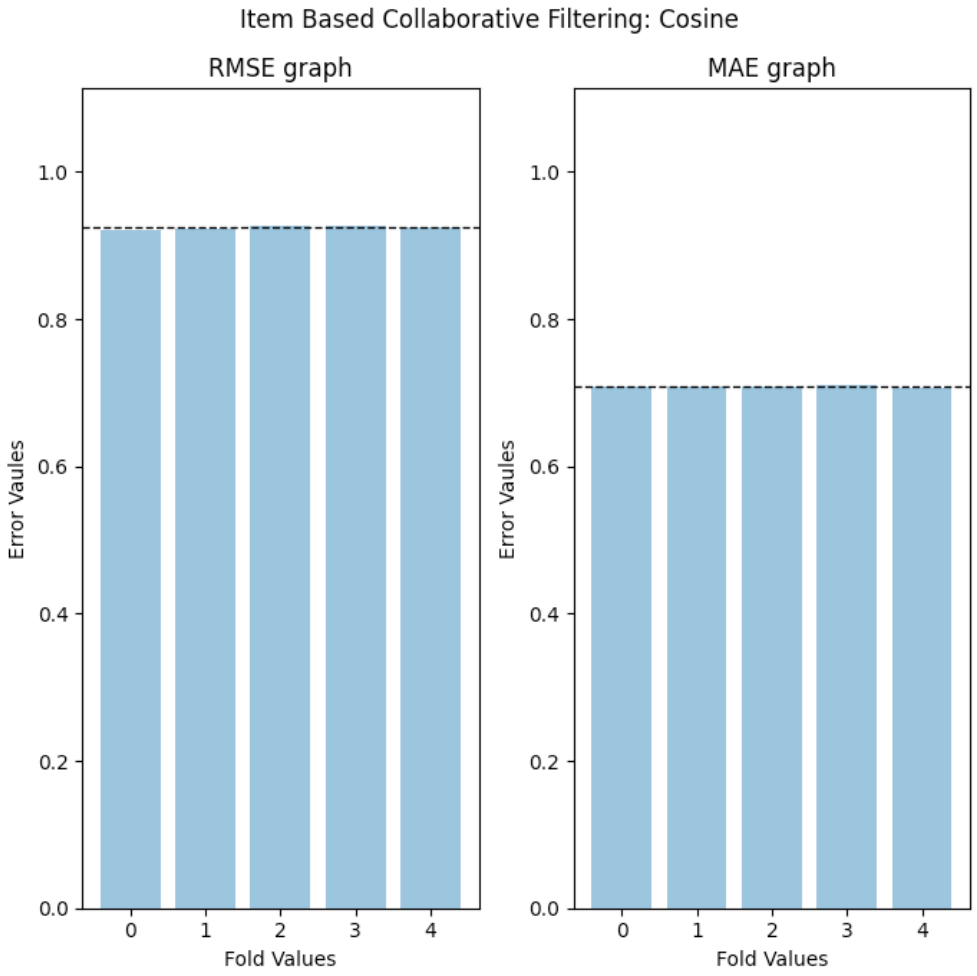
##

6) Item based Collaborative Filtering with Cosine similarity:

##

Output:

```
RMSE:  [0.92180239 0.92385006 0.92325792 0.92139516 0.93197032]
RMSE mean:  0.9244551671905237
MAE:  [0.70816245 0.70830466 0.70446769 0.7053181  0.71426998]
MAE mean:  0.7081045765964034
```



##

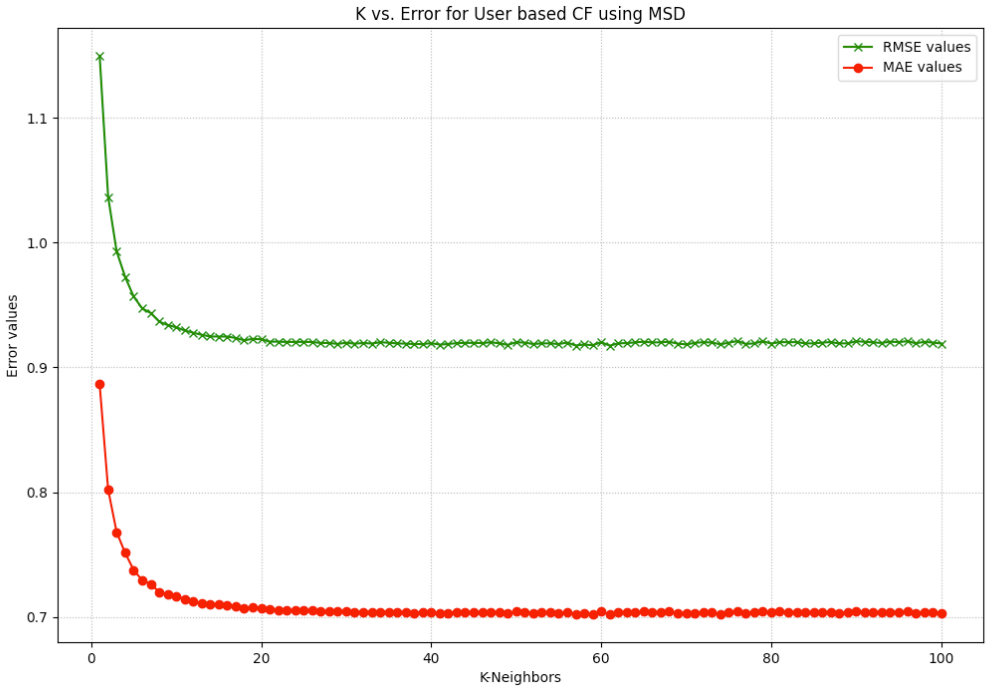
Among User based Collaborative filtering with MSD, Pearson and Cosine similarity, it was noticed that Mean Squared Distance (MSD) had the lowest mean RMSE and MAE.

Among Item based Collaborative filtering with MSD, Pearson and Cosine similarity, once again, it was noticed that Mean Squared Distance (MSD) had the lowest mean RMSE and MAE.

The results of user-based and item-based CF are consistent with each other for all similarity metrics.

f) Since MSD gave us the smallest RMSE and MAE for both user-based and item-based collaborative filtering, I decided to use this similarity metric for tasks f and g. K vs Error graphs were draw for K 1 to 100.
   1) User based Collaborative Filtering using MSD:

##
Output:

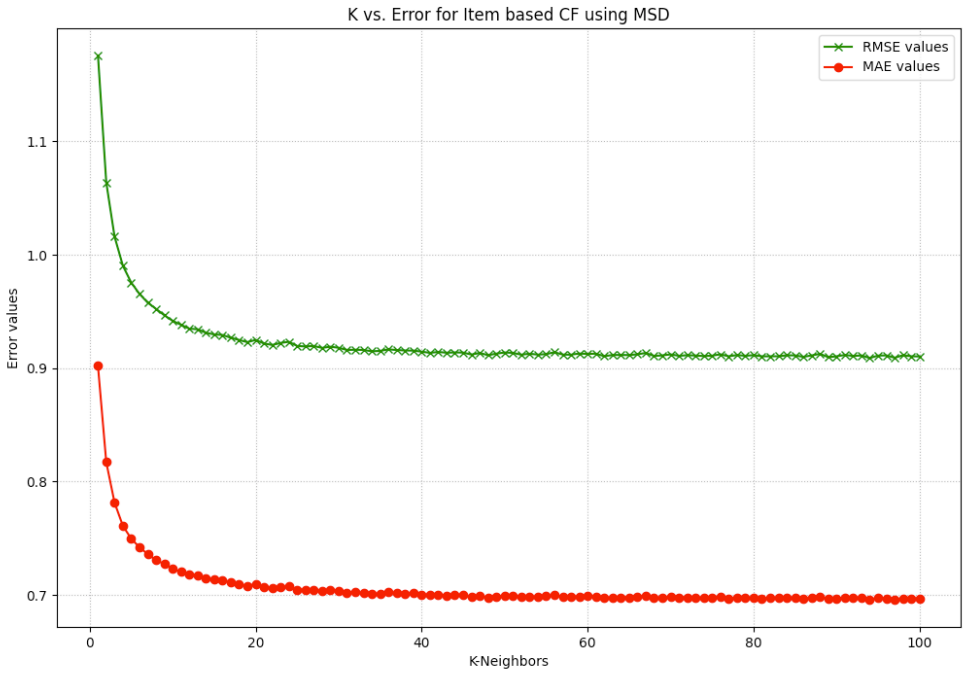


##

2) Item based Collaborative Filtering using MSD:

##

Output:



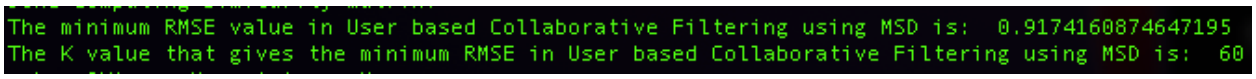
##

g) The best K was the K value that gives us the lowest RMSE and MAE value. For User based CF, the lowest RMSE value was 0.91741 and it was obtained when K = 60

##

Output:

```
The minimum RMSE value in User based Collaborative Filtering using MSD
is:  0.9174160874647195
The K value that gives the minimum RMSE in User based Collaborative
Filtering using MSD is:  60
```



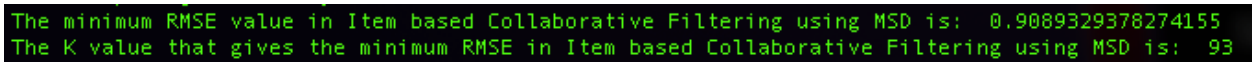
##

Similarly, for Item based CF, the lowest RMSE value was 0.908932 and it was obtained when K = 93

##

Output:

```
The minimum RMSE value in Item based Collaborative Filtering using MSD
is:  0.9089329378274155
The K value that gives the minimum RMSE in Item based Collaborative
Filtering using MSD is:  93
```



##

The best K is not the same for User-based and Item based collaborative filtering.

GitHub Link:
Assignment_06

https://github.com/monicabernard/CAP-5610_Machine-Learning.git