BigData Report

pxa170230

Q4)

Implementation

Create DataFrame using itemusermat file

Using kmeans cluster, fit the model and transform to dataset

Using moveis.dat file, predict the output and cluster the movies and print the output

Output:

```
Cluster : 0
2866 , [Buddy Holly Story, The (1978),Drama]
3749 , [Time Regained (Le Temps Retrouv�) (1999),Drama]
243 , [Gordy (1995),Comedy]
392 , [Secret Adventures of Tom Thumb, The (1993),Adventure|Children's]
540 , [Sliver (1993),Thriller]
```

Q5)

Implementation

We will use ratings file and map the data to (userid, movieid, ratings)

Split the data into 70% and 30%

Train the model for all ranks and predict the output and calculate RMSE values for them

Using RMSE values, see which model has less error and decide it is the nest model

Output:

```
For rank 4 the RMSE is 0.8768487815037069
For rank 8 the RMSE is 0.8661323150946303
For rank 10 the RMSE is 0.8645708843493792
For rank 12 the RMSE is 0.8638165717039546
For rank 16 the RMSE is 0.8630405540046656
The best model was trained with rank %s 16
```