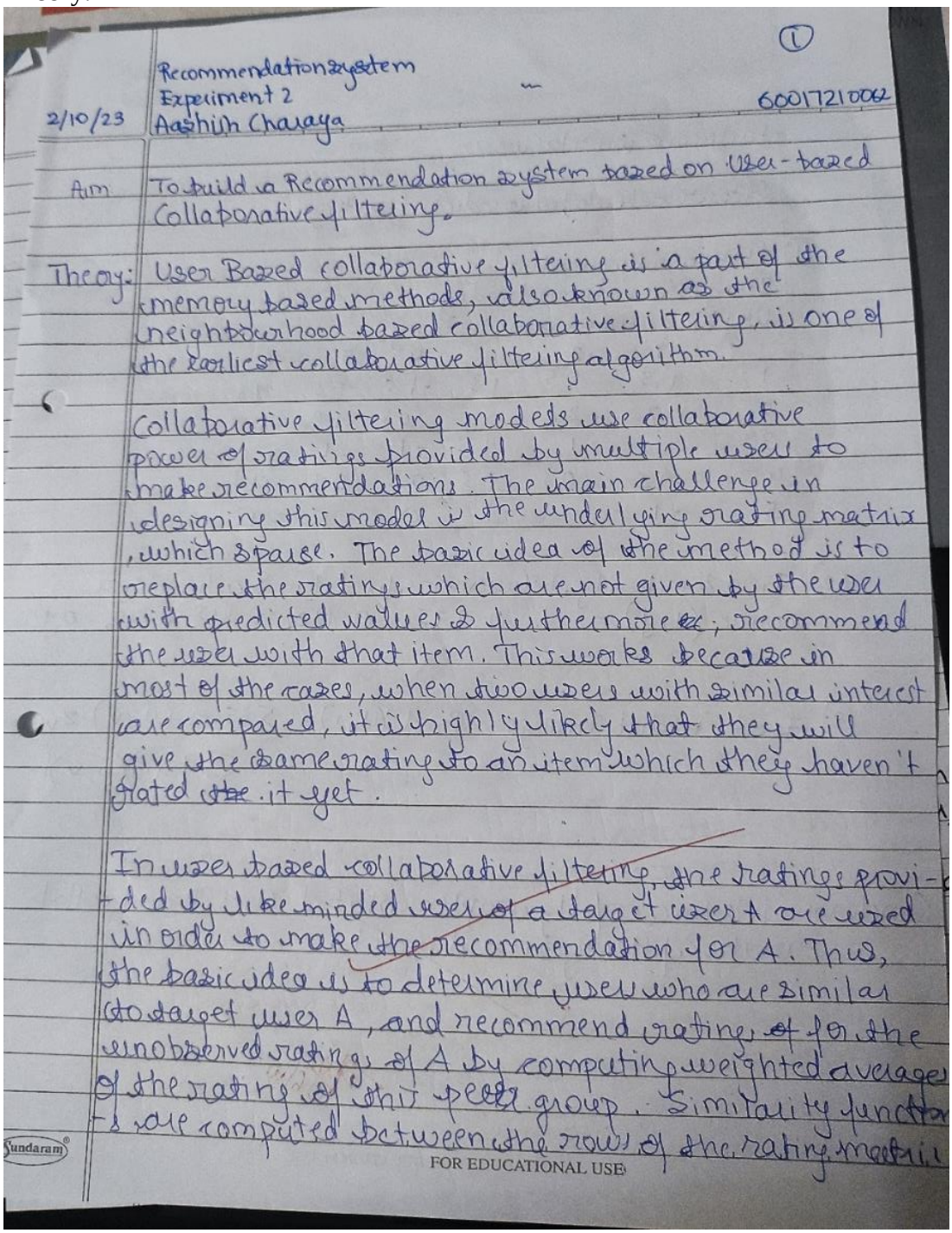
Recommendation System

Experiment 2

Aashish Charaya 60017210062 AIML

Aim: Build a user-based collaborative filtering recommendation system for different datasets.

Theory:



Recommendation System ①
Experiment 2
2/10/23 Aashish Charaya 60017210062

Aim To build a Recommendation system based on User-based Collaborative filtering.

Theory: User Based collaborative filtering is a part of the memory based methods, also known as the neighbourhood based collaborative filtering, is one of the earliest collaborative filtering algorithm.

Collaborative filtering models use collaborative power of ratings provided by multiple users to make recommendations. The main challenge in designing this model is the underlying rating matrix, which sparse. The basic idea of the method is to replace the ratings which are not given by the user with predicted values & furthermore, recommend the user with that item. This works because in most of the cases, when two users with similar interest are compared, it is highly likely that they will give the same rating to an item which they haven't rated it yet.

In user based collaborative filtering, the ratings provided by like minded users of a target user A are used in order to make the recommendation for A. Thus, the basic idea is to determine users who are similar to target user A, and recommend ratings for the unobserved ratings of A by computing weighted averages of the ratings of this peer group. Similarity functions are computed between the rows of the rating matrix.

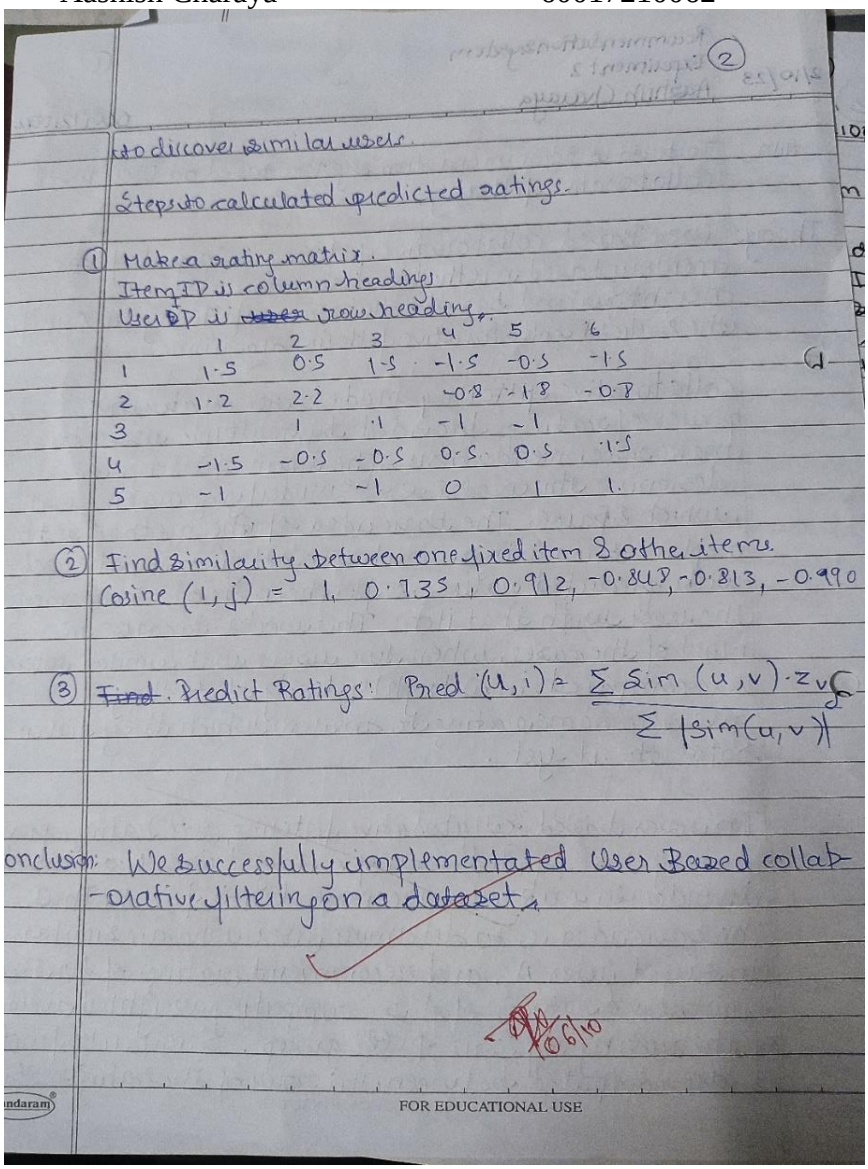
Sundaram

FOR EDUCATIONAL USE

Recommendation System

Experiment 2

Aashish Charaya 60017210062 AIML



to discover similar users

Steps to calculated predicted ratings.

① Make a rating matrix.
Item ID is column heading
User ID is row heading.

	1	2	3	4	5	6
1	1·5	0·5	1·5	-1·5	-0·5	-1·5
2	1·2	2·2		-0·8	-1·8	-0·8
3		1	·1	-1	-1	
4	-1·5	-0·5	-0·5	0·5	0·5	1·5
5	-1		-1	0	1	1.

② Find similarity between one fixed item & other items.
Cosine (i, j) = 1, 0·735, 0·912, -0·848, -0·813, -0·990

③ Predict Ratings: $Pred(u,i) = \frac{\sum sim(u,v) \cdot r_{vi}}{\sum |sim(u,v)|}$

Conclusion: We successfully implemented User Based collaborative filtering on a dataset.

10/10

FOR EDUCATIONAL USE

User Based Colaborative Filtering :

```
# Data processing
import pandas as pd
import numpy as np
import scipy.stats
# Visualization
import seaborn as sns
# Similarity
from sklearn.metrics.pairwise import cosine_similarity
```

Recommendation System

Experiment 2

Aashish Charaya 60017210062 AIML

```
# Mount Google Drive
from google.colab import drive
drive.mount('/content/drive')
# Change directory
import os
os.chdir("/content/drive/MyDrive/recommendation_system")
# Print out the current directory
!pwd
```

```
Drive already mounted at /content/drive; to attempt to forcibly remount, call
drive.mount("/content/drive", force_remount=True).
/content/drive/MyDrive/recommendation_system
```

```
# Read in data
ratings=pd.read_csv('ml-latest-small/ratings.csv')
# Take a look at the data
ratings.head()
```

```
   userId  movieId  rating  timestamp
0       1        1     4.0  964982703
1       1        3     4.0  964981247
2       1        6     4.0  964982224
3       1       47     5.0  964983815
4       1       50     5.0  964982931
```

```
# Get the dataset information
ratings.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 100836 entries, 0 to 100835
Data columns (total 4 columns):
 #   Column     Non-Null Count   Dtype
---  ------     --------------   -----
 0   userId     100836 non-null  int64
 1   movieId    100836 non-null  int64
 2   rating     100836 non-null  float64
 3   timestamp  100836 non-null  int64
dtypes: float64(1), int64(3)
memory usage: 3.1 MB
```

```
# Number of users
print('The ratings dataset has', ratings['userId'].nunique(), 'unique users')
# Number of movies
print('The ratings dataset has', ratings['movieId'].nunique(), 'unique
movies')
# Number of ratings
print('The ratings dataset has', ratings['rating'].nunique(), 'unique
ratings')
# List of unique ratings
print('The unique ratings are', sorted(ratings['rating'].unique()))
```

Recommendation System

Experiment 2

Aashish Charaya 60017210062 AIML

```
The ratings dataset has 610 unique users
The ratings dataset has 9724 unique movies
The ratings dataset has 10 unique ratings
The unique ratings are [0.5, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0, 4.5, 5.0]
```

```
# Read in data
movies = pd.read_csv('ml-latest-small/movies.csv')
# Take a look at the data
movies.head()
```

```
   movieId                               title  \
0        1                    Toy Story (1995)
1        2                      Jumanji (1995)
2        3             Grumpier Old Men (1995)
3        4            Waiting to Exhale (1995)
4        5  Father of the Bride Part II (1995)

                                        genres
0  Adventure|Animation|Children|Comedy|Fantasy
1                   Adventure|Children|Fantasy
2                               Comedy|Romance
3                         Comedy|Drama|Romance
4                                       Comedy
```

```
# Merge ratings and movies datasets
df = pd.merge(ratings, movies, on='movieId', how='inner')
# Take a look at the data
df.head()
```

```
   userId  movieId  rating   timestamp             title  \
0       1        1     4.0   964982703  Toy Story (1995)
1       5        1     4.0   847434962  Toy Story (1995)
2       7        1     4.5  1106635946  Toy Story (1995)
3      15        1     2.5  1510577970  Toy Story (1995)
4      17        1     4.5  1305696483  Toy Story (1995)

                                        genres
0  Adventure|Animation|Children|Comedy|Fantasy
1  Adventure|Animation|Children|Comedy|Fantasy
2  Adventure|Animation|Children|Comedy|Fantasy
3  Adventure|Animation|Children|Comedy|Fantasy
4  Adventure|Animation|Children|Comedy|Fantasy
```

```
#Aggregate by movie
agg_ratings = df.groupby('title').agg(mean_rating = ('rating', 'mean'),
                                      number_of_ratings =
('rating', 'count')).reset_index()
# Keep the movies with over 100 ratings
agg_ratings_GT100 = agg_ratings[agg_ratings['number_of_ratings']>100]
agg_ratings_GT100.info()
```

Recommendation System

Experiment 2

Aashish Charaya 60017210062 AIML

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 134 entries, 74 to 9615
Data columns (total 3 columns):
 #   Column             Non-Null Count  Dtype  
---  ------             --------------  -----  
 0   title              134 non-null    object 
 1   mean_rating        134 non-null    float64
 2   number_of_ratings  134 non-null    int64  
dtypes: float64(1), int64(1), object(1)
memory usage: 4.2+ KB
```

```
# Check popular movies
agg_ratings_GT100.sort_values(by='number_of_ratings', ascending=False).head()
```

	title	mean_rating	number_of_ratings
3158	Forrest Gump (1994)	4.164134	329
7593	Shawshank Redemption, The (1994)	4.429022	317
6865	Pulp Fiction (1994)	4.197068	307
7680	Silence of the Lambs, The (1991)	4.161290	279
5512	Matrix, The (1999)	4.192446	278

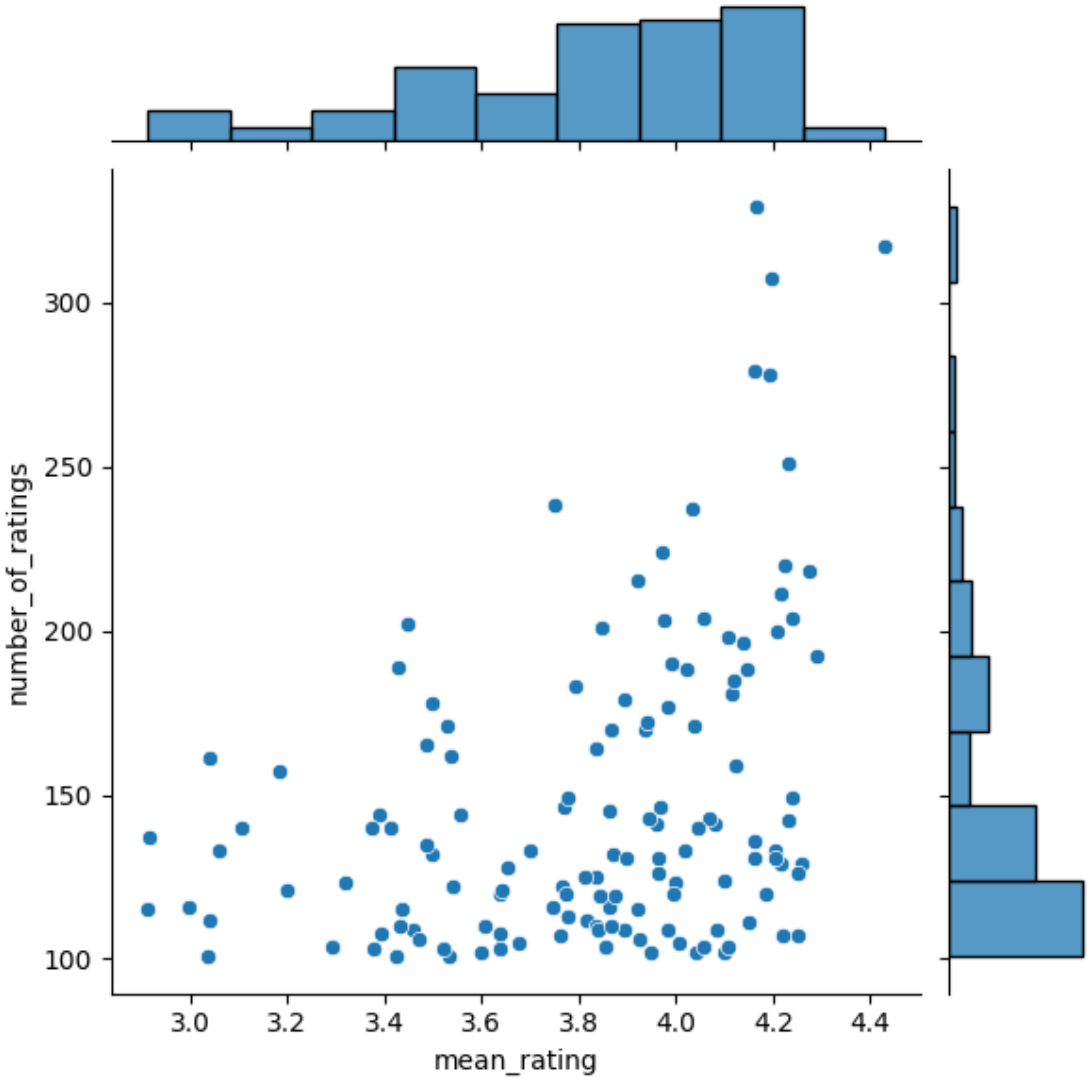
```
# Visulization
sns.jointplot(x='mean_rating', y='number_of_ratings', data=agg_ratings_GT100)
```

```
<seaborn.axisgrid.JointGrid at 0x79dc7a85bf10>
```

Recommendation System

Experiment 2

Aashish Charaya 60017210062 AIML




```
df_GT100 = pd.merge(df, agg_ratings_GT100[['title']], on='title',
how='inner')
df_GT100.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 19788 entries, 0 to 19787
Data columns (total 6 columns):
 #   Column     Non-Null Count  Dtype
---  ------     --------------  -----
 0   userId     19788 non-null  int64
 1   movieId    19788 non-null  int64
 2   rating     19788 non-null  float64
 3   timestamp  19788 non-null  int64
 4   title      19788 non-null  object
 5   genres     19788 non-null  object
```

```
dtypes: float64(1), int64(3), object(2)
memory usage: 1.1+ MB
```

```
# Number of users
print('The ratings dataset has', df_GT100['userId'].nunique(), 'unique
users')
# Number of movies
print('The ratings dataset has', df_GT100['movieId'].nunique(), 'unique
movies')
# Number of ratings
print('The ratings dataset has', df_GT100['rating'].nunique(), 'unique
ratings')
# List of unique ratings
print('The unique ratings are', sorted(df_GT100['rating'].unique()))
```

```
The ratings dataset has 597 unique users
The ratings dataset has 134 unique movies
The ratings dataset has 10 unique ratings
The unique ratings are [0.5, 1.0, 1.5, 2.0, 2.5, 3.0, 3.5, 4.0, 4.5, 5.0]
```

```
# Create user-item matrix
matrix = df_GT100.pivot_table(index='userId', columns='title',
values='rating')
matrix.head()
```

```
title   2001: A Space Odyssey (1968)  Ace Ventura: Pet Detective (1994)  \
userId
1                                  NaN                                NaN
2                                  NaN                                NaN
3                                  NaN                                NaN
4                                  NaN                                NaN
5                                  NaN                                3.0

title   Aladdin (1992)  Alien (1979)  Aliens (1986)  \
userId
1                  NaN           4.0            NaN
2                  NaN           NaN            NaN
3                  NaN           NaN            NaN
4                  4.0           NaN            NaN
5                  4.0           NaN            NaN

title   Amelie (Fabuleux destin d'Amélie Poulain, Le) (2001)  \
userId
1                                                      NaN
2                                                      NaN
3                                                      NaN
4                                                      NaN
5                                                      NaN
```

Recommendation System

Experiment 2

Aashish Charaya 60017210062 AIML

```
title   American Beauty (1999)  American History X (1998)  \
userId
1                          5.0                        5.0
2                          NaN                        NaN
3                          NaN                        NaN
4                          5.0                        NaN
5                          NaN                        NaN

title   American Pie (1999)  Apocalypse Now (1979)  ...  True Lies (1994)  \
userId                                              ...
1                      NaN                    4.0  ...               NaN
2                      NaN                    NaN  ...               NaN
3                      NaN                    NaN  ...               NaN
4                      NaN                    NaN  ...               NaN
5                      NaN                    NaN  ...               2.0

title   Truman Show, The (1998)  Twelve Monkeys (a.k.a. 12 Monkeys) (1995)  \
userId
1                           NaN                                        NaN
2                           NaN                                        NaN
3                           NaN                                        NaN
4                           NaN                                        2.0
5                           NaN                                        NaN

title   Twister (1996)  Up (2009)  Usual Suspects, The (1995)  WALL·E (2008)
\
userId
1                  3.0        NaN                         5.0            NaN
2                  NaN        NaN                         NaN            NaN
3                  NaN        NaN                         NaN            NaN
4                  NaN        NaN                         NaN            NaN
5                  NaN        NaN                         4.0            NaN

title   Waterworld (1995)  Willy Wonka & the Chocolate Factory (1971)  \
userId
1                     NaN                                         5.0
2                     NaN                                         NaN
3                     NaN                                         NaN
4                     NaN                                         4.0
5                     NaN                                         NaN

title   X-Men (2000)
userId
1                5.0
2                NaN
3                NaN
4                NaN
5                NaN
```

```
[5 rows x 134 columns]
```

```
# Normalize user-item matrix
matrix_norm = matrix.subtract(matrix.mean(axis=1), axis = 'rows')
matrix_norm.head()
```

```
title   2001: A Space Odyssey (1968)  Ace Ventura: Pet Detective (1994)  \
userId
1                                  NaN                                NaN
2                                  NaN                                NaN
3                                  NaN                                NaN
4                                  NaN                                NaN
5                                  NaN                          -0.461538

title   Aladdin (1992)  Alien (1979)  Aliens (1986)  \
userId
1                  NaN     -0.392857            NaN
2                  NaN           NaN            NaN
3                  NaN           NaN            NaN
4             0.617647           NaN            NaN
5             0.538462           NaN            NaN

title   Amelie (Fabuleux destin d'Amélie Poulain, Le) (2001)  \
userId
1                                                     NaN
2                                                     NaN
3                                                     NaN
4                                                     NaN
5                                                     NaN

title   American Beauty (1999)  American History X (1998)  \
userId
1                     0.607143                   0.607143
2                          NaN                        NaN
3                          NaN                        NaN
4                     1.617647                        NaN
5                          NaN                        NaN

title   American Pie (1999)  Apocalypse Now (1979)  ...  True Lies (1994)  \
userId                                              ...
1                      NaN              -0.392857  ...               NaN
2                      NaN                    NaN  ...               NaN
3                      NaN                    NaN  ...               NaN
4                      NaN                    NaN  ...               NaN
5                      NaN                    NaN  ...         -1.461538

title   Truman Show, The (1998)  Twelve Monkeys (a.k.a. 12 Monkeys) (1995)  \
userId
```

Recommendation System

Experiment 2

Aashish Charaya 60017210062 AIML

```
1                          NaN                                  NaN
2                          NaN                                  NaN
3                          NaN                                  NaN
4                          NaN                            -1.382353
5                          NaN                                  NaN

title   Twister (1996)  Up (2009)  Usual Suspects, The (1995)  WALL·E (2008)
\
userId
1            -1.392857        NaN                    0.607143            NaN
2                  NaN        NaN                         NaN            NaN
3                  NaN        NaN                         NaN            NaN
4                  NaN        NaN                         NaN            NaN
5                  NaN        NaN                    0.538462            NaN

title   Waterworld (1995)  Willy Wonka & the Chocolate Factory (1971)  \
userId
1                     NaN                                    0.607143
2                     NaN                                         NaN
3                     NaN                                         NaN
4                     NaN                                    0.617647
5                     NaN                                         NaN

title   X-Men (2000)
userId
1           0.607143
2                NaN
3                NaN
4                NaN
5                NaN

[5 rows x 134 columns]
```

```
# User similarity matrix using Pearson correlation
user_similarity = matrix_norm.T.corr()
user_similarity.head()
```

```
userId       1    2    3         4         5         6         7         8
\
userId
1       1.000000  NaN  NaN  0.391797  0.180151 -0.439941 -0.029894  0.464277
2            NaN  1.0  NaN       NaN       NaN       NaN       NaN       NaN
3            NaN  NaN  NaN       NaN       NaN       NaN       NaN       NaN
4       0.391797  NaN  NaN  1.000000 -0.394823  0.421927  0.704669  0.055442
5       0.180151  NaN  NaN -0.394823  1.000000 -0.006888  0.328889  0.030168

userId  9         10   ...       601       602       603       604       605
\
userId                ...
```

Recommendation System

Experiment 2

Aashish Charaya 60017210062 AIML

```
1       1.0 -0.037987  ...  0.091574  0.254514  0.101482 -0.500000  0.780020
2       NaN  1.000000  ... -0.583333       NaN -1.000000       NaN       NaN
3       NaN       NaN  ...       NaN       NaN       NaN       NaN       NaN
4       NaN  0.360399  ... -0.239325  0.562500  0.162301 -0.158114  0.905134
5       NaN -0.777714  ...  0.000000  0.231642  0.131108  0.068621 -0.245026

userId       606       607       608       609       610
userId
1       0.303854 -0.012077  0.242309 -0.175412  0.071553
2       0.583333       NaN -0.229416       NaN  0.765641
3            NaN       NaN       NaN       NaN       NaN
4       0.021898 -0.020659 -0.286872       NaN -0.050868
5       0.377341  0.228218  0.263139  0.384111  0.040582

[5 rows x 597 columns]
```

```
# User similarity matrix using cosine similarity
user_similarity_cosine = cosine_similarity(matrix_norm.fillna(0))
user_similarity_cosine
```

```
array([[ 1.        ,  0.        ,  0.        , ...,  0.14893867,
        -0.06003146,  0.04528224],
       [ 0.        ,  1.        ,  0.        , ..., -0.04485403,
        -0.25197632,  0.18886414],
       [ 0.        ,  0.        ,  0.        , ...,  0.        ,
         0.        ,  0.        ],
       ...,
       [ 0.14893867, -0.04485403,  0.        , ...,  1.        ,
         0.14734568,  0.07931015],
       [-0.06003146, -0.25197632,  0.        , ...,  0.14734568,
         1.        , -0.14276787],
       [ 0.04528224,  0.18886414,  0.        , ...,  0.07931015,
        -0.14276787,  1.        ]])
```

```
# Pick a user ID
picked_userid = 1
# Remove picked user ID from the candidate list
user_similarity.drop(index=picked_userid, inplace=True)
# Take a look at the data
user_similarity.head()
```

```
userId       1    2    3         4         5         6         7         8
\
userId
2            NaN  1.0  NaN       NaN       NaN       NaN       NaN       NaN
3            NaN  NaN  NaN       NaN       NaN       NaN       NaN       NaN
4       0.391797  NaN  NaN  1.000000 -0.394823  0.421927  0.704669  0.055442
5       0.180151  NaN  NaN -0.394823  1.000000 -0.006888  0.328889  0.030168
6      -0.439941  NaN  NaN  0.421927 -0.006888  1.000000  0.000000 -0.127385
```

Recommendation System

Experiment 2

Aashish Charaya 60017210062 AIML

```
userId  9         10   ...       601       602       603       604       605
\
userId                ...
2       NaN  1.000000  ... -0.583333       NaN -1.000000       NaN       NaN
3       NaN       NaN  ...       NaN       NaN       NaN       NaN       NaN
4       NaN  0.360399  ... -0.239325  0.562500  0.162301 -0.158114  0.905134
5       NaN -0.777714  ...  0.000000  0.231642  0.131108  0.068621 -0.245026
6       NaN  0.957427  ... -0.292770 -0.030599 -0.123983 -0.176327  0.063861

userId       606       607       608       609       610
userId
2       0.583333       NaN -0.229416       NaN  0.765641
3            NaN       NaN       NaN       NaN       NaN
4       0.021898 -0.020659 -0.286872       NaN -0.050868
5       0.377341  0.228218  0.263139  0.384111  0.040582
6      -0.468008  0.541386 -0.337129  0.158255 -0.030567

[5 rows x 597 columns]
```

```
# Number of similar users
n = 10
# User similarity threashold
user_similarity_threshold = 0.3
# Get top n similar users
similar_users =
user_similarity[user_similarity[picked_userid]>user_similarity_threshold][pic
ked_userid].sort_values(ascending=False)[:n]
# Print out top n similar users
print(f'The similar users for user {picked_userid}')
```

```
The similar users for user 1
```

```
# Movies that the target user has watched
picked_userid_watched = matrix_norm[matrix_norm.index ==
picked_userid].dropna(axis=1, how='all')
picked_userid_watched
```

```
title   Alien (1979)  American Beauty (1999)  American History X (1998)  \
userId
1          -0.392857                0.607143                   0.607143

title   Apocalypse Now (1979)  Back to the Future (1985)  Batman (1989)  \
userId
1                   -0.392857                   0.607143      -0.392857

title   Big Lebowski, The (1998)  Braveheart (1995)  \
userId
1                       0.607143          -0.392857
```

Recommendation System

Experiment 2

Aashish Charaya 60017210062 AIML

```
title   Clear and Present Danger (1994)  Clerks (1994)  ...  \
userId                                                  ...
1                            -0.392857      -1.392857  ...

title   Star Wars: Episode IV - A New Hope (1977)  \
userId
1                                        0.607143

title   Star Wars: Episode V - The Empire Strikes Back (1980)  \
userId
1                                              0.607143

title   Star Wars: Episode VI - Return of the Jedi (1983)  Stargate (1994)  \
userId
1                                              0.607143        -1.392857

title   Terminator, The (1984)  Toy Story (1995)  Twister (1996)  \
userId
1                     0.607143         -0.392857       -1.392857

title   Usual Suspects, The (1995)  \
userId
1                          0.607143

title   Willy Wonka & the Chocolate Factory (1971)  X-Men (2000)
userId
1                                         0.607143      0.607143

[1 rows x 56 columns]
```

```
# Movies that similar users watched. Remove movies that none of the similar
users have watched
similar_user_movies =
matrix_norm[matrix_norm.index.isin(similar_users.index)].dropna(axis=1,
how='all')
similar_user_movies
```

```
title   Aladdin (1992)  Alien (1979)  \
userId
9                  NaN           NaN
108                NaN           NaN
154                NaN           NaN
366                NaN           NaN
401          -0.382353           NaN
502                NaN        -0.375
511                NaN           NaN
550                NaN           NaN
595                NaN           NaN
```

```
598                  NaN            NaN

title   Amelie (Fabuleux destin d'Amélie Poulain, Le) (2001)  \
userId
9                                                       NaN
108                                                0.466667
154                                                     NaN
366                                                     NaN
401                                                     NaN
502                                                     NaN
511                                               -0.653846
550                                                     NaN
595                                                     NaN
598                                                     NaN

title   Back to the Future (1985)  Batman Begins (2005)  \
userId
9                        0.333333                   NaN
108                      0.466667                   NaN
154                           NaN                   NaN
366                           NaN             -0.205882
401                           NaN                   NaN
502                           NaN                   NaN
511                           NaN                   NaN
550                           NaN                   NaN
595                           NaN                   NaN
598                           NaN                   NaN

title   Beautiful Mind, A (2001)  Beauty and the Beast (1991)  \
userId
9                            NaN                          NaN
108                     0.466667                          NaN
154                          NaN                          NaN
366                          NaN                          NaN
401                          NaN                    -0.382353
502                          NaN                          NaN
511                          NaN                          NaN
550                          NaN                          NaN
595                          NaN                          NaN
598                          NaN                          NaN

title   Blade Runner (1982)  Bourne Identity, The (2002)  Braveheart (1995)
\
userId
9                       NaN                          NaN                NaN
108                0.466667                          NaN                NaN
154                     NaN                          NaN                NaN
366                     NaN                          NaN          -0.205882
```

```
401                    NaN                          NaN                 NaN
502                    NaN                          NaN                 NaN
511                    NaN                          NaN                 NaN
550                    NaN                          NaN                 NaN
595                    NaN                          NaN                 NaN
598                    NaN                     0.888889                 NaN

title   ...  Shrek (2001)  Silence of the Lambs, The (1991)  \
userId  ...
9       ...           NaN                             NaN
108     ...           NaN                             NaN
154     ...           NaN                             NaN
366     ...           NaN                             NaN
401     ...      0.117647                             NaN
502     ...           NaN                             NaN
511     ...           NaN                             NaN
550     ...           NaN                             NaN
595     ...           NaN                             NaN
598     ...     -2.111111                       -2.611111

title   Spider-Man (2002)  Star Wars: Episode I - The Phantom Menace (1999)
\
userId
9                     NaN                                             NaN
108              0.466667                                             NaN
154                   NaN                                             NaN
366                   NaN                                             NaN
401                   NaN                                             NaN
502                   NaN                                             NaN
511             -1.153846                                       -0.653846
550                   NaN                                             NaN
595                   NaN                                       -0.333333
598                   NaN                                             NaN

title   Terminator 2: Judgment Day (1991)  Titanic (1997)  Toy Story (1995)
\
userId
9                                     NaN             NaN               NaN
108                                   NaN       -0.533333               NaN
154                                   NaN             NaN               NaN
366                             -0.205882             NaN               NaN
401                                   NaN             NaN          0.117647
502                                   NaN             NaN               NaN
511                                   NaN             NaN               NaN
550                                   NaN             NaN         -0.277778
595                                   NaN             NaN               NaN
598                                   NaN             NaN               NaN
```

Recommendation System

Experiment 2

Aashish Charaya 60017210062 AIML

```
title   Up (2009)  Usual Suspects, The (1995)  WALL·E (2008)
userId
9             NaN                         NaN            NaN
108           NaN                         NaN            NaN
154      0.214286                         NaN            NaN
366           NaN                         NaN            NaN
401      0.617647                         NaN       0.617647
502           NaN                         NaN            NaN
511     -0.153846                         NaN            NaN
550      0.222222                         NaN      -0.277778
595           NaN                    0.666667            NaN
598           NaN                         NaN            NaN

[10 rows x 62 columns]
```

```
# Remove the watched movie from the movie list
similar_user_movies.drop(picked_userid_watched.columns,axis=1, inplace=True,
errors='ignore')
# Take a look at the data
similar_user_movies
```

```
title   Aladdin (1992)  Amelie (Fabuleux destin d'Amélie Poulain, Le) (2001)
\
userId
9                  NaN                                                  NaN
108                NaN                                             0.466667
154                NaN                                                  NaN
366                NaN                                                  NaN
401          -0.382353                                                  NaN
502                NaN                                                  NaN
511                NaN                                            -0.653846
550                NaN                                                  NaN
595                NaN                                                  NaN
598                NaN                                                  NaN

title   Batman Begins (2005)  Beautiful Mind, A (2001)  \
userId
9                        NaN                       NaN
108                      NaN                  0.466667
154                      NaN                       NaN
366                -0.205882                       NaN
401                      NaN                       NaN
502                      NaN                       NaN
511                      NaN                       NaN
550                      NaN                       NaN
595                      NaN                       NaN
598                      NaN                       NaN

title   Beauty and the Beast (1991)  Blade Runner (1982)  \
```

Recommendation System

Experiment 2

Aashish Charaya 60017210062 AIML

```
userId
9                              NaN                 NaN
108                            NaN            0.466667
154                            NaN                 NaN
366                            NaN                 NaN
401                      -0.382353                 NaN
502                            NaN                 NaN
511                            NaN                 NaN
550                            NaN                 NaN
595                            NaN                 NaN
598                            NaN                 NaN

title   Bourne Identity, The (2002)  Breakfast Club, The (1985)  \
userId
9                              NaN                         NaN
108                            NaN                   -0.533333
154                            NaN                         NaN
366                            NaN                         NaN
401                            NaN                         NaN
502                            NaN                         NaN
511                            NaN                         NaN
550                            NaN                         NaN
595                            NaN                         NaN
598                       0.888889                         NaN

title   Catch Me If You Can (2002)  Dark Knight, The (2008)  ...  \
userId                                                       ...
9                              NaN                     NaN  ...
108                       0.466667                     NaN  ...
154                            NaN                     NaN  ...
366                            NaN               -0.205882  ...
401                            NaN                     NaN  ...
502                            NaN                     NaN  ...
511                            NaN                     NaN  ...
550                      -0.277778               -0.277778  ...
595                            NaN                     NaN  ...
598                            NaN                     NaN  ...

title   Monsters, Inc. (2001)  Ocean's Eleven (2001)  \
userId
9                         NaN                   NaN
108                       NaN                   NaN
154                       NaN                   NaN
366                       NaN                   NaN
401                  0.117647                   NaN
502                       NaN                   NaN
511                       NaN                   NaN
550                       NaN                   NaN
595                       NaN                   NaN
```

```
598                          NaN                0.888889

title   Pirates of the Caribbean: The Curse of the Black Pearl (2003)  \
userId
9                                                         NaN
108                                                       NaN
154                                                       NaN
366                                                 -0.205882
401                                                  0.117647
502                                                       NaN
511                                                       NaN
550                                                       NaN
595                                                       NaN
598                                                       NaN

title   Shawshank Redemption, The (1994)  Shrek (2001)  Spider-Man (2002)  \
userId
9                                    NaN           NaN                NaN
108                                  NaN           NaN           0.466667
154                                  NaN           NaN                NaN
366                                  NaN           NaN                NaN
401                                  NaN      0.117647                NaN
502                             0.125000           NaN                NaN
511                             0.346154           NaN          -1.153846
550                             0.222222           NaN                NaN
595                                  NaN           NaN                NaN
598                                  NaN     -2.111111                NaN

title   Terminator 2: Judgment Day (1991)  Titanic (1997)  Up (2009)  \
userId
9                                     NaN             NaN        NaN
108                                   NaN       -0.533333        NaN
154                                   NaN             NaN   0.214286
366                             -0.205882             NaN        NaN
401                                   NaN             NaN   0.617647
502                                   NaN             NaN        NaN
511                                   NaN             NaN  -0.153846
550                                   NaN             NaN   0.222222
595                                   NaN             NaN        NaN
598                                   NaN             NaN        NaN

title   WALL·E (2008)
userId
9                 NaN
108               NaN
154               NaN
366               NaN
401          0.617647
```

```
502              NaN
511              NaN
550        -0.277778
595              NaN
598              NaN

[10 rows x 38 columns]
```

```
# A dictionary to store item scores
item_score = {}
# Loop through items
for i in similar_user_movies.columns:
  # Get the ratings for movie i
  movie_rating = similar_user_movies[i]
  # Create a variable to store the score
  total = 0
  # Create a variable to store the number of scores
  count = 0
  # Loop through similar users
  for u in similar_users.index:
    # If the movie has rating
    if pd.isna(movie_rating[u]) == False:
      # Score is the sum of user similarity score multiply by the movie
rating
      score = similar_users[u] * movie_rating[u]
      # Add the score to the total score for the movie so far
      total += score
      # Add 1 to the count
      count +=1
  # Get the average score for the item
  item_score[i] = total / count
# Convert dictionary to pandas dataframe
item_score = pd.DataFrame(item_score.items(), columns=['movie',
'movie_score'])

# Sort the movies by score
ranked_item_score = item_score.sort_values(by='movie_score', ascending=False)
# Select top m movies
m = 10
ranked_item_score.head(m)
```

```
                                          movie  movie_score
16  Harry Potter and the Chamber of Secrets (2002)     1.888889
13    Eternal Sunshine of the Spotless Mind (2004)     1.888889
6                     Bourne Identity, The (2002)     0.888889
29                          Ocean's Eleven (2001)     0.888889
18                               Inception (2010)     0.587491
3                      Beautiful Mind, A (2001)     0.466667
5                           Blade Runner (1982)     0.466667
```

Recommendation System

Experiment 2

Aashish Charaya 60017210062 AIML

```
12                          Donnie Darko (2001)    0.466667
10                        Departed, The (2006)     0.256727
31              Shawshank Redemption, The (1994)   0.222566
```

GitHub Repo : Alpha-131/RS_Experiments (github.com)

Conclusion: Implemented an User-based Collaborative filtering recommendation engine on different datasets.