Programming Assignment 2

Student Name: Anirudh Heda Student Session: CS6675

Problem 3: **Collaborative Recommender Systems**
Option 1: **Recommending Movies**

1. **Source Code:**
   - Github: https://github.com/hedaanirudh/RecommenderSystem
   - Code can be executed by following the commands specified in README.md file.

2. **References**
   - https://towardsdatascience.com/various-implementations-of-collaborative-filtering-100385c6dfe0
   - https://www.ethanrosenthal.com/2015/11/02/intro-to-collaborative-filtering/
   - https://realpython.com/build-recommendation-engine-collaborative-filtering/

3. **Recommender System**

This recommender system recommends movies to new users by employing cosine similarity based collaborative filtering. (python package – sklearn.metrics.pairwise.cosine_simlarity)

We use a public dataset from MovieLens to create our initial user vs movie matrix. Every row in the matrix represents a unique user and his/hers rated movies. We compute row wise cosine similarity to identify the similarity between users based on their ratings for different movies and also compute column wise cosine similarity to identify the similarly rated movies based on the different users' ratings.

Now when a new user comes in, he/she can rate a few movies that they like and the user similarity matrix is recomputed based on the new entry. We can then find users similar to the new user based on the similarity scores and can predict that, since these users are similar to each other they might like same kind of movies. Hence, we can recommend movies based on what other users like.

4. **Experimental Results**

Here are the results by varying the number of users and number of movies considered for similarity matrix calculation and subsequent recommendations.

My ratings for movies:

Movie Name	Rating
Toy Story (1995)	5
Bridget Jones's Diary (2001)	2
Star Wars: Episode IV - A New Hope (1977)	4
Dumb & Dumber (Dumb and Dumber) (1994)	2.5
Forrest Gump (1994)	4

Jurassic Park (1993)	5
Silence of the Lambs, The (1991)	4.5
Shawshank Redemption, The (1994)	4
Terminator 2: Judgment Day (1991)	4
Batman Forever (1995)	2

a) No of users = 15, No of Movies = Top 50 with maximum ratings

Movies similarly rated to: “Aladdin”

	Aladdin (1992)
Aladdin (1992)	1.00000
Beauty and the Beast (1991)	0.83828
Bridget Jones's Diary (2001)	0.66453
Terminator 2: Judgment Day (1991)	0.64238
Sixth Sense, The (1999)	0.62940
Pretty Woman (1990)	0.60159

Movies similarly rated to: “Batman Forever”

	)	Batman Forever (1995)
Batman Forever (1995)		1.00000
Braveheart (1995)		0.91026
Apollo 13 (1995)		0.86924
Speed (1994)		0.83623
Shawshank Redemption, The (1994)		0.80915
True Lies (1994)		0.80812

Movies similarly rated to: “Independence Day”

		Independence Day (a.k.a. ID4) (
Independence Day (a.k.a. ID4) (1996)		1.00000
Forrest Gump (1994)		0.86267
Terminator 2: Judgment Day (1991)		0.81437
Star Wars: Episode VI - Return of the Jedi...		0.80851
Jurassic Park (1993)		0.80358
Clear and Present Danger (1994)		0.72463

Recommendations for me based on my ratings: ["Schindler's List (1993)" 'Gladiator (2000)', 'Star Wars: Episode VI - Return of the Jedi (1983)' 'Apollo 13 (1995)', 'Sixth Sense, The (1999)']

b) No of users = 100, No of Movies = Top 500 with maximum ratings

Movies similarly rated to: "Aladdin"

	Aladdin (1992)
Aladdin (1992)	1.00000
Lion King, The (1994)	0.73715
Beauty and the Beast (1991)	0.72333
Mask, The (1994)	0.65071
Coneheads (1993)	0.64174
Fugitive, The (1993)	0.62065

Movies similarly rated to: "Batman Forever"

	Batman Forever (1995)
Batman Forever (1995)	1.00000
Cliffhanger (1993)	0.78462
Outbreak (1995)	0.78026
True Lies (1994)	0.77804
Batman (1989)	0.71396
Ace Ventura: Pet Detective (1994)	0.71050

Movies similarly rated to: "Independence Day"

	Independence Day (a.k.a. I...
Independence Day (a.k.a. ID4) (1996)	1.00000
Mission: Impossible (1996)	0.70585
Jurassic Park (1993)	0.63528
Toy Story (1995)	0.62907
Star Wars: Episode IV - A New Hope ...	0.61823
Terminator 2: Judgment Day (1991)	0.61796

Recommendations for me- ['Die Hard (1988)' 'Airplane! (1980)' 'Terminator, The (1984)', 'Raiders of the Lost Ark (Indiana Jones and the Raiders of the Lost Ark) (1981)', 'Ghostbusters (a.k.a. Ghost Busters) (1984)']

c) No of users = All (950), No of Movies = All

Movies similarly rated to: “Aladdin”

	Aladdin (1992)
Aladdin (1992)	1.00000
Beauty and the Beast (1991)	0.74706
Lion King, The (1994)	0.71791
Jurassic Park (1993)	0.61348
True Lies (1994)	0.59991
Batman (1989)	0.59672

Movies similarly rated to: “Batman Forever”

	Batman Forever (1995)
Batman Forever (1995)	1.00000
Batman (1989)	0.70564
True Lies (1994)	0.64274
Ace Ventura: Pet Detective (1994)	0.62043
Cliffhanger (1993)	0.60187
Dances with Wolves (1990)	0.58593

Movies similarly rated to: “Independence Day”

	Independence Day (a.k.a. ID4) (1...
Independence Day (a.k.a. ID4) (1996)	1.00000
Mission: Impossible (1996)	0.67735
Jurassic Park (1993)	0.63842
Twister (1996)	0.62062
Star Wars: Episode VI - Return of the...	0.58885
Terminator 2: Judgment Day (1991)	0.58841

Recommendations for me- ['X2: X-Men United (2003)' 'Finding Nemo (2003)' 'Shrek (2001)', 'Gladiator (2000)' 'Lord of the Rings: The Two Towers, The (2002)']

5. Results and Discussion

For the movie similarity – When the number of users and movies are less and there are not enough ratings for movie like 'batman forever', and hence I do not see an apt recommendation. The movies seem to be similarly rated but the data points to judge it on is not enough.
But, when the same is carried out for higher number of users (like in (C)), we get more ratings for a movie by more users and the recommendations for similar movies seem to be valid and aligned.

Hence, I agree with similarly rated movies recommendations, but at the same time I can observe that it is more relevant when the number of data points i.e., users and movies is higher.

Movie Recommendations based on my inputs -

For part a) I only like 2 out of the 5 movies that were recommended. The recommendations aren't the best because I am only being compared with only 15 other users based on only 50 movies. Hence the intersection of my interest and theirs isn't optimal.

But, for part b) and c) of the experiments, I really like the recommendations. I think it aligns with my interests of cartoons, sci-fi and action movies. I understand that these recommendations are way better than part a) because of the number of input data points. The matrix created was huge and I was compared to almost 100 different users (in b) matching me with users with a similarity score of about (0.6) (unlike part a) where I was matched to user with score 0.4).

Hence, with an increased number of users and movies data set, I feel the recommendations did get better as well as the similarly rated movies.

I also realize that there is a scope of improvement using alternative model based techniques which might have better predictions and small MSE error values.

6. My experiences and lessons learned

- I got a hands-on experience on building recommender systems using collaborative filtering.
- Learnt about different python packages and dataframes.
- Reading about collaborative filtering also helped me learn about other memory-based techniques like content-based filtering and k-near user based collaborative filtering.
- I learn more about model-based methods using neural networks during my research.
- Overall, I found the project pretty rewarding and helpful.