CAIRO UNIVERSITY

FACULTY OF ENGINEERING

HEMN454 – Data Mining and Machine Learning in Healthcare

Task 4 – Recommender Systems

Amira Mahmoud Farid 1170498

Recommender Systems

Code Description

1. read movies and ratings data
2. select the first 200 users
3. merge users with movies
4. view movies description and ratings description
5. get sum, mean, and count of ratings of all movies
6. creating dataframe with 'rating' mean values
7. plot 'num of ratings column' graph
8. plot of 'ratings' column graph
9. Construct the required TF-IDF matrix by fitting and transforming the data
10. check shape of similarity matrix
11. Sorting values according to the 'num of rating column'
12. calculate the cosine similarity matrix
13. build a reverse map of indices and movie titles
14. create get_recommendations function

    Function that takes in movie title as input and outputs most similar movies

    input parameters: title, cosine_sim = cosine_sim

    return: the top 10 most similar movies

    14.1. get the pairwsie similarity scores of all movies with that movie

    14.2. sort the movies based on the similarity scores

    14.3. get the scores of the 10 most similar movies

    14.4. get the movie indices

    14.5. return the top 10 most similar movies
15. call function to get top 10 Similar movies to Toy Story (1995):
16. call function to get top 10 Similar movies to Waiting to Exhale (1995)

References

https://www.geeksforgeeks.org/python-implementation-of-movie-recommender-system/

https://www.kaggle.com/shubham3117/recommendation-engine