Collaborative Filtering Recommender System – MovieLens 100k

# 1. Introduction

In this project, we built a movie recommender system using Collaborative Filtering techniques on the MovieLens 100k dataset. Collaborative Filtering (CF) works by recommending items based on user interactions — not based on item content. It's how platforms like Netflix and Amazon suggest products or movies.

# 2. What is Collaborative Filtering?

Collaborative Filtering makes predictions about a user’s interests by collecting preferences from many users. It assumes that if two users rated some items similarly, they will rate other items similarly too.

There are two major types of Collaborative Filtering:

* - User-Based Filtering: Finds users similar to you, and recommends items they liked.
* - Item-Based Filtering: Finds items similar to what you liked, and recommends those.

# 3. Dataset: MovieLens 100k

The MovieLens 100k dataset contains 100,000 ratings from 943 users on 1,682 movies. Each user has rated at least 20 movies. The two main files used are:  
  
- u.data: Contains ratings with user\_id, item\_id, rating, and timestamp.  
- u.item: Contains mapping from item\_id to movie titles.  
  
This data was used to create a user-item rating matrix, which is the foundation for Collaborative Filtering.

# 4. Evaluation Metric

We use RMSE (Root Mean Squared Error) to evaluate how close the predicted ratings are to the actual ratings. A lower RMSE means the model is better at predicting user preferences.

# 5. Collaborative Filtering Methods Tested

We tested the following three methods:

* - User-Based Filtering with Cosine Similarity
* - Item-Based Filtering with Cosine Similarity
* - User-Based Filtering with Pearson Correlation

Cosine Similarity compares users or items based on the angle between their rating vectors, while Pearson Correlation considers the linear relationship between ratings by subtracting each user's mean rating.

# 6. Results

Here are the results from the three tested models:

|  |  |  |
| --- | --- | --- |
| Method | RMSE | Predicted Rating (User 196 → Movie 302) |
| User-Based + Cosine | 1.0219 | 4.20 |
| Item-Based + Cosine | 0.4893 | 3.61 |
| User-Based + Pearson | 0.9442 | 3.93 |

# 7. Interpretation of Results

- Item-Based Filtering with Cosine Similarity had the lowest RMSE (0.4893), making it the most accurate method for predicting ratings.  
- While User-Based models work well, Item-Based Filtering tends to perform better in practice, especially when there are more items than users.  
- Pearson correlation can sometimes outperform cosine for highly personal rating patterns, but that wasn’t the case here.

# 8. Final Recommendations for User 196

Based on the trained model, here are the top 5 movie recommendations for User 196 (sorted by predicted rating):

- The Godfather (1972)  
- Star Wars (1977)  
- Fargo (1996)  
- Toy Story (1995)  
- L.A. Confidential (1997)

# 9. Conclusion

Collaborative Filtering is a powerful approach for recommendation systems, especially when content metadata is unavailable. This project showcased how similarity metrics like Cosine and Pearson can be applied to either users or items to generate accurate predictions. The MovieLens dataset provided an excellent base to evaluate and understand the different Collaborative Filtering strategies.