Hi Prof. Hua,

I have developed a **movie recommendation system** using two approaches:

1. **Collaborative Filtering** – Recommends movies based on user preferences.
2. **Content-Based Filtering** – Recommends movies based on their features (genres).

**EDA:**

The dataset contains missing values:

* NaN values were replaced with 0 to create a usable matrix for collaborative filtering.

To reduce noise and improve recommendation accuracy:

* Movies with less than 10 ratings were removed.
* Users who rated fewer than 50 movies were removed.

**Method 1:**

Collaborative Filtering is a technique that recommends movies based on the preferences of similar users.

1. Converting Data into a Sparse Matrix since lot of values are 0’s.
2. A KNN model is used to find similar movies based on user rating patterns. The similarity between movies is measured using cosine similarity.
3. Created a recommendation function.
4. Using Gradio created small app for user interaction.

A screenshot of a computer

AI-generated content may be incorrect.

**Method 2:**

Content-based filtering recommends movies that are similar in content (e.g., genre). Instead of looking at user ratings, it compares the attributes of the movies.

The genres column contains multiple genres separated by |. To process this:

* Genres were split into a list for each movie.
* One-hot encoding was applied to convert genres into numerical features.

A cosine similarity matrix was created to compute similarity scores between movies based on genres.

A screenshot of a computer

AI-generated content may be incorrect.

**References:**

<https://www.datacamp.com/courses/building-recommendation-engines-in-python>

<https://towardsdatascience.com/recommender-systems-a-complete-guide-to-machine-learning-models-96d3f94ea748/>

<https://www.geeksforgeeks.org/recommendation-system-in-python/>

<https://www.youtube.com/watch?v=kuC38ZCcbZI>

<https://www.youtube.com/watch?v=XfAe-HLysOM>