Milestone 1 - Proposal

Topic: Movie Recommendation Web Application

Business Problem:

One major challenge faced by movie viewers is making a choice from the large number of available movies, especially from online streaming platforms. This project aims to solve this problem by building a web application that provides movie recommendations based on a user's input of a favorite movie.

Datasets:

The dataset will be obtained from the MovieLens dataset provided by GroupLens Research. This dataset comprises of several movie ratings by various users. Specifically, we will be using two data files: Movies.csv and Ratings.csv.

Methods:

The main method used in this project is Item-Based Collaborative Filtering, leveraging cosine similarity as the measure of similarity between different movies. We also make use of Python's Flask framework for the backend, HTML and CSS for the frontend, and Pandas for data manipulation.

Ethical Considerations:

While the project doesn't deal with sensitive data, it's important to ensure that any recommendation system is fair and unbiased. It should not favor certain movies due to reasons beyond the scope of user ratings and preferences. It's also important to note that users of the application won't be supplying personal data, preserving their privacy.

Challenges/Issues:

One challenge might be dealing with sparse data. The MovieLens dataset is quite large, and many users have only rated a few movies, resulting in a sparse user-item interaction matrix. Memory allocation might be another issue due to the size of the dataset when creating a movie-user matrix. Also, handling movies with the same title but different release years might be a bit tricky.

References:

GroupLens: MovieLens Datasets (https://grouplens.org/datasets/movielens/)

Python Flask Documentation (https://flask.palletsprojects.com/en/2.0.x/)

Scikit-learn: Machine Learning in Python (https://scikit-learn.org/stable/)