Movie Recommendation System

Introduction

In today's digital age, people consume a vast amount of media content, particularly movies. With the overwhelming number of options available, selecting a movie that aligns with a user's preference becomes increasingly difficult. Recommendation systems play a vital role in enhancing user experience by filtering and suggesting content based on user data and preferences. This project focuses on building a Movie Recommendation System that can intelligently suggest movies to users using both collaborative and content-based filtering techniques. The system aims to provide a simplified, interactive, and user-friendly interface using Streamlit, making it accessible to all users regardless of technical background.

Abstract

The objective of this project is to implement a hybrid movie recommendation system using machine learning algorithms and tools. It leverages Collaborative Filtering, which recommends movies by identifying users with similar tastes, and Content-Based Filtering, which suggests movies based on genres and content similarity. The project uses the MovieLens dataset, which contains real-world data of movies and user ratings. Python is used for data preprocessing, model development, and logic implementation, while Streamlit is utilized to develop a simple web application. This system serves as an efficient tool for personalized movie discovery and demonstrates the practical application of machine learning in real-life scenarios.

Tools Used

• Python – Core programming language for backend logic and data processing.
• Pandas – For data loading, cleaning, and manipulation.
• Scikit-learn – For building machine learning models and similarity computations.
• Streamlit – To build the web interface for user interaction.
• MovieLens Dataset – A dataset containing movie metadata and user ratings.
• Jupyter Notebook / Visual Studio Code – For development and testing.

Steps Involved in Building the Project

1. Data Collection: Downloaded and extracted 'movies.csv' and 'ratings.csv' from the MovieLens dataset.
2. Data Preprocessing: Cleaned data using Pandas, merged datasets, and handled missing

values.
3. Collaborative Filtering: Constructed a user-movie rating matrix and calculated similarity between users using cosine similarity.
4. Content-Based Filtering: Processed genres using TF-IDF vectorization and computed content similarity among movies.
5. Recommendation Logic: Developed functions to generate movie recommendations based on selected method.
6. Interface Design: Built an interactive UI using Streamlit, allowing users to input preferences.
7. Testing & Validation: Tested the application for different user IDs and genres to ensure valid recommendations.
8. Documentation: Created a structured project report and usage guide.

Conclusion

The Movie Recommendation System project demonstrates the effective application of data science and machine learning techniques to build a real-world solution. The hybrid recommendation approach combines the strengths of collaborative and content-based filtering, resulting in a robust and personalized experience for users. Using Streamlit, the application was made accessible and easy to use, enhancing its practicality. This project lays the foundation for further enhancements such as incorporating sentiment analysis, user reviews, and integrating external APIs for richer movie metadata. In the future, deep learning techniques or reinforcement learning could be used to further improve the recommendation quality.