Movie Recommendation System – Project Report

1. Introduction

The current era of digital streaming makes users get confused by the enormous count of movies available on platforms such as Netflix, Amazon Prime, and more. It is a difficult decision to select a movie to watch. A content-based movie recommendation system facilitates the work by suggesting suitable movies according to user interest. The purpose of this project is to establish a content-based movie recommendation system based on machine learning. The system employs genre similarity to find and recommend comparable movies and displays the output through a simple user interface developed with Streamlit. The project illustrates the end-to-end application of machine learning and web deployment in finding a real-world solution.

2. Abstract

The project targets developing a content-based recommendation system that utilizes genre information to recommend similar movies. With the MovieLens dataset, we preprocess the data to obtain significant features and use TF-IDF vectorization to transform text of genres into numeric vectors. The cosine similarity is used to identify similarities between movies based on similarity in genres. The system displays the top 5 movie recommendations for a user-chosen movie via a web interface. The method is easy, efficient, and scalable. Streamlit interface makes it easy for the users to engage with the model, making the system usable by general users. Other methods such as collaborative filtering or sentiment analysis can be included to enhance performance in later releases.

3. Tools Used

Python: Main programming language employed in development.

Pandas: Data manipulation, filtering, and preprocessing.

Scikit-learn: Used for TF-IDF vectorization and cosine similarity calculation.

Streamlit: To develop the web-based user interface for the recommendation system.

MovieLens Dataset: Utilized public dataset for movie metadata and genres.

4. Steps Involved in Building the Project

Data Collection:

The dataset used is the “movies.csv” file from the MovieLens dataset, which contains movie titles and genre information.

Data Preprocessing:

Read the dataset using Pandas.Processed the genres column to replace pipe-separated strings with spaces for analysis.Checked for and dealt with any missing values.

Feature Extraction:

Used TF-IDF Vectorizer on the genres column to generate numerical representations.This enabled the model to learn the importance of various genre words.

Similarity Computation:

Used cosine similarity to compute movies based on genre vectors.This gave a matrix of pairwise movie similarities.

Recommendation Function:

Created a function to take a movie name as input and return the top 5 similar movies based on genre similarity scores.

Building the User Interface:

Developed a web interface using Streamlit.The user picks a movie from a dropdown and clicks the "Recommend" button.The application shows the suggested movie titles below the button.

5. Conclusion

This project successfully illustrates how machine learning algorithms can be utilized to create a simple yet effective movie recommendation system. By putting together TF-IDF vectorization and cosine similarity with a clean web interface, the system makes sense of recommendations in real-time. The genre-based approach is a good way to go and can be further enriched with extra information such as user ratings or movie reviews. Collaborative filtering, user profiling, and deep learning-based recommendations are potential areas for improvement in the future. In general, the project demonstrates how ML and data science principles can be used to address common issues in the entertainment sector.