Final Report: Movie Recommendation System

1. Introduction

With the growth in the number of movies that can be viewed online, selecting the appropriate movie becomes an intimidating task. Recommender systems assist users in finding their favorite kind of content. This project aims to construct a content-based movie recommendation system that recommends similar movies based on genre similarity, enabling users to find suitable and well-rated movies.

2. Abstract

The project leverages TF-IDF vectorization on movie genres and a Nearest Neighbors algorithm to recommend similar movies. A clean user interface built with Streamlit allows users to select a movie and receive recommendations sorted by average user ratings. Additional metadata such as IMDb links is also provided to enhance user experience.

3. Tools Used

- Python

- Pandas, Scikit-learn

- Streamlit (for UI)

- MovieLens Dataset (movies.csv, ratings.csv, links.csv)

4. Steps Involved in Building the Project

1. Data Preprocessing

   - Loaded movies.csv, ratings.csv, and links.csv.

   - Filled missing genre data and merged average ratings and IMDb IDs.

2. Feature Engineering

   - Applied TF-IDF vectorization to the genres column using TfidfVectorizer.

3. Model Building

   - Used NearestNeighbors from scikit-learn with cosine similarity.

- Fitted the model on the TF-IDF matrix.

4. Recommendation Logic

- Given a movie title, retrieved similar movies based on genre proximity.

- Sorted results by average user rating.

5. Streamlit App Development

- Created an interactive UI using selectbox, spinner, and markdown.

- Displayed each recommendation with title, genre, rating, and IMDb link.

5. Conclusion

This system demonstrates the practical application of content-based filtering for movie recommendation. It offers accurate and intuitive suggestions without requiring user profile data. This method is scalable and effective for platforms with rich genre data but limited user behavior data.