Movie Recommendation System using Nearest Neighbors

Objective

This project builds a content-based movie recommendation system using TF-IDF vectorization and the Nearest Neighbors algorithm. The goal is to suggest similar movies based on their titles.

Dataset Overview

The dataset contains a list of movies with the following attributes:

- **ID: Unique movie identifier**
- **Year: Release year of the movie**
- **Movie Name: Title of the movie**

Data Preprocessing

- **The dataset is loaded with appropriate encoding to handle errors in character formats.**
- **Missing values are removed to ensure data consistency.**
- **A TF-IDF (Term Frequency-Inverse Document Frequency) vectorizer is applied to extract features from movie titles by converting text into numerical vectors.**

Building the Recommendation Model

- **The TF-IDF matrix is generated to represent the movie titles in vector space, helping to measure similarities between movies.**
- **A Nearest Neighbors (KNN) model is trained using cosine similarity as the distance metric.**
- **Given a movie title, the system retrieves the top 5 most similar movies by computing their cosine distances from the selected movie.**

Example Recommendation

- **If a user selects a movie, the system suggests five similar movies based on title similarity.**
- **The model successfully finds recommendations that match the theme and keywords of the given movie title.**

Conclusion

This project demonstrates how text-based similarity can be used for movie recommendations. By utilizing TF-IDF and Nearest Neighbors, the system provides efficient and relevant suggestions. This approach can be expanded by incorporating user preferences, genres, and ratings for more personalized recommendations.