Description:

The aim of this project is to build a movie recommender system using content-based recommendation techniques. The system uses the metadata of movies such as genre, director, actors, and plot summary to recommend similar movies to the user. The project is implemented using Python libraries like NumPy, Pandas, Scikit-learn, Streamlit, Pickle, Requests and nltk. Further, Streamlit is used to build the web app of the following Data Science Project.

Project Steps:

- **Data Collection:** Collected the movie data from popular data source Kaggle. The data is of TMDB website, which was initially distributed in two different datasets: Movies and Credits. Containing various information about the movie and the crew worked in its development. Data Source Link: Kaggle
- **Data Cleaning and Preprocessing:** First, the two datasets are merged by the 'movie_id' column, which was same in both the datasets. Clean the data by removing duplicates, missing values, and irrelevant columns. Preprocess the data by tokenizing the plot summary, genre, cast, crew, keywords, and overview, and by removing stop words, into a new column named as "Tags". Saved the preprocessed data into a new dataframe.
- **Feature Extraction:** Extract the important features from the preprocessed data like movie_id, title, and Tags.
- **Build Recommender System:** Use the extracted features to build a content-based movie recommender system. The system recommends movies to the user based on the movie selections.

Python Libraries:

- NumPy, Pandas: Used for data manipulation and cleaning.
- Scikit-learn: Used for feature extraction and building the recommender system.
- NLTK: Used for natural language processing, preprocessing the plot summary.
- Streamlit: Used for building the web app of the project
- Request: Used to make the TMDB API call to fetch the movie's poster images from the TMDB website.
- Pickle: Used for converting Python object into a byte stream for sending the data from ipynb file to py file.

Conclusion:

In this project,I have built a movie recommender system using content-based recommendation techniques. We have used various Python libraries like pandas, scikit-learn, and nltk to implement the system, and Streamlit to make the web app of it. The system recommends movies to the user based on their movie selections.