Movie Recommendation System

A machine learning-based Movie Recommendation System built using the TMDB dataset (Movies + Credits). The system recommends movies similar to a selected movie based on content features.

Project Overview

Recommendation systems are widely used in platforms like Netflix, Amazon Prime, and YouTube. This project demonstrates a content-based recommendation engine, which suggests movies by analyzing similarities in their metadata (genres, keywords, cast, crew, and overview).

Features

- Recommend movies based on a chosen title
- Uses cosine similarity to find the closest matches
- Combines data from both movies dataset and credits dataset
- Clean feature engineering for improved accuracy

Technologies Used

- Python
- Pandas, NumPy → Data preprocessing
- Scikit-learn → Vectorization & similarity calculation
- NLTK / re → Text cleaning (if used)
- Streamlit / Flask (optional) → For deployment

Dataset

This project uses two datasets from TMDB:

1. tmdb_5000_movies.csv
    - Contains: movie id, title, overview, genres, keywords, etc.
2. tmdb_5000_credits.csv
    - Contains: cast and crew details for each movie

Dataset link: [TMDB 5000 Movie Dataset on Kaggle]

How It Works

1. Load both datasets (movies.csv & credits.csv).
2. Merge them using the **movie ID**.
3. Select important columns: **genres, keywords, overview, cast, crew**.
4. Preprocess text (cleaning + combining into a single "tags" column).
5. Convert text into vectors using **CountVectorizer / TF-IDF**.
6. Compute similarity between movies using **Cosine Similarity**.
7. When a user selects a movie, recommend the **Top N most similar movies**.

Requirements

- Python 3.x
- pandas
- numpy
- scikit-learn

Example Output

If the user selects **"The Dark Knight Rises"**, recommendations might include:

- The Dark Knight
- Batman Begins
- Man of Steel
- Superman Returns
- Avengers: Age of Ultron