🎬 Project Title: Movie Recommendation System

1. Introduction

Movie recommendation systems are widely used on streaming platforms like Netflix, Amazon Prime, and Hotstar to personalize content for users. This project demonstrates how to build a content-based recommendation system using NLP techniques to suggest movies similar to a selected one based on genres, keywords, and overview.

2. Abstract

The system uses the TMDB 5000 Movies Dataset from Kaggle. After preprocessing the dataset to combine key textual features, the text is vectorized using TF-IDF. A cosine similarity matrix is generated to measure similarity between movies. The final output is a functional web app built with Streamlit that allows users to select a movie and view top 5 similar suggestions. The model doesn't require user history, making it ideal for cold-start scenarios.

3. Tools and Technologies

- Python (Jupyter Notebook & Streamlit)
- Pandas, NumPy
- Scikit-learn (TF-IDF, cosine_similarity)
- Streamlit (UI)
- Dataset: TMDB 5000 Movies Dataset

4. Workflow

- Load and clean dataset
- Extract and combine key text features (overview, genres, keywords)
- Generate a combined tags column
- Use TF-IDF vectorizer to convert text into vectors
- Compute cosine similarity between movies
- Recommend top 5 movies based on selected title
- Build a Streamlit app for user interaction

5. Results

The application successfully recommends movies based on textual similarity. For example, selecting “Avatar” suggests science-fiction and action films with similar plots and themes.

6. Conclusion

This project demonstrates the power of content-based filtering using NLP and TF-IDF. It can be enhanced further with collaborative filtering or by integrating TMDB API for poster images and detailed movie data.

7. Future Enhancements

- Add posters and movie details using TMDB API
- Combine with collaborative filtering for hybrid recommendations
- Allow search input instead of dropdown