Movie Success Prediction and Sentiment Analysis Web App

Introduction

The movie industry generates large volumes of data related to genres, revenue, reviews, and audience behavior. Predicting whether a movie will be a *Hit*, *Flop*, or *Average* helps producers and marketers assess commercial potential. This project delivers a user-friendly web application that allows users to:

- Search for movies
- View detailed metadata
- Submit review comments for sentiment prediction
- Predict box-office outcomes using machine learning

Abstract

This application integrates data scraping, machine learning, natural language processing, and database management into one platform. It uses:

- A sentiment classifier for user reviews
- A movie success classifier based on revenue, budget, and other metadata

Real-time data is fetched using the OMDb and TMDb APIs and stored in an SQLite database. The application interface is built using Streamlit, enabling users to interact with the prediction tools easily.

Tools & Technologies Used

Category	Tools / Libraries
Programming Language	Python, SQL
Web Interface	Streamlit
APIs	OMDb API, TMDb API
ML Models	Random Forest, Logistic Regression
Data Handling	Pandas, NumPy, ast
Database	SQLite
File Formats	.pkl, .py, .ipynb

Steps Involved

1. Data Collection

- Movie metadata and posters were collected from OMDb and TMDb APIs.
- Kaggle datasets were used for model training.

2. Preprocessing

- Genre lists were processed for one-hot encoding.
- Missing values were handled or dropped.

3. Model Training

- A Random Forest Classifier was trained to predict movie success.
- A Logistic Regression model was used for sentiment analysis.

4. Sentiment Analysis

- User comments were classified as Positive or Negative.
- Scores were saved to the database for analytics.

5. Prediction & UI

- Genre-wise movie predictions (Hit/Flop/Average) were displayed with posters.
- Users received feedback and updates based on interactions.

6. Integration & Testing

- Models were saved as .pkl files and integrated into Streamlit pages.

Conclusion

This project successfully demonstrates how modern tools can be combined into a seamless movie analytics platform. The application enables:

- Access to enriched movie metadata
- Predictive modeling for commercial success
- Real-time sentiment tracking

This system can be extended to support real-world recommendation engines, production decisions, and audience behavior analytics.