Movie Success Prediction and Sentiment Study

Introduction

This project aims to explore the relationship between public sentiment, movie characteristics, and their commercial success. By combining natural language processing and machine learning, we can predict how well a movie performs and how audiences feel about it.

Abstract

The project leverages IMDB/Kaggle datasets to predict movie success using regression modeling and to study sentiment patterns in user reviews. VADER sentiment analysis was applied to extract viewer sentiment. The resulting model offers insight into which features (rating, sentiment, genre) influence box office performance.

Tools Used

-Python (Pandas, NLTK, Scikit-learn, Seaborn)

-NLTK + VADER for sentiment analysis

-Excel for data cleaning and inspection

Steps Involved in Building the Project

1. Data Collection: Movie metadata and reviews imported from IMDB/Kaggle datasets.

2. Preprocessing: Missing values were handled, and features like rating, sentiment, and genre were selected.

3. Sentiment Analysis: VADER was applied to analyze the polarity of user reviews and generate a compound score.

4. Predictive Modeling: A regression model was built using imdb rating, sentiment, and other features to predict box_office revenue.

5. Visualization: Charts were created to observe genre-based sentiment trends and compare actual vs predicted performance

Conclusion

This study showed that combining sentiment analysis with traditional movie metrics (like IMDB ratings) provides better insight into a film's potential success. Sentiment scores positively correlated with box office collections in many genres. This model can assist stakeholders in gauging movie performance early using public reaction data.