Movie Success Prediction

Introduction

The project aims at predicting the financial success of films based on the TMDB dataset. The film

industry is characterized by high capital expenditure, and the prediction of success will enable film

producers and investors to make informed decisions. The aim of the project is to examine film data

such as budget, revenue, popularity, and ratings to establish whether a film is successful or not.

Abstract

The project involved loading and combining the TMDB movie and credits datasets based on the

movie title as a common identifier. Data preprocessing was done to eliminate missing and unnecessary

data. Profit was determined by calculating revenue and budget, and a success indicator was

developed to categorize films as hits or flops. A Random Forest classification algorithm was

developed to predict the success of films based on budget, vote average, and popularity. The

performance of the algorithm was measured by accuracy and classification.

Tools Used

Python, Pandas, NumPy, Scikit-learn, Matplotlib

Steps Involved

1. Loaded TMDB movie and credits datasets.

2. Merged datasets using the title column.

3. Cleaned data and handled missing values.

4. Calculated movie profit and created success labels.

5.Selected relevant features for prediction.

6. Split data into training and testing sets.

7. Trained Random Forest machine learning model.

8. Evaluated model performance and visualized results.

Conclusion

The results indicate that budget, ratings, and popularity are important factors in determining movie

success. The machine learning model is able to predict movie success using past data. This method

can be used by film production companies and investors to make informed financial decisions