Movie Success Prediction and Sentiment Study

Prepared by M.TARUNGURURAJAN, July 20,2025

1. Project Overview

This project aims to predict the success of movies using various features such as budget, revenue, runtime, popularity, vote metrics, and genres. Additionally, it includes sentiment analysis of movie reviews to understand audience opinions and how they correlate with success.

2. Data Used

We used the 'movies_metadata.csv' dataset from TMDb. The dataset includes metadata for thousands of films such as title, budget, revenue, genres, popularity, and overview.

3. Data Preprocessing

Steps included:

- Handling missing values
- Filtering non-zero budget and revenue
- Parsing genres
- Converting release dates to year
- Creating success label where revenue > budget.

4. Predictive Modeling

We trained a Logistic Regression model using cleaned data to predict movie success. The target variable is a binary label indicating whether the movie's revenue exceeded its budget.

5. Sentiment Analysis

Movie reviews were analyzed using a pre-trained sentiment model (VADER/TextBlob). Reviews were labeled as Positive, Neutral, or Negative. We analyzed how sentiment affects success.

6. Visualizations & Dashboard

Dashboards were created using Power BI to show:

- Movie success by genre and year
- Distribution of budget and revenue
- Sentiment distribution by success.

The dashboard helps stakeholders identify patterns and trends.

7. Results

- Accuracy of predictive model: ~78%
- Most successful genres: Action, Adventure, Animation
- Positive sentiment is often correlated with higher success.

8. Tools Used

Python (pandas, numpy, sklearn, matplotlib), SQLite, Power BI, TextBlob/VADER for Sentiment Analysis.

9. Future Work

- Use deep learning models like LSTM for text analysis
- Incorporate external data like actor/crew stats
- Predict revenue range instead of binary classification.

10.🎬 Conclusion

This project explored the factors influencing movie success using a combination of predictive modelling and sentiment analysis. By processing and analysing movie metadata and user-generated reviews, we achieved the following:

- ✅ **Built a predictive model** using features like budget, revenue, runtime, and popularity, which accurately classified whether a movie was a success or not.
- ✅ **Performed sentiment analysis** on movie reviews to determine how audience sentiment correlates with movie performance.
- ✅ **Created visual dashboards** to highlight trends such as the genres most associated with blockbuster success, and how sentiment aligns with high-revenue films.
- ✅ **Stored processed data in SQLite** and enabled SQL-based querying for dynamic exploration and reporting.

Overall, the analysis demonstrated that both **objective factors** (like budget and vote count) and **subjective sentiment** (audience reviews) play a vital role in determining a film's success. The results offer valuable insights for producers, marketers, and streaming platforms in making data-driven decisions.