İsmail Sarp Erdem 32105

DSA210 Spring 2025

Movie Success Analysis Project Proposal

Problem Statement

The film industry is a multi-billion-dollar sector where predicting a movie's success remains a major challenge. Many factors contribute to a movie's financial performance, including cast, director, budget, marketing, genre, and reviews. By using data science techniques, my aim is to analyze and predict the success of movies based on IMDb ratings, box office revenue, and other relevant attributes. This analysis will help us understand which factors contribute most to a movie's success.

How Are We Going To Tackle This Problem?

I will collect movie-related data from multiple sources, including IMDb and Box Office Mojo, using web scraping and APIs. My goal is to analyze the relationships between movie attributes (such as budget, cast, director, genre, etc.) and their box office revenue. By applying data science methodologies, we will identify trends, build predictive models, and provide insights into what makes a movie successful.

Methods and Tools We Are Planning to Use

- **Data Collection**: I will use web scraping tools such as BeautifulSoup and Selenium to extract data from IMDb and Box Office Mojo.

- **Data Processing & Analysis**: Libraries such as pandas, NumPy will be used for cleaning, preprocessing, and statistical analysis.

- **Machine Learning Models**: We will apply linear regression to predict a movie's box office revenue based on various attributes.

- **Visualization**: Tools such as Matplotlib and Seaborn will be used for data visualization to illustrate correlations.

Expected Outcome

We aim to create a predictive model that estimates a movie's box office revenue based on key attributes. Additionally, we expect to identify the most influential factors that contribute to a movie's commercial success. The findings will be visualized in an interactive report or dashboard for easy interpretation.

References

- IMDb: https://www.imdb.com

- Box Office Mojo: https://www.boxofficemojo.com

- The Movie Database API: https://www.themoviedb.org