CSE343/ECE343 : MACHINE LEARNING PROJECT PROPOSAL (Monsoon 2021)

BOX OFFICE PREDICTION

Utkarsh Dubey	Robin Garg	Bhaskar Gupta	Krishna Jalan
2019213	2019092	2019237	2021001

Motivation

The motion picture industry is a multibillion-dollar business, and there is a massive amount of data related to movies available over the internet. Box office revenue prediction is an important problem in the film industry that governs financial decisions made by producers and investors. Predicting the box office profits of a movie prior to its worldwide release is a significant but difficult problem that needs an advance of intelligence.

This project proposes a decision support system for the movie investment sector using machine learning techniques.

Related Work

1. *A Survey on Prediction of Movie's Box Office Collection Using Social Media*

   Study to list the factors upon which the success of a movie can depend.

2. *Predicting Box Office Revenue for Movies*

   The study provides a computational model for movie revenue prediction using a combination of features extracted from movie database metadata.

3. *A Machine Learning Approach to Predict Movie Box-Office Success*

   The study predicts an approximate success rate of a movie based on its profitability by analyzing historical data from different sources like IMDb, Rotten Tomatoes.

   Finally figuring out that budget, IMDb votes and no. of screens are the most important features.

Expected Timeline of project completion

Week	Expected Work
1	Research more about problem statement
2	Data Mining and/or Scraping
3	Data or Feature Analysis
4	Data preprocessing/cleaning
5	Training the model and Feature Selection
6	Analysis of training using different metrics
7	Evaluation and Normalization
8	Parameter Tuning and Refinement
9	Finalisation of Project for Presentation
10	Documentation and Report Writing

Work Distribution

1) Utkarsh Dubey - Data preprocessing, data scraping, analysis of data on graphs, Parameter Selection for models.

2) Robin Garg - Finding and analysing Dataset. Training Logistic Regression, Analysis using K-Fold technique.

3) Bhaskar Gupta - Verification of Accuracy/ Parameterization, Model Selection, Parameter/ Hyperparameter Selection

4) Krishna Jalan - Reducing noise from data, Analysis of results/models using different evaluation methods

Rest of the work like research, documentation/ report etc. will be done by all four of us.

Final Outcome

At the completion of this project, we will have a better and deeper understanding of the concepts of ML and the ability to apply them in real-world projects like this one. Ability to apply suitable data processing and cleaning techniques covered but not limited to the course and train/test different models available that suits our purpose/cause; differentiate and explain the result produced by each method.

A working model for predicting estimated revenue that a movie might generate beforehand.

Evaluate the accuracy/error of all the methods used and apply available tuning techniques to enhance the model accuracy. Detect and overcome the challenges like overfitting, underfitting, etc.