Meriç Demirörs(201401004)-Emre Belikırık(191401009) 03.10.2022

Project Proposal: Predicting the movie revenues with Machine Learning

Background

Trying to realistically calculate the revenue of a movie by looking at the watch number prediction charts of a newly released movie and some other data. Or, in some cases, the revenues of the movies are kept secret or not/can't calculated, revenue estimatation with the data for movies in such situations.

Objectives

• Pulling the necessary data from the IMDB site and filling in the blanks.
• Clean up the data that does not provide the required features and make the information trainable for the model.
• Creating meaningful new features by using different features with feature engineering.
• Train Machine Learning models and estimate revenue from view counts and other relevant data.
• Increase or decrease the number of data, changing the properties if necessary.
• Making this model available to everyone

Scope

It is a public model that can make a realistic revenue estimation when movie companies make a forecast for the future, or when the data of the movies with unknown revenues are given to the model.

Methods

We will be using known machine learning models(SVMs, Regression models, trees and forests, and more) Our models will be different in some ways from the other models used in similar problems models since our data has different features and more up to date.

Evaluation

We will be waiting for numeric guesses of revenues and our metric will be "how close is our revenue estimation to real revenue". We will discuss the result using knowing error calculating functions such as MSE, RMSE and more.

Time interval

Phase No.	Description of Work	Start and End Dates
Phase One	Data extraction and cleaning(using Python libraries)	September 24- October 25
Phase Two	Setting the data for learning	October 25- November 15
Phase Three	Training Model	November 15- December 15

References

- https://www.imdb.com/interfaces/
- https://datasets.imdbws.com/
- https://www.the-numbers.com/box-office-records/worldwide/all-movies/cumulative/all-time
- https://www.boxofficemojo.com/year/world/

And Readings

- https://arxiv.org/pdf/1804.03565.pdf
- https://www.researchgate.net/profile/Dipankar-Chaki/publication/322138608_A_Machine_Learning_Approach_to_Predict_Movie_Box-Office_Success/links/5a7e42934585154d57d4f318/A-Machine-Learning-Approach-to-Predict-Movie-Box-Office-Success.pdf?origin=publication_detail
- http://snap.stanford.edu/class/cs224w-2015/projects_2015/Predicting_Box_Office_Revenue_for_Movies.pdf
- file:///home/meric/%C4%B0ndirilenler/entropy-24-00711-v2-3.pdf