RUTGERS
THE STATE UNIVERSITY
OF NEW JERSEY

Movie Success Predictor

Akhil Velagapudi, Nithin Tammishetti, Pranav Kanukollu, Ramaseshan Parthasarathy, and Zeeshan Qureshi
{av481, nt264, pvk9, rp770, zq40}@scarletmail.rutgers.edu
Advisor: Prof. Marsic

Background

The movie industry is indisputably one of the most lucrative and influential establishments in the world. There are entire news outlets and websites dedicated to movie news, ratings, and discussion. Social media platforms are even filled with adverts and discussion about movie related topics. And with the rise of large public datasets and advancements in machine learning, data mining and structuring of data has become crucial to industries everywhere in the competitive business world. Our project hopes to collect, organize, and structure the massive amounts of movie related data from various sources and use it to predict the success of movies before release. A successful predictive model would not only allow the public to make more informed decisions about their movie consumption, but it would also enable theaters and industry leaders to make better choices that lead to profitable ventures.

Motivations and Objectives

- Movie theatres have arbitrary methods for pricing of tickets and planning for audience seating purely based on how successful they "think" movie will be and how much hype there is surrounding the movie
- Objectives
  - To structure and store movie data found all across the internet
  - To quantify an unreleased movie's "hype" by predicting global box office revenue
  - To build a simple mobile application for theaters and moviegoers to view the predictions

Android Application
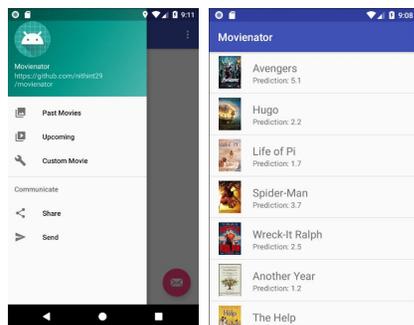
Acknowledgement

We would like to thank Professor Ivan Marsic for his continued support throughout the semester.
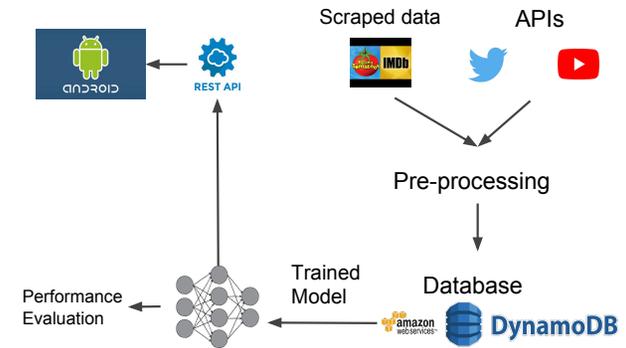
References

[1]Rhee, Travis Ginmu, and Farhana Zulkernine. "Predicting Movie Box Office Profitability: A Neural Network Approach." *2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA)*, 18 Dec. 2016, doi:10.1109/icmla.2016.0117.

Methodology

- Step 1: Obtained various data from different APIs and scraped from some sources (Wikipedia)
- Step 2: Preprocessed the data and populated the database (AWS DynamoDB)
- Step 3: Trained various ML models
- Step 4: Evaluated accuracy and chose best model for prediction
- Step 5: Displayed results on Android user interface

Our model was trained on a k-folds of historical movie data. The success measure was the revenue. Features trained upon include: budget, genre, critic reviews, google trends, trailer views, cast popularity, director popularity, etc…

Results and Performance

Distances between Predictions and Historical Box Office Revenue
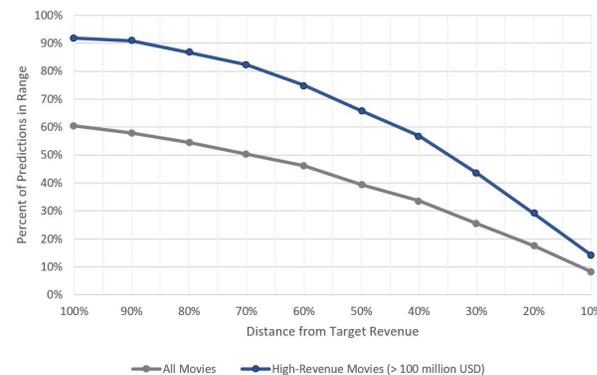
Random Sample Prediction Results

Movie Title	Prediction ($)	Target Historical Box Office Revenue ($)	Difference
Pixels	226,966,990	243,637,091	-7%
The Other Boleyn Girl	67,156,381	78,269,970	-14%
Riddick	133,129,331	98,337,295	35%
Just Go with It	245,791,375	214,918,407	14%
The Expendables	268,218,567	274,470,394	-2%
Precious	45,900,831	47,536,959	-3%
Terminator Genisys	535,177,178	440,603,537	21%
21	93,671,152	69,823,199	34%
Moana	567,128,087	643,331,111	-12%
The Shallows	85,253,966	119,100,758	-28%
Yes Man	211,448,890	225,990,978	-6%
Project Almanac	20,990,652	32,248,241	-35%