

## LITERATURE REVIEW

As searching through to internet and looking up papers in "Movie Revenue Prediction with Machine Learning" topic we can see that there is a notable amount of papers and studies which may and will inspire further works on topic. But, since we do not have time to take a look at all of this available information right now we picked four papers(which are referenced bottom of the document) and we read those ones, as we know there is more works and more experience about this topic in the internet we will continue to read and learn about other available papers. But for now we will be talking about selected four papers.

To summarize the common features for all four papers they all have some special datasets for themselves and unique methods to use, they all used IMDb data(some of them uses more than that[such as OMDB which is the second most popular data site in this four paper]). At some points each one has a difference side from others and these differences are listed below(some attributes are same but since they are not used in all four papers they are included):

- Uses ABD and China specific data only[1]
- Uses gross domestic product (GDP) and the number of movie screens (NMS) as feature[1]
- Trains the model in the LSTM way[1]
- Looked up Rotten Tomatoes, Box Office Mojo and Metacritic for data[2]
- Calculates thing called "Star Power" for Actor/Actress/Director which indicates the "persons impact on movies popularity/revenue"[2]
- Calls attention to revenue spike at specific months such as festival times[2]
- Uses "budget of the movie" as a feature, and emphasized that is important[2]
- Says that some movies have a good revenue despite their quality just because they are a sequel to another good movie[2]
- Uses A graph representation of actor-actor, actor-movie and movie-movie relations[3]
- Takes data from ShowBiz Data(a private movie database website).[3]
- Uses 771 English movies released in the USA in 2010-2015[4]
- Uses MPAA rating, and indicates its importance[4]
- Movies production stage is not very useful to predict revenue unlike the first week of screening[4]
- Uses film critic and the academy award nominations data[4]
- Uses number of screening in the first weekend[4]
- Uses budget, star power, festival times and sequel data as well as second paper[4]
- Says that sequel movies has bigger revenue[4]
- removes release date feature since they think its unimportant[4]

Mentioned Machine Learning models are NN, SVM, NLP, Random Forest, Regression models, ARIMA and some other economic regression/prediction models which are out of our scope for now. Since all four papers data are separate from each other, as expected best performed models are separate also(SVM for [1], NN for [2], two layered NN for [3], Multiple Linear Regression Model(2 separate model for 2 different prediction type:{1.Pre-production: predicts revenue during the movies production state and 2.Post-Release: predicts revenue after movies first week screening})) for [4])

After looking up these papers, we saw that our data scraping source is much narrower than some of papers, despite this our data has more features(ratings, release year, total votes, number of votes at each point) than any other data in papers. We may consider expanding our data in the future if necessary.

And in the model side, we will seeking to include all models except NLP to our study, and after results came in, we will also try embedded models of these models. The reason to excluding NLP is that our data will not contain any text represented data to train.

#### REFERENCES:

- <https://www.google.com/url?sa=t&rct=j&q=&esrc=s&source=web&cd=&cad=rja&uact=8&ved=2ahUKEwiIiNH6ND6AhVsVfEDHcB6BREQFnoECBMQAQ&url=https%3A%2F%2Fwww.mdpi.com%2F1099-4300%2F24%2F5%2F711%2Fpdf&usg=AOvVaw0dyDg4d1gy5S2UwFZ6RWGj> [1]
- [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](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) [2]
- [http://snap.stanford.edu/class/cs224w-2015/projects\\_2015/Predicting\\_Box\\_Office\\_Revenue\\_for\\_Movies.pdf](http://snap.stanford.edu/class/cs224w-2015/projects_2015/Predicting_Box_Office_Revenue_for_Movies.pdf) [3]
- <https://arxiv.org/pdf/1804.03565.pdf> [4]