Nver Kasis

Liana Minasyan

Davit Mkhitaryan

Saten Harutyunyan

**Introduction and Literature Review**

In the 21st century, when types of entertainment have become more common than ever before, going to cinemas and watching movies are still a good option for spending leisure time. Although watching movies seems easy and entertaining, the actual process that it takes until a film is delivered to the viewer is not as simple. The amount of film budgets that is usually published on the Internet may seem really scary at first sight. In reality, they are even scarier. Proven among dozens of work and stated by Zipin (2019), “major studios don’t disclose the full budgets for their films (production, development, and marketing/advertising). This is in part because it costs far more to make and market a film than it seems.” Moreover, many famous films, despite their big success, do not really cover their expenses. For example, a lot of people were shocked [when leaked financial statements showed](http://www.deadline.com/2010/07/studio-shame-even-harry-potter-pic-loses-money-because-of-warner-bros-phony-baloney-accounting/) that Harry Potter and the Order of the Phoenix had allegedly lost 167 million dollars, despite 967 million dollars in global revenues (Anders, 2011).

There is quite a lot of research regarding the topic of movies. According to a study done by Simonton, although filmmaking requires substantial capital investment, it is not known whether cinematic creativity is positively correlated with the size of the film's budget. In our dataset, we have both the capital invested and the film’s budget which are both used in our models. In spite of the fact that this analysis is somewhat similar to ours, however, their sample of feature films were released between 1997 and 2001. Let’s find results that could be useful to producers today! (Simonton, D. K., 2005) Another interesting paper illustrates an approach that assesses the contributions of movie stars to the revenues of the films in which they appear. A review of the literature suggests that this issue has received only scant and inadequate treatment in the past. We will tackle it down and uncover its level of contribution today. (Wallace, W., Seigerman, A., & Holbrook, M., 1993)

In this project, we use a dataset of 2236 movies to find the best model that predicts if the movie will cover its budget based on several characteristics. Furthermore, we will predict the genre of the movie.

**Theoretical framework**

The data was obtained from the the moodle page of the American University of Armenia provided by Mr. Madoyan. We got lucky because the data was already factored and clean. We have employed 5 different models to predict whether or not a movie’s expense would be covered. The focus of this paper is to predict whether or not the movie’s costs will be covered or not as well as its genre. First, we took out a couple of variables that were not so useful when creating our model. For example, the title, Director, Writer, Actors, Plot, Language, Country have been omitted. The reason why we took them out is because they are are descriptive texts and are not relevant to our case. Next, we dug deep into the empirical analysis and tried the following models:

* Naive Bayes
* Decision Tree
* Logistic Regression
* K Nearest Neighbours

Later on in the paper, we will analyze and compare each and everyone of these models and choose the best amongst them.

**Data**

**Independent Variables:**

year <- Year of movie published (Integer)

duration <- Movie Duration, in minutes (Integer)

gross\_adjusted <- Adjusted Gross Revenue (Numeric)

budget\_adjusted <- Adjusted Film Budget (Integer)

cast\_facebook\_likes <- Cumulative amount of likes of Cast members on facebook (Integer)

reviews <- Number of film reviews (Integer)

index <- Movie Index (Numeric)

Rated <- Movie Rating (Factor)

Metascore <- Movie Metascore (Integer)

imdbRating <- IMDB Rating (Numeric)

imdbVotes <- Number of IMDB Votes (Integer)

Production <- Movie production company (Factor)

OscarWon <- Number of times Oscars won (Integer)

OtherWin <- Number of times won in other nominations (Integer)

OscarNom <- Number of times nominated for Oscar (Integer)

OtherNom <- Number of times nominated for other nomination (Integer)

**Dependent Variables:**

genre\_first <- Movie Genre (Factor)

Covered <- Budget Covered? (Factor)

**Descriptive Analysis**

(See in .Tex file)

**Regression Results**

(See in .Tex file)

In conclusion, in the process of predicting whether the movie will cover its expenses or not, the best model is the General Logistical Model because it has the highest accuracy level of 73.54%. As for predicting the genre, we knn

**References**

1. Zipin D. (2019). “How Exactly Do Movies Make Money?”.

https://www.investopedia.com/articles/investing/093015/how-exactly-do-movies-make-money.asp

1. Anders C. J. (2011). “How much money does a movie need to make to be profitable?”. <https://io9.gizmodo.com/how-much-money-does-a-movie-need-to-make-to-be-profitab-5747305>
2. SIMONTON, D. K. (2005), Cinematic Creativity and Production Budgets: Does Money Make the Movie?. The Journal of Creative Behavior. <https://onlinelibrary.wiley.com/doi/abs/10.1002/j.2162-6057.2005.tb01246.x>
3. Wallace, W., Seigerman, A., & Holbrook, M. (1993). THE ROLE OF ACTORS AND ACTRESSES IN THE SUCCESS OF FILMS: HOW MUCH IS A MOVIE STAR WORTH? Journal of Cultural Economics, 17(1), 1-27. Retrieved from www.jstor.org/stable/41810482