IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)
An International Open Access, Peer-reviewed, Refereed Journal

Film success prediction using ML and sentimental analysis

Abhishek Sharma
I.T
S.A.K.E.C
Mumbai,India

Sahil Shetty
I.T
S.A.K.E.C
Mumbai, India

Vinit Savla
I.T
S.A.K.E.C
Mumbai, India

Abstract— **Film industry is an very well established industry which is multi billion.Lot of stake holders are involved in making of a movie.Success of movie depends upon various factors such as cast, genre of movie etc etc.But in our project we focus upon actor performance based on previous movie.Machine learning algorithm like linear regression helps us to predict approximate rating and collection of movie. It helps us to identify whether movie will be hit or flop.**

Keywords—Linear regression ,machine learning

I. Introduction

While Making a movie there are various stakeholder of movie who are involved in a movie. To avoid low success rate of movie a mechanism is developed used to predict success of movie so the stake holder can use this mechanism to get the prediction of their movie and on the basis of it they can take decision. They can make a decision before release of movie to avoid loss. Previous performance of actor, actress, director influence success of movies in todays time due to their weightage. We have used machine learning technique to reduce level uncertainity. The system is used to predict future of movie for business purpose so decision can be made without taking risk and decision maker will have information about income and rating. Film industry has a impact over million of people. In given mechanism we focus on attribute related to success of movie prediction such as actor and actress contribute in a success of movie. The proposed system report on technique used. . We also found that, the budget of a film is no indication of how well-rated it will be, there is a downward trend in the quality of films over time. Another important factors are the director and actors/actresses involved in a film. while performing data mining on IMDB it is difficult

II. Literature Survey

Various Research have been done in this field. In 2006 when Netflix declare prize money of about 1 million USD to the best team for who improvise their movie rating algorithm cinematch. Google has an application system that work on search volume of movie trailer and on basis of that gives it prediction. On basis of that it predict the opening week collection. Awad, Delarocas and Zhang (2004) analyzed over the data that make impact on movie rating on movie success. They developed statistical models based on movie ratings to estimate forecast revenue. They examined relationship of critic and consumer communication and online word of mouth. They came to the decision that professional critics, traditional consumer communication and online word-of-mouth has great influence for increasing number of movie viewers. Some research say movie review is indirectly proportional to box office success. So we chose to develop an application based on multiple attribute. We have identified following patterns:1) Popularity of star caste is crucial success to movie.(2) combination of past successful genre and a sequel movie is another pattern for success.(3) new movie in not popular genre with well less known star caste could be a pattern for a flop. We obtain data for the previous movies through Kaggle repoistry. From here we downloaded the dataset and used as an input. Kaggle.com is a website that provides dataset for free for its users. Thus we got dataset for free of cost. This dataset consists of 651 rows and 32

III. Working

Our model follows a particular methodology deals with different steps of the project which consists of data collection, data preprocessing, generating training and testing dataset, model generation, prediction and outcomes.These method prevents us from getting any irrelavent data which further keeps our outcome more relevant and accurate for prediction . Here we collect data set from kaggle which consist of 32 attribute and 651 tuples. Further steps are explained below

A. Data Collection

The dataset is collected and kept so information is sorted out. Rundown which incorporates literary informationabout the information just as a table of film rank, the number of votes and film titles. We obtain data for the previous movies through Kaggle repoistry. From here we downloaded the dataset and used as an input. Kaggle.com is a website that provides dataset for free for its users. Thus we got dataset for free of cost. This dataset consists of 651 rows and 32

B. Data Preprocessing

Before applying data mining technique, we need to apply preprocessing technique to avoid duplicate value.To avoid duplicate value we use technique such as cleaning,varaiable transformation,partion and other techniques.Since the collected data available is raw so preproceesing is necessary.It is one of the most important phasefor project.After data is preprocessed we need to do data integration and transformation.

C. Generating Traing and Test Data set

Training dataset is collection of data set of attribute used in our model.Naive bayes classifier is trained on this dataset.we consider input vector and output vector in training of dataset.output vectors is also called as target.Current models run and is compared with target for each input vector. Based on result the model is adjusted. the test dataset is a dataset used to provide an unbiased evaluation of a *final* model fit on the training dataset.it is also knows holdout dataset

We do data analysis we analyse selected attribute that might help us with most accurate results.According to our model we have noted following attributes which are very cruial imdb_rating, imdb_num_votes, critics_rating, critics_score, audience_rating and audience_score. We genaralise a graph based on following attribute for future prediction

We can see relation between critics score and audience score of kaggle which makes our data analysis complete

D. Model Generation

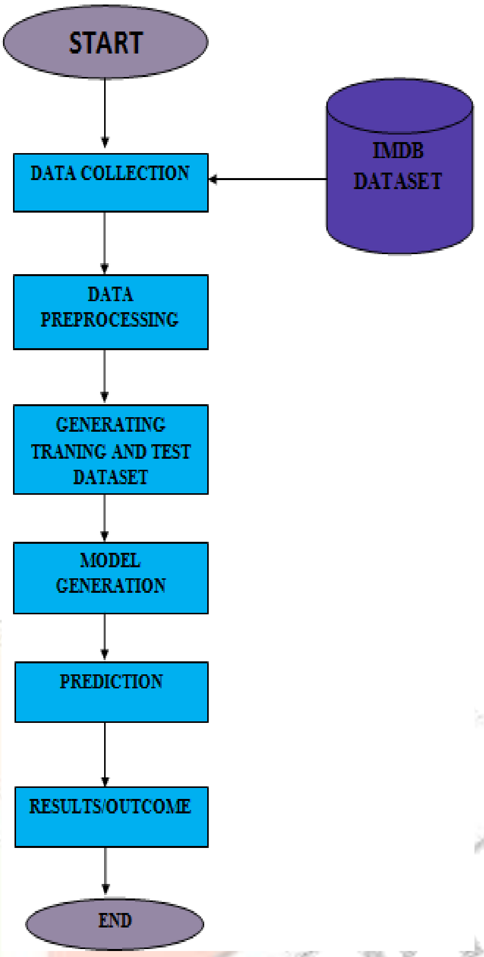
Modelling is defined as mathematically formulated way to make prediction of an approximation.In modeling we use statistical model of mathematical equation. Representing by statistical model we can use standard deviation. We use linear regression for model generation .Analyzing attribute plays crucial important to get accurate output. we analysed that the audience score and critics score are strongly correlated to each other also imdb rating and audience score are strongly correlated to each other. Thus these four attribute play very crucial role in our model

Linear regression is used in model generation .The algorithm is used as it helps us to give most accurate result .Multiple linear regression is similar to linear regression it is just an extension of linear regression . The relationship depends on two or more variable.Value we want to predict is dependent variable.The aim is to collect number of likes,dislikes, and view count of trailer, released date,star caste, their popularity .This helps us to know earning of the movie.Weka tool is used

E. Prediction

Prediction using machine learning is to identify data points on description of another data value. By using prediction we can derive a relationship between thing we know and thing we want to know.Regression analysis is used for prediction.We use critic score as input to predict audience score.Here movie name and critcs score is one of input then it will predict output as audience score. If it is close to early data then it fits in and on basis of that we can say that movie is hit or flop

IV. Workflow



V. Summary

Analysis is one of the most crucial thing in our model to develop any strategic plan .Feasibility studies help us how wide project can be used.WBS defines work done in different phase

VI. Conclusion

In this project we are trying to determine relationship between different attribute present.Here our aim is to establish relationship and how we can use that for prediction .Critic score is very crucial .It establishes relationship with audience score.Thus we can predict movie success based on critic score .Star caste involved in a movie plays an important role in a success prediction . previous performance of actors play an important role in it. we can assume that if we have movie gross score and movie net profit along with movie manufacturing cost, then we can build a more strong model for movie success prediction. In future, we can apply other machine learning algorithms for movie success prediction

Acknowledgment

We wish to express our profound gratitude to our principal Dr. Bhavesh Patel and project guide Prof. Lukesh Kadu and project co-guide Prof. Dhwaniket Kamble for allowing us to go ahead with this project and giving us the opportunity to explore this domain. We would also like to thank our I/c Head of Department Ms. Swati Nadkarni for our constant encouragement and support towards achieving this goal. We would also like to thank the Review Committee for their invaluable suggestions and feedback without whom our work would have been very difficulty. We take this opportunity to express our profound gratitude and deep regards to our guide Prof. Lukesh Kadu and co-guide Prof. Dhwaniket Kamble for him exemplary guidance, monitoring and constant encouragement throughout the course of this project. The blessing, help and guidance given by him time to time shall carry us a long way in the journey of life on which we are about to embark. No project is ever complete without the guidelines of these experts who have already established a mark on this path before and have become masters of it, and as a result, our teachers. So, we would like to take this opportunity to thank all those who have helped us in implementing this project.

References

- Basuroy, S., Chatterjee, S. and Ravid, S. A. (2003). How Critical Reviews? The Box Office Effects of Film Critics, Star Power, and Budgets. Journal of Marketing, 67(4), 103–117. https://doi.org/10.1509/jmkg.67.4.103.18692 Cerrito, P. B. (2008). The Difference between Predictive Modeling and Regression, 1–19. Chakravarty, A., Liu, Y. and Mazumdar, T. (2010).
- The Differential Effects of Online Word-of-Mouth and Critics' Reviews on Pre-release Movie Evaluation. Forthcoming at Journal of Interactive Marketing. https://doi.org/10.1016/j.intmar.2010.04.001 Chang, B.-H. D. G. Lowe, and Ki, E.-J.
- Devising a Practical Model for Predicting Theatrical Movie Success: Focusing on the Experience Good Property. Journal of Media Economics, 18(4), 247–269. https://doi.org/10.1207/s15327736me1804 Chok, N. (2010).
- Krushikanth R. Apala ; Merin Jose ; Supreme Motnam ; C.-C. Chan ; Kathy J. Liszka ; Federico de Gregorio" Prediction of Movies Box Office Performance Using Social Media", 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis (ASONAM 2013), PP 1209- 1214.
- What Determines Box Office Success of a Movie in the United States? Proceedings for the Northeast Region Decision Sciences Institute, (757), 447. Deuchert, E., Adjamah, K. and Pauly, F. (2005)