iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** V **Month of publication:** May 2023

DOI: https://doi.org/10.22214/ijraset.2023.52894

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue V May 2023- Available at www.ijraset.com

Movie Success and Target Audience Prediction

Dr. Shabnam Sayyad[1], Aditya Bangali[2], Pooja Ligade[3], Chhaya Gade[4], Aryan Khetarpal[5]
[1, 2, 3, 4, 5]*Computer Engineering Department, AISSMS College of Engineering, Pune*

Abstract: ***The Film Industry is one of the largest industries in the world that has a huge flowof capital flowing through it. Lots of movies release across the world through differentmediums whether theatrically or through OTT etc. But very few achieve the desired success. The unpredictability of the film industry and the huge amount of capital and workforce investment which is required in film production makes film makers nervous and they require amedium which will assure them about the prediction of success of their movie and also the target audience of their movie before its release, so that they can take right decisions about their project in time. In this model, the movie's success will be predicted and also the targetage group will be determined which can be targeted in attracting the audience by correct marketing and film making policies. Two Imdb datasets are utilized in the model respectively for prediction of success of the movie and its target audience. For predicting the success of themovie we have used Stochastic Gradient Descent Classifier Algorithm.***

Keywords: Stochastic Gradient Descent, Movie Success, Target Audience, Machine Learning, Classification.

I. INTRODUCTION

The film industry is a highly competitive domain, where the success or failure of a movie can significantly impact the financial returns and reputation of its creators. The ability to forecast a movie's success and identify its target audience beforehand is of paramount importance for film studios, production companies, and investors. This project aims to leverage data-driven techniques and predictive modeling to anticipate whether a movie is likely to be a hit or a flop, while also determining the target age group that it is most likely to appeal to.

By harnessing the power of machine learning and data analysis, this project seeks to offer valuable insights and aid decision-making processes within the movie industry. The ability to accurately predict a movie's performance at the box office can help filmmakers optimize their marketing strategies, allocate resources more efficiently, and increase their chances of producing successful and profitable films. Additionally, understanding the target age group for a movie is crucial for effective marketing and distribution. Different age groups have distinct preferences, tastes, and consumption patterns when it comes to movies. By identifying the appropriate age demographic, filmmakers can tailor their promotional campaigns, select appropriate theaters, and curate content that resonates with their intended audience. This project aims to provide a framework for predicting the primary target age group of a movie, offering valuable guidance to industry stakeholders seeking to maximize the reach and impact of their films. This model combines the power of predictive modeling, data analysis, and machine learning to forecast a movie's success and determine its target age group. By harnessing these insights, filmmakers, studios, and investors can make moreinformed decisions, optimize their marketing efforts, and increase the likelihood of producing financially successful movies that cater to their intended audience.

II. LITERATURE SURVEY

Sahu et al. [1] introduced a movie recommendation system that relies on content-based approaches. Their system utilizes various movie features such as genre, cast, director, keywords, and movie description. By incorporating the recommendation system output along with movie rating and voting information from similar movies, they developed a novel feature set. To predictmovie popularity across multiple classes, they proposed a deep learning model based on convolutional neural networks (CNN). Moreover, their research extended to predicting the popularity of upcoming movies among diverse age groups.

Reddy et al. [2] introduced a recommendation system focused on predicting user preferences based on movie genres. Their approach utilized content-based filtering techniques using genre correlation. The movie lens dataset was used as the basis for their system, and the data analysis tool employed was R Language. While simple recommendation systems consider only a few parameters, more complex ones take multiple parameters into account. In their research, the authors also highlighted the role of Mobile Cloud Computing (MCC) in energy conservation. However, they emphasized that security issues in MCC are significant challenges that require greater attention compared to other issues.

Abidi et al. [3] proposed a model for predicting movie popularity. Their approach involved implementing a combination of machine learning and statistical modeling techniques to investigate the problem.

©IJRASET: All Rights are Reserved

The main aim of their research was to enhance and compare the methodologies used by previous researchers. By conducting regression analysis, the model successfully predicted movie popularity and demonstrated accurate performance. It is worth noting that conducting data mining on IMDb poses challenges due to the extensive number of attributes associated with movies within a variable scope.

H.Verma and G.Verma [4] presented a prediction model for Bollywood success through a comparative analysis of performance-based supervised machine learning algorithms. The study aimed to develop and compare the performance of five different prediction models using a supervised machine learning algorithm system: Decision Tree (DT), Random Forest (RF), Support Vector Machine (SVM), and Logistic Regression (LR).One significant limitation of this study was the small sample size. The researchers were only able to obtain data from 200 movies, which was considered insufficient considering the requirements of the analysis.

Deldjoo et al. [5] presented a novel movie recommender system that aims to address the new item problem in the movie domain. Their approach involves integrating state-of-the-art audio and visual descriptors that can be automatically extracted from video content, forming what they refer to as the "movie genome." The system combines collaborative filtering (CF) and content-based filtering (CBF) techniques.However, the authors acknowledged that relying solely on CF and CBF may not be sufficient to ensure accurate weight learning. It is important to note that this method may not be applicable to all scenarios, and the authors highlighted certain drawbacks associated with its implementation.

Chen et al. [6] propose RRGAN, a generative model and framework for recommendation systems. RRGAN utilizes a minimax game between a generative model and a discriminative model. The generative model predicts ratings for a top N list of users or items based on reviews, while the discriminative model aims to distinguish between the predicted ratings and real ratings. This approach enhances the understanding of users and items using ratings and reviews, considering both explicit and implicit feedback. However, RRGAN is less efficient compared to other methods.

Yang et al. [7] propose a lambdaGAN top N recommendation model. Their approach incorporates the Lambda Strategy into generative adversarial training, optimizing the model using rank-based matrices. This allows for the application of generative adversarial training in pairwise scenarios, making it applicable for recommendation systems. The authors suggest that in future work, LambdaGAN can be extended to other scenarios such as web research and question answering, which presents a promising direction for further investigation.

Das et al. [8] introduce a novel approach to enhance movie recommendations for users. Their method builds upon matrix factorization (MF) techniques for recommendation but incorporates additional features extracted from movie posters and still frames. By integrating these visual features, they aim to predict users' movie watching interests more accurately, enabling more precise movie recommendations. The performance of their approach is further improved by leveraging enhanced convolutional neural network (CNN) features.

Van Meteren et al. [9] propose an innovative and efficient hybrid deep learning model called the Tag-Aware Recommender System based on Deep Learning-Intelligent Computing System. This model aims to enhance the tag-aware recommender system by utilizing deep learning techniques.The proposed model incorporates word embeddings to represent tags, enabling the utilization of semantic information in constructing user and item profiles. Additionally, the model integrates forward and recurrent neural networks to extract high-level latent features of users and items. This allows for the exploitation of both static and temporal features of user preferences.

Asad KI et al. [10] conducted a study that considered rating decisions in conjunction with inherent movie attributes to develop a classification approach. The research focused on the classification accuracy of decision tree and rule-based classifiers. The findings indicated that the director's rank and budget were significant factors in determining the classification outcomes. However, both classifiers did not yield substantial results. In addition, the study examined the correlation coefficient between the budget and financial returns on both domestic and worldwide scales. The analysis revealed a positive correlation, indicating that higher budgets were associated with higher financial returns, and vice versa. These findings have practical implications for film financing organizations, assisting them in making decisions regarding movie rentals, streaming services, brand sponsorship, and other related aspects.

III. OBJECTIVES

This research paper primarily concentrates on employing machine learning algorithms to forecast the Box office performance. The goals of the study are as follows:

1) To anticipate the potential success of upcoming movies.
2) To accurately determine the intended age group of audience for a movie.

3) To enhance the accuracy of predictive models through the utilization of machine learningtechniques.
4) To improve the efficiency of predictions.

IV. DATASET DESCRIPTION

1) *Dataset for Movie Success Prediction:* The dataset used for this model is IMDb dataset obtained from Kaggle website. This dataset contains information about 5043 movies and it has 28 columns containing various variables.
2) *Dataset for Target Audience Prediction:* This dataset is made by parsing the IMDb page by using ImdbPY Library. This dataset contains information about 2782 movies and ithas the variables which describe the voting demographics of various age groups like “under 18”, “from 18-29”,“from 30-44”,“45+” for every movie along with its genre.

V. METHODOLOGY

The model is developed using Python Programming Language and its many different libraries. The developed system will be available on any computer or laptop. The system will be available for the filmmaker who is interested in checking the chances of his upcoming movie‟s success for assurance and getting the target audience based on age groups so that the maker can develop marketing strategies according to the results. The filmmaker will enter the names of actors, director, genre and language as an input and the model will display the result as either Hit or Flop along with the prediction of vote average of each age group. SGD (Stochastic Gradient Descent) Classifier algorithm is used for predicting the success of movies. The SGD (Stochastic Gradient Descent) Classifier is an algorithm used for classification tasks in machine learning. Itis a linear classifier that uses stochastic gradient descent as its optimization method. The SGD Classifier is particularly well-suited for large-scale and sparse datasets, as it updates the model parameters based on small random subsets of the training data, making it computationally efficient. It is commonly used for tasks such as text classification and natural language processing.

A. *Movie Success Prediction*

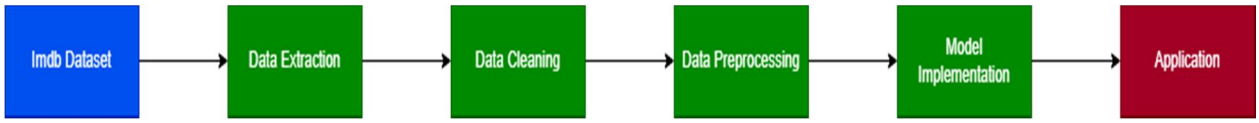


Fig. 1: Flow of system1

Step

1) *Data Extraction:* This is the first step of the model in which the data is extracted from the website called Kaggle which will be further utilised to analyse the data and then building the model which will generate correct predictions.
2) *Data Cleaning:* In this step, the cleaning of the data is done by performing various measures which deal with unwanted and corrupt data like missing values, tuples, etc so that the model will have good accuracy and little to no noisy data in the predictions. By using isna.sum() we got the total NaN values and by using dropna() we dropped them.
3) *Data Preprocessing:* In this step, we removed unwanted variables like plot_keywords, country, duration, actor_1_facebook_likes, actor_2_facebook_likes, actor_3_facebook_likes, color, from the predictions. They were so different from each other and hardly of any use for the model, hence we ignored them. The „genres‟ feature in the dataset had multiple categories in each row. So we splitted it and accepted only the first term in the entire entry and added a new column in a dataset named “main_genre” which contains the first term of every entry of „genres‟ column. We created a model that predicts either a movie is Hit or Flop, hence instead of taking imdb_score directly, we binned it into 2 buckets, they are 1~5 and 5~10, which represents Flop and Hit respectively. Thus prepared the data suitable for predictions by using SGD Classifier. Splitted the dataset into training and testing, 70% is the training data and 30% is the testing data.
4) *Model Implementation*: For the implementation purpose we used SGD Classifier, the SGDClassifier is part of the scikit-learn library, which is a popular machine learning library. TheSGD Classifier is imported from the „sklearn.linear_model‟ module within scikit-learn.

B. *Target Audience Prediction*

This subsection will work on the basis of input of the user for „main_genre‟ feature in the first model, the dataset used for this subsection has demographic data of voters based on their age groups. They are namely „under18‟, „from18-29‟, „from30to44‟ and „above45‟.

The details of total number of voters and vote average of every group for each movie are present along with the genre of the respective movie. We took genres and ignored the total number of votes and took the mean of vote average of each age group for every genre, which will be displayed when the user enters a particular genre in model first model.
Thus the user will directly come to know which age group of audience favours what genre on thebasis of historical data.

VI. RESULTS

The model developed on the SGD Classifier was able to achieve an accuracy of 90.04%



Fig. 2: Sample Input



Fig. 3: Output for Movie Success And Target Audience Prediction

VII. CONCLUSION

Coming towards the conclusion, the prediction of success of a movie and its target audience inits concept phase is an important task that can help the filmmakers and investors to properly hire the correct people and utilize appropriate marketing and promotional strategies to attract the people across age groups. The film industry is very large and it's one of the largest industries in the world. So the chances of getting failure are very high and thus the filmmakers and investors look for a reliable system that can help them in predicting the possible future results. The model developed is reliable due to its accuracy of 90.04% on testing dataset. In addition to detecting thesuccess of an unreleased movie it also predicts the target audience of the movie. This project can provide the way to filmmakers to build proper plans and develop strategies accordingly, to increase their possibilities of success.

©IJRASET: All Rights are Reserved

VIII. FUTURE SCOPE

1) Development of a mobile application can be done for movie makers to enhance feasibility.
2) Create a suggestion system for makers to receive pre-release strategy suggestions.
3) Creation of the system which will continuously update itself with the latest data, for convenience and adaptation to current trends, can be done.
4) The extension of this system can be done for television shows based on genres and TRPratings.
5) Utilisation of movie posters for result prediction.
6) Analysis of response by people to various movie promotional materials on various socialmedia platforms can be done for more reliable predictions.

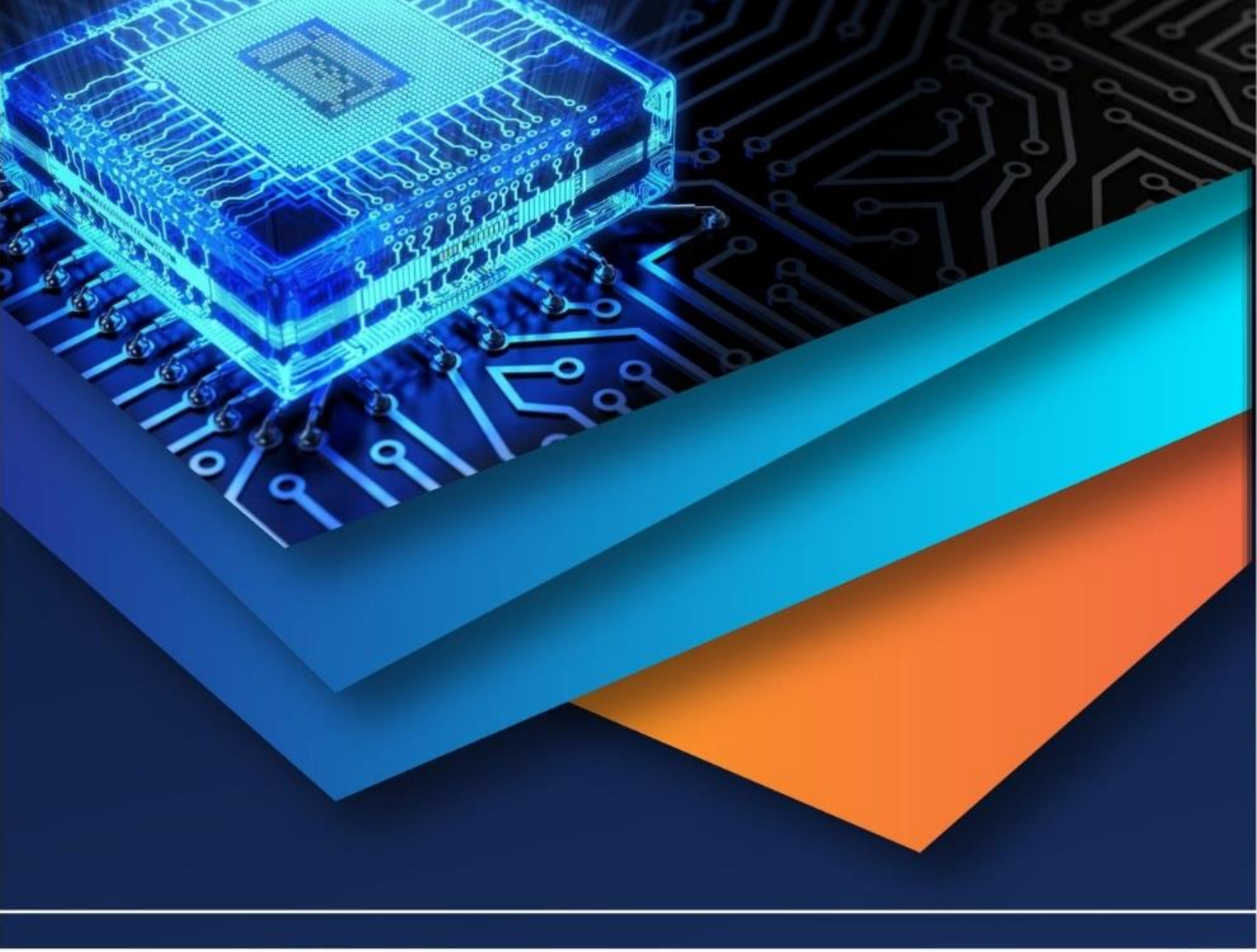
IX. ACKNOWLEDGMENT

The satisfaction and joy that accompany the successful completion of any task are greatly attributed to the individuals who have played a significant role in making it possible. Their unwavering support and encouragement have truly crowned our efforts with success. Therefore, we would like to take this opportunity to express my heartfelt appreciation to the following individuals:
First and foremost, we are sincerely grateful to Dr. D. S. Bormane, Principal of AISSMS Collegeof Engineering, for his continuous support and encouragement throughout this endeavor. His guidance has been instrumental in our achievements.
We would also like to extend my thanks to Dr. S. V. Athawale, Head of the Department of Computer Engineering, for his unwavering support and guidance. His expertise and direction have been invaluable throughout the project.
Furthermore, my gratitude goes to Dr. S. F. Sayyad, my project guide and Professor, for her constant supervision and setting precise deadlines. Her insightful suggestions have been a driving force behind the completion of this work.
Lastly, we would like to express our appreciation to the teaching and non-teaching staff of the Department of Computer Engineering for their cooperation and assistance. Their collective efforts have contributed to the smooth progress of this project. We are also thankful to our friends, whose direct or indirect assistance has been instrumental during the course of this project.
In conclusion, we are indebted to all these individuals for their unwavering support, guidance, and cooperation, which have played a pivotal role in the successful completion of this project.

REFERENCES

[1] S. Sahu, R. Kumar, M. S. Pathan, J. Shafi, Y. Kumar and M. F. Ijaz, "Movie Popularity and Target Audience Prediction Using the Content-Based Recommender System," in IEEE Access, vol. 10, pp. 42044-42060, 2022, doi: 10.1109/ACCESS.2022.3168161.
[2] S. R. S. Reddy, S. Nalluri, S. Kunisetti, S. Ashok, and B. Venkatesh, „„Content-based movie recommendation system using genre correlation,"" in Smart Intelligent Computing and Applications. Singapore: Springer, 2019, pp. 391–397.
[3] S. M. R. Abidi, Y. Xu, J. Ni, X. Wang, and W. Zhang, „„Popularity prediction of movies: From statistical modeling to machine learning techniques,"" Multimedia Tools Appl., vol. 79, nos. 47–48, pp. 35583–35617, Dec. 2020.
[4] [4]H. Verma and G. Verma, „„Prediction model for bollywood movie success: A comparative analysis of performance of supervised machine learning algorithms,"" Rev. Socionetwork Strategies, vol. 14, no. 1, pp. 1–17, Apr. 2020.
[5] Y. Deldjoo, M. F. Dacrema, M. G. Constantin, H. Eghbal-zadeh, S. Cereda, M. Schedl, B. Ionescu, and P. Cremonesi, „„Movie genome: Alleviating new item cold start in movie recommendation,"" User Model. User-Adapted Interact., vol. 29, no. 2, pp. 291–343, Apr. 2019.
[6] W. Chen, H.-T. Zheng, Y. Wang, W. Wang, and R. Zhang, „„Utilizing generative adversarial networks for recommendation based on ratings and reviews,"" in Proc. Int. Joint Conf. Neural Netw. (IJCNN), Jul. 2019, pp. 1–8.
[7] Y. Wang, H.-T. Zheng, W. Chen, and R. Zhang, „„LambdaGAN: Generative adversarial nets for recommendation task with lambda strategy,"" in Proc. Int. Joint Conf. Neural Netw. (IJCNN), Jul. 2019, pp. 1–8.
[8] N. Das, S. Borra, N. Dey, and S. Borah, „„Social networking in web based movie recommendation system,"" in Social Networks Science: Design, Implementation, Security, and Challenges. Cham, Switzerland: Springer, 2018, pp. 25–45
[9] R. Van Meteren and M. Van Someren, „„Using content-based filtering for recommendation,"" in Proc. Mach. Learn. Inf. Age, MLnet/ECML2000 Workshop, vol. 30, 2000, pp. 47–56.
[10] Asad KI, Ahmed T, Saiedur Rahman M (2012) Movie popularity classification based on inherent movie attributes using C4.5, PART and correlation coefficient. 2012 Int Conf informatics. Electron Vision, ICIEV 2012:747–752.

©IJRASET: All Rights are Reserved



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)