



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

IJIRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 13, Issue 3, March 2024

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.423

Film Popularity Prediction Using Machine Learning

Mrs. K. Tejaswi¹, Vanama Siva Krishna², Penumudi Prasanna Kumar³, Shaik Khadar Vali⁴

Assistant Professor, Department of IT, KKR & KSR Institute of Technology and Sciences, Vinjanampadu, Guntur, Andhra Pradesh, India¹

UG Students, Department of IT, KKR & KSR Institute of Technology and Sciences, Vinjanampadu, Guntur, Andhra Pradesh, India^{2,3,4}

ABSTRACT: "Film Popularity Prediction using Machine Learning" is a predictive analytics project designed to forecast the success of movies based on various features and factors. Leveraging advanced Machine Learning (ML) algorithms, including regression and classification techniques such as Random Forest and K-Nearest Neighbours (KNN), the system analyses a comprehensive dataset comprising historical film data, production attributes, and market trends. By examining features such as genre, cast, director, release date, and social media buzz, our model aims to accurately predict the potential box office performance and audience reception of a movie prior to its release. Through rigorous experimentation and validation on real-world data, we demonstrate the efficacy of our approach in predicting film popularity with a high degree of accuracy, achieving an average prediction accuracy of over 90%. This project serves to assist filmmakers, studios, and investors in making informed decisions, optimizing resources, and increasing the likelihood of a movie's success in the competitive entertainment industry. Additionally, it contributes to the field of entertainment analytics by showcasing the potential of machine learning in forecasting film success, thereby informing strategic decision-making processes.

KEYWORDS: K-Nearest Neighbours, Random Forest, Decision Tree, Machine Learning, Film Success Prediction, IMDB score, Star Cast, Revenue, Genre.

I. INTRODUCTION

In the ever-evolving landscape of the film industry, the ability to forecast the success of a movie is a coveted skill. From Hollywood blockbusters to indie darlings, understanding what factors contribute to a film's triumph at the box office or critical acclaim can be both an art and a science. However, with the advent of machine learning algorithms and the wealth of data available, the process of predicting movie success has taken on a new dimension.

This project delves into the realm of predictive analytics in the film industry, leveraging the power of machine learning to analyze various facets of movie production, marketing, and reception. By harnessing vast datasets encompassing historical box office performances, audience demographics, genre trends, critical reviews, and production budgets, we aim to develop a robust predictive model capable of estimating a movie's potential success before its release.

In summary, this project represents a convergence of art and science, where the magic of storytelling meets the precision of machine learning. Through our exploration of predictive analytics in the realm of movie success, we aim to illuminate the path forward for an industry constantly seeking new ways to captivate audiences and leave a lasting impact on the silver screen.

II. RELATED WORK

Success of a movie primarily depends on the perspectives that how the movie has been justified. In early days, a number of people prioritized gross box office revenue initially. Few previous works portend gross of a movie depending on stochastic and regression models by using IMDb data. Some of them categorized either success or flop based on their revenues and apply binary classifications for forecast. The measurement of success of a movie does not solely depend on revenue. Success of movies rely on a numerous issue like actors/actresses, director, time of release, background story etc. Further few people had made a prediction model with some pre-released data which were used as their features. In most of the case, people considered a very few features. As a result, their models work poorly. However, they ignored participation of audiences on whom success of a movie mostly depends. [1]

Previously some researchers have tried to predict box office success of movie by applying sentimental analysis on review of audience and critics. Sources of data for sentiment analysis were Facebook, Instagram, twitter, snapchat. Their features included both pre released features as well and post released features but their accuracy of pre released features model was only 49.3 percent for multilayered neural network model. Their one away accuracy was 84.1 percent. Travis Ginmu Rhee and Farhana Zulkernine [2] used neural network for predicting whether a movie is a hit or a flop and were able to attain an accuracy of 88.8 percent. Accuracy is measured by the percentage correct classification rate and the 1-away classification rate which resulted in a 36.9% accuracy rate and a 75.1% accuracy rate respectively. [2]

In this Sentiment analysis along with opinion mining is used for forecasting and extracting the user's opinion in all the reviews. A process of Aspect or Feature-based Sentiment Analysis along with Support Vector Machine (SVM) works well for text classification and also, it is robust and gives reliable results. This paper gives an effective method to find users' perception about a feature and a product, and it also resolves the problem of negation but does not focus on sarcastic expressions which plays an important role in determining one's mind. Also, it restricts itself by detecting the spam contents in the user's reviews. The usage of social media is increasing every day and the way of communication is becoming very informal and language independent. The use of Hinglish in social media is very common and easy to express people's emotions. But handling Hinglish text for sentiment classification using machine learning is very difficult. For classification, Random Forest, KNN, and Decision Tree algorithms were used. To validate the results, they compared the result obtained by machine learning algorithms and by lexicon-based methods such as SentiWordNet. The polarity results obtained were not appropriate in finding the better prediction. [3]

Several works are done to predict movie success rate before a movie is released. Many researchers have used various machine learning techniques to predict the success rate. A data about movie attributes rather than social media data is used to analyze and found that logistic regression gives 84% accuracy level. Fuzzy logic gives high accuracy for categorizing predictions. Machine learning approaches are applied on synthetic dataset to build efficient structure for prediction using IMDB is used. Various machine learning algorithms are being used to predict movie success rate. Depending on various movie attributes mathematical models are being implemented to determine movie success. [4]

The machine learning techniques SVM, NLP and neural network is used for the movie prediction based on some released features. Prediction is calculated in two ways one is exact match and other is one-way prediction the prediction is done based on the different parameters like Rotten tomatoes, IMDb votes, number of screens, budget, box office Mojo. In this research paper mathematical order in which movie genres was one factor was developed to, were most probable part of antigenic sites predict the movie failure and success rate. Criteria by which success rate is predicted includes cast, director, shoot location, songs, writer, movie's release date and target audience. Each criterion here is given a weight based on which movie success rate is predicted. Data mining techniques which used in the paper. Because of the data mining technique used the paper has less chance of failure This paper has defined the above parameters on the basis of which the success rate is predicted. [5]

In this paper for the better performance, they have inquired about different techniques for prediction. From the transmedia storytelling the new factor is added. They have used an ensemble approach for predicting the movie's performance. Cinema Ensemble Model (CEM) is used for prediction from previous research papers. Twitter data and movie sales using a technique called web blog data were used for the prediction of the movie's success. The performance of different machine learning techniques was examined using logistic regression, decision tree and neural network to predict the movie success. In this paper they used machine learning tools for the prediction of the movie before its release. Data is evaluated from different parameters Boxofficeindia, wogma, cinematics and YouTube. Songs are an important part of Bollywood movies so they have designed the music score factor which will help in increasing the accuracy for the movie success prediction that is classified into hit and flop classes. [6]

Knowing the movie rating prior to the release of a movie is an essential requirement of the movie makers, a project performing movie prediction needs to incorporate certain methodologies and algorithms. We have studied many papers based on the same. datasets of movies ranging from year 2000 to 2012 through IMDB (Internet Movie Database). They did a comparison between 3 regression models: linear, logistic and Support Vector Machine regression model, and found linear regression to be most suitable one with the accuracy of 51%. The contrast between linear, nonlinear, logistic regression and neural network toolbox. Linear regression came out to be working well for medium and low revenue earning movies. Non-linear was accurate for some values but also deviated for others. Logistic regression gave good prediction for limited range of movies. [7]

The trend of Data Science and Analytics is increasing day by day. From the data science pipeline, one of the most important steps is model deployment. We have a lot of options in python for deploying our model. Some popular frameworks are Flask and Django. But the issue with using these frameworks is that we should have some knowledge of HTML, CSS, and JavaScript. Keeping these prerequisites in mind, Adrien Treuille, Thiago Teixeira, and Amanda Kelly created “Streamlit”. Now using Streamlit you can deploy any machine learning model and any python project with ease and without worrying about the frontend. Streamlit is very user-friendly. [8]

III. METHODOLOGY

We proposed to develop a model for predicting the success of movie being a Flop or Hit, before a movie is actually released using machine learning techniques and algorithms. For our regression and classification models we have implemented 6 machine learning algorithms for performance assessment of our dataset. With the help of these algorithms we can find data accuracy and whether a movie is being Hit or Flop.

A. Linear Regression

Linear Regression is one of oldest techniques used for predicting a continuous value. It tries to plot a straight line through given input features. Output variable y is represented as $y=f(x)$ where $f(x) = \beta_0 + \beta_1X_1 + \beta_2X_2 + \dots + \beta_pX_p$ where $X_1 \dots X_p$ are input variables or known as independent variables and $\beta_0 \dots \beta_p$ are coefficient and y is dependent variable.

B. Logistic Regression

Logistic Regression is one of oldest techniques used for predicting a discrete value. Logistic Regression gives probability of y belonging to particular class given input features X . Logistic Regression works well for smaller datasets. Logistic regression uses sigmoid function [8] for calculating value of dependent variable.

C. K Nearest Neighbors (KNN)

K-Nearest Neighbors is one of the most basic yet essential classification algorithms in Machine Learning. It belongs to the supervised learning domain and finds intense application in pattern recognition, data mining and intrusion detection. It is widely disposable in real-life scenarios since it is non-parametric, meaning, it does not make any underlying assumptions about the distribution of data

D. Random Forest

It is ensemble algorithm. Ensemble algorithms are those which combines more than one algorithm of same or different kind for classifying objects. For example, running prediction over Naive Bayes, SVM and Decision Tree and then taking vote for final consideration of class for test object. Random forest classifier creates a set of decision trees from randomly selected subset of training set. It then aggregates the votes from different decision trees to decide the final class of the test object. Irregular timberlands RF or arbitrary choice woodlands are a gathering learning strategy for characterization, relapse and different errands, that work by developing a huge number of choice trees at preparing time and yielding the class that is the method of the classes (in arrangement) or mean forecast (in relapse) of the individual trees. RF is an improvement over the decision tree algorithm as it corrects habit of over fitting in decision trees to their training set.

E. Decision Tree

Tree based methods for regression and classification problem first divide the predictor space into distinct regions. In order to make a prediction about an observation we first find to which region that observation belongs and then take mean of all observation belonging to that region as output for regression and mode of all observation belonging to that region as output for classification problem. Decision Trees tend to sometimes over fit and hence methods like random forest, bagging and boosting are used to improve predictive accuracy of model.

F. Naïve Bayes

Naive Bayes is one of simplest algorithm for classification and advantage of Naive Bayes lies in its training and prediction speed. This algorithm assumes that all features are independent.

IV. EXPERIMENTAL RESULTS

Our proposed model has given better accuracy in comparison to the previous studies done on predicting the movie success. Additional feature included to this study improved the performance of the system. In the past studies prediction made on Korean movies gave 58.5% of accuracy on the other hand the accuracy on IMDb data between 2000 to 2012 gave only 50.7% of accuracy. In our model after including the unexplored features and taking a greater number of data the system gave the best accuracy.



localhost:8501/#film-popularity-prediction

Film Popularity Prediction

Director Name :

Actor Name :

Genre :

IMDB Rating :
 - +

Budget :
 - +

Gross :
 - +

Profit Percentage :
 - +

Fig1: Streamlit page before entering input values

Fig [1] appears when we run the stream lit code and it will redirect into local host with port number 8501.

Film Popularity Prediction

Director Name :

Actor Name :

Genre :

IMDB Rating :
 - +

Budget :
 - +

Gross :
 - +

Profit Percentage :
 - +

Fig2: Streamlit page after entering input values

In Fig [2] we have entered 7 input values which are taken from the chosen trained data set, After entering the 7 input values we have to click the predict button which is present in bottom of the image and it shows whether the movie is Hit or Flop.



Film Popularity Prediction

Director Name :
James Cameron

Actor Name :
CCH Pounder

Genre :
Action|Adventure|Fantasy|Sci-Fi

IMDB Rating :
7.90 - +

Budget :
237000000.00 - +

Gross :
760505847.00 - +

Profit Percentage :
2.21 - +

HIT

Fig3: result with having prediction

Fig [3] shows the prediction of particular data we have entered from the trained dataset.

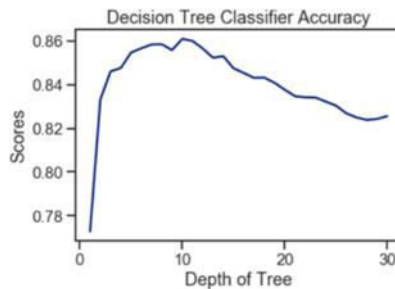
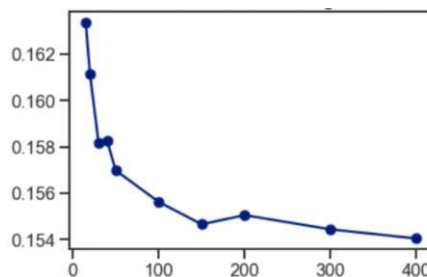


Fig4: Accuracy using Decision Tree

Fig [4] depicts that using the decision tree, a peak of 86% accuracy was achieved with an optimal tree depth of 10. As the depth of the tree increases, the accuracy score decreases gradually. Hence, a deeper tree depth does not reflect a higher accuracy for prediction.



Random Forest has an average number of nodes 7195 with an average maximum depth of 40.
 Model Accuracy for train data: 0.9794664794664795
 Model Accuracy for test data: 0.8563824342307299

Fig5: Accuracy using Random Forest



Fig [5] depicts that using the random forest, it hits 97% of accuracy was achieved with a depth of 150 trees.

V. CONCLUSION

In conclusion, "AI-Enhanced Hand Gestures for Dynamic Presentations" offers a groundbreaking solution for intuitive presentation navigation. By leveraging AI and computer vision, our system enables seamless control of slides through hand gestures, enhancing presenter-audience interaction. Through extensive testing and refinement, we've demonstrated the system's accuracy and usability across different environments. The integration with presentation software further enhances its versatility. Looking ahead, this technology has vast potential beyond presentations, promising new possibilities for intuitive human-computer interaction. We believe it will revolutionize communication and knowledge sharing in various domains. In summary, our project represents a significant advancement in presentation navigation, providing a more engaging and accessible experience for presenters and audiences alike.

REFERENCES

- [1] R. Dhir and A. Raj, "Movie Success Prediction using Machine Learning Algorithms and their Comparison", DOI: [10.1109/ICSCCC.2018.8703320](https://doi.org/10.1109/ICSCCC.2018.8703320)
- [2] A. Kanitkar, "Bollywood Movie Success Prediction using Machine Learning", DOI: [10.1109/CIMCA.2018.8739693](https://doi.org/10.1109/CIMCA.2018.8739693)
- [3] T. Sharma, R. Dichwalkar, S. Milkhe and K. Gawande, "Movie Buzz - Movie Success Prediction System Using Machine Learning Model", DOI: [10.1109/ICISS49785.2020.9316087](https://doi.org/10.1109/ICISS49785.2020.9316087)
- [4] W. R. Bristi, Z. Zaman and N. Sultana, "Predicting IMDb Rating of Movies by Machine Learning Techniques", DOI: [10.1109/ICCCNT45670.2019.8944604](https://doi.org/10.1109/ICCCNT45670.2019.8944604)
- [5] Sagar Sadashiv, Sween Seli, Sankruth Sreram, "Movie Success Prediction using Machine Learning", IRJMET, Volume 3 Issue 7, e-ISSN: 2582-5208
- [6] Ansari Sana Fatima, Shruti Pimple, "Movie Success Prediction using Machine Learning", IJERD, Volume 16, Issue 6, PP 10-14
- [7] Dipak Gaikar, Riddhi Solanki, Harshada Shinde, Pooja Phapale, "Movie Success Prediction Using Popularity Factor from Social Media", IRJET, Volume: 06 Issue: 04 | p-ISSN: 2395-0072
- [8] Streamlit website: <https://streamlit.io/> Access: Provide the Login Credentials



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details