Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 6, No. 5, June : 2023

SURVEY OF MACHINE LEARNING TECHNIQUES FOR MOVIE SUCCESS PREDICTION

Mr. Saurabh Mondal, MScIT student, Dept of Information Technology, University of Mumbai.
Ms. Vishakha Gidwani, MScIT student, Dept of Information Technology, University of Mumbai.
Mr. Nikhil K. Pawanikar, Asst. Professor, Dept of Information Technology, University of Mumbai.
Dr. Mrs. R. Srivaramangai, Head, Dept of Information Technology, University of Mumbai.

Abstract
The movie business is expanding quickly, whether it's a Bollywood or a Hollywood production. Each year, the global film business creates a sizable number of films. Due to digitization, films have undergone significant changes. For a movie to be successful, the box office must do well. However, very few films have piqued the interest of both financial specialists and economists alike. With the expansion of the film business, understanding the likely box office earnings of a film is becoming more and more crucial for stakeholders like producers and directors. The success of a movie must be predicted because it requires a lot of time and money. It is quite helpful for moviegoers who need to be aware of the calibre and success rate of forthcoming films in advance. Therefore, it is crucial to foresee events more accurately. One of the main causes of the entertainment industry's unpredictable success and failure is the movie sector. Machine Learning, in particular, may play a significant role in the advancement of computing to help forecast a movie's likelihood of success and its general direction.

Keywords: Machine Learning, Artificial Intelligence, Success Prediction, Ensemble Methods, Random Forest, Decision Tree, IMDB - Internet Movie Database

I. Introduction

Prior to the COVID-19 epidemic, Indian Cinema was the world's largest film industry in terms of the number of films produced each year. The film industry worldwide has matured, and it uses significant investment leveraging continuously evolving technology to produce the best content. Production companies make significant investments in the making of a certain movie because they believe it has a better chance of being a box office success [1]. A blockbuster movie is one that makes a lot of money and can cover the costs of unsuccessful projects from the same production company. The production company can still profit from taking a calculated risk to make money, even if the expensive movie doesn't do well. The blockbuster theory's main tenet is that a higher-budget movie has a better probability of earning more money than a lower-budget movie. Like any other industry, the film industry deserves to increase its profitability by harnessing the power of data. Consider the statistics in Figure 1, showing the top five and worst five movies at the Bollywood box office with respect to their budget and revenue. Figure 1 (left) indicates that even after significant investment, the movies turned out to be loss-making. There are various techniques and methods that can help predict the movie’s success based on parameters we already have data on. Simply put, by employing predictive analysis using machine learning techniques, it is possible to predict a movie's success. Due to the unpredictable yet lucrative nature of the film industry, numerous researchers are currently exploring the subject of predicting movie success. By developing precise models, it is possible to lessen the industry's uncertainty and avoid any potential financial losses. The basic idea behind these techniques is to train a model with historic data and predict the success or failure rate based on the information and patterns stored in the knowledge base of the model. The larger the knowledge base, the greater the accuracy drawn from the algorithm. A few characteristics are essential to the film's success. There are several entities and databases that collect statistics on movies that can be used in various algorithms to our advantage. In order to predict success, the availability of data on various attributes of a movie is

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 6, No. 5, June : 2023

necessary, and therefore the existence of platforms such as IMDB and MovieDB is beyond valuable and can be considered for our analysis.

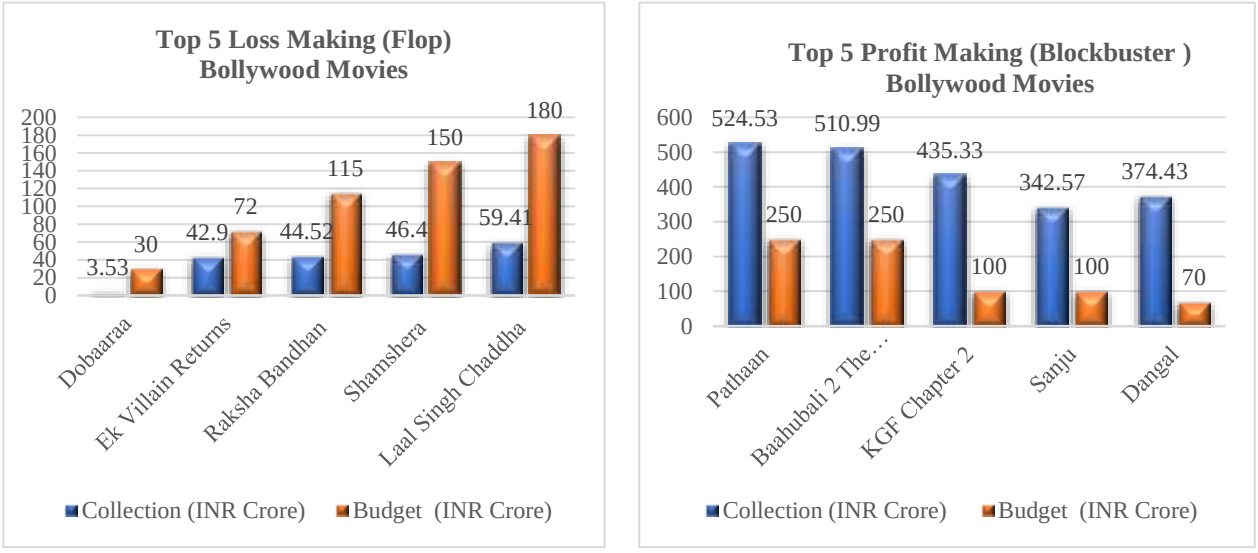

Figure 1 - Top 5 Loss Making (Flop) Bollywood Movies (left) Courtesy - studydhaba.com, Top 5 Profit Making (Blockbuster) Bollywood Movies (right) Courtesy - sacnilk.com

The remaining part of this paper is organized as follows: Section II defines the machine learning methods that will be elaborated in the rest of the paper; Section III presents the survey of the literature by researchers in the area of movie success prediction; Section IV presents our observations of the literature survey; and Section V presents the concluding remarks.

II. Overview of machine learning methods used for movie success prediction

2.1 K-Nearest Neighbor (KNN): One of the supervised machine learning algorithms used for both classification and regression is the K-nearest neighbour (KNN) algorithm. It is primarily employed in the industry for the classification and prediction of problems. It employs 'feature similarity' to forecast the values of new data points, which means that the new data point will be given a value depending on how closely it resembles the points in the training set. [2].

2.2 Linear Regression: According to the definition given in [3], linear regression is a statistical model that examines the connection between a dependent variable and a predetermined set of independent variables. It follows that as the value of one or more independent variables changes (increases or decreases), the value of the dependent variable similarly changes (increases or decreases). This is why linear regression is so named. Since the relationship is linear, it demonstrates how the dependent variable's value changes in response to the independent variable's value. The link between the variables is depicted by the model as a slanted straight line.

2.3 Naive Bayes: The classification method known as naive Bayes algorithms [4] is based on the Bayes theorem and makes the strong assumption that all predictors are independent of one another. To put it simply, the assumption is that each feature in a class is independent of every other feature in the class. Finding the posterior probabilities, or the likelihood of a label given some observable features (L | features), is the main goal of Bayesian classification.

$$P(L \mid \text{features}) = \frac{P(L)\, P(\text{features} \mid L)}{P(\text{features})}$$

In this case, the posterior probability of class is (L | features). (L) is the prior probability of the class, (features | L) is the likelihood, which is the predictor's probability given the class, and (features) is the predictor's prior probability.

2.4 Support vector machines (SVMs): The powerful and adaptable supervised machine learning approach known as Support vector machines [5] is utilised for both classification and regression. They are frequently applied to categorization issues. This model is a multidimensional hyperplane representation of several classes. An iterative process will be used to generate the hyperplane in order to reduce error. To find a maximum marginal hyperplane (MMH), it is necessary to classify the datasets.

2.5 Logistic Regression: A supervised learning technique called logistic regression [6] is used to forecast the likelihood of a target variable. Due to the dichotomous nature of the dependent variable, there are only two viable classes. As a binary variable, the dependent variable can only have data values of 1 (success/yes) or 0 (failure/no). P(Y = 1) is predicted by this model mathematically as a function of X. One of the most straightforward machine learning algorithms, it may be applied to a variety of categorization issues, including spam identification, success prediction, cancer diagnosis, etc.

2.6 Random Forest: A supervised learning method used for both classification and regression is called random forest [7]. Random forest is primarily utilized for categorization issues, though. Given that a forest is made up of trees, a forest with more trees will be more robust. Similar to that, this method builds decision trees based on data samples, obtains predictions from each one, and then votes to determine which is the best answer.

III. Related work

Following section presents the survey of work done by researchers on the topic of movie success prediction:

3.1 Random Forest and Decision Tree based Methods: Masih and Ihsan [8] in their work proposed state that when linked with director, awards have strong predictive value, and their accuracy has increased when combined with other characteristics. With the help of movie metadata, Razeen [11] seeks to estimate movie income using a variety of regression approaches. Except in one instance where Twitter data was used, the R2 values were greater when compared to other publications that were reviewed (R-Squared values measure the model's goodness of fit and show a perfect fit). Although it has been widely established that employing trend-based, real-time social media data, such as tweets from Twitter, may produce outstanding results. Although some would argue that the accuracy is insufficient for practical usage, the methods described in this study demonstrate that movie revenue can be forecast successfully. According to Jatale [12], it is definitely possible to forecast a movie's success with a high degree of accuracy. Therefore, a production firm can assess its production before the start of production by using our prediction engine. Additionally, the production company can determine whether a film is worth investing in and make decisions accordingly. When paired with real, accurate data and fuzzy logic, it can be viewed as an effective means of categorizing predictions and adds more accuracy and dynamicality to the system. Based on previous movie events and their success rates, the predictive model will forecast the movie's performance. The crucial component of the model developed by L. K. et al. [13] is the confusion matrix, which uses the Random Forest Algorithm to predict the outcome of a movie even before its release. The system is configured as a go-to model for movie investors to have trust in the amount they invest and minimize risk thanks to the machine learning approach employed in this project. Because it is so unpredictable, predicting the box office performance of new films is a crucial undertaking for the entertainment industry. Using information from IMDb, predictions are produced. It is laborious to mine IMDb data since each attribute connected

with a movie will be in a separate dimension and contain a significant amount of noisy and missing data. The random forest approach has been employed in this work to address the problems with tweets. The suggested model seeks to predict the profitability of films. 76% of forecasts are accurate. In their study, Gupta et al. [14] employed comparative analysis to obtain the most precise conclusions regarding the likelihood that a film will be a box office smash, average hit, or flop, utilising various data mining and machine learning algorithms. A movie's success depends on more than just the aspects that are associated with movies. The size of the audience plays a crucial role in how successful a movie is. Several techniques were used, including random forest, decision tree regressor, and support vector regression (SVR). The accuracy of the decision tree regressor was 100%, while the SVR success rate was around 98.81%, the prediction rate for the random forest was 100%, and the success rate for the Lasso regressor was approximately 98.45%. The effective approach was created by Sivakumar et al. [17], but it is unable to predict a new movie's anticipated rating before its debut. Additionally, utilising information gathered from internet sources such as data or features, it is impossible to forecast a new film's success before its debut. Only up to 5 attributes were chosen in order to simplify handling the massive amount of data and ensure that only the pertinent information is used and understood by the user. Consequently, the model has been taught to forecast movie success. For all of the testing data, the model's accuracy was greater than 70%. Thus, it can be inferred that the model is effective in predicting a movie's success before it is released. The influence of predictive data analysis on decision-making in the film business is highlighted by Naik [30]. Rapid Miner is used to create predictive models that make use of this data and assess the portability of the video using machine learning techniques including Decision Tree, Logistic Regression, Nave Bayes, Random Forest, and SVM. The obtained results are assessed and represented visually. Tableau and Rapid Miner both use graphic representations of these findings. From the early stages of pre-production to the movie's release, the model offers fantastic insights for investors to discover the elements that are crucial to the success of the production, including all the important facets of marketing and promotion. The Movie Review Rating Prediction (MR2P), created by Adetunji et al. [32], is an effective movie recommendation algorithm that will improve prediction accuracy. It was developed after a thorough analysis of the current movie success method. By knowing the type of movie to invest in, movie stakeholders (producers, directors, crew, and cast who are already working in the industry or who are aspiring) will benefit from larger profits. The production studios were assisted by Reddy et al. [34] in forecasting box office earnings that might be utilised to make plans for the development and distribution stages of movies. When compared to other methods, this procedure produced results with lesser Root Mean Square Error for Random Forest, and Random Forest outperformed the XGBoost and Linear Regression algorithms. How IMDB movie data was applied for movie rating prediction was described by Abidin et al. [39]. The IMDB data set was structured and made ready for data mining methods. On the WEKA application environment, these algorithms were run, and the results for movie ratings and confusion matrices were acquired. The seven machine learning algorithms utilised have average performance ratings of 73.5% to 92.7% on the data set. The performance of the Random Forest algorithm was the best, at 92.7%. This is the highest score obtained among similar studies. Instead of relying on reviews and the opinions of others to determine whether or not a film would be successful, Dhir and Raj [40] suggested a method to forecast a movie's box office performance. The suggested study offers a very effective method for predicting IMDb scores on the IMDb Movie Dataset. We shall attempt to identify the key elements affecting IMDb Movie Data's score.

3.2 K-nearest Neighbor, Support Vector Machine and Decision Tree based Methods: Latif and Afzal [9] used algorithms like simple logistic and logistic regression, neural network - multilayer perceptron, decision tree j48, proposed research that aims to predict the popularity of a movie using a Machine learning approach. Strong classification techniques for machine learning are available. The proposed study tries to advance earlier studies. When performing classification, it has been discovered that simple logistic and logistic regression yield the best results. A mathematical model was put up by

Anantharaman V et al. [16] to forecast the profitability of films based on a variety of factors, including the director, genre, actor, budget, etc. In order to classify films based on patterns found in earlier records, data mining techniques like the Decision Tree Algorithm are applied. Results can be used to forecast whether upcoming films will succeed or fail. Model accuracy is 80%, which is higher than that of previous approaches. According to Abarja and Wibowo [18], CNN can be utilised as a model for tiny textual datasets. Features play a significant role in the construction of prediction models. Additionally, the performance of rating prediction was positively impacted by the use of historical information as a predictor. Prior to a film's debut, historical values derived from earlier releases allowed for a more precise rating prediction. The model could be tuned with various optimisation techniques and architecture changes to improve performance. Better features might potentially be produced by using an improved and updated dataset. Chintu Kumar [20], using performance measuring metrics like precision, recall, F1 score, and accuracy, among others, compares the suggested method and the existing method. Python is a language that is used to simulate the suggested and current methods since it is simple to use and consumes less computing time than other languages. We can forecast the success rate of films based on these characteristics with ease. The elements that can increase the accuracy of movie success predictions must be added. Use a hybrid approach to machine learning that combines the key elements of random forest with another approach, such as logistic regression, to increase the success of movie prediction. Shahrivar and Jernbäcker [38] looked into the viability of categorizing box office receipts and movie ratings using information that was accessible prior to release. This was accomplished by creating a categorization model using internet-sourced metadata, such as information on the budget and the actors involved. Using the method with the best success rate, this study was able to accurately estimate the rating of a movie 82% of the time. Antony and Francis [19] used the k-Nearest Neighbor algorithm and proposed a method for predicting movie box office performance. The method used is the most effective for predicting the success of a film. The success of a film is not only determined by these aspects of cinema. The number of people who watch a movie is crucial to its success. The findings suggest that box office success may be predicted with reasonable accuracy by evaluating movie emotions. Sandhiya G et al. [23] used and built three separate systems that recommend movies to the users. This proves that the system is valid for recommending movies. By comparing the results, it can be concluded that each algorithm is efficient in each way to build a recommendation system. A movie recommendation system was created by Abhishek Singh et al. [28] by combining the KNN algorithm and the collaborative filtering technique. In general, hybrid-based approaches, content-based approaches, and collaborative filtering approaches are used to create recommendation systems. The collaborative filtering method used by this system includes a variety of techniques, including matrix factorization and user-based recommendations. The interrelationships between numerous traditional elements and social signals employed when developing the predictive model for estimating the overall box office receipts and critical acclaim for a particular movie were described by Darekar et al. [29]. The findings demonstrate that a prediction model that incorporates both traditional and social components can have a greater rate of accuracy. Movie stakeholders can utilise the model established to anticipate a movie's performance before it is released in order to make better decisions. Fatima and Pimple's [31] goal is to create a model that forecasts whether a movie will be a hit, flop, or super hit based on a variety of factors, with budget being a key factor. The success rate of the film is anticipated based on these factors. The degree of vaccination success is determined by a number of factors. The film producer can take into account these aspects of our project to determine their financial roadmap and gauge their comfort level with taking risks. KNN, Linear Regression, Naive Bayes, Support Vector Machine (SVM) Linear Model, and Logistic Regression were the algorithms employed. Midhun Josey [21] proposes a study that proves that features such as Genre, followers, view count, etc. positively affect the success of a movie. It also proves that the sentimental analysis of tweets and comments is directly related to public opinion and plays a huge role in the success of a movie. The algorithms used, such as Neural networks and Support Vector Machine, allow efficient prediction of the movie box office. Thus, the movie investors achieve a great deal with

this kind of research, and it helps them efficiently invest in movies that can be a success and make a profit for their investment. A system was put forth by Chaki et al. [26] to assist investors connected to this industry in minimising investment risks. The algorithm uses historical data from many sources, including IMDb, Rotten Tomatoes, Box Office Mojo, and Metacritic, to estimate a movie's success rate based on its profitability. Based on some pre-released and post-released features, the system forecasts a movie's box office earnings using Support Vector Machine (SVM), Neural Networks, and Natural Language Processing. This study demonstrates that when one-way prediction is taken into account, SVM has an accuracy of 83.44% for pre-released features and an accuracy of 88.87% for all features compared to neural networks' 84.1% and 89.27%, respectively. Before a movie is actually published, Sadashiv et al. [36] created a model to forecast whether it will be a hit or a flop. This model uses machine learning techniques and algorithms.

3.3 Linear Regression, Logistic Regression, Naïve Bayes and Neural Network based Methods: Asur and Bernardo A. [24] looked at the viability of using tweets to predict the box office performance of upcoming films. This work's key contribution is a demonstration of how transfer learning can be used to accomplish this. To the best of our knowledge, this is the first piece of work to investigate the golden mean for predicting box office success. Data from social networks and IMDb were used to create a predictive model by Kharb et al. [37] that reflected the box office performance of films. Our models allowed us to recognise the following patterns: (A) the popularity of the lead actor or actress is essential to a movie's success; (B) the pairing of a sequel with a genre that is now popular; and (C) a new picture in a genre that isn't as popular with audiences may serve as an example of a failure. Dissanayake and Vidanagama's [33] use of machine learning models allowed them to anticipate a film's success. Classification and regression models have not been applied together in the associated research efforts. However, both classification and regression models were applied in this study. The option to treat movie success prediction as a classification problem or a regression problem in subsequent research is thus provided. Topf and Ossella-Durbal's [35] investigation of the factors that contribute to a domestic hit at the box office The findings demonstrate that particular elements of a movie draw larger crowds. The most profitable films are typically animated ones with PG-13 ratings that appeal to the broadest audience. This could play a role in "Avatar's" commercial success. It is animated and has a PG-13 rating. However, the additional effect on income of a film of a specific genre and age classification is not examined in this study. We look at the genre and age rating categories separately. Apala et al. [10] employed Naive Bayesian and K-means clustering predictive models were used to estimate how well films would do at the box office based on data from social media and IMDb. The following trends are identified by the models: (1) the leading actress's popularity is essential to a film's success; (2) the pairing of a successful genre with a follow-up film is another pattern for success; and (3) a new film in an unpopular genre starring a low-profile actor could be a pattern for a flop. It has to do with how each attribute's weights are determined. The forecast also includes yet-to-be-released films. An excellent prediction performance from the model is demonstrated by the first results of tracking 13 of the films.

3.4 Other methods: By merely using the metadata of a film that has already been published, Mahmud et al. [25] attempted to investigate the issue of predicting the money generated by a film. Since sales data is not accessible in the first few weeks after the film's release, no sales information is utilised to estimate income. The investigation also shows that, despite the small size of the dataset, the MLP model outperforms other machine learning algorithms. Using the MLP model and our custom features, it achieved a respectable one class accuracy of 85.31%. In their research, Peleja et al. [27] suggested a recommendation architecture in which ratings and unrated comments from media consumers are combined to increase the accuracy of recommendations. For a cable operator looking to develop a system that uses user complaints and compliments as indicators of consumer happiness, this is the perfect application. The evaluation using actual user data serves as an example of how crucial it is to

employ text analysis methods to change users' explicit ratings. Techniques that are used to infer a customer's movie rating were recommended by R. Vignesh et al. [15]. The project chooses the MovieLens dataset, which contains 10 million ratings, and divides it into a coaching set and an examination set. For formula analysis, we employ the RMSE approach. In their research, Patel et al. [22] discovered a link between numerical type attributes that are employed as a scoring system and how this link might be used for prediction. The link between the reviewers' and the audience's scores is very favourable. Additionally, it can be said that critics' ratings are the best indicators of audience ratings. As a result, movie predictions can be made based on critics' ratings. In their study, Jain et al. [41] concentrated on coming up with and creating a mathematical model to forecast movie performance. The main concept behind the project is a mathematical method for predicting popular films.

IV. Observations

This section presents our observations of the survey performed.

4.1 Types of methods used: Figure 2 shows the different methods used by researchers to address the challenge of predicting movie success. Random Forest and Decision Tree are the two most popular machine learning algorithms. The flexible algorithm Random Forest can tackle classification and regression issues effectively.

Figure 2: Techniques used for predicting movie success

Similarly, Decision Tree is a powerful algorithm that emulates human decision-making capabilities and is easy to comprehend due to its tree-like structure. While Decision Trees are composed of a single tree, Random Forest combine several Decision Trees to produce more accurate results.

4.2 Accuracy of Methods: Figure 3 shows the accuracy of different methods used for predicting movie success. We observe that the highest accuracy is given by the 'Decision Tree' and 'Random Forest' algorithms. Both Decision Tree and Random Forest are widely used machine learning algorithms that excel in classification and regression tasks. Although each algorithm has its own set of strengths and weaknesses, both are known for their ability to make precise predictions. Through the aggregation of results from multiple Decision Trees, Random Forest mitigates the variance and overfitting problems

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 6, No. 5, June : 2023

associated with Decision Trees, resulting in more resilient and accurate predictions. As a result, both algorithms have their own unique benefits and can be applied in different scenarios to achieve precise results.

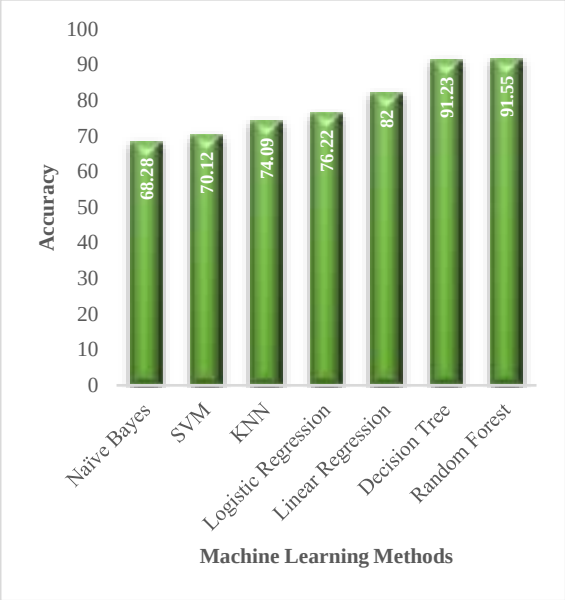

Figure 3: Accuracy of different methods

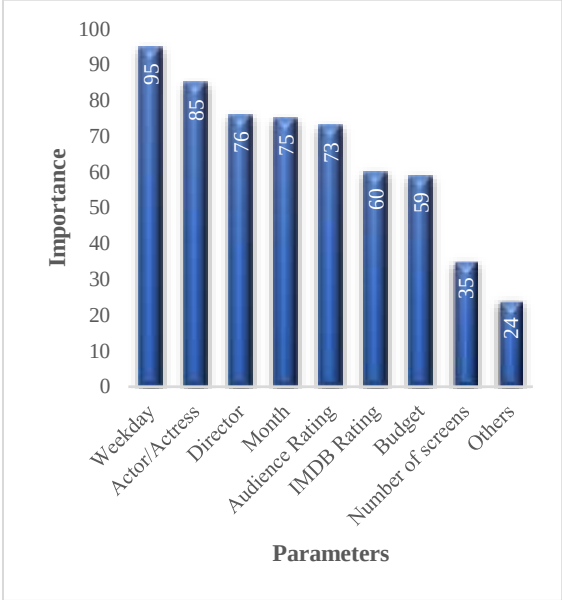

Figure 4: Different parameters used for predicting movie success

4.3 Different Parameters Used: Figure 4 shows the different parameters used in different algorithms for predicting movie success. There are numerous parameters that can impact the overall performance and success of a movie. One of these parameters is the day of release, with films that are released on weekends having a higher likelihood of performing better than those released on weekdays due to more people having free time on weekends. The casting of a movie also plays a critical role in its success, as popular and talented actors have the ability to attract more viewers and generate excitement. Similarly, the director's skill and experience can provide a distinct vision for the movie, setting it apart from other films. Furthermore, movie success prediction models can also consider other factors like the audience rating, budget, number of screens, and others, which can assist movie studios in making informed decisions about the production and marketing of their films, ultimately improving their chances of success at the box office.

4.4 Datasets: Figure 5 shows the different datasets used in different algorithms for predicting movie success. When it comes to movie prediction using machine learning, the IMDB dataset is the most popular and widely used among researchers and practitioners. This could be due to the large number of movies in the dataset the rich and diverse metadata about each movie. As a result, the IMDB dataset provides a rich source of information for developing and training machine learning models. On the other hand, the Metacritic dataset is also moderately used. Although it may not have the same scale and scope as the IMDB dataset, Metacritic still provides useful information about movies, such as user ratings, reviews, and genre, which can be leveraged. However, the other datasets, which are used least frequently, may have limited applications. This could be due to several reasons, such as the smaller size of the dataset, limited metadata, or the dataset being more specific to a certain type of movie or genre. Therefore, based on the analysis, it is concluded that IMDB and Metacritic are the most suitable datasets, while other datasets have limited applications in this regard.

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 6, No. 5, June : 2023

4.5 Publication Statistics: The analysis of research papers related to movie success prediction indicates a significant interest among researchers and practitioners in recent years. Specifically, the years 2019 and 2020 witnessed a surge in the number of papers published on this topic, indicating the growing importance of predictive analysis and machine learning techniques in the film industry. The trend demonstrates the enthusiasm and dedication of researchers towards improving the accuracy of predicting movie success, which can ultimately aid in reducing financial losses and making the film industry more profitable. The focus on this area of research is promising, as it could lead to more informed decision-making in the industry and potentially increase the number of successful films produced. Overall, the increasing attention towards this field is a positive development and highlights the potential for advancements in predicting the success of movies in the future. Figure 6 shows the frequency of different categories of papers published. Figure 7 shows the number of papers published year-wise while Figure 8 shows the year when individual paper was published.

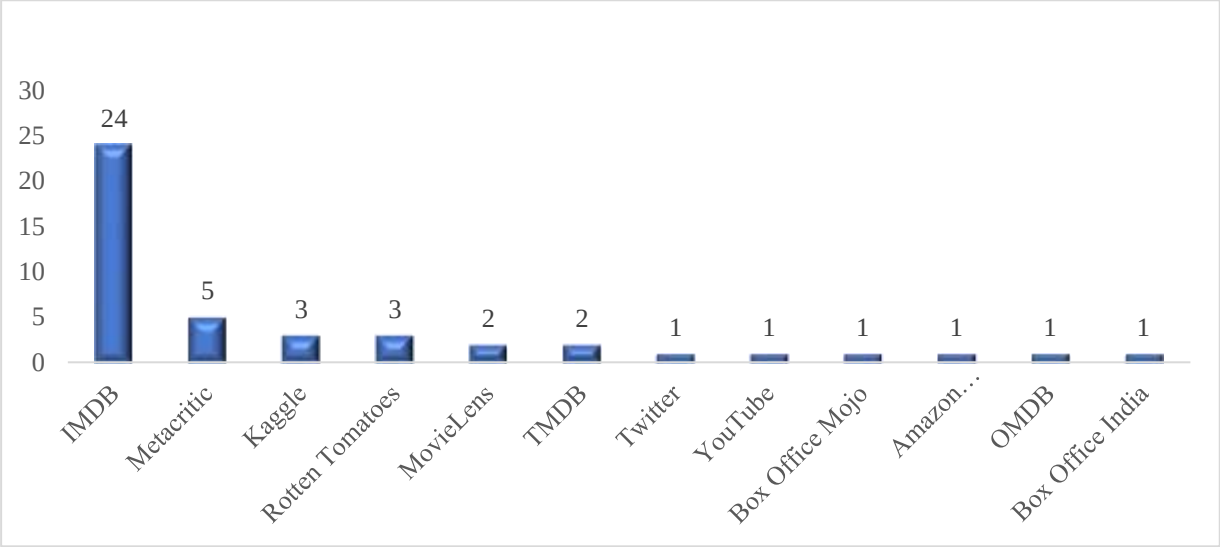

Figure 5: Datasets surveyed

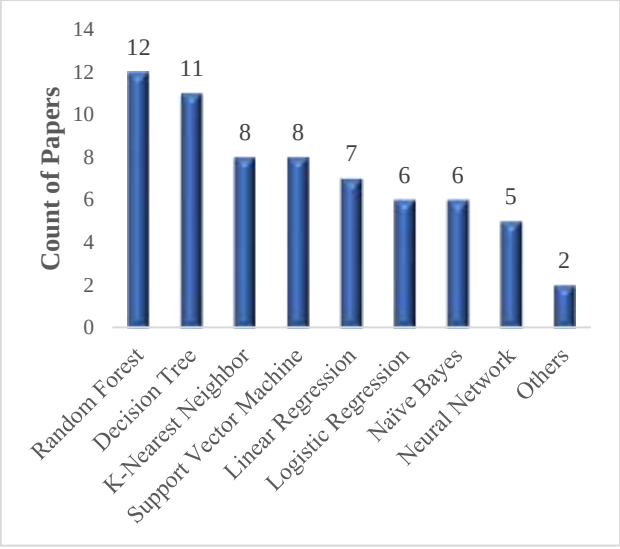

Figure 6: Frequency of different categories of papers published

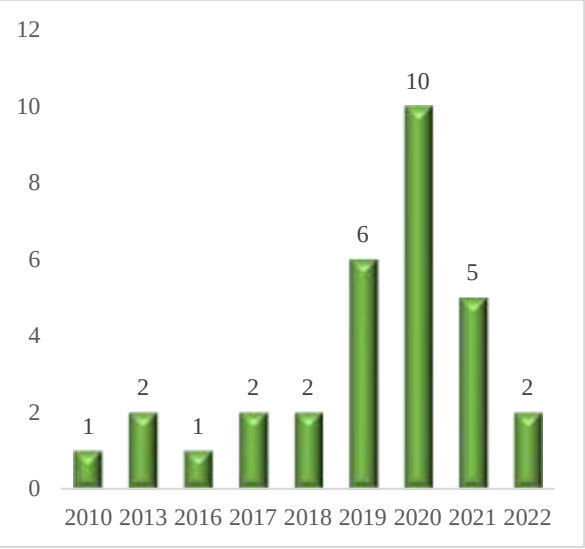

Figure 7: Papers published year wise

Industrial Engineering Journal
ISSN: 0970-2555
Volume : 52, Issue 6, No. 5, June : 2023

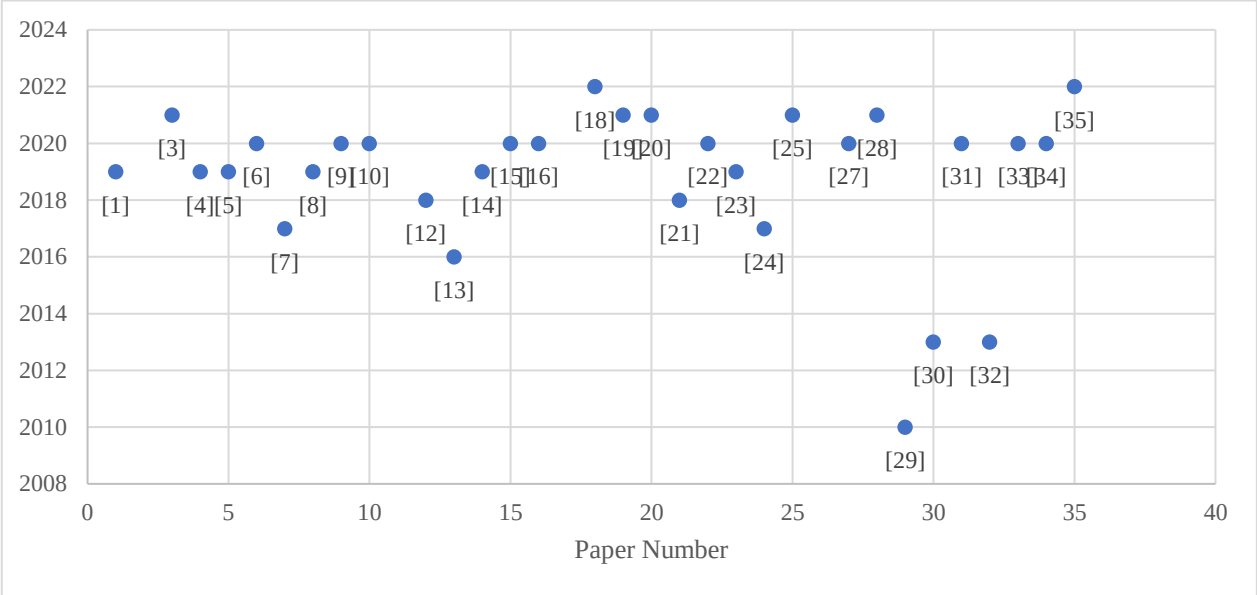

Figure 8: Year of individual paper publication

V. Conclusion

A high level of prediction accuracy is required by a machine learning-based prediction model for movie success. However, throughout time, efforts to increase the model's predictive ability have solely focused on changing the algorithms rather than identifying significant aspects that would be essential to predicting the success of a film. In the past, academics have mostly concentrated on introducing new machine learning algorithms and evaluating their performances. This has pretty much been the only goal of their study. While these initiatives have helped to raise forecast accuracy, we think there is still room for improvement by considering other factors. Future plans call for developing a framework for the movie investment market in which machine learning-based predictive modelling will be used to forecast a movie's score based on the value of its characteristics, allowing us to categorize the film as a hit, average, or flop.

References

[1] Sarantinos, J G. "What Is A Blockbuster Movie?" Https://Gideonsway.Wordpress.Com/, 12 Aug. 2012, gideonsway.wordpress.com/2012/08/12/what-is-a-blockbuster-movie/. Accessed 8 May 2023.

[2] Singh, Aishwarya. "KNN Algorithm: Introduction to K-Nearest Neighbors Algorithm for Regression." Https://Www.Analyticsvidhya.Com/, 22 Aug. 2018, www.analyticsvidhya.com/blog/2018/08/k-nearest-neighbor-introduction-regression-python/. Accessed 8 May 2023.

[3] BEERS, BRIAN . "What Is Regression? Definition, Calculation, and Example." Investopedia, 31 Mar. 2023, www.investopedia.com/terms/r/regression.asp. Accessed 8 May 2023.

[4] RAY, Sunil . "Naive Bayes Classifier Explained: Applications and Practice Problems of Naive Bayes Classifier." Https://Www.Analyticsvidhya.Com/, 11 Sept. 2017, www.analyticsvidhya.com/blog/2017/09/naive-bayes-explained/. Accessed 8 May 2023.

[5] Saini, Anshul. "Support Vector Machine(SVM): A Complete Guide for Beginners." Https://Www.Analyticsvidhya.Com/, 12 Oct. 2021, www.analyticsvidhya.com/blog/2021/10/support-vector-machinessvm-a-complete-guide-for-beginners/. Accessed 8 May 2023.

[6] Bonthu, Harika. "An Introduction to Logistic Regression." Https://Www.Analyticsvidhya.Com/, 11 Jul. 2021, www.analyticsvidhya.com/blog/2021/07/an-introduction-to-logistic-regression/. Accessed 8 May 2023.
[7] E R, Sruthi. "Understanding-random-forest." Https://Www.Analyticsvidhya.Com, 17 Jun. 2021, www.analyticsvidhya.com/blog/2021/06/understanding-random-forest/. Accessed 8 May 2023.
[8] Masih, Salman, and Imran Ihsan. "Using Academy Awards to Predict Success of Bollywood Movies Using Machine Learning Algorithms." International Journal of Advanced Computer Science and Applications, vol. 10, no. 2, 2019, pp. 438-446.
[9] Latif, Muhammad H., and Hammad Afzal. "Prediction of Movies Popularity Using Machine Learning Techniques." IJCSNS International Journal of Computer Science and Network Security, vol. 16, no. 8, 2016, pp. 127-131.
[10] Apala, Krushikanth R., et al. "Prediction of Movies Box Office Performance Using Social Media." IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, 2013, pp. 1209-1214.
[11] Razeen, Faaez, et al. "Predicting Movie Success Using Regression Techniques." IEEE, 2020.
[12] Jatale, Sanyam, et al. "Analyzing and Predicting The Success Of Box Office Collection Of A Movie Using Machine Learning." International Journal of Advanced Research in Science, Communication and Technology (IJARSCT), vol. 6, no. 1, 2021, pp. 325-331.
[13] L K, Jayashree, et al. "MOVIE SUCCESS RATE FORECAST SYSTEM." International Research Journal of Computer Science (IRJCS), vol. 6, no. 6, 2019, pp. 247-255.
[14] Gupta, Kshitij, et al. "Movie Success Prediction." International Journal of Recent Technology and Engineering (IJRTE), vol. 8, no. 3, 2019, pp. 5659-5663.
[15] Vignesh , Ranjitha , et al. "Movie Success Prediction Using Recommendation Techniques." IAEME, vol. 11, no. 4, 2020, pp. 422-426.
[16] V, Ananthraman, et al. "Movie Success Prediction Using Data Mining." "GRD Journals | Global Research and Development Journal for Engineering | National Conference on Emerging Research Trend in Electrical and Electronics Engineering (ERTE'19)", 2019, pp. 312-314.
[17] Sivakumar, Pirunthavi, et al. "Movie Success and Rating Prediction Using Data Mining Algorithms." Journal of Information Systems & Information Technology (JISIT), vol. 5, no. 2, 2020, pp. 72-80.
[18] Abarja, Rudy A., and Antoni Wibowo. "Movie Rating Prediction Using Convolutional Neural Network Based on Historical Values." International Journal of Emerging Trends in Engineering Research, vol. 8, no. 5, 2020, pp. 2156-2164.
[19] Antony, Elizabeth, and Nimmy Francis. "Movie Box Office Success Prediction Using Machine Learning." Proceedings of the National Conference on Emerging Computer Applications (NCECA), vol. 4, no. 1, 2022, pp. 621-624.
[20] Kumar, Chintu. "MOVIE SUCCESS PREDICTION USING RANDOM FOREST (RF) CLASSIFIER." Journal Of Harmonized Research (JOHR), vol. 8, no. 3, 2020, pp. 81-88.
[21] Josey, Midhun. "PREDICTING BOX OFFICE SUCCESS OF MOVIES USING SOCIAL MEDIA." JETIR, vol. 6, no. 6, 2019, pp. 191-195.
[22] Patel, Payal, et al. "MOVIE SUCCESS PREDICTION USING DATA MINING." International Research Journal of Engineering and Technology (IRJET), vol. 7, no. 4, 2020, pp. 3247-3250.
[23] G, Sandhiya, et al. "MOVIE RECOMMENDATION SYSTEM USING MACHINE LEARNING ALGORITHMS." International Research Journal of Engineering and Technology (IRJET), vol. 8, no. 3, 2021, pp. 2723-2725.
[24] Asur , Sitaram , and Bernardo A. Huberman. "Predicting the Future with Social Media." IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, vol. 1, 2010, pp. 492-499.
[25] Mahmud, Quazi I., et al. "A Machine Learning Approach to Predict Movie Revenue Based on Pre-Released Movie Metadata." Journal of Computer Science 2020 a Creative Commons Attribution (CC-BY) 3.0 License, 2020, pp. 2156-2164.
[26] Chaki, Dipankar, et al. "A Machine Learning Approach to Predict Movie Box-Office Success." 20th International Conference of Computer and Information Technology (ICCIT), 2017, pp. 754-759.

[27] Peleja, Filipa et al. “A recommender system for the TV on the web: integrating unrated reviews and movie ratings.” Multimedia Systems 19 (2013): 543-558.
[28] Singh, Abhishek, et al. "A Research Paper on Machine Learning Based Movie Recommendation System." International Research Journal of Engineering and Technology (IRJET), vol. 8, no. 3, 2021, pp. 990-997.
[29] Darekar, Sachin, et al. "Movie Success Prediction Based on Classical and Social Factors." IJESC, vol. 8, no. 2, 2018, pp. 16101-16104.
[30] Naik, Altamash. "Prediction Models for Box Office Revenue Before Theatrical Release of Movies." National College of Ireland, 2020.
[31] Fatima, Ansari S., and Shruti Pimple. "Movie Success Prediction Using Machine Learning." International Journal Of Engineering Research And Development, vol. 16, no. 6, 2020, pp. 10-14.
[32] Adetunji, Oluwatofunmi , et al. "Design of a Movie Review Rating Prediction (MR2P) Algorithm." Technoscience Academy, vol. 6, no. 4, 2020, pp. 423-432.
[33] D.M.L, Dissanayake, and Vidanagama V.G.T.N. "Early Prediction of Movie Success Using Machine Learning Models." International Journal of Computer Applications, vol. 183, no. 44, 2021, pp. 14-21.
[34] V, Gangadhar, et al. "Movie Prior Release Box Office Prediction A Machine Learning Based Approach." European Journal of Molecular & Clinical Medicine, vol. 7, no. 2, 2020, pp. 5016-5028.
[35] Topf, Patrick, and Ilaria Ossella-Durbal. "Examining Success at the Domestic Box-Oice in the Motion Picture Industry." Illinois Wesleyan University Digital Commons @ IWU, 2010.
[36] Sadashiv, Sagar, et al. "MOVIE SUCCESS PREDICTION USING MACHINE LEARNING." International Research Journal of Modernization in Engineering Technology and Science, vol. 3, no. 7, 2021, pp. 2021-2024.
[37] Kharb, Latika, et al. "Forecasting Movie Rating Through Data Analytics." Springer Nature Singapore Pte Ltd. 2020, 2020, pp. 250-256.
[38] SHAHRIVAR, POJAN, and CARL JERNBÄCKER. "Predicting Movie Success Using Machine Learning Techniques." FöRutsäGelse Av Filmers FramgåNg Med MaskininläRning School of Computer Science and Communication, 2017.
[39] D, ABİDİN, et al. "Movie Rating Prediction with Machine Learning Algorithms on IMDB Data Set." International Conference on Advanced Technologies, Computer Engineering and Science, 2018, pp. 231-235.
[40] Dhir, Rijul and Anand Arokia Raj. “Movie Success Prediction using Machine Learning Algorithms and their Comparison.” 2018 First International Conference on Secure Cyber Computing and Communication (ICSCCC) (2018): 385-390.
[41] Jain, Hrithik, et al. "MOVIE SUCCESS PREDICTION USING DATA MINING." International Journal of Creative Research Thoughts (IJCRT), vol. 10, no. 4, 2022, pp. 291-294.
[42] Verma, Hemraj, and Garima Verma. "Prediction Model for Bollywood Movie Success: A Comparative Analysis of Performance of Supervised Machine Learning Algorithms." Https://Link.Springer.Com, 27 Jun. 2019, link.springer.com/article/10.1007/s12626-019-00040-6. Accessed 8 May 2023.
[43] VATTIKUNTA, HARSHADEEP. "Predicting Movie Success..!" Https://Www.Kaggle.Com/, www.kaggle.com/code/harshadeepvattikunta/predicting-movie-success. Accessed 8 May 2023.